ORIGINAL PAPER

Structure and dynamics analysis of multi-domain putative β-1,4-glucosidase of family 3 glycoside hydrolase (*Ps*GH3) from *Pseudopedobacter saltans*

Priyanka Nath[1] · Arun Goyal[1]

Abstract

Structure and conformational behaviour of a putative β-1,4-glucosidase of glycoside hydrolase family 3 (*Ps*GH3) from *Pseudopedobacter saltans* was predicted by using *in-silico* tools. *Ps*GH3 modeled structure constructed using Phyre2 displayed multidomain architecture comprising an N-terminal $(\beta/\alpha)_8$-fold domain followed by $(\alpha/\beta)_6$-sandwich domain, PA14 domain, and a C-terminal domain resembling an immunoglobulin fold. Ramachandran plot displayed 99.3% of amino acids in the allowed region and 0.7% residues in the disallowed region. Multiple sequence alignment (MSA) and structure superposition of *Ps*GH3 with other homologues from GH3 family revealed the conserved residues, Asp274 and Glu624 present in loops LA and LB, respectively originating from N-terminal domain act as catalytic residues. The volume and area calculated for *Ps*GH3 displayed a deep active-site conformation comparable with its homologues, β-1,4-glucosidases (GH3) of *Kluyveromyces marxianus* and *Streptomyces venezuelae*. Molecular dynamic (MD) simulation of *Ps*GH3 structure for 80 ns suggested stable and compact structure. Molecular docking studies revealed deeper active site conformation of *Ps*GH3 that could house larger cellooligosaccharides up to 7° of polymerization (DP7). The amino acid residues, Ala86, Leu88, Cys275, Pro483, Phe493, Asn417, Asn491, Pro492, and Leu495 created a binding pocket near the catalytic cleft, crucial for ligand binding. MD simulation of *Ps*GH3 in the presence of cellooligosaccharides, viz., cellobiose and celloheptaose showed stability in terms of RMSD, R_g, and SASA values till 80 ns. The calculation of average number of hydrogen bond (H-bond), interaction energy, and binding free energy confirmed the stronger binding affinity of the larger cellooligosaccharides such as celloheptaose in the binding cavity of *Ps*GH3.

Keywords Family 3 glycoside hydrolase · β-1,4-glucosidase · 3D modeling · Molecular docking · *Pseudopedobacter saltans*

Introduction

The extensive consumption of fossil fuels has caused the enhancement in the rate of global warming; therefore, the research and development in the field of renewable energy production have also increased. One of the forms of renewable energy is bioethanol. Bioethanol can be produced from plant-based lignocellulosic biomass. This lignocellulosic biomass is made up of cellulose, hemicellulose, and lignin. Cellulose is the most abundant polysaccharide present on earth and is composed of glucose units forming a parallel and consecutive linear chain [1]. Cellobiose, a disaccharide of glucose in the repeating unit, is present in cellulose [2]. The hydrolysis of cellobiose can be achived by β-1,4-glucosidase that gives two molecules of glucose [3]. The cellulose content in the plant biomass is more than 50%, and it is present in the crystalline fibre form which is resistant to hydrolysis [4]. The complete hydrolysis of cellulosic chain can be brought about by the combined action of endoglucanase, cellobiohydrolase, and β-glucosidase, collectively called as cellulases, converting cellulose polymer into cello-oligosaccharides, cellobiose, and glucose, respectively [5]. Cellulolytic enzymes belong to different glycoside hydrolase families (GH) (www.cazy.org). They act synergistically, in which endo-β-1,4-glucanase (EC 3.2.1.4) acts randomly on cellulosic chains and yields larger cellooligosaccharides as a hydrolysed products [6]. Cellobiohydrolase (EC 3.2.1.91) acts at the

✉ Arun Goyal
  arungoyl@iitg.ac.in

1   Carbohydrate Enzyme Biotechnology Laboratory, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India

non-reducing end of the larger cellooligosaccharides and generates cellobiose as a primary product after the hydrolysis [7]. Finally, β-1,4-glucosidase (EC 3.2.1.21) hydrolyzes the released cellobiose moiety to form two glucose molecules [1]. Several studies suggested that cellobiose formed during the hydrolysis of lignocellulosic material for bioethanol production inhibits the cellulases during the enzymatic hydrolysis process [8–10]. Thus, the instant removal of cellobiose becomes essential for increasing the efficiency of the process. The commercial cellulolytic enzymes mixture lacks the sufficient β-1,4-glucosidase activity. Therefore, efforts were made for the production of enzyme cocktails rich in β-1,4-glucosidase activity [11, 12]. The presence of adequate β-1,4-glucosidase activity in the enzyme cocktail displayed an increase in hydrolysis performance, resulting in a 20–40% increase in total bioethanol production [13]. Therefore, the identification, production, functional, and structural characterisation of an efficient β-1-4-glucosidase, from a bacterial source, *Thermoanaerobacterium aotearoense*, was carried out [14]. The enzyme, β-1,4-glucosidase belongs to GH families 1, 3, 5, 9, 30, and 116 (www.cazy.org) [15–17]. GH3 is one of the largest families in the CAZy database [18]. It contains around 44 of β-glucosidases from bacterial, mold, and yeast origin [19]. Moreover, the GH3 family is known to display β-1,4-glucosidase, β-xylosidase, and exo-1,3-1,4-β glucanase activities [18]. The most studied fungal β-1,4-glucosidase belongs to GH3 family [19]. However, the structural data of GH3 enzymes are still limited. Therefore, for understanding, the functional aspects of family GH3 enzymes, their structural characterisation is required. In this study, computational approaches were used for modeling of β-1,4-glucosidase of family GH3 (*Ps*GH3) from *Pseudopedobacter saltans* and understanding its structural characteristics. The stability and compactness of the developed modeled structure of *Ps*GH3 was analyzed by Molecular Dynamics (MD) simulation. Furthermore, the binding interaction of *Ps*GH3 with most probable ligands, the cellooligosaccharides, was studied by molecular docking approach for identification of crucial amino acid residues involved in substrate binding and catalysis. Furthermore, the cellooligosaccharides-bound *Ps*GH3 complexes obtained from molecular docking analysis were subjected to protein-ligand complex MD simulation for analyzing the conformation and stability of *Ps*GH3 in the presence of these ligands. Moreover, the affinity of cellooliogosaccharides towards *Ps*GH3 was also determined and compared.

Materials and methods

Sequence and evolutionary investigation of *Ps*GH3

The amino acid sequence of β-1,4-glucosidase (*Ps*GH3) belonging to family 3 GH from *P. saltans* having Genbank ID-ADY51046 and UniProt ID F0S5X5 was fetched from the NCBI database (http://www.ncbi.nlm.nih). The preserved domain within the amino acid sequence of *Ps*GH3 was explored by feeding the amino acid sequence of *Ps*GH3 in the conserved domain database (http://www.ncbi.nlm.nih.gov/cdd/). The signal peptide existing in the amino acid sequence of *Ps*GH3 was recognized by utilizing SignalP 4.1 server [20]. Furthermore, to determine the homologous sequences of *Ps*GH3 from the PDB database, the amino acid sequence of *Ps*GH3 was analyzed by using BLASTp tool [21]. The homologous sequences, thus obtained after BLASTp analysis, were further aligned by using Clustal Omega [22] and were envisioned by employing the web-based tool, ESPript 3 [23]. The evolutionary correlation between *Ps*GH3 and its homologues was studied by the phylogenetic tree created in Mega 7 software by using the neighbor-joining method [24].

3-dDimensional model prediction and validation of *Ps*GH3

The 3D model structure of *Ps*GH3 was predicted by using Phyre2 web server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). Phyre 2 is web-based server used for the prediction of protein structure and function [25]. This online program functions statistically for the template-based modeling of proteins. The modeled structure of *Ps*GH3 thus obtained was further energy minimized by using YASARA energy minimization tool (http://www.yasara.org/minimizationserver.php). The energy minimized modeled *Ps*GH3 structure was analyzed by PyMol 2.0 software [26] and further validated by SAVES web server (http://servicesn.mbi.ucla.edu/SAVES/). The stereo-chemical properties of the energy-minimized *Ps*GH3 structure were further explored by the PROCHECK [27]. An integrated web-based server, the ProSA (https://prosa.services.came.sbg.ac.at/prosa.php), was used to analyze the deviation of statistical Z-score for the modeled *Ps*GH3 structure from the high-resolution PDB deposited structures. The ERRAT plot [28] and VERIFY-3D [29] were further used to validate the energy-minimized modeled *Ps*GH3 structure to analyze the compatibility of the atomic model with the amino acid sequence. The illustration of the topology for *Ps*GH3 structure was generated by employing PDBSum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html). The conserved active site amino acid residues of *Ps*GH3 involved in catalysis were recognized by the superposition of the energy-minimized *Ps*GH3-modeled structure with its template structure of β-1,4-glucosidase from *Kluyveromyces marxianus* (PDB ID: 3ABZ) by using PyMOL 2.0 [26]. β-1,4-glucosidase from *Kluyveromyces marxianus* (PDB ID: 3ABZ) was employed for structure superposition based on structural comparison obtained from DALI server.

Molecular dynamics simulation of *Ps*GH3 modeled structure

The molecular dynamics simulation of modeled *Ps*GH3 structure was performed by employing Gromacs v 5.14 [30] equipped in a supercomputing facility (Param-Ishan), available at Indian Institute of Technology Guwahati, Assam, India. The GROMOS9653a6 FF force field [31] was utilized for the calculation of the protein forces. The *Ps*GH3 modeled structure was positioned inside a box with a cubic shape of volume 1405.23 nm^3 and dimensions 6.43 × 7.12 × 8.97. The system was then solvated with a simple point charges water model (SPC) to maintain the equilibrium. The MD simulation contained 42,563 water molecules, in which 6 Cl$^-$ counter ions were added to neutralize the charges present on *Ps*GH3. Solvated systems were then minimized at a maximum force of 1000.0 kJ/mol/nm by using 50,000 steps cut-off. Furthermore, the system was equilibrated in NVT (number of atoms, volume, and temperature of the system) ensemble at 300 K over 500 ps at constant volume and temperature, by coupling with a modified Berendsen thermostat with a time constant of 0.1 ps [30]. Additionally, another 500 ps NPT (normal pressure and temperature) equilibration was performed at constant temperature (300 K) and pressure (1 atm) by coupling with modified Berendsen thermostat with a time constant of 0.1 ps and Parrinello-Rahman barostat [32] with a time constant of 2 ps, respectively. Additionally, all bonds were constrained via LINCS algorithm [33]. Finally, the 80-ns production run was setup at a constant temperature of 300 K by using modified Berendsen thermostat and pressure 1 atm by using Parrinello-Rahman barostat with a time step of 2 fs, The LINCS algorithm was used to constrain the lengths of the hydrogen atom containing bonds. Additionally, the *Ps*GH3 modeled structure was analyzed throughout the process of simulation, as a time-dependent function to validate its stability in the solvent system. After 80 ns of simulation, the root mean square deviation (RMSD) and root-mean-square fluctuation (RMSF) of the simulated system was analyzed by using gmx rms and gmx rmsf programs, respectively. The radius of gyration (R_g), intramolecular hydrogen bonds (H-bond), and solvent accessible surface area (SASA) were determined by using the program gmx gyrate, gmx h-bond, and gmx sasa, respectively. The simulated structure after 80 ns was extracted and visualized by ChimeraX [34].

The active site investigation of *Ps*GH3

The existence of the accessible pockets for ligand binding on the surface of *Ps*GH3 was determined by CASTp, a web-based server (http://cast.engr.uic.edu) [35]. The area and volume of those accessible pockets for ligand binding present on the surface of *Ps*GH3 were calculated by employing a probe radius with a default value of 1.4 Å, by

using the molecular surface model (Connolly's surface) method and accessible surface model (Richards surface) as described previously [36]. Additionally, the dissemination of electrostatic potential over the substrate binding pocket and surface of *Ps*GH3 structure was calculated by using a tool known as Adaptive Poisson-Boltzmann Solver (APBS) equipped PyMOL 2.0 software [26]. The structures were then further envisaged and examined by PyMOL 2.0 software [26].

Ligand-binding analysis of *Ps*GH3 by molecular docking approach

The molecular docking of *Ps*GH3 with the most probable ligands (cellooligosaccharides) was executed in a web-based server, SwissDock, (http://www.swissdock.ch/docking). In the initial step, the *Ps*GH3 structure was prepared for docking, by employing the Dock Prep tool available in UCSF Chimera 1.10.1 [37]. In this step, the hydrogen atoms were added, and partial charges were allocated by using AMBER99 FF force field. Moreover, the cellooligosaccharides were designed and downloaded from the GLYCAM server [38]. The docking analysis of *Ps*GH3 modeled structure with cellooligosaccharides in the SwissDock was executed by submitting the protein in PDB and ligand in Mol2 file format in the server. The SwissDock utilizes the EADock DSS software in which after accomplishment of each docking run, the output clusters are attained and ranked by a fixed SwissDock algorithm [39, 40]. This algorithm ranks the cluster based on scoring function known as FullFitness (FF). Moreover, the distinct conformer obtained from each cluster was further arranged according to the FF score. A more favorable binding mode has a higher negative FF score and hence, a better fit. The UCSF-Chimera 1.10.1 was used for analyzing the results obtained from SwissDock [37]. Furthermore, different docking poses were screened based on binding free energy value (deltaG, kcal/mol). The best docked protein-ligand complex displaying the highest negative binding free energy was considered and was further selected. The selected ligand-bound protein structure of *Ps*GH3 was extracted using ChimeraX [34] and envisioned in PyMol 2.0 [26]. The illustration of amino acid residues interacting with ligand in the protein-ligand complex was created by using the Ligplot software [41].

Molecular dynamics of protein-ligand complexes

The two docked complexes (*Ps*GH3 + cellobiose and *Ps*GH3 + celloheptaose) with the highest negative binding free energy obtained from the molecular docking study were further subjected to MD simulation to understand the stability and conformational changes of the protein in the presence of ligands. The MD simulations were performed by using GROMACSv 5.14 [30], while the protein forces were

calculated by GROMOS96 43A1 FF force field [31]. The ligand topologies for cellobiose and celloheptaose were generated by using PRODRG server [42]. The two MD simulations systems were solvated with a simple point charges water model (SPC) with a dodecahedron box configuration with a volume 555 nm^3. The MD simulation system, *Ps*GH3 + cellobiose contained 33,671 water molecules, in which 8 Cl$^-$ counter ions were added to neutralize the charges while *Ps*GH3 + celloheptaose contained 33,647 water molecules and 11 Cl$^-$ counter ions. The MD systems were further subjected to energy minimization with the steepest descent algorithm for the removal of any steric clashes and bad contacts and at maximum force of 1000.0 kJ/mol/nm by using 50,000 steps cutoff. After energy minimization, the equilibration with position restraint was carried out under NVT at 300 K over 500 ps at constant volume and temperature, by coupling with a modified Berendsen thermostat with a time constant of 0.1 ps. Additionally, another 500 ps NPT equilibration was performed by using modified Berendsen thermostat coupling (300 K) and isotropic Brendsen pressure coupling (1 atm) [30]. For both equilibration steps, the LINCS algorithm was used to constrain the lengths of the hydrogen-containing bonds [33]. Finally, the 80- ns production run at a constant temperature of 300 K by coupling with a modified Berendsen thermostat and pressure 1 atm with a time step of 2 fs employing Parrinello-Rahman for isotropic pressure coupling [32] using LINCS constraints for H-bonds were carried out in a supercomputing facility described in the previous section (Molecular dynamics simulation of *Ps*GH3 modeled structure). Thereafter, the stability of the protein-ligand complexes during the course of simulation was analysedanalyzed by using various parameters by employing in-built "gmx" commands of GROMACS. The plots for root-mean-square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), number of hydrogen bonds (H-bond), and solvent accessible surface area (SASA) of protein-ligand complexes upon simulation were generated using GraphPad prism 9.0.0. Moreover, the average energies comprising of short-range interactions between protein and the ligand represented in the terms of Lennard Jones and coulombic interaction energy were also analyzed using gmx energy command. Post MD simulations, the protein-ligand complexes were further subjected to the total binding free energy, potential energy (electrostatic and Van der Waals interactions), and solvation free energy (polar and non-polar solvation energies) calculations by using the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method by using g_mmpba tool [43].

Calculation of binding free energy using Poisson-Bolzmann surface area (MM-PBSA)

The Poisson-Boltzmann or generalized Born and surface area continuum solvation (MM-PBSA or MM-GBSA) is a method for the calculation of the ligand affinity for proteins [44]. The binding free energy of the protein-ligand complexes (*Ps*GH3 + cellobiose and *Ps*GH3 + celloheptaose), was calculated from their last stable 20 ns trajectories. The tool "g_mmpbsa" with default parameters [43] was utilized for potential energy and solvation free energy calculations.

Results and discussion

Sequence and evolutionary investigation of *Ps*GH3

The blastp analysis of amino acid sequence of *Ps*GH3 having GenBank ID-ADY51046 from *Pseudopedobacter saltans* exhibited the sequence identity for amino acid of *Ps*GH3 with earlier biochemically characterized β-1,4-xylosidase, viz., Xyl3A (38.9%) from *Trichoderma reesei* and XlnD (37.8%) from *Aspergillis nidulans,* respectively (Table 1). However, *Ps*GH3 also exhibited the sequence identity with earlier biochemically characterized β-1,4-glucosidase,viz., *Em*GH1 (34.4%) from *Erythrobacter marinus*, *Km*BglI (33.5%) from *Kluyveromyces marxianus*, DesR (33.2%) from *Streptomyce venezuelae*, and *Bl*BG3 (30.9%) from *Bifidobacterium longum*, respectively (Table 1). Moreover, the SignalP 4.1 server inferred the presence of an N-terminal signal peptide sequence comprising 20 amino acids with a cleavage site between Ala19 and Gln20. The sequence analysis of *Ps*GH3 by BLAST and conserved domain database displayed that the N-terminal domain of *Ps*GH3 comprises 375 amino acids for β-glucosidase (BG1X) and 434–747 amino acid for Glyco Hydro 3_C superfamily domain (GH3-C domain) that also contained an inserted PA14 domain between amino acids, 460–581 (Fig. 1a). The PA14 domain was named after its location that was identified in the protective antigen of anthrax toxin [45]. The PA14 domain has a binding function rather than a catalytic role though it is shown to be located in the catalytic core [18]. The rest of the amino acids 747–854 constituted the C-terminal domain. The amino acid residues position 20–62 and 375–434 were not assigned in the conserved domains search for *Ps*GH3 sequence; therefore, they were considered as unknown (Fig. 1a). This distinct domain architecture was also reported for β-1,4-glucosidase (GH3) structure from *K. marxianus* [18].

The phylogenetic tree analysis gave the relatedness of *Ps*GH3 with other homologous proteins from different organisms. The *Ps*GH3 sequence was closely related to an exo- β-1-4-xylosidase of GH3 family from *A. nidulans*, whereas the GH3 enzyme from *Bacteroides intestinalis* was found distantly related to *Ps*GH3 (Fig. 1b). Furthermore, the alignment of sequence of *Ps*GH3 with *T. reesei* (PDB ID-5A7M), *A. nidulans* (PDB ID-6Q7I), *B. longum* (PDB ID-5Z9S), *S. venezuelae* (PDB ID- 4I3G), *E. marinus* (PDB ID-5Z87), and *K. marxianus* (PDB ID-3 AC0) by using MSA, suggested that Asp274 and Glu624 are the catalytic amino acid residues,

Table 1 BLASTp investigation of *Ps*GH3 amino acid residue sequence against PDB database

Name of organism	Name of enzyme	Query cover (%)	Identity (%)	E-value	Total score	Resolution (Å)	PDB ID
T. reesei	Xyl3A	88	38.9	8e-44	158	1.8	5A7M
A. nidulans	XlnD	89	37.8	1e-43	158	1.4	6Q7I
B. longum	BlBG3	91	30.9	1e-27	112	2.3	5Z9S
S. venezuelae	DesR	87	33.2	2e-27	112	1.4	4I3G
E. marinus	Em GH1	80	34.4	6e-27	110	2.3	5Z87
K. marxianus	KmBglI	84	33.5	2e-26	108	2.5	3 AC0

while the residues Lys193, His194, and Tyr242 might be involved in binding and were found conserved in family 3 GH proteins (Fig. 2).

## 3-Dimensional structure modeling of *Ps*GH3 and structural analysis

The 3-dimensional modeled structure of *Ps*GH3 protein was created by Phyre2 by employing multiple templates, of which the β-1,4-xylosidase (GH3) structure of *Hypocrea jecorina* showed 100% confidence with a percentage identity of 40%, while β-1,4-glucosidase (GH3) structures of *K. marxianus* (PDB ID: 3 AC0) and *S. venezuelae* (PDB ID-4I3G) showed 100% confidence with percentage identity of 38% and 34%, respectively. The *Ps*GH3 modeled structure generated was energy minimized by the energy minimization tool, YASARA. The energy of *Ps*GH3 modeled structure, initially was 1.5×10^5 kJmol^{-1}, which after minimization, showed lower energy, $- 4.96 \times 10^5$ kJmol^{-1}. Thus, the energy-minimized *Ps*GH3 structure was obtained from the server and used for further structure characterization. The modeled structure of *Ps*GH3 was composed of four distinct domains,
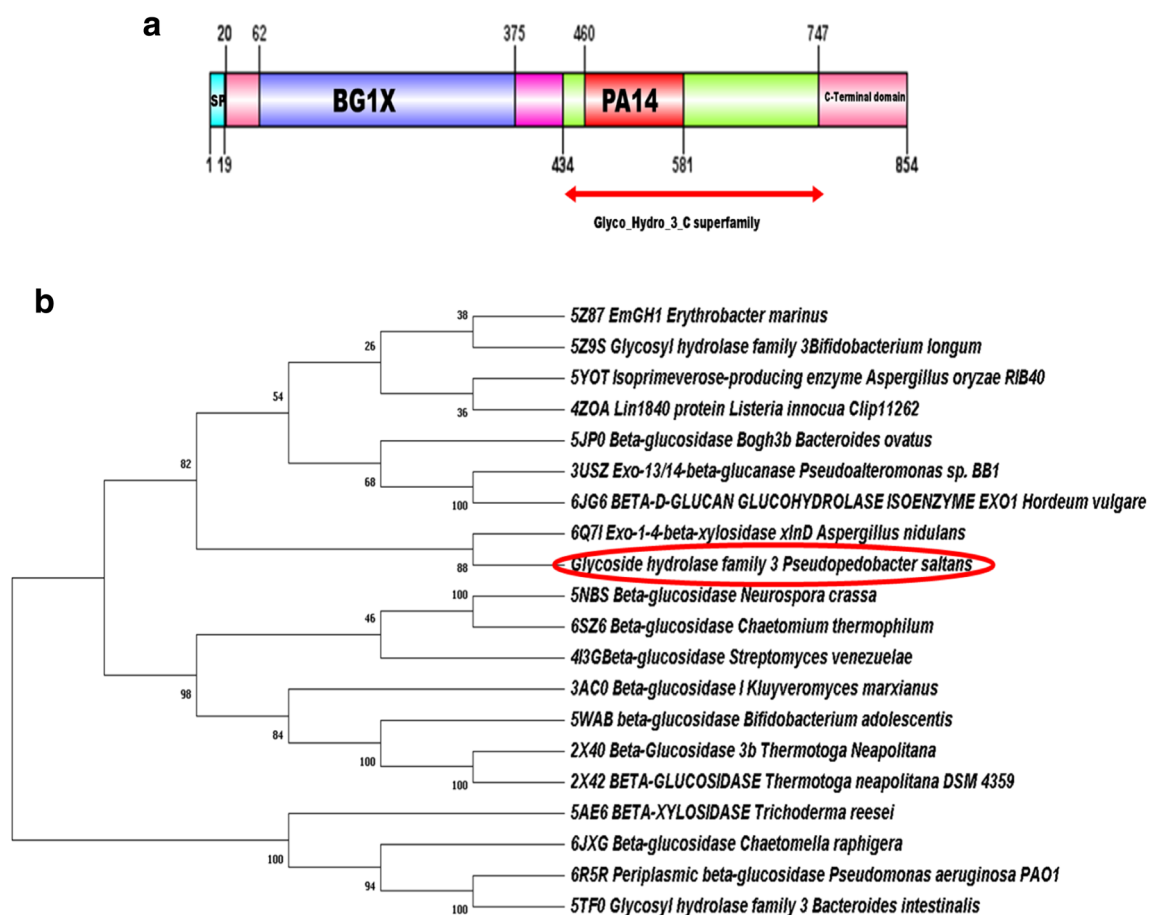
Fig. 1 a Molecular architecture of *Ps*GH3 amino acid sequence. **b** Phylogenetic analysis for elucidating the evolutionary correlation between *Ps*GH3 and its homologues by utilizing the neighbor-joining method by using MEGA 7
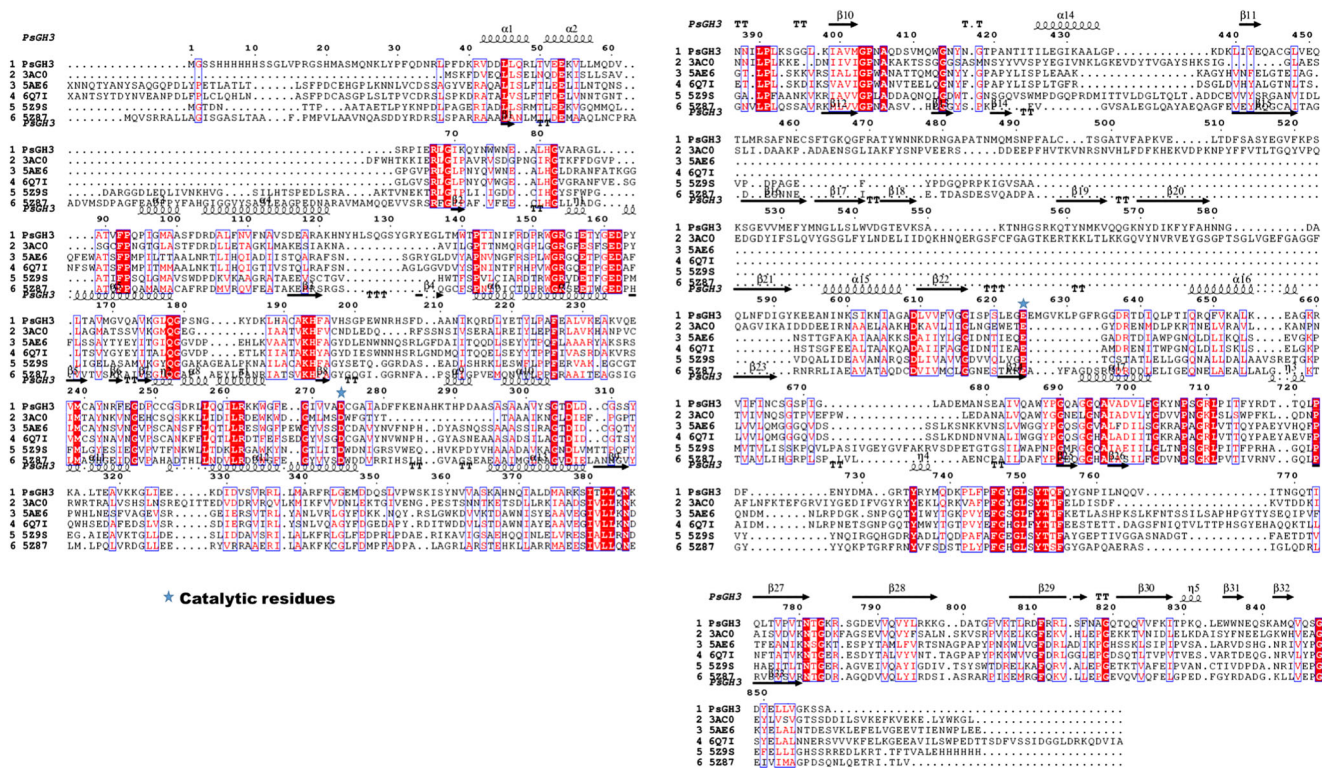
Fig. 2 Multiple sequence alignment of *Ps*GH3 with its homologous proteins, viz., Xyl3A (PDB ID-5AE6), XlnD (PDB-ID 6Q7I), *Em*GH1 (PDB ID 5Z87), *Km*BglI (PDB ID 3 AC0), , and *Bl*BG3 (PDB ID-5Z9S)

as also described in Table 1. In MSA, the conserved amino acid residues are shown in black with the red background and semi-conserved amino acid residues are depicted in red with white background

viz., an N-terminal $(\beta/\alpha)_8$-fold-like domain 1 (grey) comprising 357 amino acids from (21–378), an N- terminal $(\alpha/\beta)_6$-sandwich domain 2 (yellow) comprising 341 amino acids from (382–723), the domain 1 and domain 2 were connected by a short linker C1, comprising amino acids Ser 379, Ile380, and Thr 381 (Fig. 3a). In a previous study, it was reported that the two N-terminal domains of β-1,4-glucosidase (GH3) from *Hypocrea jecorina* were connected by a 16 residue linker [46]. In domain 2 of *Ps*GH3, a PA14 i.e. domain 3 (blue) was found to be inserted between the α helix D2α4 and D2α6 by linkers LC and LD, respectively (Fig. 3a). The domain 3 (PA14) comprising 133 amino acids from (459–592) have a jelly-roll fold with 10 antiparallel β-strands forming a two-layered β-sheet (Fig. 3a). In an earlier study, a similar fold of PA14 domain was reported to be inserted between a β-strand and an α-helix of domain 2 in the β-glucosidase (GH3) structure of *K. marxianus* [18]. The N-terminal domain was followed by a C-terminal domain 4 (green) (Fig. 3a). The C-terminal domain 4 comprised 124 amino acids, (735–859). The C-terminal domain 4 was linked to domain 2 through a linker C2 comprising, 12 amino acids (722–734) (Fig. 3a). The domain 4, forms 9 β-strands resembling an immunoglobin fold. A similar C-terminal domain exhibiting immunoglobin fold was reported earlier in β-1,4-glucosidase *Km*BglI and DesR structures of *K. marxianus* [18] and *S. venezuelae* [47], respectively. The

overall topology diagram of multi-domain *Ps*GH3 is shown in Fig. 3c. The overall fold of *Ps*GH3 strongly resembles the *Km*BglI, β-1,4-glucosidase (GH3) structure from *K. marxianus* [18]. The structure comparison using DALI server also demonstrated that *Ps*GH3 is closely related to β-1,4-glucosidase (GH3) structures from *K. marxianus* (PDB ID: 3 AC0) and *S. venezuelae* (PDB ID- 4I3G) with RMSD values of 0.3 Å and 1.1 Å, respectively. The *Ps*GH3 active site yields a deep pocket conformation (Fig. 3b). A similar pattern of the active site was previously reported in β-1,4-glucosidase (GH3) structure from *K. marxianus* [18] and *H. jecorina* [46]. In *Ps*GH3, near the active-site, the loops LA and LB, originating from domains 1 and 2, respectively are present (Fig. 3a, b). The loop LA comprises 14 amino acids (274–288), and LB is comparatively longer, comprising 22 amino acids (624–646). The longer active-site loop LB, where the catalytic acid/base residue Glu624 is present, causes the formation of a deeper active-site pocket in *Ps*GH3 structure. Similarly, the deeper conformation of active site was reported due to longer and structurally flexible loops near the active site in β-1-4-glucosidase (GH3) structure of *H. jecorina* [46] and *B. longum* [48]. The superposition of *Ps*GH3 structure with its homologue, β-1,4-glucosidase (PDB ID-3 AC0) structure from *K. marxianus* revealed that the key active-site residues are spatially aligned and found to be conserved (Fig. 3d, e).

Fig. 3 3-Dimensional modeled structure analysis of *Ps*GH3. **a** Cartoon view showing the different domains. **b** Surface view highlighting the catalytic site for binding. **c** Illustration topology diagram of *Ps*GH3 modeled structure demonstrating the positioning of the secondary structures (cylinders signify α-helix and thick arrows signify the β-strand). **d** Superposition of *Ps*GH3 structure (Cyan) with homologous GH3 structure (PDB ID: 3 AC0) from *K. marxianus* (Green). **e** The superposed binding cavity of *Ps*GH3 structure with homologous GH3 structure (PDB ID: 3 AC0) from *K. marxianus*, the conserved active site shown in red, while subsite residues are shown in blue for *Ps*GH3 and **f** distance between acid/base and nucleophile residue
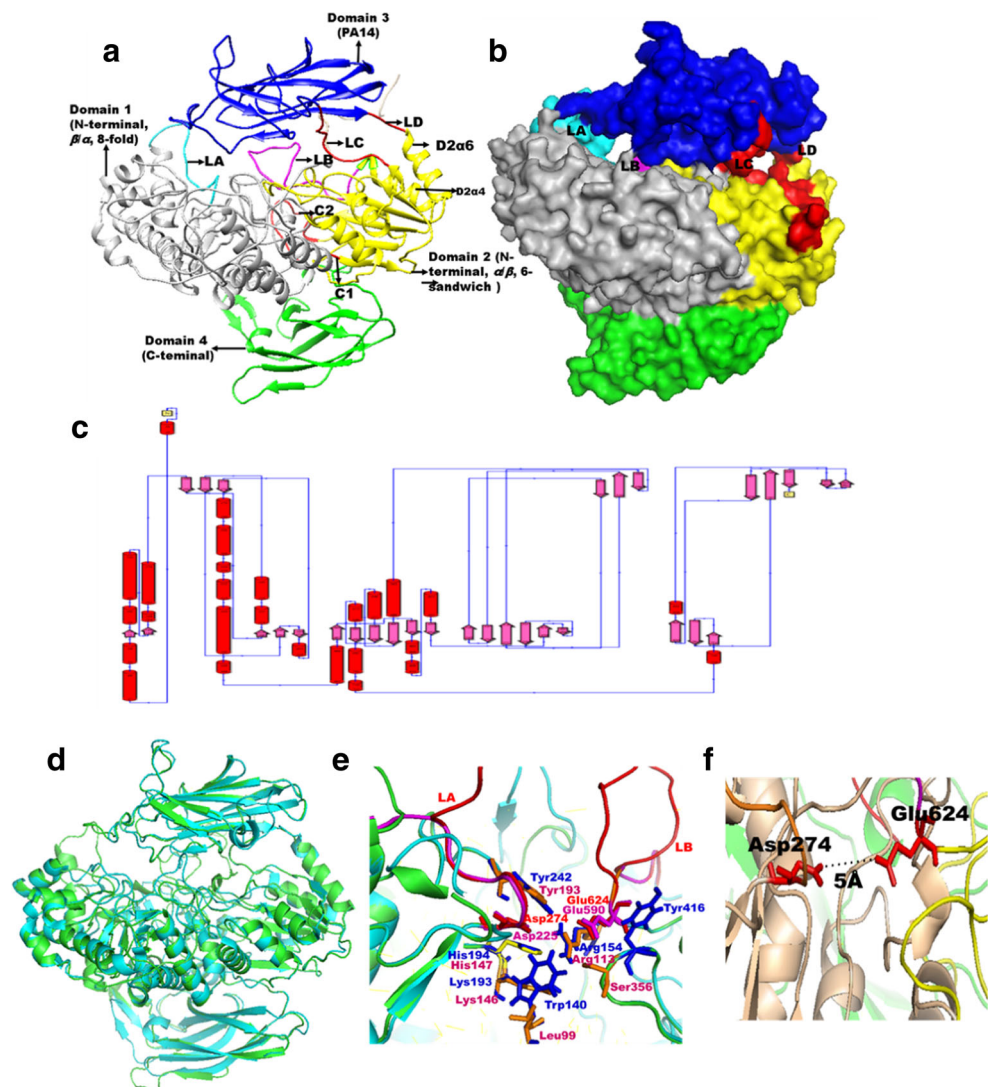
Asp274 and Glu624 are the nucleophile and acid/base catalytic amino acid residues, respectively (Fig. 3e), which are also conserved in family 3 GH proteins. The hydrolytic mechanism for catalysis (Retaining or Inverting) can be speculated by measuring the distance between the carboxyl group of catalytic residues through in silico process as described earlier [49]. In an earlier study, these mechanisms for hydrolysis were experimentally determined, which suggested that the average distances for inverting and retaining mechanism should be 10.5 Å (± 2 Å) and 5.5 Å, respectively [50]. The distance between the carboxyl group Asp274 (nucleophile) and Glu624 (acid/base) for *Ps*GH3 determined by in silico approach was 5 Å (Fig. 3f). These results suggested the retaining type of hydrolytic mechanism for *Ps*GH3. Additionally, the residues Lys193, His194, and Tyr242, in the substrate binding site were found highly conserved between *Ps*GH3 and its homologous protein structure of β-1,4-glucosidase (PDB ID- 3 AC0) from *K. marxianus* (Fig.

3e). These residues were found conserved also in MSA analysis as described earlier in the previous section.

Quality assessment of *Ps*GH3

The energy minimized modeled *Ps*GH3 structure was validated by the Ramachandran plot through PROCHECK server. The modeled *Ps*GH3 structure showed 83.1% of amino acid residues in the most favored region, 14.1% in the additionally allowed region, 2.1% in the generously allowed region, and only 0.7% amino acid residues (Asp186, Ala445, and Asp589) in the disallowed region, out of non-proline and non-glycine residues (Fig. 4a). This result indicated that the *Ps*GH3 modeled structure follows the backbone phi (φ) and psi (ψ) dihedral angles of the Ramachandran plot. ProSA results also showed that the *Ps*GH3 modeled structure is accurate having a *Z*-score of − 8.7 and belongs to the X-ray zone (Fig. 4b). VERIFY 3D result showed that 80.7% amino acid

residues of the modeled *Ps*GH3 structure have 3D-1D score ≥ 0.2 indicating the compatibility of amino acid residues with the modeled structure (Fig. 4c). The ERRAT plot of modeled *Ps*GH3 structure suggested an overall quality factor of 82.8% establishing that the structure has an acceptable quality (Fig. 4d), since a good 3D model should have a quality factor > 50% [51].

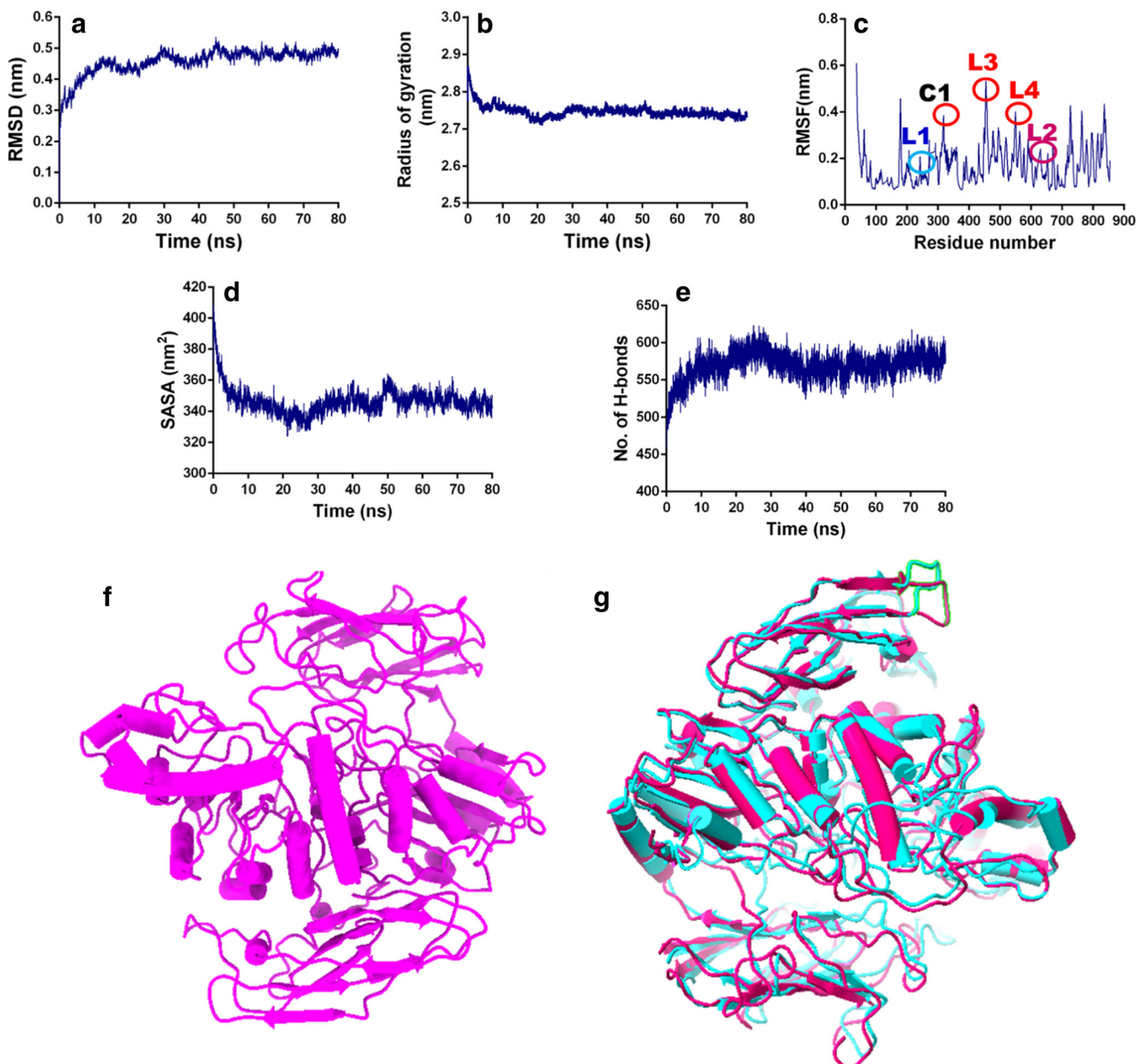## Molecular dynamics simulation of *Ps*GH3

Molecular dynamics simulations of the modeled structure of *Ps*GH3 was performed to analyze its conformational behaviour, stability, as well as compactness at the nanosecond level. The values for RMSD for the complete trajectory were calculated by gmx rms program to analyze the deviation from the primary structure. The RMSD profile depicted that at 50 ns, the RMSD stretched a plateau phase of value 0.5 nm and continued to be stable up to 80 ns with less oscillations of ~ 0.01 nm (Fig. 5a). These less oscillations and stable RMSD depicted that the *Ps*GH3 structure achieved stable conformation during MD simulation. The estimation of the overall compactness of *Ps*GH3 throughout the MD simulation was assessed by utilizing the gmx gyrate program. These analyses indicated the fluctuation of radius of gyration (R_g), between 2.75 and 2.89 nm till 30 ns due to the flexibility; however, after 30 ns, the *Ps*GH3 attained stability as well as

compactness at 2.7 nm and continued to be stable till 80 ns (Fig. 5b). These observations demonstrated a stable conformation of the *Ps*GH3 modeled structure. The structural variations of *Ps*GH3 after MD simulation were analyzed by root mean square fluctuation (RMSF) by using the gmx rms program. RMSF computes the movement of a specific atom, or group of atoms, comparative to the reference structure used in a MD simulation. The higher values of RMSF were observed for the flexible amino acid residues with maximum α-carbon flexibility between all the residues in a protein molecule. However, lower values of RMSF were observed for the rigid amino acid residues with the lowest flexibility. In *Ps*GH3, the starting amino acid residues of N-terminal and the end residues of C-terminal displayed higher flexibility (Fig. 5c). The higher fluctuation in RMSF was also found evident among residues, 379–391(Fig. 5c). In this particular region, the linker (C1) was involved in connecting the domain 1 and domain 2 of *Ps*GH3 as described in the previous section (3D modeling of *Ps*GH3 and structural analysis). This higher fluctuation in RMSF for the region demonstrated the flexibility of the linker connecting the two domains of *Ps*GH3. In an earlier study, the flexibility of the linker connecting two domains was described in cellulase (Cel45) from *Humicola insolens* [52]. Similarly, the loops L3 (435–458) and L4 (592–594), where domain 3 (PA14) was inserted in domain 2, showed comparatively higher fluctuation in RMSF (Fig. 5c) suggesting flexibility
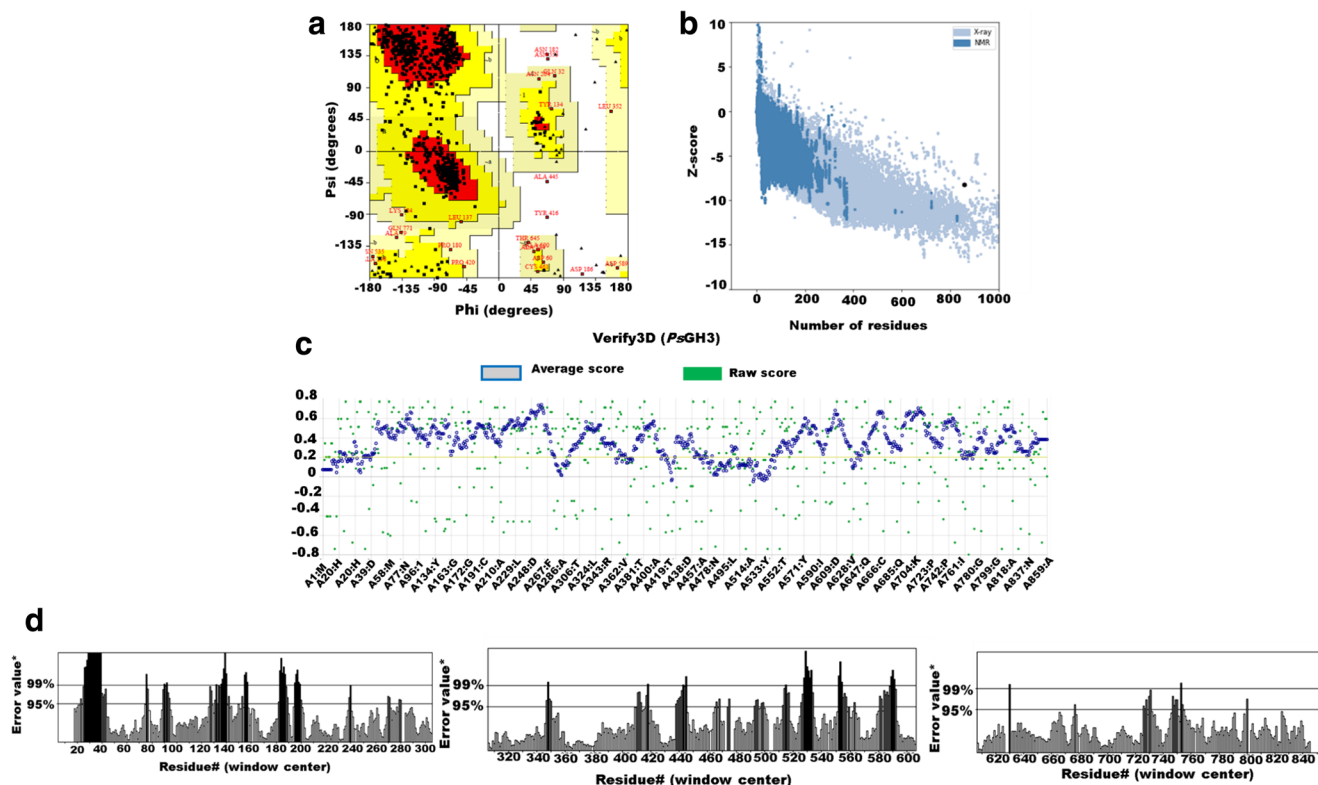
Fig. 4 The structure validation and exploration of 3D modeled structure *Ps*GH3. **a** The Ramachandran plot examination by employing PROCHECK in web-based UCLA server, SAVES. **b** ProSA plot displaying Z-score. **c** VERIFY 3D. **d** ERRAT plot

Fig. 5 Molecular dynamics simulation of *Ps*GH3 modeled structure. **a** RMSD plot. **b** Radius of gyration plot. **c** RMSF plot. **d** SASA plot and **e** h-bond plot, **f** *Ps*GH3 simulated structure (magenta). **g** Structure superposition of simulated *Ps*GH3 (magenta) with modeled *Ps*GH3 structure (cyan)

of these loops. In an earlier study, the flexibility of loops extending between two domains was reported in a multidomain β-1,4-glucosidase 3B from *Thermotoga Neapolitana* [53]. The amino acid residues of *Ps*GH3 residing within the active site core formed between the interface of domain 1 and domain 2 and existing in loops L1 (274–289) and L2 (624–634), showed less fluctuation (Fig. 5c). A previous study also proposed that the catalytic core is the principle factor for enzyme activity and stability [54]. Therefore, the lower fluctuation in the catalytic core suggested stability of the *Ps*GH3 for catalytic function. The gmx sas programme was used to calculate the solvent accessible surface area.

The average SASA of the *Ps*GH3 calculated was 350 nm^2 and it remained stable till 80 ns (Fig. 5d). This also proposed that the global contact of *Ps*GH3 towards the solvent area remained unaffected. Similarly, this also proposed that the availability of the substrate remained unaffected in the catalytic core of *Ps*GH3. Similar observation for SASA was made for the catalytic cores in modular chimeric enzyme [55]. The gmx hbond program was used to estimate the intramolecular hydrogen bond analysis of *Ps*GH3. The analysis showed that 581 average intramolecular hydrogen bond of *Ps*GH3 were formed till 80 ns (Fig. 5e). The hydrogen bonds thus formed further aided the *Ps*GH3 structure in attaining its stable state.

However, the superposition of *Ps*GH3 initial structure and MD simulated *Ps*GH3 structure using ChimeraX (Fig. 5f) displayed the deviation in the loop areas (Fig. 5g) and gave a RMSD of 0.2 Å. The MD simulated *Ps*GH3structure was found stable and thus utilized for further analysis.

Active site examination of *Ps*GH3 structure

The active site pocket of *Ps*GH3 was examined by computing the area and volume by CASTp server and was compared with its homologues, viz., β-1,4-glucosidase (GH3) structure from *K. marxianus* (PDB ID: 3 AC0) and *S. venezuelae* (PDB ID-4I3G). The computed active-site solvent accessible mouth surface area and volume of *Ps*GH3 were 3810.2 Å^2 and 5352.1 Å^3 (Fig. 6a), respectively. A similar analysis performed for the homologous β-1,4-glucosidase structures of 3 AC0 and 4I3G revealed the computed active site solvent accessible mouth surface area of 3899.8 Å^2 and 2461.3 Å^2, respectively (Fig. 6d, g) and the volume 4820 Å^3 and 2489.9 Å^3, respectively (Fig. 6d, g). These analyses confirmed that the active site structure of *Ps*GH3 possesses deeper conformation similar to its homologous β-1,4-glucosidase (GH3) structures from *K. marxianus* [18] and *S. venezuelae* [47]. The charge distribution for *Ps*GH3 and its homologous β-glucosidase (GH3) structures 3 AC0 and 4I3G structures displayed that their active site is a mixture of positively and negatively charged amino acid residues (Fig. 6b, e, h). These patterns of mix charges in active site were demonstrated for GH3 β-1,4-glucosidase from the thermophilic fungus *Chaetomium thermophilum* and was found beneficial for structural stability [56]. Furthermore, the active site analysis of *Ps*GH3 showed the presence of some conserved (Arg154, Lys192, His193, and Tyr242) and semi-conserved (Trp140 and Tyr416) amino acid residues which along with active site residues Asp274 and Glu624 might be important in substrate binding and catalysis (Fig. 6c). Correspondingly, in the active site analysis of β-1,4-glucosidase (GH3) structure of *K. marxianus* (PDB ID: 3 AC0) conserved residues, such as Arg113, Lys146, His147, and Tyr193 (Fig. 6f) are present for substrate binding along with active site residues Asp225 and Glu590 as described earlier [18]. Moreover, in the active site analysis of β-1,4-glucosidase (GH3) structure of *S. venezuelae* (PDB ID: 4I3G) along with conserved active site residues, Asp273 and Glu579, the amino acid residues, such as Gly99, Pro150, His196 and Ala 198 (Fig. 6i) are present, which were earlier described to be involved in substrate binding [47].

Ligand-binding analysis of *Ps*GH3 by molecular docking approach

The molecular docking of *Ps*GH3 with different cellooligosaccharides was performed for analysis of different interactions involved between the ligands (cellooligosaccharides) and the amino acid residue present in the catalytic site of *Ps*GH3. The docking analysis performed on the server SwissDock displayed more than 24 probable docking resolutions for *Ps*GH3. The docking results of *Ps*GH3 with cellooligosaccharides displayed information about different amino acid residues participating in protein-ligand interactions (Table 2). *Ps*GH3 exhibited binding with various cellooligosaccharides, viz., glucose, cellobiose, cellotriose, cellotetraose, cellopentaose, cellohexaose, and celloheptaose with free energy of binding (ΔG) of − 7.26, − 7.51, − 9.15, − 9.51, − 10.74, − 11.62, and − 14.04 kcal/mol, respectively (Table 2). The molecular docking analysis of xylooligosaccharides with catalytic clefts of the *Ps*GH3 was also performed but no significant changes were observed in binding free energy upon docking (data not shown). Therefore, the binding interaction studies of *Ps*GH3 with most probable ligands, i.e., cellooligosaccharides was continued. The binding studies of cellooligosaccharides with *Ps*GH3 revealed that the catalytic cleft of *Ps*GH3 efficiently binds glucose and cellobiose and can also house cellooligosaccharides up to 7 degree of polymerization (DP7) (Table 2). The possible accommodation of larger cellooligosaccharides in the catalytic cleft of *Ps*GH3 could be because of the deep and large size of the catalytic pocket as described in the earlier section of active site examination of *Ps*GH3 structure. Similarly, in an earlier study, it was demonstrated that a β-1,4-glucosidase from family 3 glycoside hydrolase (*Mt*Bgl3b) of *Myceliophthore thermophile* could accommodate large cellooligosaccharides such as cellotetraose in its catalytic cleft due to its large catalytic pocket [57]. The fitting of glucose, cellobiose and celloheptaose inside the catalytic cleft of the modeled *Ps*GH3 is shown in Fig. 7a, d g.

The docking interaction of *Ps*GH3 showed that the amino acid residues, Gly87, Gly418, Thr419, and Met489 are involved in polar interactions or building hydrogen bonds with the ligand glucose, while, Glu78, Ala84, Ala86, Tyr416, Asn417, Pro420, and Pro492 are making hydrophobic non-polar interactions with glucose (Table 2, Fig. 7b, c). The molecular docking interaction of *Ps*GH3 with cellobiose displayed the involvement of amino acid residues Gln25, Lys27, Pro420, Thr485, Met489 in hydrogen bond formation or polar interaction, while, Thr419, Ala421, Pro492 were involved in hydrophobic non-polar interactions (Table 2, Fig. 7e, f). The docking interaction of *Ps*GH3 with celloheptaose revealed amino acid residues Glu25, Ala84, Glu78, Gly87, Tyr416, Gly418, Thr419, Thr485, Arg479, Met489, and Ala494, involved in hydrogen bond formation or polar interactions, while, Ala86, Leu88, Cys275, Pro483, Phe493, Asn417, Asn491, Pro492, and Leu495 are involved in hydrophobic non-polar interactions (Table 2, Fig. 7h, i). Thus, from this docking analysis, the deduced amino acid residues found near the catalytic cleft of *Ps*GH3 could be involved in binding the ligands and therefore crucial for catalysis.
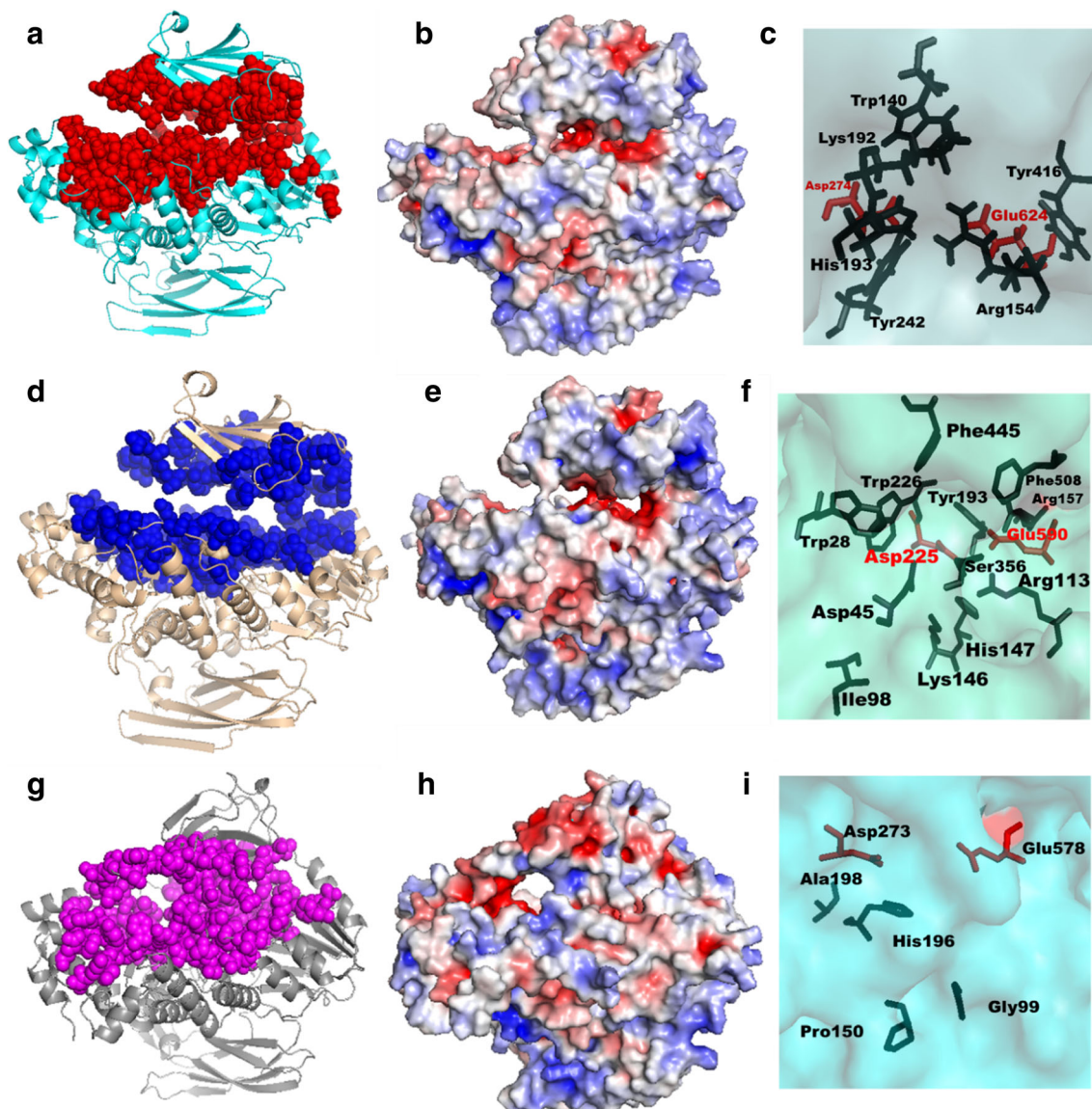
Fig. 6 Active-site pocket analysis of *Ps*GH3 and its closest homologues, viz., *Km*BglI (PDB ID 3 AC0) from *K. marxianus* and DesR (PDB ID-4I3G). Analysis of active-site volume by employing the CASTp web-based server. **a** *Ps*GH3. **d** 3 AC0. **g** 4I3G. The calculation of electrostatic potential distribution for **b** *Ps*GH3, **e** 3 AC0, **h** 4I3G, and the active-site organization of **c** *Ps*GH3, **f** 3 AC0, and **i** 4I3G

Molecular dynamics of protein-ligand complexes

After the MD simulation of protein-ligand complexes (*Ps*GH3 + cellobiose and *Ps*GH3 + celloheptaose), the stability and conformational changes of the protein-ligand complexes were compared with the only protein (*Ps*GH3) without the ligands by analysing the parameters, RMSD, RMSF, R_g, H-bond, and SASA. The Fig. 8 depicts that all MD simulation systems including the only *Ps*GH3 as well as the protein-ligand complexes attained stability after 20 ns and remained stable till 80 ns (Fig. 8a). The RMSD plot of *Ps*GH3 + cellobiose and *Ps*GH3 + celloheptaose after attaining stability at 20 ns remained stable till 80 ns, with an average RMSD of 0.56 nm and 0.72 nm, respectively (Fig. 8a). Similarly, the RMSD plot of

only *Ps*GH3 also showed stability at 80 ns with an average RMSD of 0.45 nm (Fig. 8a). A similar pattern of RMSD was obtained for native as well as the protein-ligand complexes of peptide deformylase from *Xanthomonas oryzae* upon MD simulation [58]. Therefore, from the RMSD plot, it was evident that the protein as well as the protein-ligand complexes reached equilibrium at 80 ns and thus could be considered for further analysis. The structural compactness of only the protein and the protein-ligand complexes during MD simulation were analyzed by Radius of gyration (R_g) as shown in Fig. 8b. The fluctuation of R_g values for both only *Ps*GH3 and *Ps*GH3 + cellobiose were in the between 2.75 and 2.89 nm till 30 ns; however, after 30 ns, both the systems attained stability as well as compactness at 2.7 nm and continued to be stable till 80 ns. The fluctuation in

Table 2 Molecular docking analysis of *Ps*GH3 with cellooligosaccharides

| Cello-oligosacharides	Free energy of binding (kcal/mol)	Amino acid involved in hydrogen bond formation	Amino acid involved in hydrophobic bond formation
Glucose	− 7.26	Gly87, Gly418, Thr419, Met489	Glu78, Ala84, Ala86, Tyr416, Asn417, Pro420, Pro492
Cellobiose	− 7.51	Gln25, Lys27, Pro420, Thr485, Met489	Thr419, Ala421, Pro492
Cellotriose	− 9.15	Glu78, Gly87 Tyr416, Met489	Gln25, Ala84, Arg85, Leu88, Ala86, Asn417, Thr419, Pro420, Pro492
Cellotetraose	− 9.51	Ala86, Thr419, Pro492, Met489	Gly87, Leu88, Ala421, Pro483, Ala484,Thr485, Phe493
Cellopentaose	− 10.74	Asp60, Cys309, Asn491, Asp583	Trp75, Trp76, Glu78, Cys275, Tyr416, Asn417,Val507, Leu509, Pro492 Gly582, Met626
Cellohexaose	− 11.62	Ala84, Arg85, Ala86, Gly87, Thr419, Arg479, Ala482, Thr485, Ala494, Met489, Pro492	Trp75, Leu88, Pro420, Ser490, Phe493
Celloheptaose	− 14.04	Glu25, Ala84, Glu78, Gly87, Tyr416, Gly418, Thr419, Thr485, Arg479, Met489, and Ala494	Ala86, Leu88, Cys275, Pro483, Phe493, Asn417, Leu495, Pro492, and Asn491

the R_g values (2.87–2.77 nm) for *Ps*GH3 + celloheptaose was observed till 40 ns, while after 40 ns, it achieved stability and compactness at 2.75 nm and remained stable till 80 ns (Fig. 8b). The fluctuation in R_g can be due to the flexibility in the protein-ligand complex upon MD simulation. Similar observation was reported earlier in the R_g during protein-ligand complex simulation of a chimeric enzyme [55]. Therefore, the only protein, as well as the protein-ligand complexes, showed a relatively consistent and similar value of R_g confirming that all the three systems have stable conformation. Similarly, in an earlier study, the stability and compactness were observed for both native protein as well as protein-ligand complexes of peptide deformylase from *Xanthomonas oryzae* upon MD simulation [58]. The RMSF plot indicated that most of the residues in the only *Ps*GH3 and protein-ligand complexes showed a similar pattern of fluctuations throughout the simulation (Fig. 8c). The amino acid residues of *Ps*GH3 residing within the active site core existing in the loops L1 (274–289) and L2 (624–634), showed less fluctuation (0.2–0.3 nm) depicting rigidity and intactness of the binding cavity. Similarly, in a previous study, the rigidity and compactness of the binding site of ligand-protein complexes upon MD simulation were reported for mitogen-activated protein kinase 4 in *Leishmania* sp. [59]. The Solvent Accessible Surface Area (SASA) parameter computes the fraction of the protein surface accessible to the water solvent. Thus, the range of conformational changes that occurred in the protein during these interactions could be predicted by calculation of SASA [59]. Figure 8d depicts the plot of SASA during the course of MD simulation for only the protein and the protein-ligand complexes. The average value of SASA for only *Ps*GH3 was 350 nm^2, and it remained stable till 80 ns (Fig. 8d), while the average value calculated for the protein-ligand complexes, *Ps*GH3 + cellobiose and *Ps*GH3 + celloheptaose were 324 nm^2 and 333 nm^2, respectively. The protein-ligand complexes showed a slight decrease in SASA values which indicated that

in the protein-ligand complexes, the surface of the protein exposed to the aqueous solvent is reduced due to the ligand binding. Similarly, in a previous study on the remdesivir-protein system, a decrease in SASA values for protein-ligand complexes was observed [60]. The number of hydrogen bonds (H-bond) analysis indicates the strength and specificity of ligand for binding to the active site of protein as mentioned earlier [59]. Figure 8e depicts an average number of hydrogen bonds formed between protein and the ligands. The analysis showed that an average of 4 hydrogen bonds between *Ps*GH3 and cellobiose till 80 ns (Fig. 8e), while an increase in the average number of hydrogen bond to 11 was observed between *Ps*GH3 and celloheptaose. Hence, the average number of hydrogen bonds between the protein and ligand indicated that both the ligands, cellobiose, and celloheptaose were bound to *Ps*GH3. However, celloheptaose bound more strongly to the active site of *Ps*GH3. Additionally, for further validation of affinity of ligands for *Ps*GH3, the average short-range interaction energies were estimated for protein-ligand complexes.

The average short-range Lennard-Jones energy calculated for *Ps*GH3 + cellobiose and *Ps*GH3 + celloheptaose were − 115.4 kJ/mol and − 181 kJ/mol, respectively. Moreover, the average short-range coulombic energy calculated for *Ps*GH3 + cellobiose and *Ps*GH3 + celloheptaose were − 90 kJ/mol and − 678 kJ/mol, respectively. Thus, the interaction energies calculated for both protein-ligand complexes confirmed the molecular docking results, where it suggested that the catalytic cleft of *Ps*GH3 could accommodate larger cellooligosaccharides and efficiently act upon them.

Binding energy of protein-ligand complexes using MM-PBSA

The binding free energy is the overall energy comprising the polar, nonpolar or SASA energy, and non-bonded
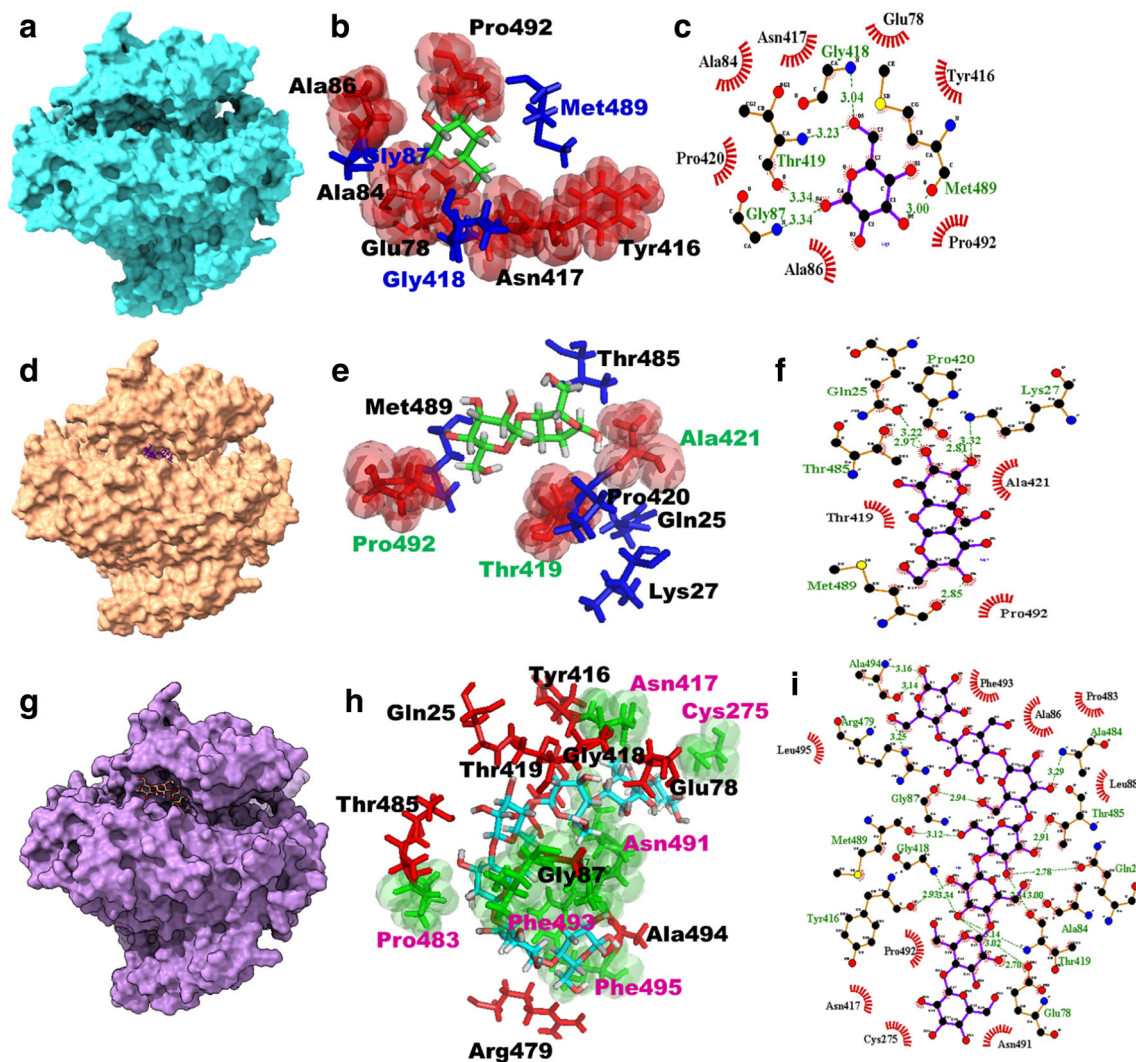
Fig. 7 Molecular docking analysis of *Ps*GH3 with cellooligosaccharides, viz., glucose, cellobiose and celloheptaose. Surface view by using ChimeraX. **a** Glucose. **d** Cellobiose. **g** Celloheptaose. Active site showing ligand orientation. **b** Glucose, **e** Cellobiose. **h** Celloheptaose and two-dimensional schematic presentation. **c** Glucose. **f** Cellobiose, **i** Celloheptaose

interaction energies (electrostatic and Van der Waals interaction) and can be calculated for the protein-ligand complexes by employing the MM-PBSA method [44]. The calculated binding free energies are presented in Table 3. The protein-ligand complex, *Ps*GH3 + cellobiose displayed binding free energy of − 227.4 ± 12.5 kJ/mol (Table 3), while *Ps*GH3 + celloheptaose displayed much higher binding free energy of − 1347.4 ± 38.8 kJ/mol (Table 3). Hence, from these values, it was found that, though, both the cellooligosaccharides showed binding affinity for *Ps*GH3, but celloheptaose showed excellent binding affinity in terms of binding free energy. Thus, the affinity of celloheptaose calculated by MM-PBSA corroborated with the results of molecular docking (Table 2) and interaction energy calculations, as described in previous sections.

Conclusions

The three-dimensional modeled structure of a putative β-1,4-glucosidase of family GH3 (*Ps*GH3) from *Pseudopedobacter saltans* was developed by computational modeling. *Ps*GH3 modeled structure demonstrated multi-domain structure comprising an N-terminal $(\beta/\alpha)_8$-fold-like domain followed by an $(\alpha/\beta)_6$-sandwich domain, a PA14 domain, and a C-terminal domain resembling an immunoglobulin fold, which is a distinctive feature of the GH3 family. The modeled structure showed 99.3% of amino acid residues in the favoured region and only 0.7% amino acid residues in the disallowed region. MSA revealed that Asp274 and Glu624 are the nucleophiles and acid/base catalytic amino acid residues and are conserved across the family 3 glycoside hydrolase. The distance between the carboxyl group of catalytic residues was 5 Å, which

Fig. 8 Molecular dynamics simulation analysis of *Ps*GH3 and *Ps*GH3-ligand complexes, viz., *Ps*GH3 + cellobiose and *Ps*GH3 + celloheptaose. **a** RMSD plot, **b** R$_g$ plot, **c** RMSF plot, **d** SASA plot, and **e** H-bond plot
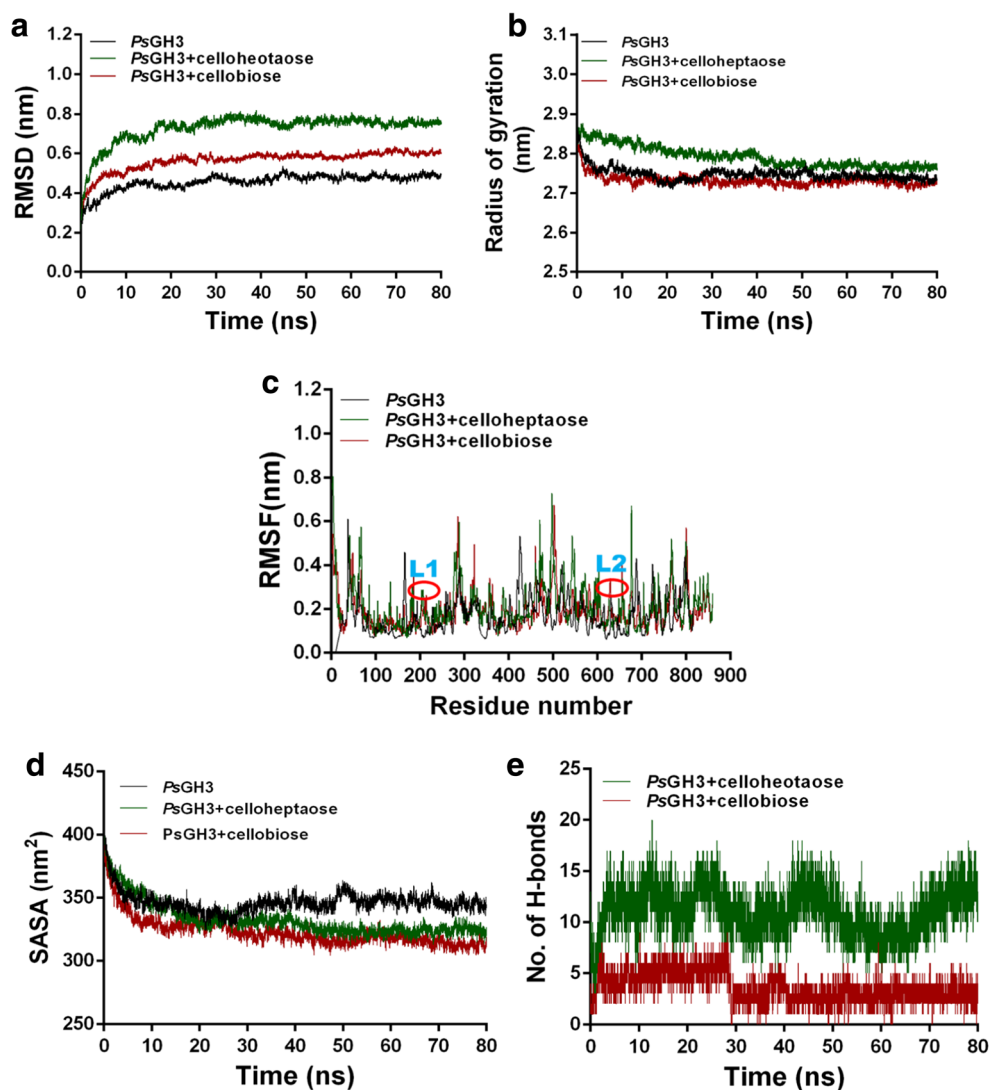
indicated the retaining type of hydrolytic mechanism of *Ps*GH3. MD simulation of the modeled *Ps*GH3 structure for 80 ns showed stable conformation. The active-site analysis of *Ps*GH3 displayed a mixture of positively and negatively charged amino acid residues present in the catalytic cleft which is made up of some conserved amino acid residues, viz., Arg154, Lys192, His193, along with Asp274 and Glu624 active site residues. The molecular docking analysis revealed that the catalytic cleft of *Ps*GH3 could house larger cellooligosaccharides up to DP7, because of its large and deep

size catalytic pocket. The binding interaction of *Ps*GH3 after docking analysis revealed the putative amino residues Ala86, Leu88, Cys275, Pro483, Phe493, Asn417, Asn491, Pro492, and Leu495, near active-site might be involved in substrate binding. The MD simulation of *Ps*GH3 in presence of cellobiose and celloheptaose displayed stable RMSD, R$_g$ and SASA values till 80 ns. Thus the results projected towards the comfortability of the cellooligosaccharides in the binding cavity of *Ps*GH3. The low fluctuation of RMSF in presence of cellobiose and celloheptaose in the binding cavity indicated

Table 3   Energy parameters calculated by MM-PBSA method for *Ps*GH3-ligand complexes

Protein-ligand complexes	Van der Waals energy (kJ/mol)	Electrostatic energy (kJ/mol)	Polar solvation energy (kJ/mol)	SASA energy (kJ/mol)	Binding free energy (kJ/mol)
*Ps*GH3 + cellobiose	− 130.1±− 8.7	− 205.5±31.5	122.1±43.3	− 13.9±1.1	− 227.4±12.5
*Ps*GH3 + celloheptaose	− 235.7±23.1	− 1727.9±71.1	644.6±31.6	− 28.3±1.7	− 1347.4±38.8

preserved rigidity and compactness of the binding site residues in the protein-ligand complexes. The calculations of average number of hydrogen bonds (H-bond), interaction energy, and binding free energy confirmed the stronger affinity of *Ps*GH3 towards larger cellooligosaccharides.

Declarations

Ethics approval   This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate   The authors declare the consent to participate.

Consent for publication   The authors declare the consent for publication.

Conflict of interest   The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Grabnitz F, Seiss M, Rucknagel KP, Staudenbauer WL (1991) Structure of the β-glucosidase gene bglA of *Clostridium thermocellum*: sequence analysis reveals a superfamily of cellulases and β-glycosidases including human lactase/phlorizin hydrolase. Eur J Biochem 200:301–309

2. McDonald JE, Rooks DJ, McCarthy AJ (2012) Methods for the isolation of cellulose-degrading microorganisms. Methods Enzymol 510:349–374

3. Sharma K, Thakur A, Kumar R, Goyal A (2019) Structure and biochemical characterization of glucose tolerant β-1, 4 glucosidase (*ht*bgl) of family 1 glycoside hydrolase from *Hungateiclostridium thermocellum*. Carbohydr Res 483:107750

4. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Mielenz JR (2006) The path forward for biofuels and biomaterials. Science 311:484–489

5. Nath P, Dhillon A, Kumar K, Sharma K, Jamaldheen SB, Moholkar VS, Goyal A (2019) Development of bi-functional chimeric enzyme (*Ct*GH1-L1-*Ct*GH5-F194A) from endoglucanase (*Ct*GH5) mutant F194A and β-1, 4-glucosidase (*Ct*GH1) from *Clostridium thermocellum* with enhanced activity and structural integrity. Bioresour Technol 282:494–501

6. Urbanowicz BR, Bennett AB, del Campillo E, Catala C, Hayashi T, Henrissat B, Teeri TT (2007) Structural organization and a standardized nomenclature for plant endo-1, 4-β-glucanases (cellulases) of glycosyl hydrolase family 9. Plant Physiol 144:1693–1696

7. Barr BK, Hsieh YL, Ganem B, Wilson DB (1996) Identification of two functionally different classes of exocellulases. Biochemistry 35:586–592

8. Bezerra RM, Dias AA (2005) Enzymatic kinetic of cellulose hydrolysis. Appl Biochem Biotechnol 126:49–59

9. Gruno M, Valjamae P, Pettersson G, Johansson G (2004) Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate. Biotechnol Bioeng 86:503–511

10. Zhao Y, Wu B, Yan B, Gao P (2004) Mechanism of cellobiose inhibition in cellulose hydrolysis by cellobiohydrolase. Sci China Ser C 47:18–24

11. Zhang M, Su R, Qi W, He Z (2010) Enhanced enzymatic hydrolysis of lignocellulose by optimizing enzyme complexes. Appl Biochem Biotechnol 160:1407–1414

12. Gurgu L, Lafraya Á, Polaina J, Marín-Navarro J (2011) Fermentation of cellobiose to ethanol by industrial Saccharomyces strains carrying the β-glucosidase gene (BGL1) from *Saccharomycopsis fibuligera*. Bioresour Technol 102:5229–5236

13. Han Y, Chen H (2008) Characterization of β-glucosidasefrom corn stover and its application in simultaneous saccharification and fermentation. Bioresour Technol 99:6081–6087

14. Yang F, Yang X, Li Z, Du C, Wang J, Li S (2015) Overexpression and characterization of a glucose-tolerant β-glucosidase from *Thermoanaerobacterium aotearoense* with high specific activity for cellobiose. Appl Biochem Biotechnol 99:8903–8915

15. Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280:309–316

16. Opassiri R, Pomthong B, Akiyama T, Nakphaichit M, Onkoksoong T, Cairns MK, JRK C (2007) A stress-induced rice (*Oryza sativa L.*) β-glucosidaserepresents a new subfamily of glycosyl hydrolase family 5 containing a fascin-like domain. Biochem J 408:241–249

17. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2008) The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res Spec Publ 37:D233–D238

18. Yoshida E, Hidaka M, Fushinobu S, Koyanagi T, Minami H, Tamaki H, Kumagai H (2010) Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β-glucosidase from *Kluyveromyces marxianus*. Biochem J 431:39–49

19. Singhania RR, Patel AK, Sukumaran RK, Larroche C, Pandey A (2013) Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. Bioresour Technol 127:500–507

20. Petersen TN, Brunak S, Von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786

21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

22. Sievers F, Higgins DG (2014) Clustal omega. Curr Protoc Bioinformatics 48:3–13

23. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res Spec Publ 31:3320–3323

24. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

25. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845

26. Schrodinger LLL (2019) The PyMOL Molecular Graphics System, Version 2.0

27. Laskowski RA, MW MA, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291

28. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519

29. Eisenberg D, Luthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404

30. Berendsen HJ, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 91:43–56

31. Van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger PKHC, Mark AE, Tironi IG (1996). Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag AG an der ETH Zurich, Zurich 86

32. Parrinello M, Rahman A (1980) Crystal structure and pair potentials: a molecular-dynamics study. Phys Rev Lett 45:1196

33. Hess B, Bekker H, Berendsen HJ, Fraaije JG (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472

34. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE (2018) UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci 27:14–25

35. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res Spec Publ 34:W116–W118

36. Binkowski TA, Joachimiak A, Liang J (2005) Protein surface analysis for function annotation in high-throughput structural genomics pipeline. Protein Sci 14:2972–2981

37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera-a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

38. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem 29:622–655

39. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98:10037–10041

40. Grosdidier A, Zoete V, Michielin O (2011) Fast docking using the CHARMM force field with EADock DSS. J Comput Chem 32:2149–2159

41. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng Des Sel 8:127–134

42. Schuttelkopf AW, Van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. Acta Crystallogr D 60:1355–1363

43. Kumari R, Kumar R (2014) C. Open Source Drug Discovery and A. Lynn. J Chem Inf Model 54:1951–1962

44. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discovery 10:449–461

45. Petosa C, Collier RJ, Klimpel KR, Leppla SH, Liddington RC (1997) Crystal structure of the anthrax toxin protective antigen. Nature 385:833–838

46. Karkehabadi S, Helmich KE, Kaper T, Hansson H, Mikkelsen NE, Gudmundsson M, Walton JD (2014) Biochemical characterization and crystal structures of a fungal family 3 β-glucosidase, Cel3A from Hypocrea jecorina. J Biol Chem 289:31624–31637

47. Zmudka MW, Thoden JB, Holden HM (2013) The structure of DesR from Streptomyces venezuelae, a β-glucosidase involved in macrolide activation. Protein Sci 22:883–892

48. Yan S, Wei PC, Chen Q, Chen X, Wang SC, Li JR, Gao C (2018) Functional and structural characterization of a β-glucosidase involved in saponin metabolism from intestinal bacteria. Biochem Biophys Res Commun 496:1349–1356

49. Goyal D, Kumar K, Sharma K, Goyal A (2019) Small-angle X-ray scattering based structure, modeling and molecular dynamics analyzes of a family 5 glycoside hydrolase first endo-mannanase named as RfGH5_7 from Ruminococcus flavefaciens. J Biomol Struct Dyn 38:1–14

50. Rye CS, Withers SG (2000) Glycosidase mechanisms. Curr Opin Chem Biol 4:573–580

51. Pourseif MM, Yousefpour M, Aminianfar M, Moghaddam G, Nematollahi A (2019) A multi-method and structure-based in silico vaccine designing against Echinococcus granulosus through investigating enolase protein. BioImpacts: BI 9:131

52. Receveur V, Czjzek M, Schulein M, Panine P, Henrissat B (2002) Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. J Biol Chem 277:40887–40892

53. Pozzo T, Pasten JL, Karlsson EN, Logan DT (2010) Structural and functional analyzes of β-glucosidase 3B from Thermotoga neapolitana: a thermostable three-domain representative of glycoside hydrolase 3. J Biol Chem 397:724–739

54. Akasako A, Haruki M, Oobatake M, Kanaya S (1997) Conformational stabilities of Escherichia coli RNase HI variants with a series of amino acid substitutions at a cavity within the hydrophobic core. J Biol Chem 272:18686–18693

55. Nath P, Sharma K, Kumar K, Goyal A (2020) Combined SAXS and computational approaches for structure determination and binding characteristics of Chimera (CtGH1-L1-CtGH5-F194A) generated by assembling β-glucosidase (CtGH1) and a mutant endoglucanase (CtGH5-F194A) from Clostridium thermocellum. Int J Biol Macromol 148:364–377

56. Imran M, Poudel N, Li DM, Papageorgiou AC (2019) Crystal structure of a GH3 β-glucosidase from the thermophilic fungus Chaetomium thermophilum. Int J Mol Sci 20:1–15

57. Zhao J, Guo C, Tian C, Ma Y (2015) Heterologous expression and characterization of a GH3 β-glucosidase from thermophilic fungi Myceliophthora thermophila in Pichia pastoris. Appl Biochem Biotechnol 177:511–527

58. Joshi T, Joshi T, Sharma P, Chandra S, Pande V (2020) Molecular docking and molecular dynamics simulation approach to screen natural compounds for inhibition of Xanthomonas oryzae pv. Oryzae by targeting peptide deformylase. J Biomol Struct Dyn:1–18

59. Raj S, Sasidharan S, Dubey VK, Saudagar P (2019) Identification of lead molecules against potential drug target protein MAPK4 from L. donovani: an in-silico approach using docking, molecular dynamics and binding free energy calculation. PLoS One 14:e0221331

60. Sharma A, Vora J, Patel D, Sinha S, Jha PC, Shrivastava N (2020) Identification of natural inhibitors against prime targets of SARS-CoV-2 using molecular docking, molecular dynamics simulation and MM-PBSA approaches. J Biomol Struct Dyn:1–16