COMPUTATIONAL ANALYSIS OF ACTION MECHANISM AND EVOLUTIONARY INSIGHT OF CHITIN DEGRADING ENZYMES IN *BACILLUS CEREUS*

AMMARAH FATEEN

School of Biochemistry and Biotechnology, University of the Punjab, 54590 Lahore, Pakistan.

MUHAMMAD ZEESHAN AHMED

Department of Biochemistry, Bahauddin Zakariya University, Multan 60800, Pakistan.
School of Biochemistry and Biotechnology, University of the Punjab, 54590 Lahore, Pakistan.

NAEEM MAHMOOD ASHRAF

School of Biochemistry and Biotechnology, University of the Punjab, 54590 Lahore, Pakistan.

MUHAMMAD KHURSHID

School of Biochemistry and Biotechnology, University of the Punjab, 54590 Lahore, Pakistan.

MOAZZAM ALI

School of Biochemistry and Biotechnology, University of the Punjab, 54590 Lahore, Pakistan.

ZEESHAN MUTAHIR*

School of Biochemistry and Biotechnology, University of the Punjab, 54590 Lahore, Pakistan.
*Corresponding Author Email: zeeshan.ibb@pu.edu.pk, ORCID ID: 0000-0001-9550-8170

Abstract

Lytic Polysaccharide Monooxygenases (LPMOs) and chitinases constitute pivotal chitinolytic enzymatic components in *Bacillus cereus*, exerting significant influence on the degradation of chitin. This investigation used bioinformatics tools to endeavor the evolution and action mechanisms of enzymes (chitinases and BcLPMO10B) influencing chitin and cellulose degradation in *B. cereus*, employing computational methods for phylogenetic analysis, 3D structure modeling, validation, and molecular dynamics simulations (MDS), offering insights into their potential applications in biofuels, medicine, and agriculture. Phylogenetic analysis of chitinases and BcLPMO10B exhibit strong associations with other bacteria, indicating a potential history of shared genetic material over time. To achieve the objective of analyzing action mechanism, the homology modeling technique was employed to construct three-dimensional structures of BcLPMO10B and chitinases using SwissModel, I-TASSER, and ROBETTA Baker Laboratory. Following the comparative validation of predicted structures utilizing various servers, 3D structures generated by the Swiss Model were chosen for subsequent computational analysis. Subsequently, docking studies elucidated the functional significance of target proteins; through the prediction of ligand binding modes. The research also explored the dynamic behaviors of BcLPMO10B and chitinases by MDS. AutoDock Vina revealed robust binding affinities and unveiled key interacting amino acids of chitinases with N-acetyl-D(+)-glucosamine (GLcNAc) and BcLPMO10B with 2,6-dimethoxyphenol. Furthermore, MDS validated the stability of these complexes by assessing the temporal motion of individual atoms within these complexes, employing Root Mean Square Deviation (RMSD) analysis, recording low RMSF values, and evaluating the extent of hydrogen bonding interactions. The implications of these complexes extend the enzymatic efficiencies, functionalities and substrate binding affinities. In sum, this computational analysis augments our comprehension of these enzymes and highlights their potential applications in biotechnology and pharmaceutical research.

Keywords: *Bacillus cereus*, Chitin-Degrading Enzymes, *In-Silico* Characterization, Molecular Docking, Molecular Dynamic Simulation, Phylogenetic Analysis.

1. INTRODUCTION

Chitin is a versatile biopolymer that ranks second to cellulose among the most abundant polymers on Earth, found in algae, crustaceans, insects, and fungi. It is principally composed of (1→4)-β-linked N-Acetyl-D-glucosamine (GLcNAc) residues. Interestingly, chitin exists in three distinct structural forms, known as alpha, beta, and gamma. Among these, the alpha form predominates as the most commonly encountered variation (Beier and Bertilsson 2013). Chitin, when degraded into simple monomers, can be utilized in various medical, agricultural and industrial applications (Bhattacharya, Nagpure, and Gupta 2007). Chitinases are chitinolytic enzymes which can hydrolyze glycosidic bond (Oyeleye and Normi 2018). Lytic Polysaccharide Monooxygenases (LPMOs) have also been reported to be involved in chitin degradation through an oxidative mechanism, they can enhance chitinolytic activity in synergy with other chitinases (Mutahir et al. 2018). Microorganisms particularly bacteria are major source of chitinases and LPMOs (Bhattacharya et al. 2007). *Bacillus cereus*, a Gram-positive, rod-shaped bacterium, is commonly encountered in terrestrial environments, including soil, food matrices, and marine sponges. It exhibits a diverse repertoire of virulence determinants, encompassing phospholipase C, cereulide, sphingomyelinase, metalloproteases, and cytotoxin K (El-Arabi and Griffiths 2021; Esmkhani and Shams 2022; Yang et al. 2023). Notably, certain strains of *B. cereus* pose a risk to human health because of their ability to generate spores, which can lead to foodborne illnesses (Bottone 2010; Esmkhani and Shams 2022). Conversely, other strains demonstrate probiotic attributes in animal hosts and establish mutualistic symbioses with select plant species. It is noteworthy that *B. cereus* exhibits facultative anaerobic behavior and proficiency in the formation of robust endospores. These dual traits, coupled with the bacterium's propensity for spore and biofilm production, present notable challenges within the context of food safety, particularly regarding the potential for contamination. Additionally, flagellar motility augments biofilm development (Roberto et al. 2020).

The ecological niches where *B. cereus* thrives encompass decomposing organic matter, freshwater and marine ecosystems, phytobiomes, fomites, and the gastrointestinal tracts of various animals. This bacterium is an indigenous constituent of soil ecosystems, thereby contributing to soil contamination (Deka et al. 2022). *B. cereus* spores are ubiquitously dispersed throughout the environment, comprising dust, soil, cereal crops, aqueous environments, and other reservoirs. Consequently, *B. cereus* emerges as a recurrent contaminant in raw agricultural commodities, with starchy foods like rice and potatoes being particularly associated with emetic (vomiting) toxin-mediated outbreaks. *B. cereus*, and its broader genus Bacillus, exhibit remarkable versatility in biotechnological and industrial domains, exerting profound influences across diverse facets of human activity. These impacts extend to both pathogenic and beneficial bacterial species within the Bacillus genus, as well as numerous industrial sectors (El-Arabi and Griffiths 2021; Kowalska et al. 2023; Rodrigo, Rosell, and Martinez 2021).

B. cereus holds significant importance in various industrial applications. Bacillus species, including *B. cereus*, are utilized for the production of extracellular enzymes, such as proteases and amylases, which have wide-ranging applications in industries such as detergent manufacturing and food processing (Demirkan, Aybey Çetinkaya, and Abdou 2021; Vaikundamoorthy et al. 2018). *B. cereus* demonstrates potential as a producer of primary metabolites, including vitamins and ribonucleosides. Bacillus species, including *B. cereus*, exhibit the capability to synthesize secondary metabolites like bacteriocins and biosurfactants, showcasing their versatility across diverse industrial sectors (Feliatra et al. 2021). Select strains of *B. cereus* possess the ability to promote plant growth, making them valuable as inoculants for enhancing crop yields and overall productivity. *B. cereus* is under investigation for its effectiveness in bioremediation applications, notably in the decolorization of industrial effluents generated by pulp and paper production. These efforts have the potential to mitigate environmental pollution associated with waste disposal practices (Jan et al. 2019; Kulkova et al. 2023; Saleem, Ahmad, and Ahmad 2014).

LPMOs and chitinases represent pivotal enzymes with substantial roles in the degradation of biomass including both cellulose and chitin, playing essential roles in ecological carbon cycling and industrial applications (Eijsink et al. 2019; Li et al. 2023; Zhang et al. 2023). LPMOs constitute a class of copper-dependent enzymes that execute polysaccharide cleavage through oxidative mechanisms (Eijsink et al. 2019; Li et al. 2023; Zhang et al. 2023). Some LPMOs can oxidize C1 carbon and while others can break C4 carbon of the glycoside. However, some of the LPMOs can oxidize both of the C4 and C1 carbon of the glycosidic linkage (Li et al. 2012). LPMOs often collaborate synergistically with cellulases to augment the deconstruction of biomass (Eijsink et al. 2019; Li et al. 2023; Zhang et al. 2023). LPMOs are comprised of AA9-AA10 and AA13-AA15 families of the CAZy database (Levasseur et al. 2013) that have different regioselectivities and substrate specificities. The discovery of AA9 LPMOs has notably reshaped our comprehension of cellulose degradation with oxidizing activity by fungi. In contrast, AA10 LPMOs family can be derived from bacteria, archaea, viruses and fungi that can degrade the chitin (Zhou et al. 2019). Moreover, LPMOs have been instrumental in facilitating the production of cellulose nanofibrils, bacterial pathogenicity, and viral virulence expanding their practical relevance (Agostoni, Hangasky, and Marletta 2017; Chiu et al. 2015; Eijsink et al. 2019; Li et al. 2023; Zhang et al. 2023).

In contrast, chitinases represent a class of hydrolytic enzymes responsible for catalyzing chitin degradation and are synthesized by a diverse spectrum of microorganisms with GH18 and GH19 classes (Levasseur et al. 2013). In natural contexts, microbial chitinases shoulder the primary responsibility for chitin decomposition, thereby exerting a pivotal role in maintaining the carbon-to-nitrogen ratio equilibrium within ecosystems (Adrangi and Faramarzi 2013; Gomaa 2021; Rathore and Gupta 2015). Beyond ecological implications, chitinases have found multifaceted applications in human healthcare, including the production of pharmaceutically significant agents, the isolation of protoplasts from fungi and yeast, the control of pathogenic fungi, and the management of chitinous waste (Singh

et al. 2020). Notably, the generation of chitooligosaccharides during chitin degradation has been associated with a range of human health benefits, including antimicrobial, antioxidant, anti-inflammatory, and antitumor properties, underscoring their relevance in medical and therapeutic contexts (Anil 2022; Azuma et al. 2015).

In-silico approaches are computational methods that are used to study biological systems. These methods have become increasingly important in the field of structural biology and enzyme studies. Computational analysis can help in predicting as well as determining the structural, functional, and even interaction of different proteins with other molecules (Bhat et al. 2022; Kęska, Gustaw, and Stadnik 2021). In this respect, with the help of structural analysis of the proteins using different computational methods, their possible ways of interactions with the atoms and molecules of ligands can facilitate in designing of new and novel drugs, along with mechanism of action (Agnihotry et al. 2021; Batool, Ahmad, and Choi 2019; Dhiman et al. 2023; Hasan, Rony, and Ahmed 2021; Kingdon and Alderwick 2021). In parallel, *in-silico* studies can also help in understanding the enzymatic mechanism of action by using computational methods like molecular docking and simulation to model the interactions between enzymes and their substrates. In this way, alterations in an enzyme's structure can be predicted to influence the enzyme's catalytic activity. Computational approaches have explored a new gateway to study the structural biology of proteins and other biomolecules for the prediction of their behavior to understand the complex biological systems that would have been difficult to provide experimentally (Agnihotry et al. 2021; Batool et al. 2019; Dhiman et al. 2023; Hasan et al. 2021; Kingdon and Alderwick 2021).

In a previous study, tetra-modular LPMO [BcLPMO10A, composed of a catalytic domain, two fibronectin-type III (FnIII)-like domains, and a carbohydrate-binding module (CBM5)] from *B. cereus* was studied to validate its crucial role in promoting substrate binding and protecting the enzyme from inactivation, ultimately enhancing chitinase activity during α-chitin degradation (Mutahir et al. 2018). This study aimed to explore the evolutionary insights and mechanism of action of enzymes (chitinases and LPMO) that exert significant influence on the degradation of chitin and cellulose present in *B. cereus,* using Bioinformatics tools. For the evaluation of evolutionary relation of *B. cereus* chitinases and LPMO with other related bacterial species, we constructed phylogenetic trees using computational methodology.

For the determination of action mechanism of chitin degrading enzymes, three-dimensional (3D) structures of chitin binding single domain LPMO (called herein as BcLPMO10B) and chitinases (N-acetylglucosaminidase, and Endochitinase) were constructed using template-based and artificial intelligence (AI)-based homology modeling, that provides insight into the architectural arrangements of these enzymes. These predicted models of BcLPMO10B and chitinases were compared with different validation methods to select the best 3D models which further analyze their interactions with native substrates and to predict their potential mode of action. Using docking studies, this investigation predicted ligand binding modes for these target proteins and sheds light

on their functional roles. This study also explored the dynamic behavior of chitinases and BcLPMO10B, using molecular dynamics simulations (MDS). By simulating the temporal motion of individual atoms, we endeavored to unravel the impact of structural alterations on the enzymatic functionalities centered on addressing research queries about the structural and functional attributes of BcLPMO10B and chitinases sourced from *B. cereus*, employing *in-silico* methodologies. The insights garnered from this investigation are poised to advance our comprehension about the mechanism and potential roles played by these enzymes in biological processes, potentially bearing relevance for diverse applications, such as in the production of next-generation biofuels, medicine and agriculture.

2. MATERIAL AND METHODS

Sequences of Chitin-Degrading Enzymes

The sequences of chitinases and LPMO were identified from the GH18 and AA10 families of CAZy database (Drula et al. 2022). While, the sequences of chitin-degrading enzymes, specifically N-acetylglucosaminidase (Chi1), endochitinase (Chi2), and a single domain LPMO (BcLPMO10B) originating from the *B. cereus*, were retrieved from the UniProt database using the respective accession IDs: Q81AG3, Q81IF9, and Q81CE4.

Physiochemical Properties

Expasy ProtParam web server (https://web.expasy.org/protparam/) was employed to forecast the physicochemical characteristics of the Chi1, Chi2, and BcLPMO10B proteins (Gasteiger et al. 2003, 2005; Wilkins et al. 1999). This analysis yielded various physicochemical parameters, including molecular weight, extinction coefficient (Ec), theoretical isoelectric point (pI), the total count of positive (+R) and negative (-R) residues, aliphatic index (AI), grand average of hydropathy (GRAVY), and instability index (II) for these proteins.

Structural and Functional Characterization

For a comprehensive understanding of the structural and functional characterization of Chi1, Chi2, and BcLPMO10B proteins, we turned to the InterPro online server (https://www.ebi.ac.uk/interpro/) (Paysan-Lafosse et al. 2023). InterPro delves deep into protein analysis by providing structural and functional insights through a multifaceted approach that categorizes proteins into families, predicts domains, and identifies critical sites using predictive models called signatures. It integrates a wide array of databases, including NCBIfam, SFLD, PANTHER, HAMAP, PROSITE profiles, PROSITE patterns, SMART, CDD, PRINTS, Pfam, PIRSF, SUPERFAMILY, and CATH-Gene3D, along with tools like Phobius, SignalP, Coils, MobiDBLite, and TMHMM, to facilitate the identification of various protein features. InterPro's comprehensive approach extends to classifying proteins into different categories using databases like SignalP_EUK, SignalP_GRAM_POSITIVE, SignalP_GRAM_NEGATIVE, AntiFam, FunFam, and PIRSR, forming the InterPro consortium, a robust resource for protein analysis that

capitalizes on the strengths of multiple databases. Additionally, Argot2 is used to extract functional information from Chi1, Chi2, and BcLPMO10B proteins, employing semantic clustering of Gene Ontology (GO) terms and a weighting scheme for efficient inference of biological attributes. This is accomplished through methods such as BLAST and HMMER, providing valuable insights into the functional characteristics of these proteins.

Phylogenetic Analysis

In order to determine the evolutionary origin, functions and relationships of Chi1, Chi2, and BcLPMO10B proteins in *B. cereus* with other different but related bacterial species, the phylogenetic analysis was performed. The homologs sequences of Chi1, Chi2, and BcLPMO10B proteins in different bacterial species were searched by using basic local alignment search tool for proteins (BLASTp) (Altschul et al. 1997, 2005) from experimental clustered nr database. The clustered nr database is more compact and smaller that shows sequence length and identity at 90% with more taxonomic depth (Steinegger and Söding 2017). The obtained sequences of Chi1, Chi2, and BcLPMO10B proteins were aligned separately in MEGA 11 software (Tamura, Stecher, and Kumar 2021) using Clustal W. Prior to phylogenetic analysis, best substitution models were required to be identified for the phylogenetic tree construction. Maximum likelihood method (MLM) was used for the neighbor joining tree. The best substitution model was selected on the bases of lowest BIC (Bayesian Information Criterion) value. The phylogenetic tree was then constructed applying Bootstrap method with 1000 replicates by setting best substitution model and default setting (Keane et al. 2006; Tamura et al. 2011; Uddin et al. 2022; Zhang et al. 2019).

Secondary Structure Prediction

We employed the SOPMA tool (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) for forecasting the secondary structures of the Chi1, Chi2, and BcLPMO10B proteins. In the analysis, default settings were used, which included four conformational states (helix, sheet, turn, and coil), a window width of 17, and a similarity threshold set at 8 (Geourjon and Deléage 1995). Additionally, to depict the secondary structure in a graphical cartoon format, we harnessed the PSI-blast-based secondary structure prediction tool known as PSIPRED (McGuffin, Bryson, and Jones 2000), accessible through the online server at http://bioinf.cs.ucl.ac.uk/psipred/. This was applied to the Chi1, Chi2, and BcLPMO10B proteins.

Homology Modeling

The ab initio modeling of the 3D structures of Chi1, Chi2, and BcLPMO10B proteins was necessitated by the absence of experimentally validated structures for these proteins in both the Protein Data Bank (PDB) and the UniProt database (UniProtKB). The UniProt database provided the protein IDs for Chi1 (Q81AG3), Chi2 (Q81IF9), and BcLPMO10B (Q81CE4). Subsequently, these protein IDs underwent submission to the SWISS-MODEL web server, a highly automated tool utilizing homology-based modeling for

structure generation (Waterhouse et al. 2018). A 40% or higher sequence identity with proteins of known structures allowed for confident comparative modeling of Chi1, Chi2, and BcLPMO10B.

However, incomplete structures due to missing residues and gaps in the pairwise alignment with experimental templates (PDB codes 4S3J, 6BT9, and 5WSZ) necessitated the synthesis of the complete structures for each protein. This was achieved through both the I-TASSER server (http://zhanglab.dcmb.med.umich.edu/I-TASSER) (Yang and Zhang 2015; Zheng et al. 2021; Zhou et al. 2022) and the ROBETTA Baker server (http://robetta.bakerlab.org), employing the RoseTTAFold method (Kim, Chivian, and Baker 2004), a default option utilizing a deep learning-based modeling approach known for its superior performance. The selection of the most reliable 3D structures was based on confidence values, which range from 0.00 (low confidence) to 1.00 (high confidence), with higher values indicating greater reliability. Furthermore, the AlphaFold-predicted structures of Chi1, Chi2, and BcLPMO10B can be accessed on UniProt and were compared with the predicted structures obtained from the I-TASSER server and ROBETTA Baker server.

Structure Validation

The 3D structures of Chi1, Chi2, and BcLPMO10B proteins generated by SwissModel, I-TASSER, ROBETTA Baker Laboratory, and AlphaFold underwent thorough validation using PROCHECK, ERRAT, and Verify3D on the UCLA-DOE LAB—SAVES v6.0 platform (https://saves.mbi.ucla.edu/). PROCHECK (Laskowski et al. 1993; Rodriguez et al. 1998) assessed structural conformation through Ramachandran plots, requiring over 90% of amino acid residues to fall within favored regions for high-quality models. ERRAT (Colovos and Yeates 1993) analyzed non-bonded interactions and overall structure quality, expressed as the percentage of residues with error values below the 95% rejection limit. Verify3D (Bowie, Lüthy, and Eisenberg 1991; Lüthy, Bowie, and Eisenberg 1992) determined compatibility by comparing 3D atomic models with their amino acid sequences.

We also employed the QMEAN server (https://swissmodel.expasy.org/qmean/) to calculate a composite scoring function for the 3D structures. This allowed us to estimate both local (per-residue) and global (entire structure) errors (Benkert, Künzli, and Schwede 2009). The QMEANDisCo global score is the average per-residue score, and the provided error estimate is based on the standard deviation between QMEANDisCo global score and lDDT (ground truth). It is calculated using models of similar size to enhance prediction reliability. A QMEAN score below -4.0 indicates a lower-quality predicted structure (Kryshtafovych, Fidelis, and Tramontano 2011; Olowosoke et al. 2023).

For an overarching evaluation of each model's quality, we utilized the ProSA tool (https://prosa.services.came.sbg.ac.at/prosa.php). The results were expressed as Z-scores, which indicated the quality of the modeled proteins compared to all other

experimental protein structures determined through methods such as X-ray crystallography and NMR (Wiederstein and Sippl 2007).

The best predicted structures of Chi1, Chi2, and BcLPMO10B proteins were further validated by utilizing MolProbity (http://molprobity.biochem.duke.edu/) to evaluate model quality (Williams et al. 2018).

Active Site Prediction

We employed the Computed Atlas of Surface Topography of proteins, known as CASTp 3.0, to anticipate the active sites within the 3D Swiss Models of Chi1, Chi2, and BcLPMO10B proteins. CASTp (http://sts.bioe.uic.edu/castp/index.html?4jii) is an online tool used for identifying and measuring empty spaces within 3D protein structures (Tian et al. 2018). We submitted the modeled Chi1, Chi2, and BcLPMO10B proteins to this server, which then predicted the specific amino acids responsible for binding interactions.

Molecular Docking

Protein Preparation

To prepare the 3D-modeled structures of Chi1, Chi2, and BcLPMO10B proteins for molecular docking, a series of steps were undertaken using MGL tools (Ahmed et al. 2021). These steps included removing water molecules, ligands, non-polar hydrogen charges, and side chains other than the standard 20 amino acids from each protein structure. Any missing atoms on the amino acids within these structures were meticulously restored. Furthermore, polar hydrogens, hydrogen atoms for histidine residues, and Kollman charges were meticulously added to all protein structures. Lastly, a thorough check was conducted to ensure the balanced distribution of charges across the entire set of protein residues. The protein structures were then subjected to optimization and energy minimization using Discovery Studio.

Ligand Preparation

We obtained the structures of the ligands, GLcNAc (PubChem CID: 439174) for Chi1 and Chi2, and 2,6-Dimethoxyphenol (PubChem CID: 7041) for BcLPMO10B proteins, from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/). These ligand structures underwent a minimization process, and their torsion angles were carefully defined in preparation for the molecular docking procedure.

Docking

The process of molecular docking involving protein-ligand complexes was facilitated using AutoDock (version 4.2.6) (Ahmed et al. 2021). Both proteins and ligands were meticulously prepared and saved in. pdbqt files. Atom-specific affinity maps for various ligand atom types, along with electrostatic and desolvation potentials, were generated through the Auto-grid tool. During the molecular docking procedure, the protein molecule remained rigid. The outcomes of the molecular docking experiments were visualized using PyMol and Discovery Studio. To ascertain the nature and extent of protein-ligand

interactions, we utilized the Protein-Ligand Interaction Profiler, available at https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index (Adasme et al. 2021). Additionally, we assessed the drug score of the docked files through ProteinsPlus (https://proteins.plus/) (Schöning-Stierand et al. 2020).

Molecular Dynamics Simulation

Over a 100-nanosecond timeframe, MDS were executed using the Desmond software, developed by Schrödinger LLC (Bowers et al. 2006). The initial phase of MDS involved generating receptor and ligand complexes through docking experiments. Molecular docking studies predict the static binding state of ligands within a protein's active site, offering valuable insights into the spatial orientation of molecules in the binding site (Ferreira et al. 2015). To facilitate MDS, which track atomic movements over time using Newton's classical equations of motion (Hildebrand, Rose, and Tiemann 2019; Rasheed et al. 2021), the receptor-ligand complexes (specifically, Chi1-GLcNAc, Chi2-GLcNAc, and BcLPMO10B-2,6-dimethoxyphenol) underwent preprocessing through the Protein Preparation Wizard within Maestro. This involved complex optimization and minimization steps, with all systems constructed using the System Builder tool. The simulation environment adopted the Transferable Intermolecular Interaction Potential 3 Points (TIP3P) solvent model within an orthorhombic box, with counter ions introduced for model neutrality (Maiorov and Crippen 1994; Shivakumar et al. 2010). To replicate physiological conditions, a 0.15 M sodium chloride (NaCl) concentration was incorporated. The NPT ensemble was chosen for the entire simulation, maintaining a temperature of 300 K and a pressure of 1 atm. Before the simulation, the models underwent relaxation. Trajectory data were saved at regular 100-picosecond intervals throughout the simulation. The simulation's stability was evaluated by monitoring the root mean square deviation (RMSD) of both protein and ligand structures over time.

3. RESULTS

Sequences of Chitin-Degrading Enzymes

The amino acid sequences of the Chi1, Chi2, and BcLPMO10B proteins in *B. cereus*, consisting of 430, 674, and 221 amino acid residues, respectively, were acquired and employed as primary data in the present investigation.

Physiochemical Properties

The Expasy server's ProtParam tool predicted physicochemical parameters of chitin-degrading enzymes in *B. cereus*. Physicochemical attributes play a pivotal role in elucidating the structural and functional characteristics of proteins, facilitating the design and prediction of optimized protein variants tailored for specific molecular mechanisms and applications (Wang and Zou 2023). Table 1 presents the computed values of these physicochemical parameters for the Chi1, Chi2, and BcLPMO10B proteins.

Table 1: Physiochemical parameters of Chi1, Chi2, and BcLPMO10B proteins

Sr. No.	Physiochemical parameters	Chi1	Chi2	BcLPMO10B
1	Length in base pairs	430	674	221
2	Mol. Wt. in kD	48.19	74.23	24.11
3	Ec (M^{-1} cm^{-1}, at 280nm)	64290	161160	39545
4	pI	9.38	5.77	9.24
5	R^+	46	64	22
6	R^-	35	71	16
7	AI	84.79	68.89	79.37
8	GRAVY	-0.284	-0.548	-0.299
9	II	35.49	21.03	38.3

Structural and Functional Characterization

The structural and functional characterization of protein sequence can help in understanding the properties, biological functions, molecular interactions, evolution and biological pathways. Bioinformatics tools compare the gene or protein homologs to characterize their biological insights and application (María Hernández-Domínguez et al. 2020). In the current study. chi1, chi2 and BcLPMO10B proteins sequences are characterized for their structural and functional annotation using various databases such as InterPro, PANTHER, Pfam, and others. The detailed characterized information of domains, homologous super-families, and related features of chi1, chi2 and BcLPMO10B proteins are represented in Supplementary Material Tables S1, S2, and S3. Figure 1 illustrates the domains of Chi1, Chi2, and BcLPMO10B proteins, highlighting their respective locations and sequences. Similarly, another computational analysis, GO, used the enrichment analysis, network analysis, or machine learning for identifying the biological processes, molecular functions, and cellular components of genes and proteins (Sengupta et al. 2022). Table 2 provides molecular function, biological process, and cellular components of Chi1, Chi2, and BcLPMO10B proteins. Detailed GO annotations (molecular function, biological process, and cellular components) of Chi1, Chi2, and BcLPMO10B proteins including their respective names, total scores, and internal confidence values are presented in Supplementary Material Table S4.

Supplementary Material

Table S1: Functional Annotation and Domain Prediction of the Chi1 protein using InterPro

Section	Database	ID/Annotation	Domain/Region Identified	Protein Region
1. Domain	InterPro	IPR018392	LysM domain	3-48
	CDD/ PROSITE	cd00118/ PS51782	LysM/ LysM domain profile	3-47
	SMART/ SUPERFAMILY	SM00257/ SSF54106	LysM_2/ LysM domain	4-48
	Pfam	PF01476	LysM domain	5-47
	InterPro	IPR018392	LysM domain	52-97
	CDD	cd00118	LysM	54-96

	PROSITE	PS51782	LysM domain profile	52-96
	SMART	SM00257	LysM_2	53-97
	Pfam/ SUPERFAMILY	PF01476/ SSF54106	LysM domain	54-97
	InterPro/ PROSITE	IPR001223/ PS51910	Glycoside hydrolase family 18, catalytic domain	104-430
	Pfam	PF00704	Glycosyl hydrolases family 18	177-410
	InterPro/ SMART	IPR011583/ SM00636	Chitinase II/ 2g34	104-412
	InterPro/CDD	IPR041704, cd02874	CFLE, GH18 catalytic domain/ GH18_CFLE_spore_hydrolase	104-420
2. Homologous superfamily	InterPro/ CATH-Gene3D	IPR029070/ G3DSA:3.10.50.10	Chitinase insertion domain superfamily	324-382
	InterPro/ CATH-Gene3D	IPR036779/ G3DSA:3.10.350.10	LysM domain superfamily	3-50
	InterPro/ CATH-Gene3D	IPR036779/ G3DSA:3.10.350.10	LysM domain superfamily	51-103
	InterPro	IPR017853	Glycoside hydrolase superfamily	103-427
	SUPERFAMILY	SSF51445	(Trans)glycosidases	130-324
	SUPERFAMILY	SSF51445	(Trans)glycosidases	384-427
3. Unintegrated	CATH-Gene3D	G3DSA:3.20.20.80	Glycosidases Model: 4s3jB02	104-321, 383-430
	PANTHER	PTHR46066	Chitinase domain-containing protein 1 family member Model: PTHR46066:SF2	64-420
4. Residues	CDD	cd02874	Active site	108 (A), 138 (F), 139 (S), 180 (F), 215 (H), 217 (D), 219 (E), 281 (M), 283 (Y), 284 (D), 324 (Y), and 407 (W)

Table S2: Functional Annotation and Domain Prediction of the Chi2 protein using InterPro

Section	Database	ID/Annotation	Domain/Region Identified	Protein Region
1. Domain	InterPro/ PROSITE profiles	IPR001919/ PS51173	Carbohydrate-binding type-2 domain/ CBM2 (Carbohydrate-binding type-2) domain profile	572-674
	SMART	SM00637	cbd_2	579-672
	InterPro/ PROSITE profiles	IPR003961/ PS50853	Fibronectin type III/ Fibronectin type-III domain profile	484-576
	Pfam	PF00041	Fibronectin type III domain	488-559
	CDD	cd00063	FN3	485-567
	SMART	SM00060	FN3_2	485-557

	InterPro/ PROSITE profiles	IPR001223/ PS51910	Glycoside hydrolase family 18, catalytic domain/ Glycosyl hydrolases family 18 (GH18) domain profile	39-476
	Pfam	PF00704	Glycosyl hydrolases family 18	40-450
	InterPro/ SMART	IPR011583/ SM00636	Chitinase II/ 2g34	39-450
2. Homologous superfamily	InterPro/ CATH-Gene3D/ FUNFAM G3DSA	IPR013783/ G3DSA:2.60.40.10/ 2.60.40.10:FF:001114	Immunoglobulin-like fold/ Immunoglobulins/ Chitinase A1 Model: 2.60.40.10-FF-001114	483-567
	InterPro/ SUPERFAMILY	IPR008965/ SSF49384	CBM2/CBM3, carbohydrate-binding domain superfamily/ Carbohydrate-binding domain	574-672
	InterPro/ CATH-Gene3D/ FUNFAM G3DSA	IPR029070/ G3DSA:3.10.50.10/ 3.10.50.10:FF:000010	Chitinase insertion domain superfamily/ Chitinase A1 Model: 3.10.50.10-FF-000010	338-421
	SUPERFAMILY	SSF54556	Chitinase insertion domain	339-421
	InterPro/ SUPERFAMILY	IPR036116/ SSF49265	Fibronectin type III superfamily/ Fibronectin type III	481-561
	InterPro/ CATH-Gene3D	IPR012291/ G3DSA:2.60.40.290	CBM2, carbohydrate-binding domain superfamily	578-672
	InterPro	IPR017853	Glycoside hydrolase superfamily	34-451
	SUPERFAMILY	SSF51445	(Trans)glycosidases	34-339
	SUPERFAMILY	SSF51445	(Trans)glycosidases	422-451
3. Active site	InterPro/ PROSITE patterns	IPR001579/ PS01095	Glycosyl hydrolases family 18 (GH18) active site/ Glycosyl hydrolases family 18 (GH18) active site signature	201-209
4. Unintegrated	CATH-Gene3D	G3DSA:3.20.20.80	Glycosidases Model: 1itxA01	41-337, 422-450
	PANTHER	PTHR11177	Chitinase Model: PTHR11177:SF317	31-505
	CDD	cd06548 GH18_chitinase Model: cd06548	GH18_chitinase	41-450
5. Other Features	PHOBIUS	SIGNAL_PEPTIDE_H_REGION	Hydrophobic region of a signal peptide	6-21
	PHOBIUS/ SIGNALP_GRAM _POSITIVE/ SIGNALP_EUK	SIGNAL_PEPTIDE/ SignalP-TM/ SignalP-noTM	Signal peptide region/ SignalP-TM/ SignalP-noTM	1-32

	FUNFAM G3DSA	3.20.20.80:FF:000153	Chitinase A1 Model: 3.20.20.80-FF-000153	41-348
	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane	7-26
	PHOBIUS	SIGNAL_PEPTIDE_N_REGION	N-terminal region of a signal peptide	1-5
	PHOBIUS	SIGNAL_PEPTIDE_C_REGION	C-terminal region of a signal peptide	22-32
	PHOBIUS	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region	33-674
6. Residues	CDD	cd00063	Cytokine receptor motif	Residue: 556 (N), 557 (K), 559 (Q), 560 (P)
	CDD	cd00063	Interdomain contacts	Residue: 485 (P), 540 (P), 555 (G)
	CDD	cd06548	Active site	Residue: 44 (Y), 72 (F), 205 (D), 207 (D), 209 (E), 282 (M), 284 (Y), 285 (D), 339 (Y), 445 (W)

Table S3: Functional Annotation and Domain Prediction of the BcLPMO10B protein using InterPro

Section	Database	ID/Annotation	Domain/Region Identified	Protein Region
1. Domain	InterPro/ Pfam	IPR004302/ PF03067	Cellulose/chitin-binding protein, N-terminal/ Lytic polysaccharide mono-oxygenase, cellulose-degrading	35-198
2. Homologous Superfamily	InterPro/ SUPERFAMILY	IPR014756/ SSF81296	Immunoglobulin E-set/ E set domains	35-200
3. Unintegrated	PANTHER	PTHR34823	GLcNAc -binding protein A, Model: PTHR34823:SF1	1-215
	CDD	cd21177	LPMO_AA10	35-199
	CATH-Gene3D	G3DSA:2.70.50.50	Model: 5wszA00	35-203
4. Other Features	PHOBIUS	SIGNAL_PEPTIDE_C_REGION	C-terminal region of a signal peptide	27-34
	PHOBIUS/ SIGNALP_GRAM_POSITIVE	SIGNAL_PEPTIDE/ SignalP-TM	Signal peptide region/ SignalP-TM	1-34

	PHOBIUS	SIGNAL_PEPT IDE_ N_REGION	N-terminal region of a signal peptide	1-11
	PHOBIUS	NON_CYTOPL ASMIC_ DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region	35-221
	FUNFAM G3DSA	2.70.50.50:FF: 000001	Chitin-binding protein, Model: 2.70.50.50-FF-000001	35-201
	PHOBIUS	SIGNAL_PEPT IDE_H_ REGION	Hydrophobic region of a signal peptide	12-26
	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane	12-29

Table S4: Gene Ontology (GO) annotations for molecular function, biological process, and cellular components of Chi1, Chi2, and BcLPMO10B proteins

Proteins	GO ID	Molecular Function	Biological Process	Cellular Component	Total Score	Internal Confidence
Chi1	GO:0004568	Chitinase Activity	-	-	4858.34	0.362861
	GO:0004553	Hydrolase Activity (O-glycosyl)	-	-	3001.14	0.729386
	GO:0016787	Hydrolase Activity	-	-	1428.11	0.948207
	GO:0016798	Hydrolase Activity (Glycosyl)	-	-	8.06718	0.749436
	GO:0006032	-	Chitin Catabolic Process	-	15738.2	0.440835
	GO:0005975	-	Carbohydrate Metabolic Process	-	8601.19	0.444233
	GO:0008152	-	Metabolic Process	-	46.8849	0.939180
	GO:0016020	-	-	Membrane	65.3916	0.275806
	GO:0016021	-	-	Integral Component of Membrane	45.3957	0.082190
Chi2	GO:0030247	Polysaccharide Binding	-	-	2152.86	0.157708
	GO:0030246	Carbohydrate-Binding	-	-	1617.13	0.315633
	GO:0004568	Chitinase Activity	-	-	1444.75	0.158592

	GO:0016787	Hydrolase Activity	-	-	1247.83	0.651293
	GO:0004553	Hydrolase Activity (O-glycosyl)	-	-	916.097	0.324551
	GO:0016798	Hydrolase Activity (Glycosyl)	-	-	824.976	0.484358
	GO:0006032	-	Chitin Catabolic Process	-	9766.03	0.272583
	GO:0005975	-	Carbohydrate Metabolic Process	-	5256.27	0.285015
	GO:0008152	-	Metabolic Process	-	537.789	0.862511
	GO:0016021	-	-	Integral Component of Membrane	82.9439	0.098301
	GO:0016020	-	-	Membrane	61.5155	0.311784
	GO:0005618	-	-	Cell Wall	5.48319	0.038457
BcLPMO10B	GO:0030246	Carbohydrate-Binding	-	-	2795.65	0.4431
	GO:0004553	Hydrolase Activity	-	-	2059.3	0.4553
	GO:0016787	Hydrolase Activity	-	-	36.3418	0.4968
	GO:0016740	Transferase Activity	-	-	33.5986	0.0208
	GO:0004568	Chitinase Activity	-	-	11.5253	0.0197
	GO:0016798	Hydrolase Activity	-	-	6.4162	0.4750
	GO:0030246	-	Carbohydrate Metabolic Process	-	5223.38	0.8652
	GO:0004553	-	Metabolic Process	-	39.107	0.9829
	GO:0030246	-	-	Extracellular Region	4313.47	0.7265
	GO:0016021	-	-	Integral Component of Membrane	208.224	0.1296
	GO:0016020	-	-	Membrane	144.748	0.2591
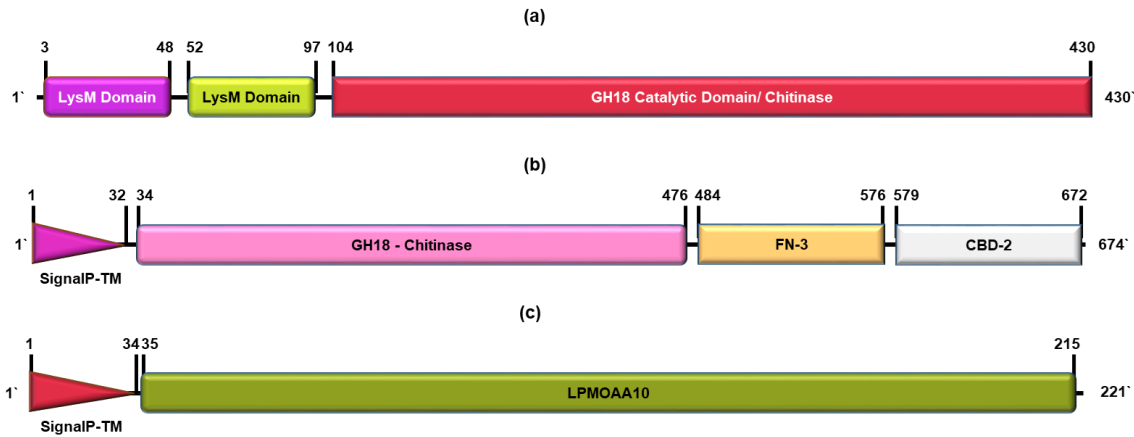
Figure 1: The structural characterization of Chi1 (a), Chi2 (b), and BcLPMO10B (c) proteins in *B. cereus*, highlighting domains with their respective locations and sequences

Table 2: Utilizing Argot2 for the Functional Annotation of Chi1, Chi2, and BcLPMO10B Proteins

GO	Chi1	Chi2	BcLPMO10B
Molecular Function	Chitinase, and hydrolase activity (O-glycosyl)	Polysaccharide binding, chitinase, and hydrolase activity (O-glycosyl)	Carbohydrate-binding, chitinase, hydrolase, and Ttransferase activity
Biological Process	Chitin catabolic, and carbohydrate metabolic Process	Chitin catabolic, and carbohydrate metabolic Process	Carbohydrate Metabolic Process
Cellular Component	Integral component of membrane	Integral component of membrane and cell wall	Extracellular region and integral component of membrane

Phylogenetic Analysis

A phylogenetic tree helps characterize bacterial proteins by revealing their evolutionary origin, function, and relationships among different species. It shows how a protein has evolved, its connections to other proteins, and its distribution across bacterial groups, aiding in inferring its function based on similarities or dissimilarities with known proteins, such as involvement in metabolic pathways, cellular processes, or virulence factors (Gabaldón 2005; Zardoya 2005).

The ancestral lineage of the Chi1 protein was elucidated through the application of the MLM, employing the Le Gascuel 2008 model (Le and Gascuel 2008), which was chosen based on its minimal BIC value of 57712.8. The construction of the phylogenetic tree for the Chi1 protein involved the utilization of respective protein sequences, yielding the highest log likelihood of -30537.63 (Figure 2). The resultant phylogenetic tree of the Chi1 protein exhibited a tripartite division into distinct clusters designated as cluster I, cluster II, and cluster III. The branches of the phylogenetic tree were supported by bootstrap

values, indicating the robustness of the clustering of associated taxa. The tree was depicted to scale, with branch lengths representing the number of substitutions per site. The analysis incorporated a total of 102 amino acid sequences, spanning 423 positions in the final dataset. Notably, the Chi1 protein demonstrated a close phylogenetic association with the Bacillus genus, particularly highlighting a distinct relationship with EAO55006 and CJD61848. This close clustering with specific members within the Bacillus lineage, as highlighted in yellow in Figure 2, underscores the unique evolutionary trajectory of Chi1, setting it apart from other members within the same lineage.
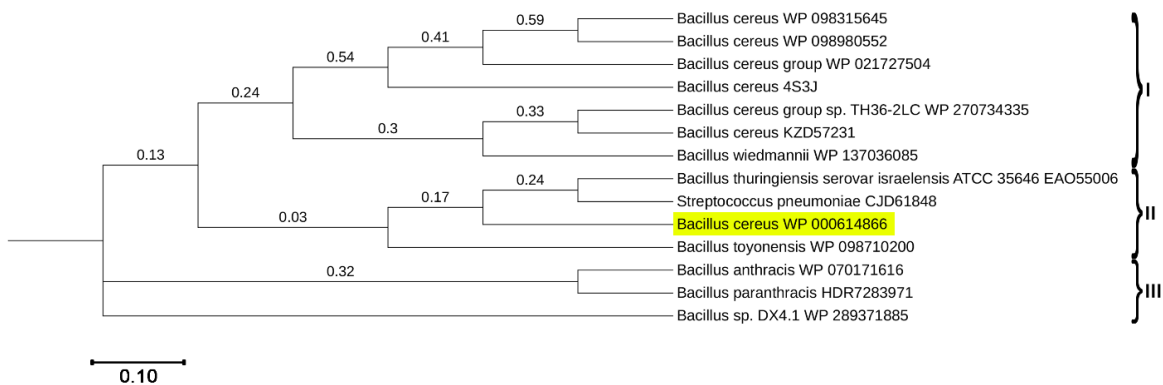
Figure 2: The construction of a phylogenetic tree elucidating the evolutionary relationships among Chi1 protein and its homologs employed rigorous methodology. Utilizing the bootstrap method with 1000 replications, we applied the p-distance as the substitutional model, specifically designating amino acids as the substitution type, while ensuring uniform rates and patterns. Retrieval of homologous Chi1 protein sequences from related bacteria for phylogenetic tree reconstruction was conducted via BLASTp from the experimental clustered nr database. Notably, our protein sequence exhibited a close clustering with Bacillus genus, indicating a distinct relationship with EAO55006 and CJD61848, distinguishing it from other members within the same lineage

The evolutionary lineage of Chi2 protein was elucidated using the MLM with the Whelan and Goldman model (Whelan and Goldman 2001), chosen for its minimal BIC value of 7156.36. The Chi2 protein's family tree was built using its protein sequences, resulting in the highest log likelihood of -3529.95 (Figure 3). The resulting phylogenetic tree, based on 28 amino acid sequences, revealed four distinct clusters (I-IV), supported by bootstrap values. Chi2 exhibited a close phylogenetic association with *B. thuringiensis*, highlighting its unique evolutionary path within the Bacillus lineage.
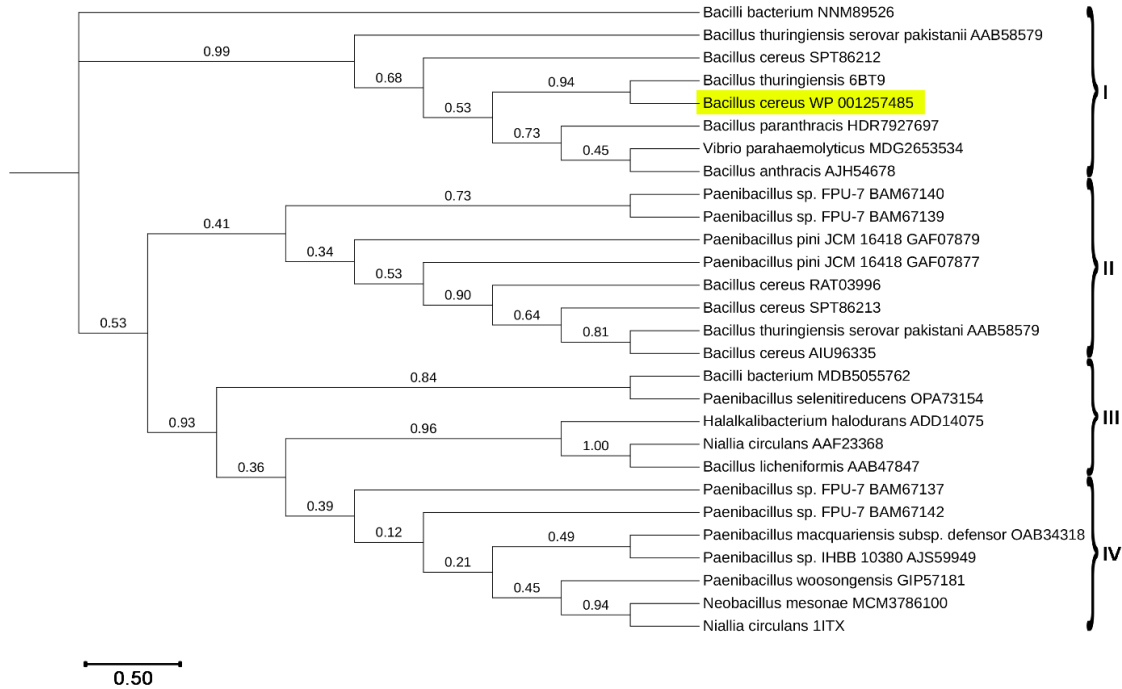
Figure 3: Phylogenetic analysis depicting the evolutionary relationships of Chi2 protein and its homologs. The tree was constructed using the bootstrap method with 1000 replications, employing the p-distance as the substitutional model with amino acids as the substitution type. Homologous Chi2 protein sequences from related bacteria were retrieved through BLASTp from the experimental clustered nr database. Notably, our protein sequence clustered closely with *B. thuringiensis*, indicating a distinctive relationship within the same lineage

The evolutionary history of the BcLPMO10B protein was discerned employing the MLM with the Whelan and Goldman model (Whelan and Goldman 2001), selected for its minimized BIC value of 3402.7. Phylogenetic tree construction involved the use of corresponding protein sequences, yielding the highest log likelihood of -1710.54 (Figure 4). The resulting phylogenetic tree displayed a trisection into clusters—cluster I and cluster II—supported by robust bootstrap values, depicting the stability of taxonomic clustering. The tree, drawn to scale, represented substitutions per site along branch lengths. The analysis encompassed 10 amino acid sequences, spanning 134 positions. Notably, the BcLPMO10B protein exhibited a proximate clustering with *B. cereus*, specifically with TK187278 and PFJ17621, followed by *B. gaemokensis*. This distinct relationship within the Bacillus lineage, highlighted in yellow in Figure 4, underscores the unique evolutionary trajectory of BcLPMO10B compared to other members within the same lineage.
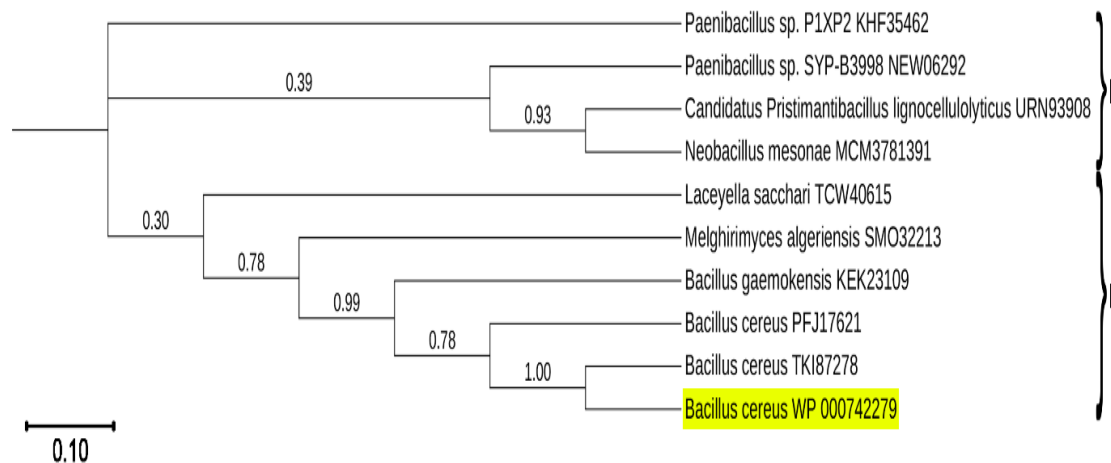
Figure 4: Phylogenetic analysis illustrating the evolutionary relationships of BcLPMO10B protein and its homologs. The tree was constructed using the bootstrap method with 1000 replications, employing the p-distance as the substitutional model with amino acids as the substitution type. Homologous BcLPMO10B protein sequences from related bacteria were obtained via BLASTp from the experimental clustered nr database. Notably, our protein sequence exhibited a close clustering with *B. cereus*, specifically TK187278 and PFJ17621, followed by *B. gaemokensis*, indicating a distinctive relationship within the same lineage

Table S5: Secondary structure parameters prediction of Chi1, Chi2, and BcLPMO10B proteins

Secondary structure parameters	Chi1	Chi2	BcLPMO10B
Alpha helix	33.95%	20.62%	19.00%
3_{10} helix	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%
Extended strand	23.26%	20.03%	19.91%
Beta turn	9.07%	6.68%	5.88%
Bend region	0.00%	0.00%	0.00%
Random coil	33.72%	52.67%	55.20%
Ambiguous states	0.00%	0.00%	0.00%
Other states	0.00%	0.00%	0.00%
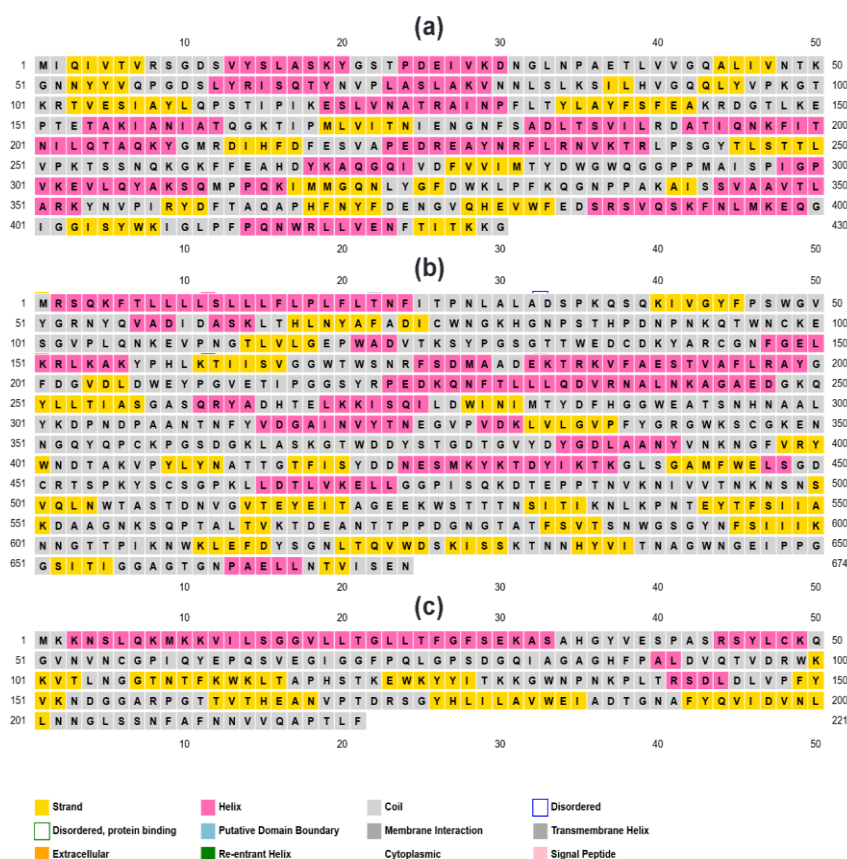
Figure S1: Sequence plot of secondary structure prediction (a) Chi1, (b) Chi2 and (c) BcLPMO10B proteins by PSIPRED

The SOPMA server was employed to predict the secondary structures of the Chi1, Chi2, and BcLPMO10B proteins. In Supplementary Material Table S5, you can find the predicted values for various parameters essential in determining the secondary structures of these proteins. To provide a visual representation of these secondary structures, Supplementary Material Figure S1 was created using PSIPRED.

Homology Modeling

In the comparative analysis of protein structures, our investigation focused on three distinct proteins: Chi1, Chi2, and BcLPMO10B. The Swiss models of these proteins were aligned with their respective templates, providing valuable insights into sequence identity, similarity, and structural features. For the Chi1 protein, the pairwise sequence alignment with the sequence of 4S3J template unveiled a remarkable 97.2% sequence identity, 98.6% sequence similarity, and minimal gaps (0.7%), yielding a substantial alignment score of 2202 (Figure 5a). Further examination of the 3D structure revealed a high degree of conservation, with a sequence identity of 98.14% and a sequence similarity of 61% in comparison to the 3D structure of 4S3J template.

The superimposed 3D structures of Chi1 protein and its template (TM score = 0.99), as illustrated in Figures 5b-c, highlight the structural congruence.Moving to the Chi2 protein, the pairwise sequence alignment with the sequence of 6BT9 template displayed a sequence identity of 92.1%, a sequence similarity of 92.7%, and a modest gap percentage of 6.3%, resulting in an alignment score of 3475 (Figure 6a). Concurrently, the 3D structure comparison with the 6BT9 template demonstrated a substantial sequence identity of 99.38% and a sequence similarity of 63%. Figures 6b-c visually depict the superimposed 3D structures of Chi2 protein and its template (TM score = 0.65). In the case of BcLPMO10B protein, the pairwise sequence alignment with the 5WSZ template indicated a sequence identity of 76.5%, a sequence similarity of 76.5%, and a notable gap content of 23.5%, with an alignment score of 931 (Figure 7a). The 3D structural comparison with the 5WSZ template showcased a striking sequence identity of 100% and a sequence similarity of 63%. Figures 7b-c provide a visual representation of the superimposed 3D structures of BcLPMO10B protein and its template (TM score = 0.15).
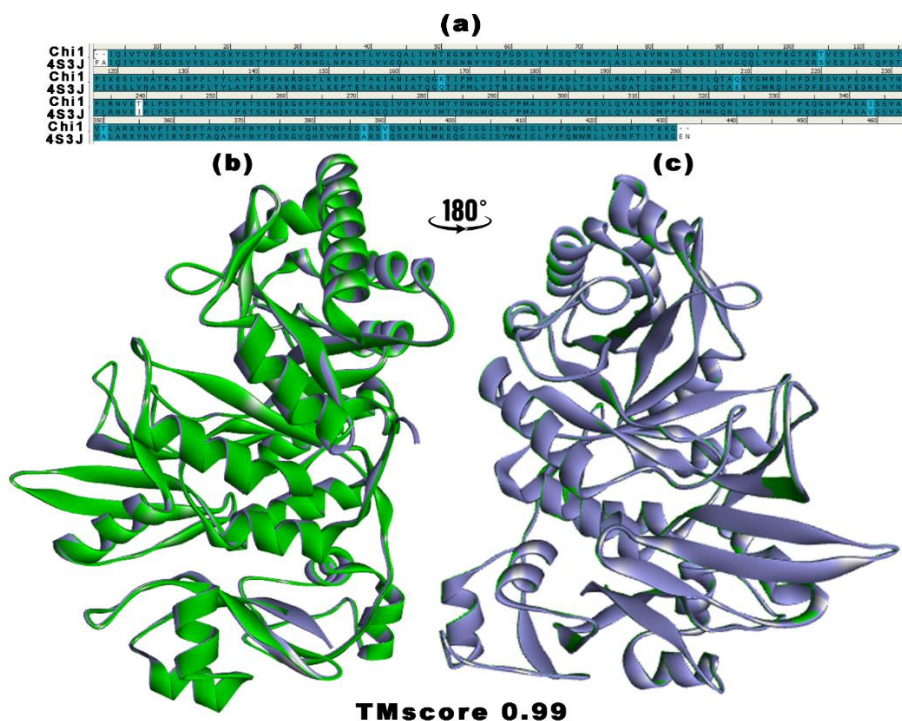

Figure 5: Representation of the Swiss model for the Chi1 protein from *B. cereus*: (a) Pairwise sequence alignment illustrating the comparison of Chi1 sequence with the template (PDB: 4S3J); (b, c) Superimposed 3D structures depicting Chi1 (in green) aligned with its template (in grey)
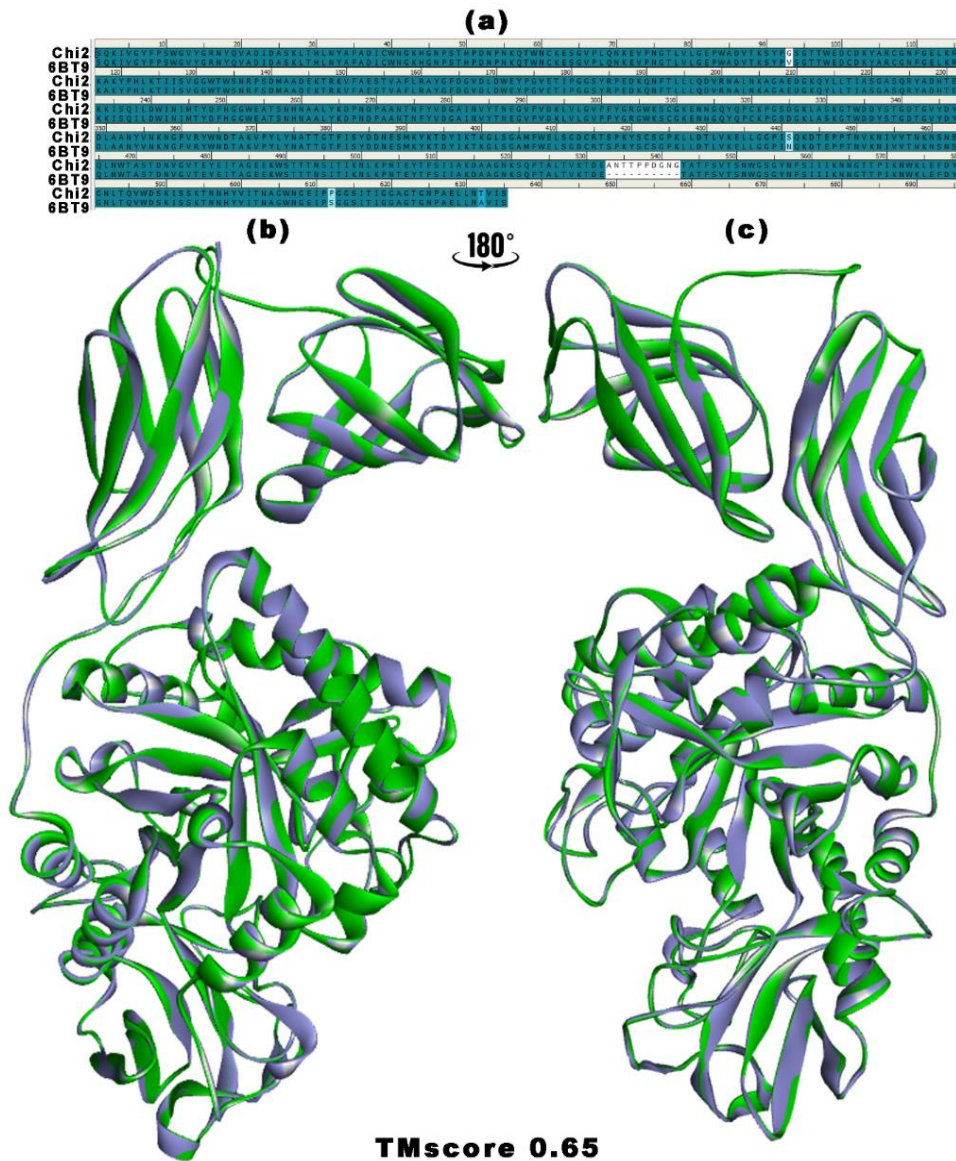
Figure 6: Depiction of the Swiss model for the Chi2 protein from *B. cereus*: (a) Pairwise sequence alignment showcasing the comparison between the Chi2 sequence and the template (PDB: 6BT9); (b, c) Superimposed 3D structures illustrating the alignment of Chi2 (in green) with its template (in grey)
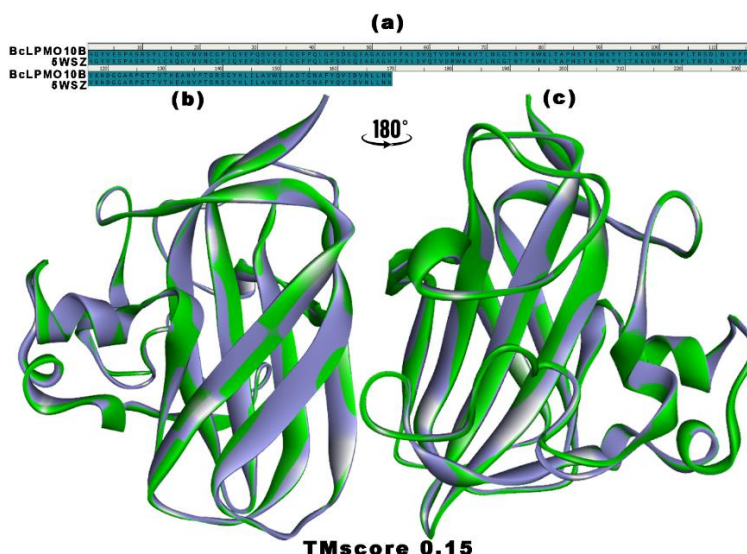
Figure 7: Illustration of the Swiss model for the BcLPMO10B protein from *B. cereus*: (a) Pairwise sequence alignment presenting the comparison of the BcLPMO10B sequence with the template (PDB: 5WSZ); (b, c) Superimposed 3D structures providing a visual representation of the alignment of BcLPMO10B (in green) with its template (in grey)

The 3D structural predictions for Chi1, Chi2, and BcLPMO10B proteins were conducted using the I-TASSER online server and the ROBETTA Baker server, as depicted in Figure 8. I-TASSER models were generated based on C-score, estimated TM-Score, and estimated RMSD, with the outcomes for each model presented in Figure 8. Concurrently, the ROBETTA Baker server employed RoseTTAFold for predicting the 3D structures of Chi1, Chi2, and BcLPMO10B proteins, with confidence values displayed in Figure 8. Additionally, the AlphaFold-predicted structures for Chi1, Chi2, and BcLPMO10B were obtained from UniProt and included in Figure 8.

To assess the reliability of the predicted structures, comprehensive structure validation was performed on all models generated by the Swiss-Model, I-TASSER server, ROBETTA Baker server, and AlphaFold. This rigorous validation process aimed to identify the most accurate and well-supported 3D structures among the predictions for Chi1, Chi2, and BcLPMO10B proteins. The results of this structure validation process contribute valuable insights into the quality and reliability of the predicted protein structures.
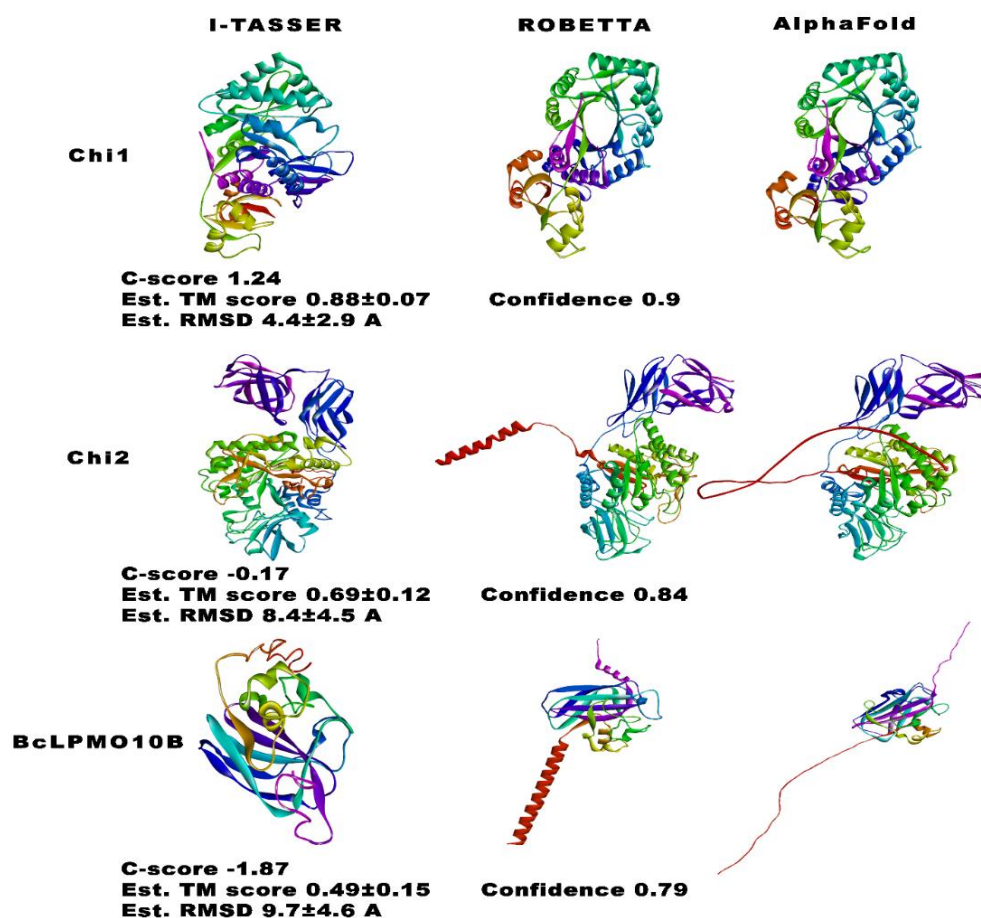
Figure 8: The 3D structure predictions for Chi1, Chi2, and BcLPMO10B proteins from I-TASSER and ROBETTA Baker servers, presenting I-TASSER models' C-score, estimated TM-Score, and RMSD, ROBETTA Baker's RoseTTAFold predictions with confidence values, and AlphaFold-predicted structures obtained from UniProt

3D Structure Validation of Modeled Proteins

The structural integrity of the 3D models of Chi1, Chi2, and BcLPMO10B proteins generated by SwissModel, I-TASSER, and ROBETTA Baker Laboratory, and AlphaFold from Uniprot underwent thorough validation using multiple criteria, including the Ramachandran plot, MolProbity, ERRAT, Verify3D, QMEAN score, and Z-score. The Ramachandran plots (Figure 9 and Table 3) analysis reveal distinctive patterns in the stereochemical quality of protein structures predicted for Chi1, Chi2, and BcLPMO10B across diverse modeling servers—Swiss Model, I-TASSER, RobettaFold, and AlphaFold. For Chi1, Chi2, and BcLPMO10B, both Swiss Model and AlphaFold consistently yield models with significantly high percentages of residues in the most favored region, demonstrating strong agreement with favorable torsional angles.

While I-TASSER and RobettaFold maintain acceptable predictions, they exhibit slightly lower accuracy in capturing preferred torsional angles across all three proteins. The distribution of residues in additionally allowed, generously allowed, and disallowed regions further refines the assessment, emphasizing the nuanced variations in prediction quality among the modeling servers. The consistent trend of higher accuracy in Swiss Model and AlphaFold models suggests their reliability in generating protein structures with favorable stereochemical characteristics.

Table 3: Ramachandran plot statistics of Chi1, Chi2, and BcLPMO10B proteins predicted by different servers

| Ramachandran plot parameters	Chi1				Chi2				BcLPMO10B			
	Swiss Model	I-TASSER	Robetta Fold	AlphaFold	Swiss Model	I-TASSER	Robetta Fold	Alpha Fold	Swiss Model	I-TASSER	Robetta Fold	Alpha Fold
Residues in the most favored region (%)	94.1	86.90	93.1	93.1	91.8	79.5	87.3	87.5	92.7	63.9	90.7	75.4
Residues additionally allowed region (%)	5.9	10.9	6.1	6.9	8	17.4	11.3	11	7.3	28.4	8.7	20.2
Residues generously allowed region (%)	0	1.6	0.5	0	0.2	1.7	0.9	0.5	0	4.9	0	3.8
Residues in the disallowed region (%)	0	0.5	0.3	0	0	1.4	0.5	1	0	2.7	0.5	0.5
Total number of non-glycine and non-proline residues (%)	100	100	100	100	100	100	100	100	100	100	100	100
Numbers of end residues (Excluding Glycine and Proline)	1	1	1	1	3	2	2	2	3	2	2	2
Numbers of Glycine residues	28	28	28	28	62	62	62	62	18	23	23	23
Numbers of Proline residues	26	26	26	26	32	35	35	35	12	13	13	13
Total number of residues	429	430	430	430	636	674	674	674	170	221	221	221
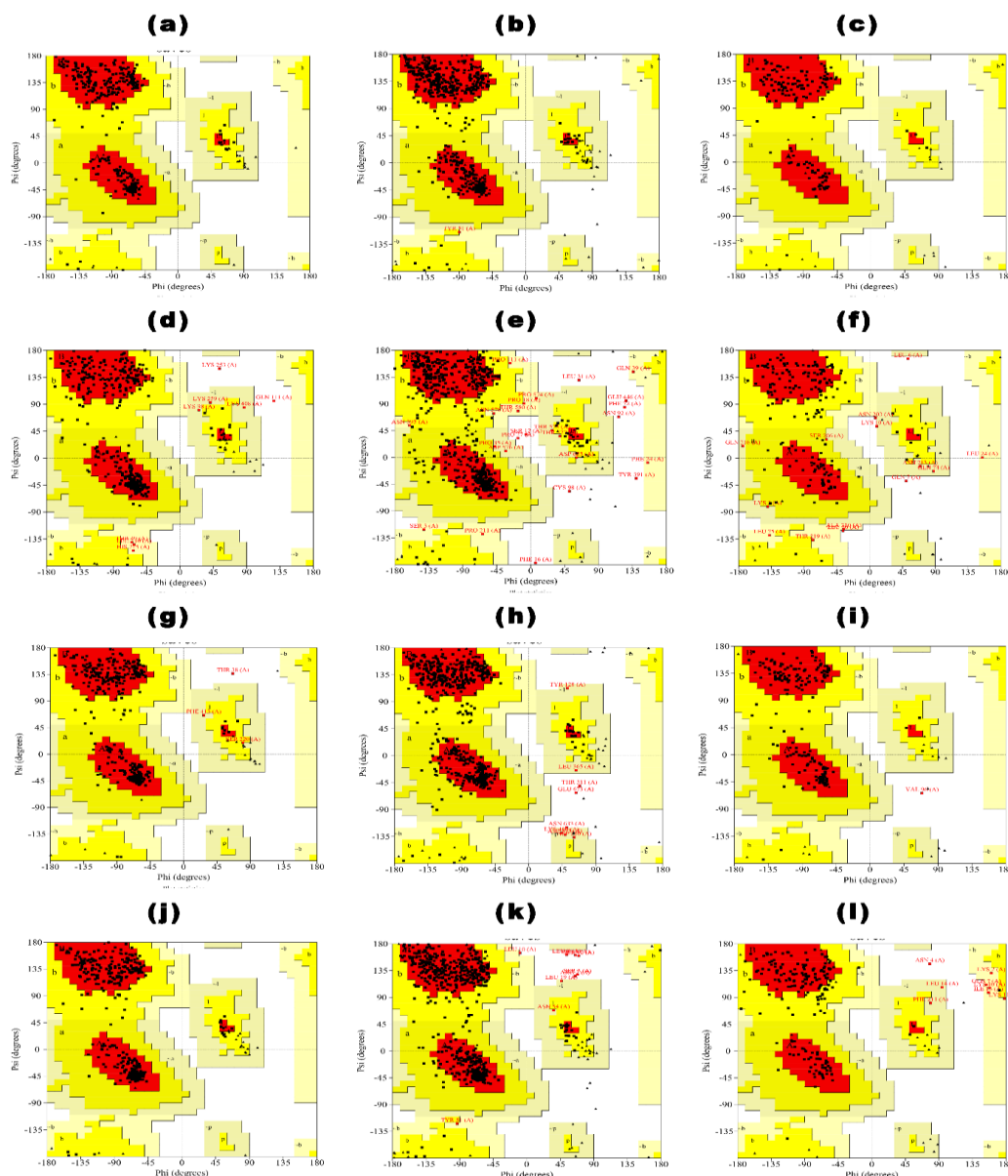
Figure 9: Ramachandran plots serve as a validation tool for the 3D modeled structures of Chi1, Chi2, and BcLPMO10B proteins generated by diverse computational servers. The plots visually represent the 3D structures of each protein as predicted by four different servers: (a, b, c) Swiss Model, (d, e, f) I-TASSER, (g, h, i) RobettaFold, and (j, k, l) AlphaFold. The X-axis (Phi) and Y-axis (Psi) depict the backbone conformation angles of amino acid residues, providing insights into the stereochemical quality and torsional angle distributions within the predicted protein structures

The comparative validation of 3D structures for Chi1, Chi2, and BcLPMO10B proteins, generated by Swiss Model, I-TASSER, RobettaFold, and AlphaFold from Uniprot, reveals detailed insights into the accuracy and reliability of structural prediction methods (Table 4). Swiss Model consistently delivers high ERRAT scores, indicative of favorable model geometry, across all proteins. AlphaFold exhibits reliable structural predictions but demonstrates a lower Z-Score for Chi1. I-TASSER presents acceptable ERRAT scores but shows a slightly lower Z-Score for both Chi2 and BcLPMO10B. RobettaFold, while promising, lacks Z-Score calculations, limiting a comprehensive evaluation. Verify3D scores consistently indicate favorable 3D-1D profile compatibility, and QMEANDisCo Global scores provide insights into the models' reliability. This comparative analysis underscores the necessity of considering multiple validation metrics for a thorough assessment of structural prediction methods, highlighting Swiss Model's consistent performance and the varying degrees of success observed with AlphaFold and I-TASSER. Further evaluation of RobettaFold is warranted to ascertain its reliability in the absence of Z-Score calculations.

Table 4: Comparative validation of 3D structures of Chi1, Chi2, and BcLPMO10B proteins generated by Swiss Model, I-TASSER, and ROBETTA Baker Laboratory, and AlphaFold from Uniprot

Protein	Servers	ERRAT	Verify3D	QMEANDisCo Global	Z-Score
Chi1	Swiss Model	94.5238	80.65%	0.94 ± 0.05	-10.61
	I-TASSER	91.1905	84.65%	0.92 ± 0.05	-10.94
	RobettaFold	95.4976	85.12%	0.87 ± 0.05	Not Calculated
	AlphaFold	93.5867	82.09%	0.93 ± 0.05	-11.18
Chi2	Swiss Model	94.7195	97.95%	0.95 ± 0.05	-10.51
	I-TASSER	87.8378	91.54%	0.87 ± 0.05	-9.23
	RobettaFold	92.1922	91.54%	0.84 ± 0.05	Not Calculated
	AlphaFold	94.4	92.58%	0.90 ± 0.05	-9.6
BcLPMO10B	Swiss Model	95	95.27%	0.96 ± 0.07	-6.23
	I-TASSER	81.2207	77.83%	0.65 ± 0.06	-4.54
	RobettaFold	96.2264	66.97%	0.79 ± 0.06	Not Calculated
	AlphaFold	88.1356	69.68%	0.74 ± 0.06	-4.87

Based on the validation by Ramachandran plot, ERRAT, Verify3D, QMEAN score, and Z-score, Swiss-Models of 3D structures of Chi1, Chi2, and BcLPMO10B proteins were selected for further validation with MolProbity and computational analysis. With the MolProbity, the structural assessment of Swiss Models of Chi1, Chi2, and BcLPMO10B proteins, as represented in the table 5, provides a comprehensive overview of various quality metrics, including clash scores, protein geometry, Ramachandran outliers, MolProbity scores, Cβ deviations, peptide omegas, and low-resolution criteria. Notably, all three proteins exhibit low all-atom clash scores, suggesting minimal steric hindrance. In terms of protein geometry, Chi1 and Chi2 meet the goals for favored rotamers and Ramachandran favored regions, but BcLPMO10B falls slightly below the target for Ramachandran favored. The Rama distribution Z-score, indicating the distribution of

Ramachandran outliers, meets the set goal for all three proteins. MolProbity scores, which evaluate overall structure quality, are within acceptable percentiles. The analysis of Cβ deviations, bad bonds, and bad angles further underscores the high-quality structural predictions for all three proteins. Examination of peptide omegas reveals relatively low percentages of cis Prolines, with BcLPMO10B showing a higher percentage. Low-resolution criteria, including CaBLAM outliers and CA Geometry outliers, are generally within acceptable limits, though Chi2 slightly exceeds the goal for CA Geometry outliers. Overall, the structural evaluation indicates robust predictions for Chi1 and Chi2, with BcLPMO10B exhibiting slightly more deviations in certain criteria.

Table 5: Structural Evaluation Metrics using MolProbity for Chi1, Chi2, and BcLPMO10B Proteins Modeled by Swiss Model
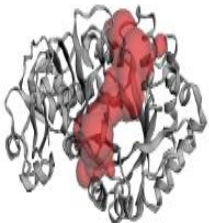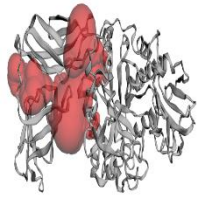
		Chi1	Chi2	BcLPMO10B	Expected Range
All-Atom	**Clashscore, all atoms:**	0.29	0.31	0	99th percentile* (N=1784, all resolutions)
Protein Geometry	**Poor rotamers**	1.34%	0.56%	0.00%	Goal: <0.3%
	Favored rotamers	97.31%	96.24%	96.45%	Goal: >98%
	Ramachandran outliers	0.23%	0.00%	0.00%	Goal: <0.05%
	Ramachandran favored	97.66%	96.84%	95.81%	Goal: >98%
-	**Rama distribution Z-score**	0.58 ± 0.40	-0.16 ± 0.30	-0.23 ± 0.64	Goal: abs(Z score) < 2
	MolProbity score	0.78	0.81	0.79	100th percentile (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	1.25%	0.70%	0.66%	Goal: 0
	Bad bonds:	0.00%	0.02%	0.00%	Goal: 0%
	Bad angles:	0.42%	0.32%	0.59%	Goal: <0.1%
Peptide Omegas	**Cis Prolines:**	7.69%	0.00%	16.67%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	0.50%	0.50%		Goal: <0.05%
Low-resolution Criteria	**CaBLAM outliers**	1.90%	2.40%	1.20%	Goal: <1.0%
	CA Geometry outliers	0.71%	0.79%	1.21%	Goal: <0.5%

Where green color represents good results, yellow represents caution, and red shows warning

Active Site Prediction

The identification of binding pockets within the Swiss-modeled Chi1, Chi2, and BcLPMO10B proteins was conducted using the CASTp 3.0 online server. Table 6 offers a concise overview of the predicted binding pockets for these proteins. Notably, Chi1 displayed a promising binding pocket characterized by a surface area of 845.103 $Å^2$ and a surface volume of 1076.248 $Å^3$. Likewise, Chi2 exhibited an active site pocket with a surface area of 2014.126 $Å^2$ and a surface volume of 2591.475 $Å^3$. Similarly, the BcLPMO10B proteins featured an active site with a surface area of 20.129 $Å^2$ and a surface volume of 4.496 $Å^3$.

Table 6: Active site prediction of the modeled 3D structure of Chi1, Chi2, and BcLPMO10B proteins using the CastP server

Sr. No.	Protein	Structure	Surface Area ($Å^2$)	Surface Volume ($Å^3$)	Active Site Amino Acids
1	Chi1		845.1	1076.2	TYR109, LEU110, GLN111, PRO112, SER113, THR114, LYS118, SER120, LEU121, ALA124, PHE138, ILE175, GLY178, ASN179, PHE180, ALA182, ASP217, GLU219, SER220, VAL221, ALA222, PRO223, THR249, VAL251, PRO252, LYS253, THR254, GLN258, GLY260, LYS261, PHE262, PHE263, GLU264, ALA265, HIS266, VAL279, MET281, TYR283, ASP284, TRP285, TRP287, GLN288, PRO300, VAL304, GLY320, TYR324, PHE326, LYS340, ALA341, ILE342, SER343, ALA346, SER405, TYR406, TRP407, LYS408, ILE409, GLY410, LEU411, PRO412
2	Chi2		2014.1	2591.5	PRO104, LEU105, GLN106, LYS108, VAL110, THR114, LEU115, VAL116, LEU117, TYR128, SER131, GLY132, THR133, ASP137, TYR141, ALA142, ARG143, CYS144, GLY145, PHE147, GLY148, GLU149, LEU150, LYS151, ARG152, LYS154, ALA155, THR162, SER191, ALA194, PHE195, LEU196, ARG197, ARG197, ALA198, TYR199, GLY200, PHE201, ASP202, ALA245, GLU246, ASP247, LYS249, TYR251, GLU483, PRO484, THR486, ASN487, VAL488, LYS489, ASN490, ILE491, VAL492, VAL493, THR494, ASN495, LYS496, THR506, ALA507, SER508, THR509, ASP510, ASN511, VAL512, GLY513, VAL514, THR515, GLU516, TYR517, THR529, THR530, ASN531, ASN541, GLU543, GLN559, PRO560, THR561, ALA562, LEU563, THR564, VAL565, LYS566, THR567, ASP568,

					GLU569, ALA570, THR572, THR573, PRO574, PRO575, ASP576, GLY 577, ASN578, GLY579, THR580, ALA581, THR582, PHE583, SER584, VAL585, THR586, SER587, ASN588, TRP589, GLY590, GLY592, SER596, ILE597, ILE598, ASP615, TYR616, SER617, GLY661, ASN662, PRO663, ALA664, GLU665, LEU666, LEU667, ASN668, THR669
3	BcLPMO10B		20.1	4.5	SER43, ARG44, SER45, VAL66, ALA83, LEU91, TYR193, GLN194, VAL195

Molecular Docking

Figure 10 illustrates the 3D and 2D structures resulting from the AutoDock Vina docking of Chi1/Chi2 and BcLPMO10B proteins with GLcNAc and 2, 6-dimethoxyphenol, respectively. These visual representations provide a comprehensive view of the interactions between the proteins and their respective ligands.

For the Chi1 protein and GLcNAc, the docking analysis revealed a binding affinity of -5.2 kcal/mol and a drug score of 0.52. Key amino acids involved in this interaction include Phe138, Glu219, Thr249, Tyr283, and Trp407 (refer to Table 7 and Figures 10c and d).

Similarly, the docking of Chi2 protein with GLcNAc exhibited a binding affinity of -5.2 and a drug score of 0.83, involving critical amino acids such as Ile491, Val493, Phe583, Ala664, Glu665, and Leu666 (see Table 7 and Figures 10g and h). Finally, the docking of BcLPMO10B protein with 2,6-dimethoxyphenol demonstrated a binding affinity of -3.9 and a drug score of 0.57, with key interacting amino acids including His35, Glu62, Glu67, His87, and Phe88 (refer to Table 7 and Figures 10k and l).
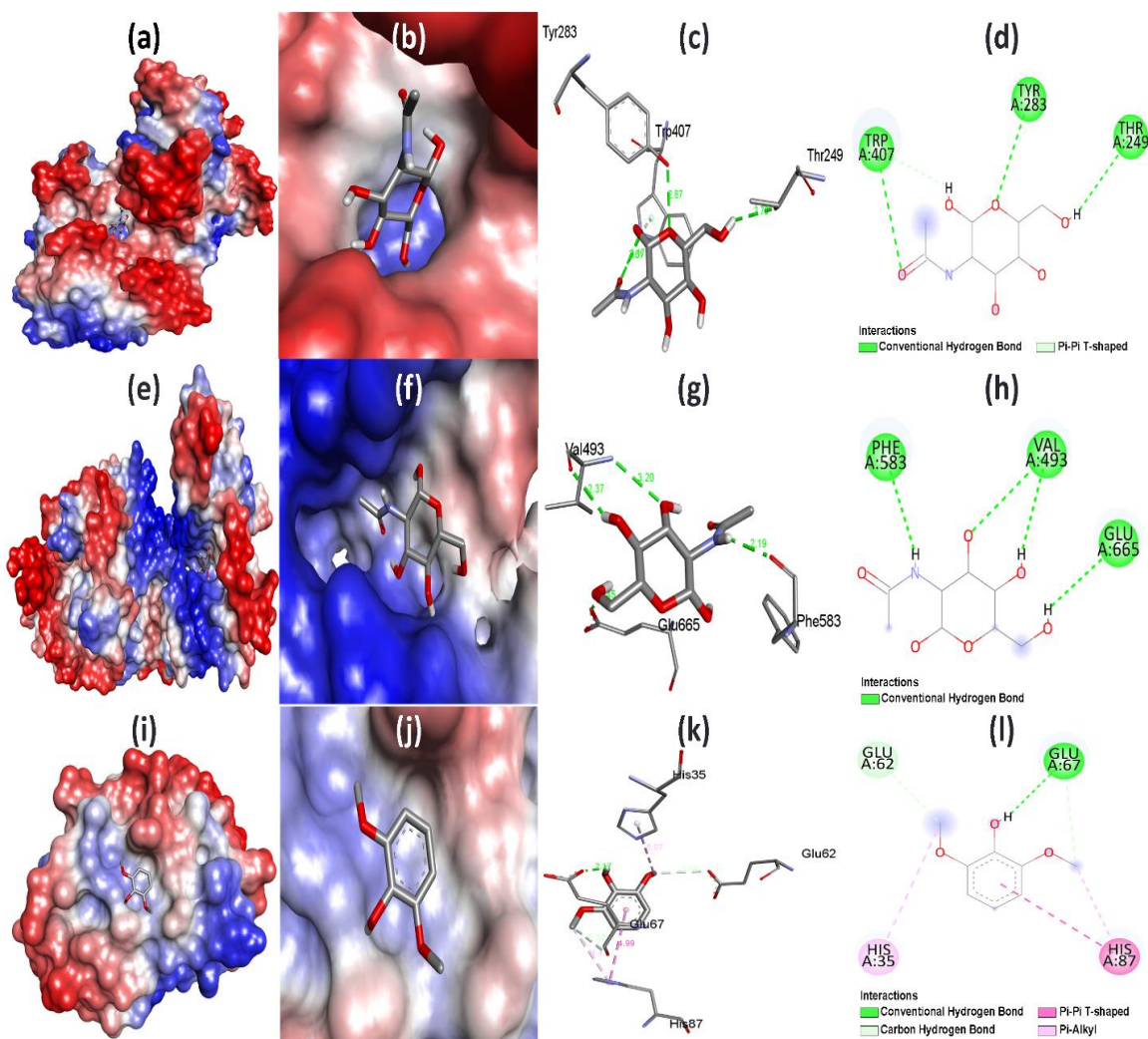
Figure 10: **Molecular docking outcomes for the interactions of Chi1/Chi2 and BcLPMO10B proteins with N-acetyl-D(+)-glucosamine and 2,6-dimethoxyphenol, respectively, are presented as Chi1 Protein Docking Results (a) Solvent ribbon size surface, (b) Active site, (c) 3D docking, and (d) 2D docking with ligand N-acetyl-D(+)-glucosamine. Chi2 Protein Docking Results (e) Solvent ribbon size surface, (f) Active site, (g) 3D docking, and (h) 2D docking with ligand N-acetyl-D(+)-glucosamine. BcLPMO10B Protein Docking Results (i) Solvent ribbon size surface, (j) Active site, (k) 3D docking, and (l) 2D docking with ligand 2,6-dimethoxyphenol**

Table 7: Systematic examination of molecular docking outcomes involving Chi1/Chi2 and BcLPMO10B proteins with N-Acetyl-D (+)-glucosamine and 2,6-dimethoxyphenol, respectively

Sr. No.	Protein	Ligand	Binding affinity (kcal/mol)	Drug score	Amino acid residues	Number of bonds	Types of bond interaction	Bond length (A°)
1	Chi1	N-Acetyl-D(+)-glucosamine	-5.2	0.52	Phe138	1	Hydrophobic Interaction	3.8
					Glu219	1	Conventional hydrogen bond	2.78
					Thr249	1	Conventional hydrogen bond	2.7
					Tyr283	1	Conventional hydrogen bond	2.87
					Trp407	2	Conventional hydrogen bond	3.17
							Pi-donor hydrogen bond	2.47
2	Chi2	N-Acetyl-D(+)-glucosamine	-5.2	0.83	Ile491	1	Conventional hydrogen bond	3.44
					Val493	2	Conventional hydrogen bond	2.37
							Conventional hydrogen bond	3.2
					Phe583	2	Conventional hydrogen bond	2.19
							Hydrophobic Interaction	3.66
					Ala664	1	Conventional hydrogen bond	3.15
					Glu665	1	Conventional hydrogen bond	2.45
					Leu666	1	Conventional hydrogen bond	1.9
3	BcLPMO10B	2, 6-Dimethoxy phenol	-3.9	0.57	His35	1	Pi-Alkyl	5.07
					Glu62	1	Carbon hydrogen bond	3.59
					Glu67	2	Conventional hydrogen bond	2.17
							Carbon hydrogen bond	3.58
					His87	2	Pi-Pi T shaped	4.99
							Pi-Alkyl	4.43
					Phe88	2	Hydrophobic Interaction	3.64
							Hydrophobic Interaction	3.77

Molecular Dynamics Simulation

In our present research, Molecular Dynamics Simulations (MDS) were carried out for significant molecular complexes, namely, the Chi1-GLcNAc complex, Chi2-GLcNAc, and BcLPMO10B-2,6-dimethoxyphenol (refer to Figures 11-14).

Upon meticulous examination of the simulation data, it was noted that the average Root Mean Square Deviation (RMSD) of the Chi1-GLcNAc complex was 2.2 Å. This value exhibited fluctuations within the acceptable range of 1–4 Å for a duration of up to 35 nanoseconds within the specific binding pocket of Chi1 (see Figure 11a). Additionally, key protein-ligand interactions within this context involved hydrogen bonding with Glu219, Thr249, Tyr283, and Trp407 (refer to Figures 10c and d). Similarly, the average RMSD of the Chi2-GLcNAc complex was determined to be 3.9 Å, displaying fluctuations within the acceptable range of 1–4 Å. This complex remained stable, albeit to a lesser extent compared to the Chi1-GLcNAc complex, within the specific binding pocket of Chi2 (see Figure 11b). Predominant protein-ligand interactions in this case included hydrogen bonding with Ile491, Val493, Phe583, Ala664, Glu665, and Leu666 (as depicted in Figures 10g and h). Lastly, the average RMSD of the BcLPMO10B-2,6-dimethoxyphenol complex was identified as 1.37 Å, with fluctuations falling within the acceptable range of 1–4 Å. Although more stable compared to the Chi1-GLcNAc and Chi2-GLcNAc complexes, this complex exhibited a slightly higher RMSD value within the specific binding pocket of Chi2 (as illustrated in Figure 11c). Notably, principal protein-ligand interactions in this instance were established through hydrogen bonding with Glu62 and Glu67 (as elucidated in Figures 10k and l).

Root Mean Square Fluctuations (RMSF) play a pivotal role in characterizing localized variations within the protein structure. These fluctuations serve as valuable indicators for pinpointing specific residues contributing to structural changes within the complex. It is essential to highlight that reduced fluctuations typically correlate with heightened structural stability. Upon comparative scrutiny, Figure 12(c), illustrating the BcLPMO10B protein complexed with 2,6-dimethoxyphenol, displayed significantly lower fluctuations compared to Figure 12(a), representing the Chi1-GLcNAc complex. Additionally, both of these complexes exhibited diminished fluctuations when compared to Figure 12(b), depicting the Chi2 protein in the complex with GLcNAc. This observation underscores the superior stability of the BcLPMO10B-2,6-dimethoxyphenol complex in contrast to the Chi1-GLcNAc and Chi2-GLcNAc complexes.

Analysis of the MD trajectories unveils that residues with elevated fluctuations are predominantly situated within loop regions or the N- and C-terminal regions of the protein structure, as depicted in Figure 13. This pattern highlights the dynamic nature of these specific protein segments. Furthermore, the stability of ligand binding to the protein is evident in the low Root Mean Square Fluctuation (RMSF) values observed for residues involved in the attachment of the ligand. These findings support the robustness of ligand-protein interactions.

In addition to RMSF, we closely monitored the distribution of secondary structure elements (SSE) throughout the simulation. The graphical representation in Figure 13 illustrates the distribution of alpha-helices and beta-strands across the protein structures, with SSE plotted against the residue index. Concerning alpha-helix content, Chi1 exhibited the highest proportion at 33.95%, followed by Chi2 at 20.62%, and BcLPMO10B

at 19.00%. Meanwhile, the extended strand conformation prevailed in Chi1, constituting 23.26% of its secondary structure, followed closely by Chi2 at 20.03%, and BcLPMO10B at 19.91%. As for beta-turn structures, Chi1 accounted for 9.07%, Chi2 for 6.68%, and BcLPMO10B for 5.88%. Notably, there is a significant difference in the random coil conformation, with Chi1 displaying 33.72%, while Chi2 exhibited a substantially higher percentage of 52.67%, and BcLPMO10B recorded the highest at 55.20%. These distinct secondary structure parameters provide valuable insights into the structural characteristics of Chi1, Chi2, and BcLPMO10B, shedding light on their conformational preferences and potential functional roles within the context of this study.

Figure 14(b) illustrates a greater number of hydrogen bonding interactions in the Chi2-GLcNAc complex compared to the Chi1-GLcNAc complex, as depicted in Figure 14(a), and surpasses the hydrogen bonding interaction observed in the BcLPMO10B-2,6-dimethoxyphenol complex, as shown in Figure 14(c).
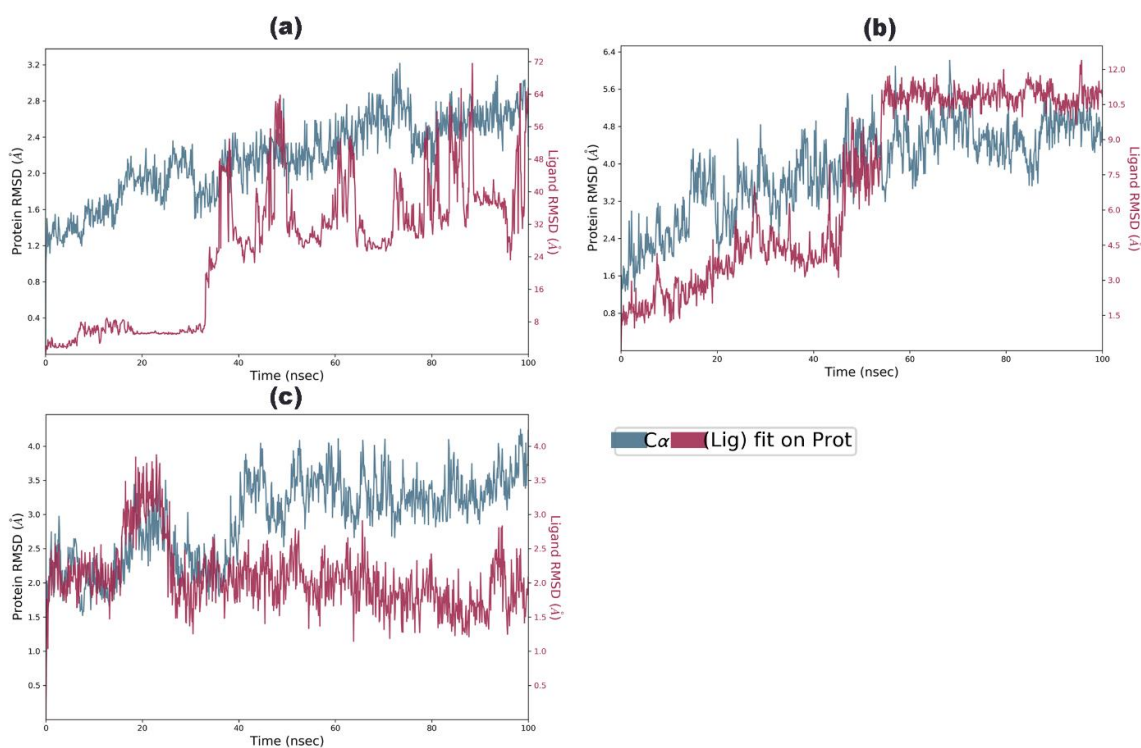

Figure 11: Changes in the trajectory analysis of the root mean square distance (RMSD) between C-alpha atoms of proteins and lead compounds were observed across time in (a) the Chi1-N-acetyl-D(+)-glucosamine complex, (b) Chi2 protein complexed with N-acetyl-D(+)-glucosamine, and (c) BcLPMO10B protein complexed with 2,6-dimethoxyphenol. In these analyses, the color pink signifies the RMSD values of the lead compound, while the color green represents the RMSD values of the protein target, both monitored over time
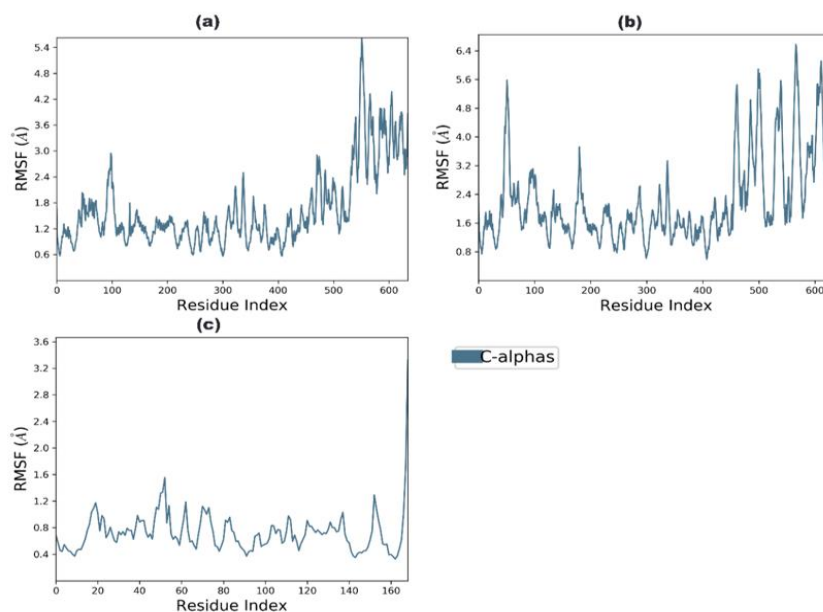
Figure 12: RMSF plots of (a) Chi1-N-acetyl-D(+)-glucosamine complex, (b) Chi2 protein complexed with N-acetyl-D(+)-glucosamine, (c) BcLPMO10B protein complexed with 2,6-dimethoxyphenol
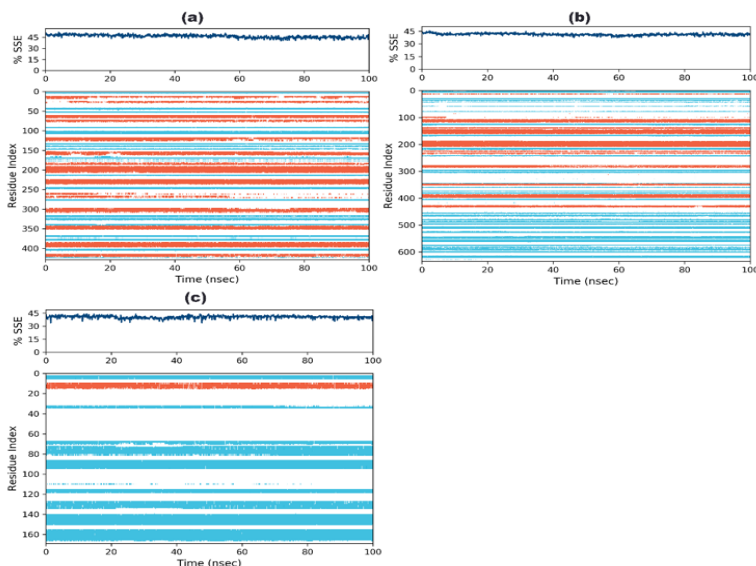

Figure 13: The spatial arrangement of protein secondary structure elements, specifically alpha helices (illustrated as red columns) and beta strands (depicted as blue columns), is observed in different scenarios throughout the simulation, including the (a) Chi1-N-acetyl-D(+)-glucosamine complex, the (b) Chi2-N-acetyl-D(+)-glucosamine complex, and the (c) BcLPMO10B-2,6-dimethoxy phenol complex, all occurring during the course of the simulation
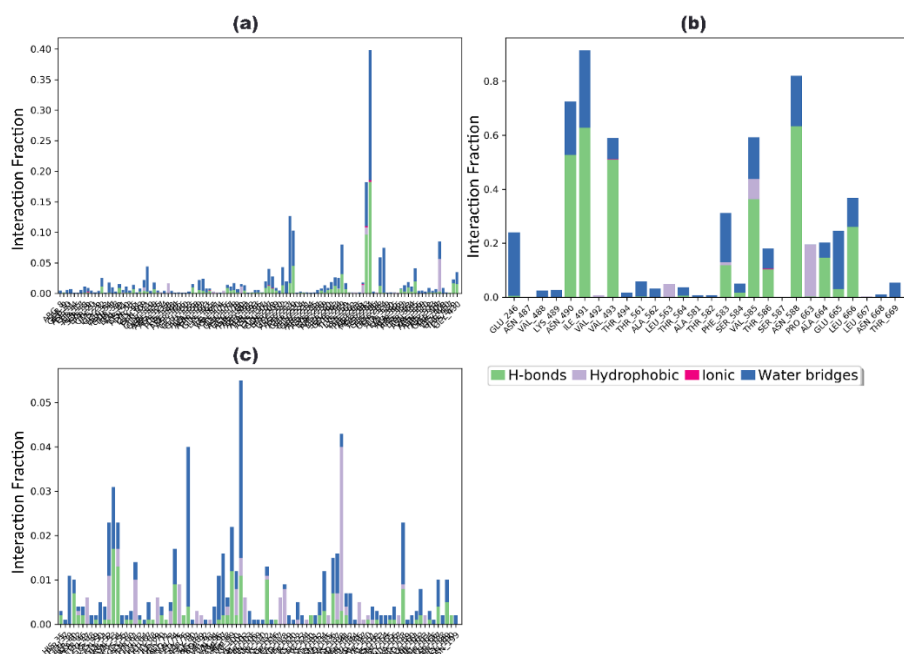
Figure 14: The contact heatmap throughout the simulation trajectory is depicted in the histogram for the Chi1-N-acetyl-D(+)-glucosamine complex (a), the Chi2-N-acetyl-D(+)-glucosamine complex (b), and the BcLPMO10B-2,6-dimethoxy phenol complex (c)

4. DISCUSSION

Modern advancements require computational protein analysis for drug design, protein-protein interactions, and structure prediction. Molecular docking and simulations, leveraging protein sequences and structures, facilitate drug discovery and predictions for uncharacterized proteins (Sen and Verma 2020). Computational simulations elucidate atom-level behavior, aiding in the prediction of protein expression, evolution, and structural trends. These studies, crucial for proteomics like vaccine development and disease understanding, present data in a simple, interpretable manner, supporting logical biological insights (Branco and Choupina 2021; Gangotia, Gupta, and Mani 2021; María Hernández-Domínguez et al. 2020). This study explored the evolutionary insights and action mechanisms of chitinolytic enzymes from B. cereus i.e., BcLPMO10B and chitinases using computational techniques.

The phylogenetic analysis presented in Figures 2-4 affirms the close evolutionary association of B. cereus chitinase proteins with those of other Bacillus species and, specifically for Chi1, with Streptomyces. Our results align with previous studies suggesting the acquisition of chitinase proteins through horizontal gene transfer (HGT) from bacteria and fungi, enhancing genetic diversity and adaptability (Honda et al. 2016; Moshe et al. 2023; Vandhana et al. 2022). Notably, environments rich in chitin, such as

soil and marine ecosystems, may drive HGT dynamics. Comparative studies with Bacillus species and Streptomyces revealed similarities in chitinase genes, supporting the hypothesis that B. cereus acquired these genes via HGT, potentially conferring a competitive advantage for chitin resource utilization and adaptation to diverse habitats. Additional analyses, including comparisons with gram-positive bacteria like Clostridium and Lactobacillus, further indicate a complex evolutionary history for B. cereus chitinase genes, potentially involving gene duplication and mutation. While these findings underscore the intricate evolutionary mechanisms shaping the chitinase protein of B. cereus, further research is warranted to elucidate its precise ecological and pathogenic roles, necessitating a comprehensive understanding of its function and regulatory mechanisms (Andreou et al. 2018; Drewnowska et al. 2020; Dutta et al. 2021; Martínez-Zavala et al. 2020; Moshe et al. 2023).

For this detailed study of action of mechanism of enzymes, we firstly constructed the 3D structures of BcLPMO10B and chitinases (Chi1, and Chi2). In evaluating the comparative analysis of protein structures (Chi1, Chi2, and BcLPMO10B), alignment with Swiss Models, I-TASSER, ROBETTA Baker Laboratory, and AlphaFold from UniProt provided insights into sequence identity, similarity, and structural features (Figures 5-8). 3D structural predictions and comprehensive structure validation using multiple criteria, including the Ramachandran plot, MolProbity, ERRAT, Verify3D, QMEAN score (Waterhouse et al. 2018), and Z-score, were conducted. Notably, Swiss Models and AlphaFold consistently exhibited high accuracy in the Ramachandran plot, showcasing favorable torsional angles for Chi1, Chi2, and BcLPMO10B (as per standard requirement (Hollingsworth and Karplus 2010)). Swiss Model consistently delivered high ERRAT scores, indicative of favorable model geometry, while AlphaFold demonstrated reliable structural predictions, albeit with a lower Z-Score for Chi1. I-TASSER presented acceptable ERRAT scores but showed slightly lower Z-Scores for Chi2 and BcLPMO10B. The comparative analysis underscores the importance of considering multiple validation metrics, with Swiss Model consistently performing well. Further evaluation of RobettaFold is warranted. Subsequent validation with MolProbity confirmed robust predictions for Chi1 and Chi2, with BcLPMO10B exhibiting slight deviations in specific criteria, providing a comprehensive overview of structural quality metrics.

The analysis of the Chi1 protein's structure and function using InterPro and Argot2 (Figure 1 and Table 2) revealed the presence of a LysM domain, which was further supported by additional annotations from CDD/PROSITE, SMART/SUPERFAMILY, and Pfam. These annotations consistently identified the LysM domain, with an additional LysM domain identified at residues 52-97, emphasizing its presence within the Chi1 protein. Additionally, the InterPro analysis indicated the existence of a glycoside hydrolase family 18 catalytic domain, suggesting Chi1's role as a glycoside hydrolase. The CDD analysis revealed specific active site residues (108, 138, 139, 180, 215, 217, 219, 281, 283, 284, 324, and 407) within the protein, indicating their importance in catalytic activity. For Chi2, InterPro annotations identified a carbohydrate-binding type-2 domain, and a fibronectin type III domain was found from residues 484-576, supported by Pfam annotations.

InterPro also identified a glycoside hydrolase family 18 catalytic domain within Chi2, spanning residues 39-476, which was further confirmed by Pfam. The active site residues (44, 72, 205, 207, 209, 282, 284, 285, 339, and 445) detected through CDD analysis are crucial for Chi2's glycoside hydrolase activity. The BcLPMO10B protein exhibited a unique domain profile, with InterPro and Pfam identifying a cellulose/chitin-binding protein N-terminal domain from residues 35-198, and a homologous superfamily annotation pointing to immunoglobulin E-set domains within residues 35-200, potentially contributing to BcLPMO10B's interaction with polysaccharides like cellulose and chitin.

Functional annotations using Argot2 for all three proteins provided insights into their molecular functions and biological processes, as summarized in Table 2. Both Chi1 and Chi2 exhibited a strong association with chitin catabolic processes, emphasizing their involvement in chitin degradation. Chi2 also demonstrated potential participation in cell wall-related processes. BcLPMO10B was implicated in carbohydrate binding and hydrolase activities, suggesting its role in the breakdown of carbohydrate substrates. The identified structural features and motifs in Chi1, Chi2, and BcLPMO10B offer valuable insights into their functions. The LysM domains in Chi1 and Chi2 likely facilitate their binding to chitin, a crucial component of fungal cell walls. Additionally, these enzymes feature glycoside hydrolase family 18 catalytic domains, essential for hydrolyzing glycosidic bonds in chitin and other carbohydrates. The presence of active site residues further supports their enzymatic activities. BcLPMO10B's N-terminal domain suggests its ability to bind cellulose and chitin, aligning with its role as a lytic polysaccharide mono-oxygenase. Immunoglobulin E-set domains within BcLPMO10B may play a vital role in substrate recognition or binding. The carbohydrate-binding and hydrolase activities of this enzyme contribute to the degradation of carbohydrates.

For validation of their predicted functions, we investigated their interactions with 2, 6-dimethoxyphenol and GLcNAc as substrates for BcLPMO10B and chitinases, respectively. To reveal the action mechanism of BcLPMO10B and chitinases, docking studies were utilized to predict the ligand binding modes. Our research also included MDS. We examined the dynamics of BcLPMO10B and chitinases by simulating the temporal motion of individual atoms. To uncover the effect of structural alterations on these proteins' enzymatic function, we conducted this study.

Exploration of the binding modes of GLcNAc and 2,6-dimethoxyphenol with Chi1 and Chi2 was conducted using AutoDock Vina, as illustrated in Figure 10. These investigations provided valuable insights into the catalytic mechanisms, substrate preferences, and binding affinities of these enzymes. Following docking, GLcNAc exhibited similar binding affinities with Chi1 and Chi2 proteins (-5.2 kcal/mol). However, the interaction of Chi1-GLcNAc involved key amino acids with three conventional hydrogen bonds and two additional bonds, while the Chi2-GLcNAc interaction featured critical amino acids with seven hydrogen bonds and one other bond, as outlined in Table 7 and Figures 10c, d, g, and h.

The predicted binding modes of substrates/ligands to Chi1/Chi2 and BcLPMO10B proteins offer insights into their enzymatic activities, revealing how these proteins interact with specific substrates and shedding light on their catalytic mechanisms.

Molecular Dynamics Simulations (MDS) were subsequently conducted on the resulting molecular complexes (Chi1-GLcNAc, Chi2-GLcNAc, and BcLPMO10B-2,6-dimethoxyphenol) obtained from docking. The temporal motion of individual atoms within these complexes was examined using Root Mean Square Deviation (RMSD) analysis (Figure 11). The Chi1-GLcNAc complex demonstrated stability within the binding pocket of Chi1, featuring prominent protein-ligand interactions involving hydrogen bonding with Glu219, Thr249, Tyr283, and Trp407.

While the Chi2-GLcNAc complex showed slightly less stability than Chi1, it maintained stability within the binding pocket of Chi2. The BcLPMO10B-2,6-dimethoxyphenol complex exhibited the highest stability (Figure 11c). Analyzing the RMSDs of these complexes provided insight into their stability and specific protein-ligand interactions contributing to it.

Root Mean Square Fluctuation (RMSF) calculations for individual residues within the protein structures were performed to understand the impact of structural changes on dynamics (Figure 11).

Residues with high fluctuations were primarily found at the N- or C-terminal regions or within loop regions. A low RMSF value for residues involved in ligand attachment indicated a robust ligand-protein interaction, emphasizing the critical role of ligand binding in stabilizing specific protein segments.

Secondary Structure Elements (SSE) were closely monitored throughout the simulation in addition to RMSD and RMSF analyses (Figure 12). The findings shed light on the functional roles of BcLPMO10B, Chi1, and Chi2 during catalysis, based on conformational changes during the process. BcLPMO10B exhibited the highest proportion of alpha helices (19.00%), followed by Chi2 (20.62%).

Distinct conformational preferences among proteins based on their secondary structures were observed, including differences in extending strands, beta-turn structures, and random coil structures among the complexes. These variations may imply different catalytic mechanisms employed by these enzymes, providing insights into their structural characteristics and potential functional roles.

Comparative analysis of the hydrogen bonding interactions revealed that Chi2-GLcNAc exhibited a greater degree of hydrogen bonding than Chi1-GLcNAc and BcLPMO10B-2,6-dimethoxyphenol (Figure 13). As hydrogen bonding interactions are crucial in enzyme-substrate interactions, Chi2 may possess a more efficient catalytic mechanism. The hydrogen bonding interactions offer insights into the respective enzymatic functions of these complexes, reflecting differences in enzyme efficiencies and substrate binding affinities.

5. CONCLUSION

Our study used computer methods to explore how certain enzymes in *B. cereus*, called BcLPMO10B and chitinases (Chi1 and Chi2), have evolved and work. These enzymes have close connections with other bacteria, suggesting they may have shared genes over time. We also created and validated 3D models of these enzymes, showing they are of high quality. By simulating their movements, we learned more about how they interact with specific substances. Additionally, we identified specific parts in Chi1, Chi2, and BcLPMO10B that help them break down carbohydrates. Our findings highlight the diversity in how these enzymes work. By studying their behaviors in detail, we gained insights into their stability and how they change shape during their activities. This study would lead to further experimental validation that may be done in relevant biological contexts to elucidate the potential applications of these enzymes in biotechnology and pharmaceutical research.

Author Contribution

Zeeshan Mutahir conceptualized and supervised the study. Ammarah Fateen and Muhammad Zeeshan Ahmed analyzed, curated and initially prepared the manuscript. Naeem Mahmood Ashraf, Muhammad Khurshid and Moazzam Ali proofread the manuscript. All authors contributed in comprehensive review, editing and granted their approval for the final manuscript.

Acknowledgements

Disclosure Statement

The authors declare no potential conflict of interest.

References

1) Adasme, Melissa F., Katja L. Linnemann, Sarah Naomi Bolz, Florian Kaiser, Sebastian Salentin, V. Joachim Haupt, and Michael Schroeder. 2021. "PLIP 2021: Expanding the Scope of the Protein-Ligand Interaction Profiler to DNA and RNA." *Nucleic Acids Research* 49(W1):W530–34. doi: 10.1093/nar/gkab294.

2) Adrangi, Sina, and Mohammad Ali Faramarzi. 2013. "From Bacteria to Human: A Journey into the World of Chitinases." *Biotechnology Advances* 31(8):1786–95.

3) Agnihotry, Shikha, Rajesh Kumar Pathak, Dev Bukhsh Singh, Apoorv Tiwari, and Imran Hussain. 2021. "Protein Structure Prediction." Pp. 177–88 in *Bioinformatics: Methods and Applications*. Academic Press.

4) Agostoni, Marco, John A. Hangasky, and Michael A. Marletta. 2017. "Physiological and Molecular Understanding of Bacterial Polysaccharide Monooxygenases." *Microbiology and Molecular Biology Reviews* 81(3):e00015-17. doi: 10.1128/mmbr.00015-17.

5) Ahmed, Muhammad Zeeshan, Shahzeb Hameed, Mazhar Ali, and Ammad Zaheer. 2021. "In Silico Molecular Docking Analysis of Limonene with The Fat Mass and Obesity-Associated Protein by Using Autodock Vina." *Scientific Journal of Informatics* 8(1):154–60.

6) Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25(17):3389–3402.

7) Altschul, Stephen F., John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schäffer, and Yi Kuo Yu. 2005. "Protein Database Searches Using Compositionally Adjusted Substitution Matrices." *The FEBS Journal* 272(20):5101–9. doi: 10.1111/J.1742-4658.2005.04945.X.

8) Andreou, Athena, Petros Giastas, Elias Christoforides, and Elias E. Eliopoulos. 2018. "Structural and Evolutionary Insights within the Polysaccharide Deacetylase Gene Family of Bacillus Anthracis and Bacillus Cereus." *Genes* 9(8). doi: 10.3390/genes9080386.

9) Anil, Sukumaran. 2022. "Potential Medical Applications of Chitooligosaccharides." *Polymers* 14(17):3558.

10) Azuma, Kazuo, Tomohiro Osaki, Saburo Minami, and Yoshiharu Okamoto. 2015. "Anticancer and Anti-Inflammatory Properties of Chitin and Chitosan Oligosaccharides." *Journal of Functional Biomaterials* 6(1):33–49. doi: 10.3390/jfb6010033.

11) Batool, Maria, Bilal Ahmad, and Sangdun Choi. 2019. "A Structure-Based Drug Discovery Paradigm." *International Journal of Molecular Sciences* 20(11):2783.

12) Beier, Sara, and Stefan Bertilsson. 2013. "Bacterial Chitin Degradation-Mechanisms and Ecophysiological Strategies." *Frontiers in Microbiology* 4(JUN):52995.

13) Benkert, Pascal, Michael Künzli, and Torsten Schwede. 2009. "QMEAN Server for Protein Model Quality Estimation." *Nucleic Acids Research* 37(SUPPL. 2):W510–14. doi: 10.1093/nar/gkp322.

14) Bhat, Gh Rasool, Itty Sethi, Bilal Rah, Rakesh Kumar, and Dil Afroze. 2022. "Innovative in Silico Approaches for Characterization of Genes and Proteins." *Frontiers in Genetics* 13:865182.

15) Bhattacharya, Debaditya, Anand Nagpure, and Rajinder K. Gupta. 2007. "Bacterial Chitinases: Properties and Potential." *Critical Reviews in Biotechnology* 27(1):21–28.

16) Bottone, Edward J. 2010. "Bacillus Cereus, a Volatile Human Pathogen." *Clinical Microbiology Reviews* 23(2):382–98.

17) Bowers, Kevin J., Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw. 2006. "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters." in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06*.

18) Bowie, James U., Roland Lüthy, and David Eisenberg. 1991. "A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure." *Science* 253(5016):164–70. doi: 10.1126/science.1853201.

19) Branco, Iuliia, and Altino Choupina. 2021. "Bioinformatics: New Tools and Applications in Life Science and Personalized Medicine." *Applied Microbiology and Biotechnology* 105(3):937–51.

20) Chiu, Elaine, Marcel Hijnen, Richard D. Bunker, Marion Boudes, Chitra Rajendran, Kaheina Aizel, Vincent Oliéric, Clemens Schulze-Briese, Wataru Mitsuhashi, Vivienne Young, Vernon K. Ward, Max Bergoin, Peter Metcal, and Fasséli Coulibaly. 2015. "Structural Basis for the Enhancement of Virulence by Viral Spindles and Their in Vivo Crystallization." *Proceedings of the National Academy of Sciences of the United States of America* 112(13):3973–78. doi: 10.1073/pnas.1418798112.

21) Colovos, Chris, and Todd O. Yeates. 1993. "Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions." *Protein Science* 2(9):1511–19. doi: 10.1002/pro.5560020916.

22) Deka, Mark A., Chung K. Marston, Julia Garcia-Diaz, Rahsaan Drumgoole, and Rita M. Traxler. 2022. "Ecological Niche Model of Bacillus Cereus Group Isolates Containing a Homologue of the PXO1 Anthrax Toxin Genes Infecting Metalworkers in the United States." *Pathogens* 11(4):470. doi: 10.3390/pathogens11040470.

23) Demirkan, Elif, Aynur Aybey Çetinkaya, and Maoulida Abdou. 2021. "Lipase from New Isolate Bacillus Cereus Ata179: Optimization of Production Conditions, Partial Purification, Characterization and Its Potential in the Detergent Industry." *Turkish Journal of Biology* 45(3):287–300. doi: 10.3906/biy-2101-22.

24) Dhiman, Vedikaa, Soham Biswas, Rajveer Singh Shekhawat, Ayan Sadhukhan, and Pankaj Yadav. 2023. "In Silico Characterization of Five Novel Disease-Resistance Proteins in Oryza Sativa Sp. Japonica against Bacterial Leaf Blight and Rice Blast Diseases." *BioRxiv* 2023.06.05.543715. doi: 10.1101/2023.06.05.543715.

25) Drewnowska, J. M., A. Fiodor, J. E. Barboza-Corona, and I. Swiecicka. 2020. "Chitinolytic Activity of Phylogenetically Diverse Bacillus Cereus Sensu Lato from Natural Environments." *Systematic and Applied Microbiology* 43(3):126075. doi: 10.1016/j.syapm.2020.126075.

26) Drula, Elodie, Marie Line Garron, Suzan Dogan, Vincent Lombard, Bernard Henrissat, and Nicolas Terrapon. 2022. "The Carbohydrate-Active Enzyme Database: Functions and Literature." *Nucleic Acids Research* 50(D1):D571–77. doi: 10.1093/nar/gkab1045.

27) Dutta, Bhramar, Jan Deska, Rajib Bandopadhyay, and Salem Shamekh. 2021. "In Silico Characterization of Bacterial Chitinase: Illuminating Its Relationship with Archaeal and Eukaryotic Cousins." *Journal of Genetic Engineering and Biotechnology* 19(1):1–11. doi: 10.1186/s43141-021-00121-6.

28) Eijsink, Vincent G. H., Dejan Petrovic, Zarah Forsberg, Sophanit Mekasha, Åsmund K. Røhr, Anikó Várnai, Bastien Bissaro, and Gustav Vaaje-Kolstad. 2019. "On the Functional Characterization of Lytic Polysaccharide Monooxygenases (LPMOs)." *Biotechnology for Biofuels* 12(1):1–16.

29) El-Arabi, Tarek F., and Mansel W. Griffiths. 2021. "Bacillus Cereus." Pp. 431–37 in *Foodborne Infections and Intoxications*. StatPearls Publishing.

30) Esmkhani, Mohammad, and Saeed Shams. 2022. "Cutaneous Infection Due to Bacillus Cereus: A Case Report." *BMC Infectious Diseases* 22(1):1–4. doi: 10.1186/s12879-022-07372-9.

31) Feliatra, Feli, Ummi Mardhiah Batubara, Yuana Nurulita, Iesje Lukistyowati, and Jarot Setiaji. 2021. "The Potentials of Secondary Metabolites from Bacillus Cereus SN7 and Vagococcus Fluvialis CT21 against Fish Pathogenic Bacteria." *Microbial Pathogenesis* 158:105062. doi: 10.1016/j.micpath.2021.105062.

32) Ferreira, Leonardo G., Ricardo N. Dos Santos, Glaucius Oliva, and Adriano D. Andricopulo. 2015. "Molecular Docking and Structure-Based Drug Design Strategies." *Molecules* 20(7):13384. doi: 10.3390/MOLECULES200713384.

33) Gabaldón, Toni. 2005. "Evolution of Proteins and Proteomes: A Phylogenetics Approach." *Evolutionary Bioinformatics* 1:117693430500100. doi: 10.1177/117693430500100004.

34) Gangotia, Disha, Aeshna Gupta, and Indra Mani. 2021. "Role of Bioinformatics in Biological Sciences." Pp. 37–57 in *Advances in Bioinformatics*. Springer Nature.

35) Gasteiger, Elisabeth, Alexandre Gattiker, Christine Hoogland, Ivan Ivanyi, Ron D. Appel, and Amos Bairoch. 2003. "ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis." *Nucleic Acids Research* 31(13):3784–88. doi: 10.1093/nar/gkg563.

36) Gasteiger, Elisabeth, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. 2005. "Protein Identification and Analysis Tools on the ExPASy Server." Pp. 571–607 in *The Proteomics Protocols Handbook*. Humana Press.

37) Geourjon, C., and G. Deléage. 1995. "Sopma: Significant Improvements in Protein Secondary Structure Prediction by Consensus Prediction from Multiple Alignments." *Bioinformatics* 11(6):681–84. doi: 10.1093/bioinformatics/11.6.681.

38) Gomaa, Eman Zakaria. 2021. "Microbial Chitinases: Properties, Enhancement and Potential Applications." *Protoplasma* 258(4):695–710.

39) Hasan, Rajnee, Md Nazmul Haq Rony, and Rasel Ahmed. 2021. "In Silico Characterization and Structural Modeling of Bacterial Metalloprotease of Family M4." *Journal of Genetic Engineering and Biotechnology* 19(1):1–20. doi: 10.1186/S43141-020-00105-Y/FIGURES/11.

40) Hildebrand, Peter W., Alexander S. Rose, and Johanna K. S. Tiemann. 2019. "Bringing Molecular Dynamics Simulation Data into View." *Trends in Biochemical Sciences* 44(11):902–13. doi: 10.1016/J.TIBS.2019.06.004.

41) Hollingsworth, Scott A., and P. Andrew Karplus. 2010. "A Fresh Look at the Ramachandran Plot and the Occurrence of Standard Structures in Proteins." *Biomolecular Concepts* 1(3–4):271–83.

42) Honda, Shotaro, Satoshi Wakita, Yasusato Sugahara, Masao Kawakita, Fumitaka Oyama, and Masayoshi Sakaguchi. 2016. "Characterization of Two Listeria Innocua Chitinases of Different Sizes That Were Expressed in Escherichia Coli." *Applied Microbiology and Biotechnology* 100(18):8031–41. doi: 10.1007/s00253-016-7546-0.

43) Jan, Mehmood, Gulmeena Shah, Sadaf Masood, Kamran Iqbal Shinwari, Rashida Hameed, E. S. Rha, and Muhammad Jamil. 2019. "Bacillus Cereus Enhanced Phytoremediation Ability of Rice Seedlings under Cadmium Toxicity." *BioMed Research International* 2019. doi: 10.1155/2019/8134651.

44) Keane, Thomas M., Christopher J. Creevey, Melissa M. Pentony, Thomas J. Naughton, and James O. McInerney. 2006. "Assessment of Methods for Amino Acid Matrix Selection and Their Use on Empirical Data Shows That Ad Hoc Assumptions for Choice of Matrix Are Not Justified." *BMC Evolutionary Biology* 6(1):1–17. doi: 10.1186/1471-2148-6-29.

45) Kęska, Paulina, Waldemar Gustaw, and Joanna Stadnik. 2021. "Trends in in Silico Approaches to the Prediction of Biologically Active Peptides in Meat and Meat Products as an Important Factor for Preventing Food-Related Chronic Diseases." *Applied Sciences (Switzerland)* 11(23):11236.

46) Kim, David E., Dylan Chivian, and David Baker. 2004. "Protein Structure Prediction and Analysis Using the Robetta Server." *Nucleic Acids Research* 32(WEB SERVER ISS.):W526. doi: 10.1093/nar/gkh468.

47) Kingdon, Alexander D. H., and Luke J. Alderwick. 2021. "Structure-Based in Silico Approaches for Drug Discovery against Mycobacterium Tuberculosis." *Computational and Structural Biotechnology Journal* 19:3708–19.

48) Kowalska, Joanna, Elzbieta Maćkiw, Dorota Korsak, and Jacek Postupolski. 2023. "Prevalence of Bacillus Cereus in Food Products in Poland." *Annals of Agricultural and Environmental Medicine*. doi: 10.26444/AAEM/168580.

49) Kryshtafovych, Andriy, Krzysztof Fidelis, and Anna Tramontano. 2011. "Evaluation of Model Quality Predictions in CASP9." *Proteins: Structure, Function and Bioinformatics* 79(SUPPL. 10):91–106. doi: 10.1002/prot.23180.

50) Kulkova, Iryna, Jakub Dobrzyński, Paweł Kowalczyk, Grzegorz Bełżecki, and Karol Kramkowski. 2023. "Plant Growth Promotion Using Bacillus Cereus." *International Journal of Molecular Sciences* 24(11):9759.

51) Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. "PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures." *Journal of Applied Crystallography* 26(2):283–91. doi: 10.1107/s0021889892009944.

52) Le, Si Quang, and Olivier Gascuel. 2008. "An Improved General Amino Acid Replacement Matrix." *Molecular Biology and Evolution* 25(7):1307–20. doi: 10.1093/molbev/msn067.

53) Levasseur, Anthony, Elodie Drula, Vincent Lombard, Pedro M. Coutinho, and Bernard Henrissat. 2013. "Expansion of the Enzymatic Repertoire of the CAZy Database to Integrate Auxiliary Redox Enzymes." *Biotechnology for Biofuels* 6(1):1–14. doi: 10.1186/1754-6834-6-41.

54) Li, Fei, Honglu Zhao, Yuxin Liu, Jiaqi Zhang, and Hongbo Yu. 2023. "Chitin Biodegradation by Lytic Polysaccharide Monooxygenases from Streptomyces Coelicolor In Vitro and In Vivo." *International Journal of Molecular Sciences* 24(1):275. doi: 10.3390/ijms24010275.

55) Li, Xin, William T. Beeson IV, Christopher M. Phillips, Michael A. Marletta, and Jamie H. D. Cate. 2012. "Structural Basis for Substrate Targeting and Catalysis by Fungal Polysaccharide Monooxygenases." *Structure* 20(6):1051–61. doi: 10.1016/j.str.2012.04.002.

56) Lüthy, Roland, James U. Bowie, and David Eisenberg. 1992. "Assessment of Protein Models with Three-Dimensional Profiles." *Nature* 356(6364):83–85. doi: 10.1038/356083a0.

57) Maiorov, Vladimir N., and Gordon M. Crippen. 1994. "Significance of Root-Mean-Square Deviation in Comparing Three-Dimensional Structures of Globular Proteins." *Journal of Molecular Biology* 235(2):625–34. doi: 10.1006/JMBI.1994.1017.

58) María Hernández-Domínguez, Edna, Laura Sofía Castillo-Ortega, Yarely García-Esquivel, Virginia Mandujano-González, Gerardo Díaz-Godínez, and Jorge Álvarez-Cervantes. 2020. "Bioinformatics as a Tool for the Structural and Evolutionary Analysis of Proteins." in *Computational Biology and Chemistry*. IntechOpen.

59) Martínez-Zavala, Sheila A., Uriel E. Barboza-Pérez, Gustavo Hernández-Guzmán, Dennis K. Bideshi, and José E. Barboza-Corona. 2020. "Chitinases of Bacillus Thuringiensis: Phylogeny, Modular Structure, and Applied Potentials." *Frontiers in Microbiology* 10:499289.

60) McGuffin, Liam J., Kevin Bryson, and David T. Jones. 2000. "The PSIPRED Protein Structure Prediction Server." *Bioinformatics* 16(4):404–5. doi: 10.1093/bioinformatics/16.4.404.

61) Moshe, Maya, Chhedi Lal Gupta, Noa Sela, Dror Minz, Ehud Banin, Omer Frenkel, and Eddie Cytryn. 2023. "Comparative Genomics of Bacillus Cereus Sensu Lato Spp. Biocontrol Strains in Correlation to in-Vitro Phenotypes and Plant Pathogen Antagonistic Capacity." *Frontiers in Microbiology* 14:996287. doi: 10.3389/fmicb.2023.996287.

62) Mutahir, Zeeshan, Sophanit Mekasha, Jennifer S. M. Loose, Faiza Abbas, Gustav Vaaje-Kolstad, Vincent G. H. Eijsink, and Zarah Forsberg. 2018. "Characterization and Synergistic Action of a Tetra-Modular Lytic Polysaccharide Monooxygenase from Bacillus Cereus." *FEBS Letters* 592(15):2562–71. doi: 10.1002/1873-3468.13189.

63) Olowosoke, Christopher Busayo, Tope Abraham Ibisanmi, Chioma Joy Eze, Abayomi Abiodun Shofunde, Tomiwa Lois Olubena, and Olalekan Akadiri. 2023. "Investigation of Polymorphism Role in Protein Structure and Function for Selected Cancer and Diabetes Disease; a Rationale to Selection of Targets for Insilico Drug Screening." *Informatics in Medicine Unlocked* 42:101342. doi: 10.1016/j.imu.2023.101342.

64) Oyeleye, Ayokunmi, and Yahaya M. Normi. 2018. "Chitinase: Diversity, Limitations, and Trends in Engineering for Suitable Applications." *Bioscience Reports* 38(4):BSR2018032300.

65) Paysan-Lafosse, Typhaine, Matthias Blum, Sara Chuguransky, Tiago Grego, Beatriz Lázaro Pinto, Gustavo A. Salazar, Maxwell L. Bileschi, Peer Bork, Alan Bridge, Lucy Colwell, Julian Gough, Daniel H. Haft, Ivica Letunić, Aron Marchler-Bauer, Huaiyu Mi, Darren A. Natale, Christine A. Orengo, Arun P. Pandurangan, Catherine Rivoire, Christian J. A. Sigrist, Ian Sillitoe, Narmada Thanki, Paul D. Thomas, Silvio C. E. Tosatto, Cathy H. Wu, and Alex Bateman. 2023. "InterPro in 2022." *Nucleic Acids Research* 51(D1):D418–27. doi: 10.1093/nar/gkac993.

66) Rasheed, Muhammad Asif, Muhammad Nasir Iqbal, Salina Saddick, Iqra Ali, Falak Sher Khan, Sumaira Kanwal, Dawood Ahmed, Muhammad Ibrahim, Umara Afzal, and Muhammad Awais. 2021. "Identification of Lead Compounds against Scm (Fms10) in Enterococcus Faecium Using Computer Aided Drug Designing." *Life (Basel, Switzerland)* 11(2):1–15. doi: 10.3390/LIFE11020077.

67) Rathore, Abhishek Singh, and Rinkoo D. Gupta. 2015. "Chitinases from Bacteria to Human: Properties, Applications, and Future Perspectives." *Enzyme Research* 2015.

68) Roberto, Adame Gómez, Cruz Facundo Itzel-Maralhi, García Díaz Lilia-Lizette, Ramírez Sandoval Yesenia, Pérez Valdespino Abigail, Ortuño Pineda Carlos, Santiago Dionisio Maria-Cristina, and Ramírez Peralta Arturo. 2020. "Biofilm Production by Enterotoxigenic Strains of Bacillus Cereus in Different Materials and under Different Environmental Conditions." *Microorganisms* 8(7):1–13. doi: 10.3390/microorganisms8071071.

69) Rodrigo, Dolores, Cristina M. Rosell, and Antonio Martinez. 2021. "Risk of Bacillus Cereus in Relation to Rice and Derivatives." *Foods* 10(2):302.

70) Rodriguez, Rolando, Glay Chinea, Nelia Lopez, Tirso Pons, and Gert Vriend. 1998. "Homology Modeling, Model and Software Evaluation: Three Related Resources." *Bioinformatics* 14(6):523–28. doi: 10.1093/bioinformatics/14.6.523.

71) Saleem, Mahjabeen, Sumaira Ahmad, and Mubashir Ahmad. 2014. "Potential of Bacillus Cereus for Bioremediation of Pulp and Paper Industrial Waste." *Annals of Microbiology* 64(2):823–29. doi: 10.1007/s13213-013-0721-y.

72) Schöning-Stierand, Katrin, Konrad Diedrich, Rainer Fährrolfes, Florian Flachsenberg, Agnes Meyder, Eva Nittinger, Ruben Steinegger, and Matthias Rarey. 2020. "ProteinsPlus: Interactive Analysis of Protein–Ligand Binding Interfaces." *Nucleic Acids Research* 48(W1):W48–53. doi: 10.1093/NAR/GKAA235.

73) Sen, Tanuka, and Naresh K. Verma. 2020. "Functional Annotation and Curation of Hypothetical Proteins Present in a Newly Emerged Serotype 1c of Shigella Flexneri: Emphasis on Selecting Targets for Virulence and Vaccine Design Studies." *Genes* 11(3):340. doi: 10.3390/genes11030340.

74) Sengupta, Kaustav, Sovan Saha, Anup Kumar Halder, Piyali Chatterjee, Mita Nasipuri, Subhadip Basu, and Dariusz Plewczynski. 2022. "PFP-GO: Integrating Protein Sequence, Domain and Protein-Protein Interaction Information for Protein Function Prediction Using Ranked GO Terms." *Frontiers in Genetics* 13:969915. doi: 10.3389/fgene.2022.969915.

75) Shivakumar, Devleena, Joshua Williams, Yujie Wu, Wolfgang Damm, John Shelley, and Woody Sherman. 2010. "Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the Opls Force Field." *Journal of Chemical Theory and Computation* 6(5):1509–19. doi: 10.1021/CT900587B.

76) Singh, Raj, Sushil Kumar Upadhyay, Manoj Singh, Indu Sharma, Pooja Sharma, Pooja Kamboj, Adesh Saini, Reena Voraha, Anil Kumar Sharma, Tarun Kumar Upadhyay, and Fahad Khan. 2020. "Chitin, Chitinases and Chitin Derivatives in Biopharmaceutical, Agricultural and Environmental Perspective." *Biointerface Research in Applied Chemistry* 11(3):9985–10005.

77) Steinegger, Martin, and Johannes Söding. 2017. "MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets." *Nature Biotechnology* 35(11):1026–28.

78) Tamura, Koichiro, Daniel Peterson, Nicholas Peterson, Glen Stecher, Masatoshi Nei, and Sudhir Kumar. 2011. "MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods." *Molecular Biology and Evolution* 28(10):2731–39. doi: 10.1093/molbev/msr121.

79) Tamura, Koichiro, Glen Stecher, and Sudhir Kumar. 2021. "MEGA11: Molecular Evolutionary Genetics Analysis Version 11." *Molecular Biology and Evolution* 38(7):3022–27. doi: 10.1093/molbev/msab120.

80) Tian, Wei, Chang Chen, Xue Lei, Jieling Zhao, and Jie Liang. 2018. "CASTp 3.0: Computed Atlas of Surface Topography of Proteins." *Nucleic Acids Research* 46(W1):W363–67. doi: 10.1093/nar/gky473.

81) Uddin, Md Bashir, Emran Hossain Sajib, Syeda Farjana Hoque, Md Nazmul Islam Bappy, Fazle Elahi, Arabinda Ghosh, Samuel Muhit, Mohammad Mahmudul Hassan, Mahmudul Hasan, Ramachandran Chelliah, Se Jin Park, Tamanna Jahan Mony, Deog Hwan Oh, and Syed Sayeem Uddin Ahmed. 2022. "Genomic Diversity and Molecular Dynamics Interaction on Mutational Variances among RB Domains of SARS-CoV-2 Interplay Drug Inactivation." *Infection, Genetics and Evolution* 97:105128. doi: 10.1016/j.meegid.2021.105128.

82) Vaikundamoorthy, Ramalingam, Rajaram Rajendran, Ananth Selvaraju, Kaviyarasan Moorthy, and Santhanam Perumal. 2018. "Development of Thermostable Amylase Enzyme from Bacillus Cereus for Potential Antibiofilm Activity." *Bioorganic Chemistry* 77:494–506. doi: 10.1016/j.bioorg.2018.02.014.

83) Vandhana, Theruvothu Madathil, Jean Lou Reyre, Dangudubiyyam Sushmaa, Jean Guy Berrin, Bastien Bissaro, and Jogi Madhuprakash. 2022. "On the Expansion of Biological Functions of Lytic Polysaccharide Monooxygenases." *New Phytologist* 233(6):2380–96.

84) Wang, Chao, and Quan Zou. 2023. "Prediction of Protein Solubility Based on Sequence Physicochemical Patterns and Distributed Representation Information with DeepSoluE." *BMC Biology* 21(1):1–11. doi: 10.1186/s12915-023-01510-8.

85) Waterhouse, Andrew, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T. Heer, Tjaart A. P. De Beer, Christine Rempfer, Lorenza Bordoli, Rosalba Lepore, and Torsten Schwede. 2018. "SWISS-MODEL: Homology Modelling of Protein Structures and Complexes." *Nucleic Acids Research* 46(W1):W296–303.

86) Whelan, Simon, and Nick Goldman. 2001. "A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach." *Molecular Biology and Evolution* 18(5):691–99. doi: 10.1093/oxfordjournals.molbev.a003851.

87) Wiederstein, Markus, and Manfred J. Sippl. 2007. "ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins." *Nucleic Acids Research* 35(SUPPL.2):W407–10. doi: 10.1093/nar/gkm290.

88) Wilkins, M. R., E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser. 1999. "Protein Identification and Analysis Tools in the ExPASy Server." *Methods in Molecular Biology (Clifton, N.J.)* 112:531–52.

89) Williams, Christopher J., Jeffrey J. Headd, Nigel W. Moriarty, Michael G. Prisant, Lizbeth L. Videau, Lindsay N. Deis, Vishal Verma, Daniel A. Keedy, Bradley J. Hintze, Vincent B. Chen, Swati Jain, Steven M. Lewis, W. Bryan Arendall, Jack Snoeyink, Paul D. Adams, Simon C. Lovell, Jane S. Richardson, and David C. Richardson. 2018. "MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation." *Protein Science* 27(1):293–315. doi: 10.1002/pro.3330.

90) Yang, Jianyi, and Yang Zhang. 2015. "I-TASSER Server: New Development for Protein Structure and Function Predictions." *Nucleic Acids Research* 43(W1):W174–81. doi: 10.1093/nar/gkv342.

91) Yang, Shuo, Yating Wang, Yangtai Liu, Kai Jia, Zhen Zhang, and Qingli Dong. 2023. "Cereulide and Emetic Bacillus Cereus: Characterizations, Impacts and Public Precautions." *Foods* 12(4):833.

92) Zardoya, Rafael. 2005. "Phylogeny and Evolution of the Major Intrinsic Protein Family." *Biology of the Cell* 97(6):397–414. doi: 10.1042/bc20040134.

93) Zhang, Yang, Delong Pan, Peiyao Xiao, Qianqian Xu, Fan Geng, Xinyu Zhang, Xiuling Zhou, and Hong Xu. 2023. "A Novel Lytic Polysaccharide Monooxygenase from Enrichment Microbiota and Its Application for Shrimp Shell Powder Biodegradation." *Frontiers in Microbiology* 14:1097492. doi: 10.3389/fmicb.2023.1097492.

94) Zhang, Ying Bo, Yuan Yuan, Yu Xin Pang, Fu Lai Yu, Chao Yuan, Xuan Hu, and Dan Wang. 2019. "Phylogenetic Reconstruction and Divergence Time Estimation of Blumea Dc. (Asteraceae: Inuleae) in China Based on Nrdna Its and Cpdna Trnl-f Sequences." *Plants* 8(7):210. doi: 10.3390/plants8070210.

95) Zheng, Wei, Chengxin Zhang, Yang Li, Robin Pearce, Eric W. Bell, and Yang Zhang. 2021. "Folding Non-Homologous Proteins by Coupling Deep-Learning Contact Maps with I-TASSER Assembly Simulations." *Cell Reports Methods* 1(3):100014. doi: 10.1016/j.crmeth.2021.100014.

96) Zhou, Xiaogen, Wei Zheng, Yang Li, Robin Pearce, Chengxin Zhang, Eric W. Bell, Guijun Zhang, and Yang Zhang. 2022. "I-TASSER-MTD: A Deep-Learning-Based Platform for Multi-Domain Protein Structure and Function Prediction." *Nature Protocols* 17(10):2326–53. doi: 10.1038/s41596-022-00728-0.

97) Zhou, Xiaoli, Xiaohua Qi, Hongxia Huang, and Honghui Zhu. 2019. "Sequence and Structural Analysis of AA9 and AA10 LPMOs: An Insight into the Basis of Substrate Specificity and Regioselectivity." *International Journal of Molecular Sciences* 20(18):4594. doi: 10.3390/ijms20184594.