Arabian Journal for Science and Engineering (2023) 48:81–91
https://doi.org/10.1007/s13369-022-07153-1

RESEARCH ARTICLE-BIOLOGICAL SCIENCES

Homology Modeling and Analysis of Vacuolar Aspartyl Protease from a Novel Yeast Expression Host *Meyerozyma guilliermondii* Strain SO

Okojie Eseoghene Lorrine[1,3] · Raja Noor Zaliha Raja Abd. Rahman[1,4,5] · Joo Shun Tan[2] · Abu Bakar Salleh[1] · Siti Nurbaya Oslan[1,3,4]

Received: 29 December 2021 / Accepted: 19 July 2022 / Published online: 9 August 2022

Abstract

In eukaryotes, aspartyl protease (PEP4) is a localized hydrolase. PEP4 was recently identified from *Meyerozyma guilliermondii* strain SO (*Mg*PEP4), a novel yeast expression host. But little is known about the structural properties and its catalytic mechanism. Multiple sequence alignment with other yeast aspartyl proteases revealed the conserved regions of *Mg*PEP4 which belongs to the pepsin/proteinase_A_fungi superfamily. Two catalytic aspartic acid residues (Asp112 and Asp297) existed as a single copy at the DTG motif, a pattern of short conserved amino acids sequence. Homology modeling of *Mg*PEP4 was done using *Saccharomyces cerevisiae* PEP4 (PDB ID: 1DPJ) as template. Using in silico analysis, we aim to reveal its stability by way of disulfide bridge formation and the catalytic mechanism of *Mg*PEP4 with a universal protease inhibitor (pepstatin A). Structurally, only two out of the four conserved cysteine residues of the polypeptide were involved in intramolecular disulfide bridges in the validated structure as opposed to two disulfide bridges present in the template which conferred a critical stabilizing role in the protein structures. Pepstatin A (pepA) was docked at the substrate-binding site and showed hydrophilic interactions with the essential catalytic aspartic residues, which preliminarily proved the catalytic mechanism of *Mg*PEP4. In conclusion, understanding of the structure and catalytic mechanism of *Mg*PEP4 at the molecular level have given an insight about its role in the degradation of recombinant proteins in *M. guilliermondii* strain SO as an expression host as well as its potential applications in food, beverages, baking, leather and pharmaceutical industries. Further development of a new yeast strain could be done using *Mg*PEP4 as the target protein.

Keywords Acidic protease · Hydrolase · Molecular docking · Protein structure and Phyre2

1 Introduction

Aspartyl protease (PEP4) is classified as pepsin-like aspartic proteinase with a proteolytic activity at acidic pH [1, 2]. It is a remarkable enzyme produced by various organisms including plants and animals, mainly as coagulating agents [3, 4]. Likewise, it involves in a loss and gain function of proteins, by undergoing both limited and digestive proteolysis [5]. In yeast cells, PEP4 enzymes inhabits vacuole [6]. They are basically identified by the presence of two adjacent and coplanar aspartic acid side chains at their active site. Furthermore, they are categorized generally into two subfamilies which are the retro-pepsins (retroviral proteinases) that are dimers consisting of two identical subunits and the pepsin-like proteinases with two non-identical but similar lobes [7].

✉ Siti Nurbaya Oslan
snurbayaoslan@upm.edu.my

1 Enzyme and Microbial Technology Research Centre, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

2 School of Industrial Technology, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

3 Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

4 Enzyme Technology and X-Ray Crystallography Laboratory, VacBio 5, Institute of Bioscience, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

5 Department of Microbiology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Interests in the structure and function of members of this group of proteolytic enzymes are due to clarification and elucidation of their mechanism in degradation and in part, to their importance in several pathological processes [8–11]. An example is the therapeutic target for inhibitor drugs for a membrane aspartyl protease that function in the cleavage of brain β-amyloid precursor protein (APP) that leads to the production of β-amyloid, where excess of β-amyloid is the major leading factor in Alzheimer's disease [9, 12]. In the vacuolar proteolytic system of yeast cells, PEP4 is regarded as essential enzyme as it is involved in the activities of other hydrolases which include proteinase B (PrB), carboxypeptidase (CPY) and aminopeptidase [13, 14].

M. guilliermondii strain SO is a novel yeast expression host isolated from spoilt orange and used for production of heterologous proteins under the regulation of P_{AOX1} and P_{FLD1} promoter systems [15]. Its aspartyl protease (*Mg*PEP4) has been identified in its proteome which was deposited in the GenBank (BioProject: PRJNA547962) as a pepsin-like proteinase [16].

Protein folding is a process by which protein molecules fold into its unique and functional 3-dimensional structure. In modern biotechnology, modeling techniques aided by computation techniques have been developed to bridge the gap between sequence and structure databases [17, 18]. A development in protein structure prediction has been the threading approach, which employs techniques of identifying the whole folds of a protein from the amino acid sequence by aligning the sequence with 3D structure in the PDB library [19–22]. It is a profound novel approach for assessing the fitness of a protein sequence with a given protein structural fold [23]. Phyre2 is one of the several protein threading tools on the web that is utilized to predict and analyze 3D protein structure [24].

In addition, conserved motifs of protein structure can elucidate more on the functional clues or confirm tentative functional assignments inferred from the sequence. Disulfide bridges formed by the intramolecular bonds of cysteine residues usually serve as additional covalent linkages in protein sequences, which contributes significantly to the protein stability [25]. A single disulfide bridge can stabilize the protein by 2–5 kcal/mol [25–27]. In 2006, Nakka et al., [28], reported the crucial role of a single disulfide bridge in the dimer interface strengthening, thus contributing to the thermal stability of a thermostable glucose-6-phosphate dehydrogenase (tG6PDH) from the hyperthermophilic bacterium *Aquifex aeolicus*.

In *M. guilliermondii* strain SO, the structure of the native *Mg*PEP4 has not been discussed, studied or released in any detail to understand the role of this enzyme in strain SO as well as its potential industrial applications as an expression host for heterologous protein production. Thus, in this study, we aimed to describe the protein modeling of aspartyl protease of *M. guilliermondii* strain SO (*Mg*PEP4) and the presence of one intramolecular disulfide bridge observed in the predicted structure as well as other conserved essential features using Phyre2 software.

2 Materials and Methods

2.1 Sequential Analysis of Aspartyl Protease (*Mg*PEP4)

Nucleotide sequence of *M. guilliermondii* strain SO aspartyl protease (*Mg*PEP4) (accession number: VMS101000005.1) was obtained from GenBank (BioProject: PRJNA547962) and was used to deduce the primary amino acid sequence using ExPASy translate server (https://web.expasy.org/translate/). Subsequently, multiple sequence alignment (MSA) of aspartyl protease amino acid sequence with PEP4 from other yeasts was carried out using Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/). Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id; Protein Homology/ analogy Recognition Engine V 2.0), was used to predict the secondary structure for *Mg*PEP4 [24]. Analysis of the molecular weight and physiochemical properties were estimated and computed using ProtScale (http://www.expasy.org/tools/protscale.html) and the ExPASy ProtParam Tool (http://web.expasy.org/protparam/), respectively. Prediction of the N-terminal signal peptide was also performed using SignalP v4.1 [29].

2.2 Structural Prediction of *Mg*PEP4 via Protein Threading

The mature amino acid sequence of *Mg*PEP4 was used to predict the three-dimensional (3D) protein structure by uploading the sequence into the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id; Protein Homology/ analogy Recognition Engine V 2.0) was based on evolutionary variation patterns [24], which searched for templates against PDB entries. Furthermore, information of the secondary structure was predicted and mapped preparatory onto the alignment with the template model. Graphical presentation of all structures was performed using PyMOL Molecular Graphics system, version 1.7.4.5, Schrodinger, LLC [30] and UCSF Chimera [31]. Further template model analysis was conducted via BLAST by uploading the mature amino acids sequence of *Mg*PEP4, where templates with evolutionary related protein structures are searched against protein database bank using protein BLAST with the protein-specific iterative (PSI-BLAST) algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with functional annotations inferred as well [32].

2.3 Validation of the Predicted *Mg*PEP4 Structure

The structure with the best confidence value as well as percentage identity was subjected to validation phase. It was assessed via ERRAT (servicesn.mbi.ucla.edu/ERRAT/) [33], Verify3D (servicesn.mbi.ucla.edu/Verify3D/) [34, 35], PROCHECK (servicesn.mbi.ucla.edu/PROCHECK/) [36, 37] and plotted in a Ramachandran diagram [38]. The results were classified based on the protein structures stretching from analysis of the overall fold of the proteins to the identification of highly specific clusters of functional residues.

2.4 Superimposition of Predicted *Mg*PEP4 Structure with Template 1DPJ

Superimposition of the predicted aspartyl protein structure with the template (PDB ID:1DPJ) was performed using PyMOL Molecular Graphics system, version 1.7.4.5, Schrodinger, LLC [30], to determine and compare the active sites and substrate-binding site in both structures. Then, the root-mean-square deviation (RMSD) was recorded.

2.5 Molecular Docking

The predicted *Mg*PEP4 structure was docked with the ligand that was co-crystallized with a protease molecule (PDB ID: 2RMP) [39]. The template (PDB ID: 1DPJ) was used as positive control for the analysis. The grid box was adjusted in the substrate-binding cleft. In addition, the pre-process requirements which included preparation of the protein and ligand, addition of hydrogen and energy minimization were performed before carrying out the docking experiments [40]. The torsional bonds of the ligand were free to rotate, while the protein molecules (macromolecule) were set to be rigid. The protein–ligand interaction was viewed using PyMOL [30] and LIGPLOT v 2.2 [41].

3 Results and Discussion

3.1 Sequence and Structural Analysis of Aspartyl Protease of strain SO

The *PEP*4 gene encoding aspartyl protease spans 1,227 bp in length and encodes a protein of 408 amino acids residues, with a predicted size of 44.2 kDa. Aspartyl protease of *M. guilliermondii* strain SO (*Mg*PEP4) revealed a high percentage identity (more than 55%) of amino acid sequence with all the sequences analyzed (Fig. 1). As shown in the multiple sequence alignment (MSA) analysis of the *Mg*PEP4, it suggested that *Mg*PEP4 shared complementary primary features with other reported yeast PEP4 and high conservation in the functional residues and elements of the family enzymes were also revealed. The active site of a protein contained the catalytic residues which were responsible for the bio-catalytic activity and were usually conserved [42]. The two aspartic acid residues, which served as catalytic essential residues in the active site of proteinase_A_fungi superfamily [43], were well conserved in *Mg*PEP4 in a DTG motifs in each lobe (Asp112 and Asp297) (Fig. 1).

With the aid of Phyre2 server, the secondary structure of *Mg*PEP4 was predicted utilizing the functions of SOPMA (Self-Optimized Prediction method With Alignment) techniques that generate a better and improved rate of the second-order forecast based on the primary sequence [24]. The predicted structure constructed showed that the *Mg*PEP4 was composed of 17% α-helix, 42% β-sheet (Fig. 2) which was similar in comparison to other yeasts PEP4 secondary structures (Table S1). At the N-terminal of *Mg*PEP4 sequence, contains the predicted signal peptides computed with SignalP v4.1 between position 21 and 22: ADA-AV with a probability of Probability: 0.7896 which indicated that the protein can be cleaved and secreted extracellularly (Fig. 2) [44].

3.2 Structure Prediction and Validation of *Mg*PEP4

This approach has been confirmed to be efficient in analyzing the relative importance of amino acids residues to protein function and structure [45]. The 3D structural prediction of *Mg*PEP4 by utilizing the intensive mode of Phyre2 [24], was modeled at > 90% confidence score where 3 templates (PDB ID: 3PSG; 1QDM and 1DPJ) were selected based on heuristics to maximize confidence score, alignment coverage and percentage identity. Meanwhile, the template PDB ID: 3PSG and PDB ID: 1QDM shared lower identity with the predicted model of *Mg*PEP4 (37% and 46%, respectively), whereas PDB ID: 1DPJ, *Saccharomyces cerevisiae* PEP4 [46] had a higher identity of 75% and was found to be sufficient for further analysis as validated by ERRAT, Verify3D and PROCHECK (Table S2).

Furthermore, the Ramachandran plots revealed the occupancy of residues of all the main chain angles of the predicted structure in the most favored region was 91.1% occupancy of residues in the most favored region (Figure S1). Thus, these results inferred that the validations of the predicted model were confirmed to be a suitable model for further analysis as the Ramachandran plot can detect any structural gross errors, with the most crucial signal for quality structures of the proteins which depends on the dihedral angles (φ and ψ torsion angles) [47, 48]. In the disallowed region, the occupancy of residue was found to be 0.0% (Figure S1). Therefore, the validation results depict a high-quality model predicted and confirmed by the percentage of the amino acid residues in the core which presents a better guide to stereochemical quality [49].

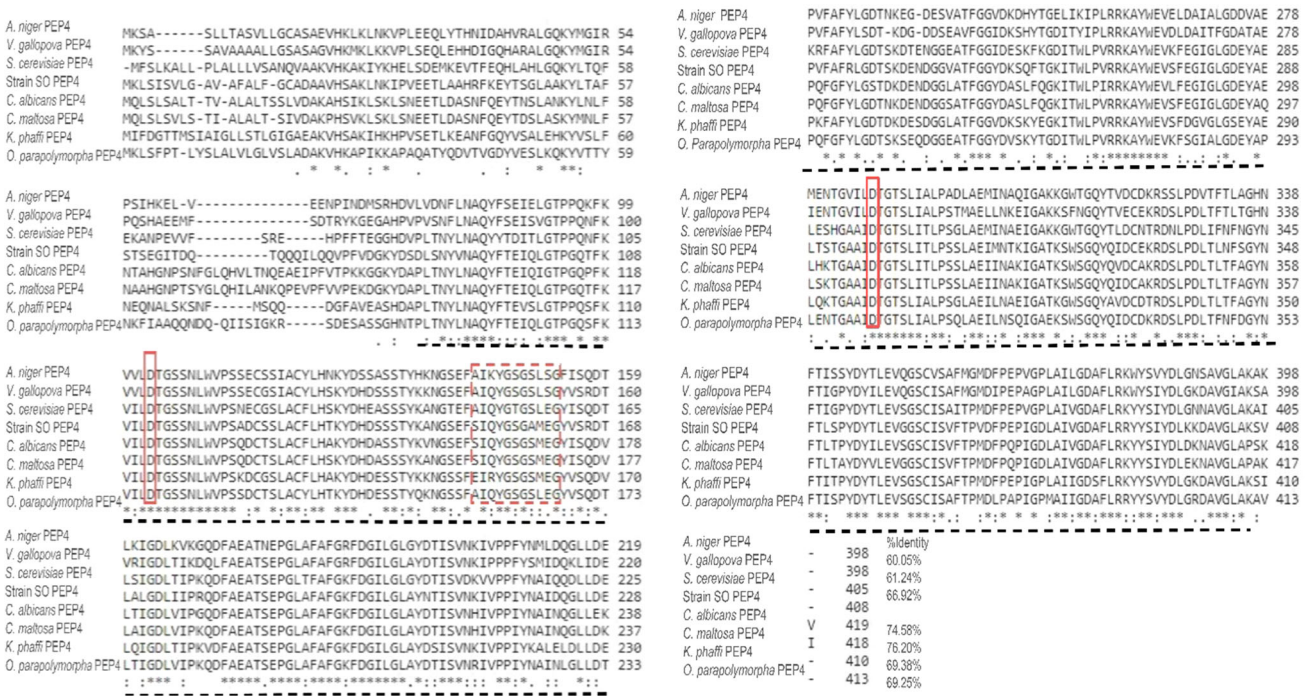

Fig. 1 Multiple sequence alignment analysis of the deduced amino acid sequence of *M. guilliermondii* strain SO aspartyl protease (*Mg*PEP4) aligned with *Komagotaella phaffii* (XP_002493333.1), *Saccharomyces cerevisiae* (NP_015171.1), *Candida albicans* (Xp_713148.1), *Aspergillus niger* (XP_001399855.1), *Candida maltose* (EMG47219.1), *Ogataea parapolymorpa* (XP_013935327.1), *Verruconis gallopava* (XP_016212833.1). The active flap sequences are indicated by a red dotted line box. The conserved superfamily domain is indicated by the black dotted lines and the red box indicates the catalytic aspartic acid residues

More so, *Mg*PEP4 (Fig. 3) structurally appeared as a kidney-shaped bilobed globular protein majorly consisting of mainly β- strands divided into N-terminal and C-terminal domains as previously reported from other studies [50, 51]. This presents a large substrate-binding pocket between the two lobes and each having the catalytic aspartic acid residues in the conserved DTG motifs at position Asp112 and Asp297 (Fig. 3). Furthermore, the mechanism of enzyme catalysis is majorly dependent on an enzyme (substrate-binding pocket) as well as a specific substrate which forms an intermediate complex and afterward, leads to the formation of a product, due to the decomposition of the activated complex.

In most aspartyl proteases, a unique structural motif termed the active flap consisting of a β-hairpin loop [52, 53] and the polyproline loop which extends over the active site had been implicated in catalysis [54], thus facilitating the interaction of substrate binding in the binding pocket, is conserved in *Mg*PEP4 predicted structure (Fig. 3). It is known to be the most flexible element of the structure in aspartyl proteinases and among different crystal structures of an enzyme, it has been reported to localize up to 8.7 Å [55]. The active flap region (green) and the polyproline loop (purple) appeared to cover the pocket of the substrate-binding site. Gly156 residue was located at the tip of the active flap as well and Glu376 was located as the hinge residue on the polyproline loop (Fig. 3).

Furthermore, in proteins, free cysteine residues do occur but most are covalently bonded to other cysteine residues to form disulfide bonds, which play a crucial role in some proteins folding and stability, especially extracellularly secreted proteins [56]. It was observed that four cysteine residues (Cys125, Cys130, Cys331 and Cys364) were conserved but only two were involved in the formation of the intramolecular disulfide bridge which built up the N-terminal loop (Fig. 3) as opposed to two or three disulfide bonds present in PEP4 in yeasts [57–59] and template 1DPJ [55]. More so, the absence of the linkage between Cys331 and Cys364 with 5.7 Å apart, contributed to the lower stability of the predicted structure on the C-terminal region which was worth investigating through superimposition.

Additionally, a fundamental element that underlies the activity of aspartyl proteinases is the presence of the active flap residue Tyr75 which is said to be a conserved residue for pepsin family [60]. It was also conserved in *Mg*PEP4 sequence at Tyr155 (Fig. 1 & 3), and also the key residues that were associated with the binding pocket formation in the active site.

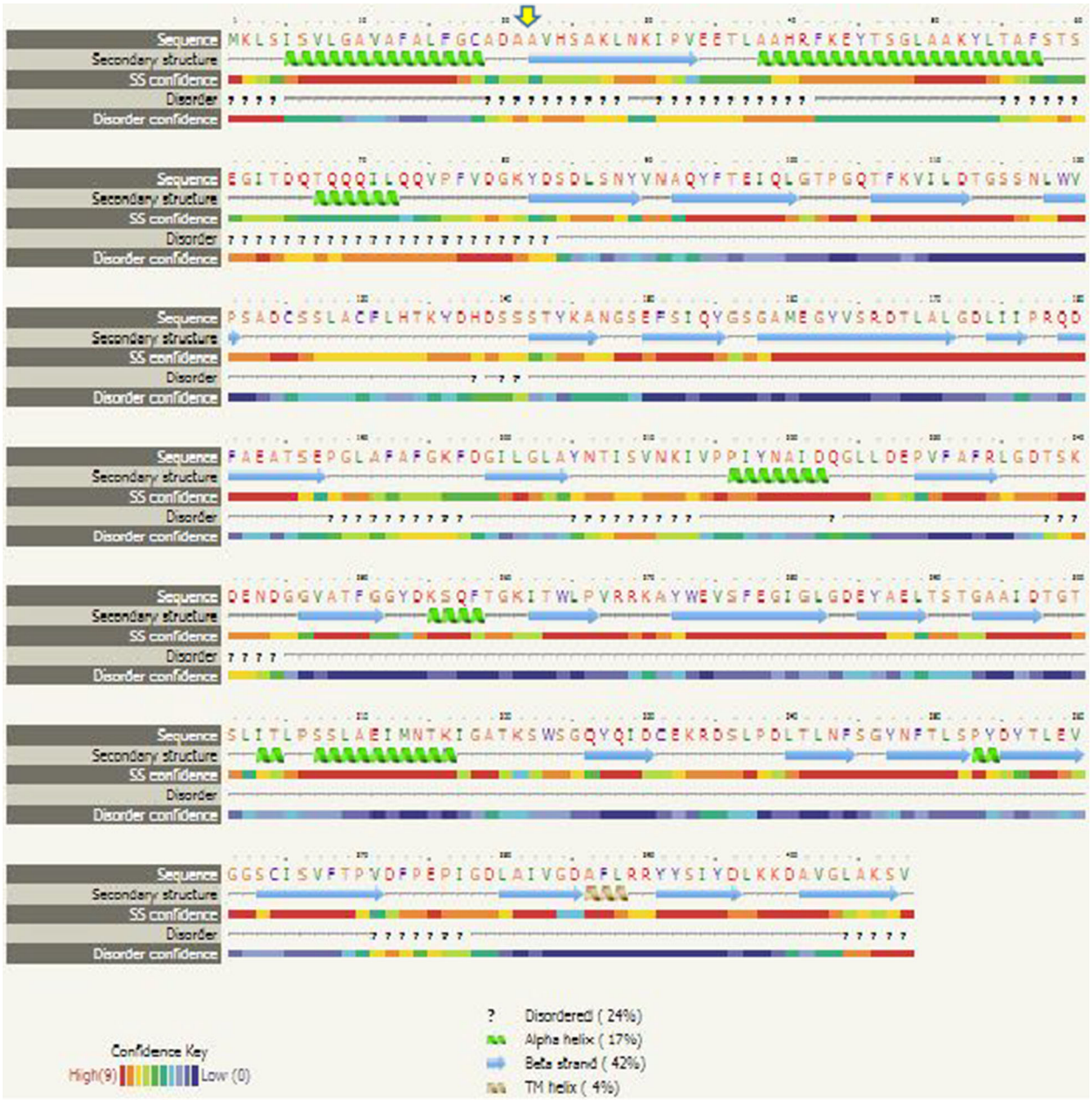

Fig. 2 Secondary structure prediction of *Mg*PEP4 using Phyre2, showing β-sheet and α-helix. The signal peptide predicted using SignalP v4.1 is indicated by the yellow arrow between position 21 and 22: ADA-AV

Some studies postulated that the orientation of the conserved tyrosine residue of the active flap of pepsin family might have a negative effect on aspartyl proteases catalytic capabilities (i.e., self-inhibition) [60]. The term "self-inhibition" of proteinase A described may represent only a transitional structural feature rather than a true inhibition. It could be different from the operation that aspartic proteinases have evolved for modulating or stabilizing their own intrinsic activity in the yeast cells. Interestingly, studies have reported the interaction mediated via the conserved water molecules within the crystallized structure of aspartyl proteases and the tyrosine residue of the active flap possesses functional significance. The authors suggested that upon ligand binding, a stronger network of hydrogen bond was formed, thus implicating the catalytic mechanism of aspartic proteases [61, 62]. Hence, the position of Tyr155 side chain in *Mg*PEP4 could exhibit an operational significance as a mechanism for stabilizing the intrinsic activity, which was potentially connected to substrate capture and cleavage [54, 63]. However, this in silico prediction and hypothesis could only be verified through the in vitro protein characterization.

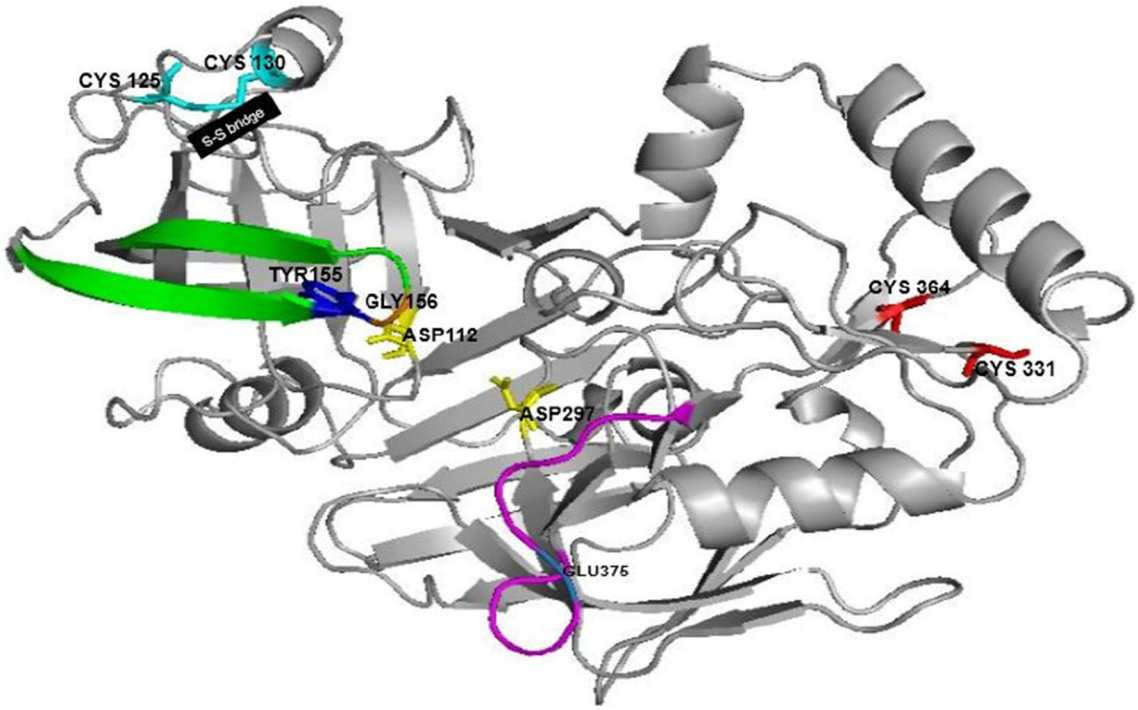

Fig. 3 The predicted 3D structure of *Mg*PEP4. The unlinked disulfide bridge between Cys331and Cys364 (red) in the C-terminal domain and single S–S bridge between Cys125 and Cys130 (cyan) in the N-terminal domain were illustrated. The two catalytic aspartic residues (Asp112 and Asp297; yellow sticks) were seen in the active site. The position of residue Tyr155 (dark blue) can be seen near the active site. The active flap is shown in green and the polyproline region is shown in purple where residues Gly156 and Glu375 are indicated with orange and light blue respectively. The structural figure was generated using the PyMOL Molecular Graphics system, version 1.7.4.5, Schrodinger, LLC

3.3 Superimposition of *Mg*PEP4 and Template 1DPJ

The superimposition between *Mg*PEP4 and the template (PDB ID: 1DPJ) showed that the RMSD value was 2.499 (Fig. 4). The structures were superimposed using the backbone carbon atoms. From the superimposition, the folds of the proteases backbone were essentially identical, with the exception of the N- and C- terminals and some loops which were regarded as the surface exposed regions. Relatively large conformational differences in these regions are supposition of mutations (deletions and/or insertions of residues) and usually present a major issue for direct RMSD calculation which regularly gives an unrealistic measure of the similarity [64, 65]. To further justify the RMSD value computed, the secondary structural fold of *Mg*PEP4 and template 1DPJ were predicted and mapped using Phyre2 tool which aided in the positional equivalence and deviations were indicated as β-strands (twenty-four), α-helices (four) and coils (Figure S2).

The spatial proximity of the two catalytic Asp residues in the complex presented an equivalent position as reported in nonliganded mature aspartic proteinases [8, 46]. Conversely, the proximity of the active site pocket was highly conserved but variation was observed in the polyproline and terminal regions. Furthermore, differences in some amino acids base pair within the superfamily domain could lead to the differences in the secondary structure prediction (Figure S2).

3.4 Molecular Docking

AutoDock Vina which was compatible with MGLTools was used to compute the binding energy and interaction by molecular docking of the predicted protein structure and ligand [40].

To elucidate more on the binding affinity of ligand toward the predicted protein receptors, molecular docking [66], was performed using the classical aspartic proteinase inhibitor pepstatin A (PepA), a microbial hexa-peptide produced by *Streptomyces* sp. [67], complexing with *Rhizomucor miehei* aspartic proteinase (PDB ID:2RMP). PepA is an aspartic protease inhibitor and also known to inhibit proteases such as chymosin, renin, pepsin, HIV protease and cathepsins D and E [68]. It presents a competitive inhibition mechanism [69, 70]. Its chemical structure composes mainly of two residues of an unusual amino acid, 4-amino-3-hydroxy-6-methylheptanoic acid (statine), having the sequence isovaleryl-L-valyl-L-valyl-statyl-L-alanyl-statine (Iva-Val-Val-Sta-Ala-Sta) [69]. The protein–ligand

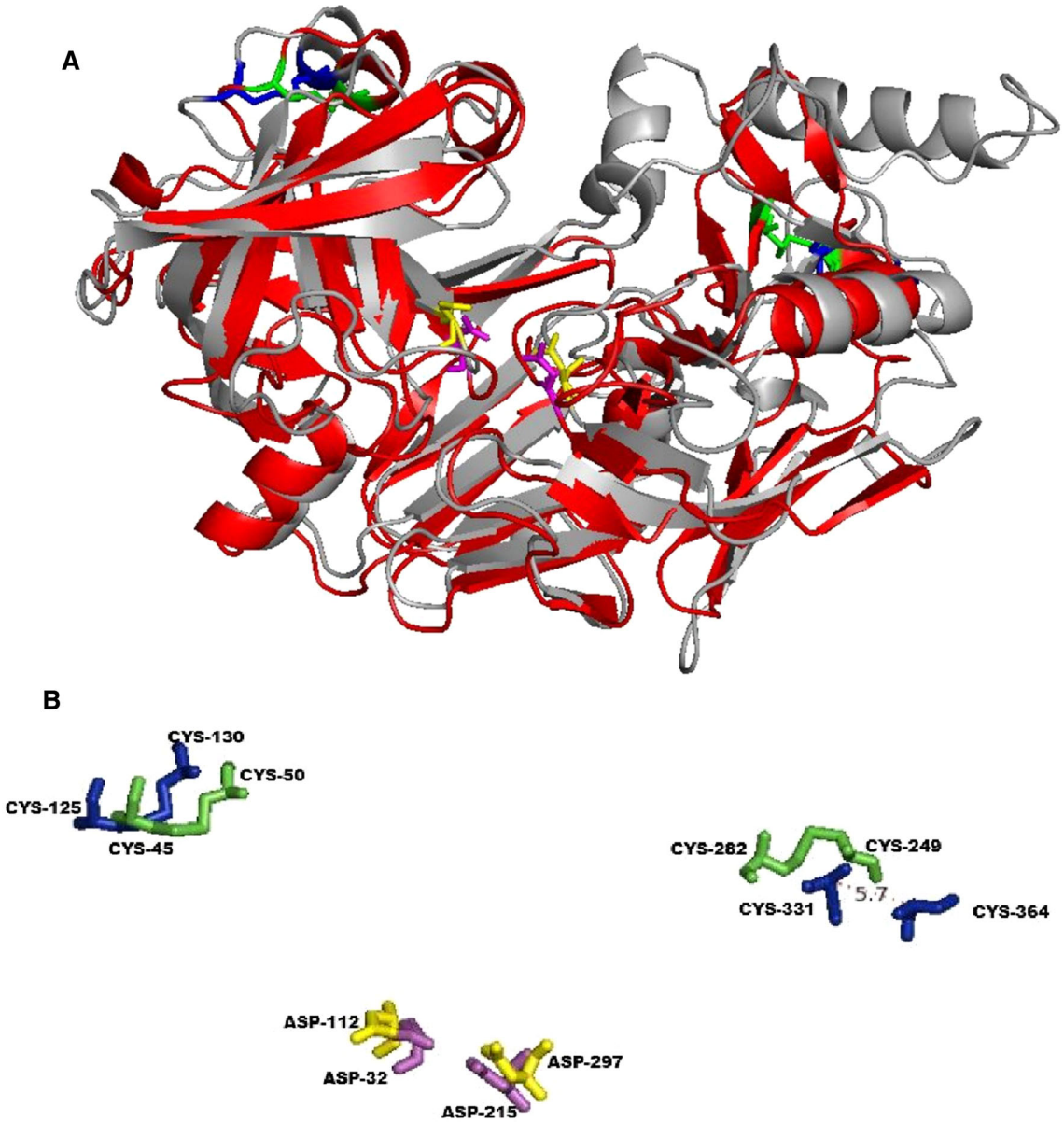

Fig. 4 Superimposition of protein structures. **A** Structural alignment of the *Mg*PEP4 and template 1DPJ in gray and red, respectively. **B** Close-up view comparison of the conserved catalytis aspartic acid residues, where strain SO catalytic residues are indicated as yellow and template 1DPJ as magenta. Conserved cysteine residues are also aligned for both structures, where strain SO residues are indicated as blue with only one disulfide bond formation. While template cysteine residues are indicated as green with two disulfide bonds formation. The structural figure was generated using the PyMOL Molecular Graphics system, version 1.7.4.5, Schrodinger, LLC

interaction was viewed using the PyMOL Molecular Graphics system, version 1.7.4.5, Schrodinger, LLC which generated intermolecular interactions and their strengths, hydrogen bonds, hydrophobic interactions and atom accessibilities [30].

From the docked complex (Fig. 5), the ligand (PepA) interacted with the residues in the substrate-binding pocket by both hydrogen bonds and hydrophobic interactions where the contacting residues gave a binding affinity of -8.0 (kcal/mol). Furthermore, the docking analysis revealed the fitting of the PepA into the large substrate-binding cleft between the two domains of *Mg*PEP4.

The hydroxyl group (OH) of the statin (STA) molecule of the PepA formed hydrogen bonds with the catalytic aspartic residues (Asp112 and Asp297) present in both domain N- and C- terminals with distances of 3.6 Å, respectively. The hydroxyl group at the This phenomenon was justified by the study reported by Yang and Quail, [41] where the statin

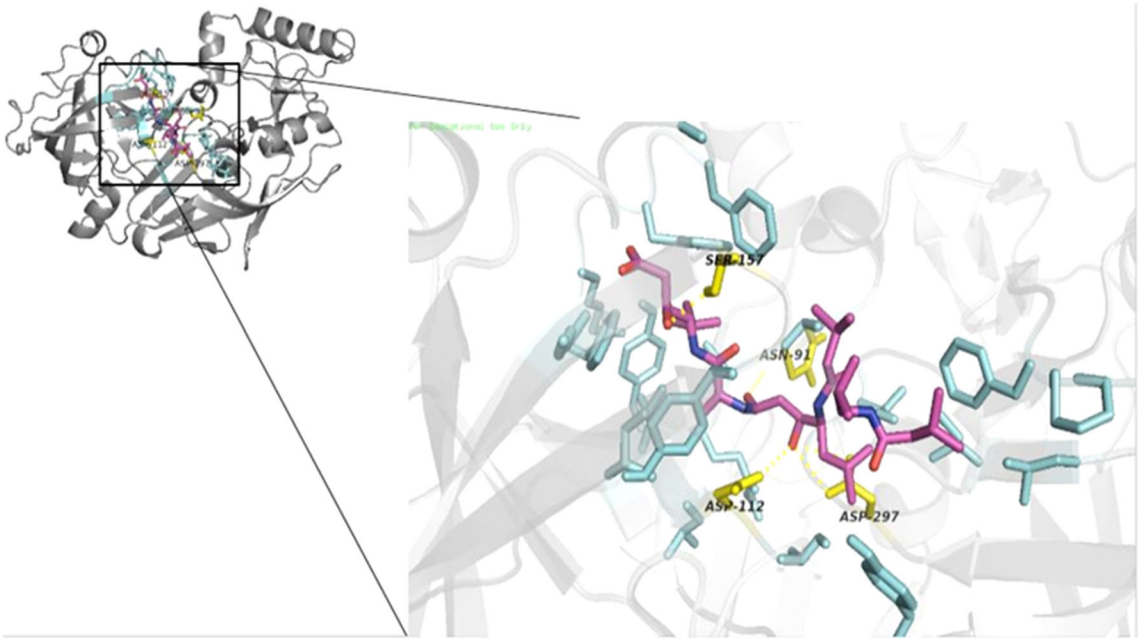

Fig. 5 Molecular docking analysis of predicted structure of *Mg*PEP4 and pepstatin A. The ligand (pink) was observed to be within the substrate-binding site of *Mg*PEP4. The catalytic aspartic acid residues (Asp112 and Asp297) were involved in the hydrophilic interactions with PepA via the hydrogen bonds (yellow dotted lines) with distances of 3.6 Å respectively, besides other hydrophilic residues (Asn91 and Ser57 with distances of 3.5 Å and 2.5 Å respectively) in the binding cleft. The ligand was also observed to be surrounded by the hydrophobic residues (cyan) which were mainly made up of phenylalanine and glycine residues. All structural figures were generated using the PyMOL Molecular Graphics system, version 1.7.4.5, Schrodinger, LLC

residue hydroxyl group formed hydrogen bonds with the two catalytic Asp38 and Asp237 in *Rhizomucor miehei* aspartic proteinase (RMP) when complexed with PepA. Moreover, this conformation mimicked the anticipated transition state of the enzyme–substrate interaction which infers the industrial application potential of *Mg*PEP4 in the degradation of heterologous proteins when *M. guilliermondii* strain SO is used as an expression host. In addition, there was no large distortion of the active site because of the binding of the ligand to the enzyme.

Besides the hydrophilic interactions, other hydrophobic contacting residues of the substrate-binding cleft were Gly189, Phe194, Met160, Lys196, Ala97, Leu380, Asn91, Gly259, Leu111, Thr113, Val190, Phe192, Phe373, Ala295, Gly114, Gly156, Gln154, Tyr155, Ile208, Ala295, Ile382 and Thr300. These residues were also found to be similar to those of 1DPJ except for the difference in position. Furthermore, an in silico site directed-mutagenesis (SDM) conducted on the catalytic residue present at the N-terminal domain to leucine in *Mg*PEP4 observed a conformational change in the substrate-binding pocket, thus suggesting the high impact of mutation on the strictly conserved and catalytically essential residues in the protein catalytic function (Figure S3) in the degradation of recombinant proteins when *M. guilliermondii* strain SO is used as an expression host.

4 Conclusion

In conclusion, the aspartyl protease (*Mg*PEP4) from *M. guilliermondii* strain SO was a typical proteinase A that belongs to the pepsin-like/proteinase A superfamily, involved in the hydrolysis of peptide bonds. The *Mg*PEP4 structure predicted via protein threading was highly similar with the template 1DPJ, possessing a large substrate-binding cleft between the two lobes. A single disulfide bridge was formed with two out of the four cysteine residues from the predicted structure which conferred stability on protein structures. Nevertheless, further analysis at the molecular and biochemical levels could also be done towards revealing the stability of *Mg*PEP4. The protein–ligand interaction justified the significance of the conserved catalytic residues which could be considered a bottleneck in the *M. guilliermondii* strain SO as an expression host. This finding has also given an insight about the potential application of *Mg*PEP4 as potential degradative agent in heterologous protein production as well as its potential applications in cheese-making, baking, leather, food, beverage and therapeutic target for drug inhibitors in biopharmaceutical industries. It is also recommended to develop aspartyl protease mutant of *M. guilliermondii* strain SO to reduce the degradative function in future expressions of recombinant proteins.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13369-022-07153-1.

Acknowledgements We would like to thank all the principle investigators in Enzyme and Microbial Technology Research Centre, Universiti Putra Malaysia for providing constructive comments and advice for this project. The first author, OEL would like to thank the Ministry of Education, Nigeria for the opportunity to further her study.

Funding This research work was funded by the Fundamental Research Grant Scheme Projects, (Code: FRGS/1/2019/STG05/UPM/02/1) from the Ministry of Higher Education Malaysia which was awarded to SNO, the corresponding author.

Declarations

Conflict of Interest The authors declare no conflict of interest.

References

1. Brett, C.L.; Kallay, L.; Hua, Z.; Green, R.; Chyou, A.; Zhang, Y.; Graham, T.R.; Donowitz, M.; Rao, R.: Genome-wide analysis reveals the vacuolar pH-stat of *Saccharomyces cerevisiae*. PLoS ONE **6**, e17619 (2011). https://doi.org/10.1371/journal.pone.0017619
2. Stanforth, K.J.; Wilcox, M.D.; Chater, P.I.; Brownlee, I.A.; Zakhour, M.I.; Banecki, K.M.R.M.; Pearson, J.P.: Pepsin properties, structure, and its accurate measurement: a narrative review. Ann. Esophagus (2021). https://doi.org/10.21037/aoe-20-95
3. Yegin, S.; Dekker, P.: Progress in the field of aspartic proteinases in cheese manufacturing: structures, functions, catalytic mechanism, inhibition and engineering. Dairy Sci. Technol. **93**, 565–594 (2013). https://doi.org/10.1007/s13594-013-0137-2
4. Hsiao, N.W.; Chen, Y.; Kuan, Y.C.; Lee, S.K.; Chan, H.H.; Kao, C.H.: Purification and characterization of aspartic protease from the Rhizopus oryzae protease extract, Peptidase R. Electron J. Biotechnol. **17**, 89–94 (2014). https://doi.org/10.1016/j.ejbt.2014.02.002
5. Minina, E.A.; Moschou, P.N.; Bozhkov, P.V.: Limited and digestive proteolysis: crosstalk between evolutionary conserved pathways. New Phytol. **215**, 958–964 (2017)
6. Li, S.C.; Kane, P.M.: The yeast lysosome-like vacuole: endpoint and crossroads. Biochim. Biophys. Acta **1793**, 650–663 (2009). https://doi.org/10.1016/j.bbamcr.2008.08.003
7. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D.: The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. **46**, D624–D632 (2018). https://doi.org/10.1093/nar/gkx1134
8. Davies, D.R.: The structure and function of aspartic proteinases. Annu. Rev. Biophys. Biophys. Chem. **19**, 189–215 (1990)
9. Hong, L.; Koelsch, G.; Lin, X.; Wu, S.; Terzyan, S.; Ghosh, A.; Zhang, X.C.; Tang, J.: Structure of memapsin 2 (b-secretase) complexed with inhibitor: a template to design drugs for Alzheimer's disease. Science **290**, 150–153 (2000). https://doi.org/10.1126/science.290.5489.150
10. Craik, C.S.; Page, M.J.; Madison, E.L.: Proteases as therapeutics. Biochem. J. **435**(1), 1–16 (2011). https://doi.org/10.1042/bj20100965
11. De Souza, P.M.; de Assis Bittencourt, M.L.; Caprara, C.C.; de Freitas, M.; de Almeida, R.P.C.; Silveira, D.; Fonseca, Y.M.; Filho, E.X.F.; Pessoa Junior, A.; Magalhães, P.O.: A biotechnology perspective of fungal proteases. J. Microbiol. **46**, 337–346 (2015). https://doi.org/10.1590/S1517-838246220140359
12. Tiwari, S.; Atluri, V.; Kaushik, A.; Yndart, A.; Nair, M.: Alzheimer's disease: pathogenesis, diagnostics, and therapeutics. Int. J. Nanomed. **14**, 5541–5554 (2019). https://doi.org/10.2147/IJN.S200490
13. Jones, E.W.: Vacuolar proteases and proteolytic artifacts *in Saccharomyces cerevisiae*. Meth. Enzymol (2002). https://doi.org/10.1016/s0076-6879(02)51844-9
14. Hecht, K.A.; O'Donnell, A.F.; Jeffrey, L.B.: The proteolytic landscape of the yeast vacuole. Cell Logist. (2014). https://doi.org/10.4161/cl.28023
15. Oslan, S.N.; Salleh, A.B.; Rahman, R.A.; Zaliha, R.N.; Leow, T.C.; Sukamat, H.; Basri, M.: A newly isolated yeast as an expression host for recombinant lipase. Cell. Mol. Biol. Lett. **20**(2), 279–293 (2015). https://doi.org/10.1515/cmble-2015-0015
16. Lorrine, O.E.; Rahman, A.N.Z.R.A.; Tan, J.S.; Farhana, R.K.; Salleh, A.B.; Oslan, S.N.: Determination of putative vacuolar proteases, PEP4 and PRB1 in a novel yeast expression host *Meyerozyma guilliermondii* strain SO using bioinformatics tools. Pertanika J. Sci. Technol. **30**(1), 777–797 (2022). https://doi.org/10.47836/pjst.30.1.42
17. Petrey, D.; Honig, B.: Protein structure prediction: inroads to biology. Mol. Cell. **20**, 811–819 (2005). https://doi.org/10.1016/j.molcel.2005.12.005
18. Porto, W.F.; Pires, A.S.; Franco, O.L.: Computational tools for exploring sequence databases as a resource for antimicrobial peptides. Biotechnol. Adv. **35**(3), 337–349 (2017). https://doi.org/10.1016/j.biotechadv.2017.02.001
19. Xu, J.; Li, M.; Kim, D.; Xu, Y.: Raptor: Optical protein threading by linear programming. J. Bioinform. Comput. Biol. **01**(01), 95–117 (2003). https://doi.org/10.1142/s0219720003000186
20. Sommer, I.; Toppo, S.; Sander, O.: Improving the quality of protein structure models by selecting from alignment alternatives. BMC Bioinformatics **7**, 364 (2006). https://doi.org/10.1186/1471-2105-7-364
21. Ma, J.; Peng, J.; Wang, S.; Jinbo, Xu.: A conditional neural fields model for protein threading. Bioinformatics **28**(2), 59–66 (2012). https://doi.org/10.1093/bioinformatics/bts213
22. Zhang, C.; Wei Zheng, S.M.; Mortuza, Y.L.; Zhang, Y.: DeepMSA: constructing deep multiple sequence alignment to improve contact prediction and fold-recognition for distant-homology proteins. Bioinformatics **36**(7), 2105–2112 (2020). https://doi.org/10.1093/bioinformatics/btz863
23. Xu, Y.; Liu, Z.; Cai, L.; Xu, D.: Protein structure prediction by protein threading. In: Xu, Y.; Xu, D.; Liang, J. (Eds.) Computational methods for protein structure prediction and modeling. Biological and medical physics, Biomedical engineering. Springer, New York (2007)
24. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E.: The Phyre2 web portal for protein modelling, prediction and analysis. Nat Protoc **10**, 845–858 (2015). https://doi.org/10.1038/nprot.2015.053
25. Galgóczy, L.; Borics, A.; Virágh, M.; Ficze, H.; Váradi, G.; Kele, Z.; Marx, F.: Structural determinants of Neosartorya fischeri antifungal protein (NFAP) for folding, stability and antifungal activity. Sci. Rep. (2017). https://doi.org/10.1038/s41598-017-02234-w
26. Creighton, T.E.: Disulphide bonds and protein stability. BioEssays **8**(2–3), 57–63 (1988). https://doi.org/10.1002/bies.950080204
27. Tidor, B.; Karplus, M.: The contribution of cross-links to protein stability: a normal mode analysis of the configurational entropy of the native state. Proteins Struct. Funct. Genet. **15**(1), 71–79 (1993). https://doi.org/10.1002/prot.340150109
28. Nakka, M.; Iyer, R.B.; Bachas, L.G.: Intersubunit disulfide interactions play a critical role in maintaining the thermostability of

glucose-6-phosphate dehydrogenase from the hyperthermophilic bacterium *Aquifex aeolicus*. Proteins **25**(1), 17–21 (2006). https://doi.org/10.1007/s10930-006-0015-3
29. Petersen, T.N.; Brunak, S.; Von Heijne, G.; Nielsen, H.: SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods. **8**, 785–786 (2011). https://doi.org/10.1038/nmeth.1701
30. Delano, W. (2002). The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA, USA. http://www.pymol.org.
31. Huang, C.C.; Meng, E.C.; Morris, J.H.; Petterson, E.F.; Ferrin, T.E.: Enhancing UCSF Chimera through web services. Nucleic Acids Res. **42**, 478–484 (2014)
32. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. **25**, 3389 (1997). https://doi.org/10.1093/nar/25.17.3389
33. Colovos, C.; Yeates, T.O.: Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. **2**(9), 1511–1519 (1993). https://doi.org/10.1002/pro.5560020916
34. Bowie, J.U.; Lüthy, R.; Eisenberg, D.A.: Method to identify protein sequences that fold into a known three-dimensional structure. Science **253**(5016), 164–170 (1991). https://doi.org/10.1126/science.1853201
35. Lüthy, R.; Bowie, J.U.; Eisenberg, D.: Assessment of protein models with three-dimensional Profiles. Nature **356**, 83–85 (1992). https://doi.org/10.1038/356083a0
36. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M.: PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26**, 283–291 (1993). https://doi.org/10.1107/S0021889892009944
37. Laskowski, R.A.; Rullmann, J.A.C.; MacArthur, M.W.; Kaptein, R.; Thornton, J.M.: AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR. **8**(4), 477–486 (1996). https://doi.org/10.1007/BF00228148
38. Ramachandran, G.N.; Sasisekharan, V.: Conformation of polypeptides and proteins. Adv. Protein Chem. **23**, 283–437 (1968). https://doi.org/10.1016/s0065-3233(08)60402-7
39. Yang, J.; Quail, J.W.: Structure of the Rhizomucor miehei aspartic proteinase complexed with the inhibitor pepstatin A at 2.7 Å resolution. Acta Crystallogr. **55**(3), 625–630 (1999). https://doi.org/10.1107/s0907444998013961
40. Trott, O.; Olson, A.J.: AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J. Comput. Chem. **31**, 455–461 (2010). https://doi.org/10.1002/jcc.21334
41. Laskowski, R.A.; Swindells, M.B.: LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. **51**, 2778–2786 (2011). https://doi.org/10.1021/ci200227u
42. Skolnick, J.; Fetrow, J.; Klolinski, A.: Structural genomics and its importance for gene function analysis. Nat. Biotechnol. **18**, 283–287 (2000). https://doi.org/10.1038/73723
43. Parr, C.L.; Keates, R.A.B.; Bryksa, B.C.; Masahiro, O.; Rickey, Y.Y.: The structure and function of *Saccharomyces cerevisiae* proteinase A. Yeast **24**, 467–480 (2007). https://doi.org/10.1002/yea.1485
44. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A.: Protein identification and analysis tools on the ExPASy server. In: Walker, J.M. (Ed.) The Proteomics Protocols Handbook, pp. 571–607. Humana Press, New Jersey (2005)
45. Zitare, U.A.; Habib, M.H.; Rozeboom, H.; Mascotti, M.L.; Todorovic, S.; Fraaije, M.W.: Mutational and structural analysis of an ancestral fungal dye-decolorizing peroxidase. FEBS J. **288**(11), 3602–3618 (2021). https://doi.org/10.1111/febs.15687
46. Kay, J.; Gustchina, A.; Li, M.; Phylip, L.H.; Lees, W.E.; Winther, J.R.; Wlodawer, A.: The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix. Nat. Struct. Biol. **7**(2), 113–117 (2000)
47. Kleywegt, G.J.; Jones, T.A.: Phi/Psi-Chology: Ramachandran revisited. Structure. **4**(12), 1395–1400 (1996). https://doi.org/10.1016/S0969-2126(96)00147-5
48. Laskowski, R.A.; MacArthur, M.W.; Thornton, J.M.: PROCHECK : validation of protein-structure coordinates. In: Arnold, E.; Himmel, D.M.; Rossmann, M.G. (Eds.) International tables for crystallography, pp. 684–687. Wiley, Hoboken (2012)
49. Morris, A.L.; MacArthur, M.W.; Hutchinson, E.G.; Thornton, J.M.: Stereochemical quality of protein structure coordinates. Proteins **12**, 345–364 (1992). https://doi.org/10.1002/prot.340120407
50. Borelli, C.; Ruge, E.; Schaller, M.; Monod, M.; Korting, H.C.; Huber, R.; Maskos, K.: The crystal structure of the secreted aspartic proteinase 3 from *Candida albicans* and its complex with pepstatin A. Proteins **68**(3), 738–748 (2007). https://doi.org/10.1002/prot.21425
51. Parr, C.L.; Keates, R.A.B.; Bryksa, B.C.; Ogawa, M.; Yada, R.Y.: The structure and function of *Saccharomyces cerevisiae* proteinase A. Yeast **24**(6), 467–480 (2007). https://doi.org/10.1002/yea.1485
52. Barman, A.; Prabhakar, R.: Computational insights into substrate and Site specificities, catalytic mechanism, and protonation states of the catalytic Asp dyad of β-Secretase. Scientifica (2014). https://doi.org/10.1155/2014/598728
53. McGillewie, L.; Ramesh, M.; Soliman, M.E.: Sequence, structural analysis and metrics to define the unique dynamic features of the flap regions among aspartic proteases. Proteins **36**, 385–396 (2017). https://doi.org/10.1007/s10930-017-9735-9
54. Tang, J.; Koelsch, G.: A possible function of the flaps of aspartic proteases: the capture of substrate side chains determines the specificity of cleavage positions. Protein Pept. Lett. **2**, 257–266 (1995)
55. Li, M.; Phylip, L.H.; Lees, W.E.; Winther, J.R.; Dunn, B.M.; Wlodawer, A.; Kay, J.; Gustchina, A.: The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix. Nat. Struct. Biol. **7**, 113–117 (2000). https://doi.org/10.1038/72378
56. Sevier, C.S.; Kaiser, C.A.: Formation and transfer of disulphide bonds in living cells. Nat. Rev. Mol. Cell Biol. **3**(11), 836–847 (2002). https://doi.org/10.1038/nrm954
57. Sepulveda, P.; Marciniszyn, J.; Liu, D.; Tang, J.: Primary structure of porcine pepsin. J. Biol. Chem. **250**, 5082–5088 (1975). https://doi.org/10.1016/S0021-9258(19)41281-7
58. Bech, A.M.; Foltmann, B.: Partial primary structure of *Mucor miehei* protease. Neth. Milk Dairy J. **35**, 275–280 (1981)
59. Dreyer, T.; Halkier, B.; Svendsen, I.; Ottesen, M.: Primary structure of the aspartic proteinase A from *Saccharomyces cerevisiae*. Carlsberg Res. Commun. **51**(1), 27–41 (1986). https://doi.org/10.1007/bf02907993
60. Gustchina, A.; Li, M.; Phylip, L.H.; Wendy, E.L.; Kay, J.; Wlodawer, A.: An unusual orientation for Tyr75 in the active site of the aspartic proteinase from *Saccharomyces cerevisiae*. Biochem. Biophys. Res. Commun. **295**, 1020–1026 (2002). https://doi.org/10.1016/s0006-291x(02)00742-8
61. Andreeva, N.S.; Rumsh, L.D.: Analysis of crystal structures of aspartic proteinases: on the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. Protein Sci. **10**, 2439–2450 (2001). https://doi.org/10.1110/ps.25801
62. Prasad, B.V.; Suguna, K.: Role of water molecules in the structure and function of aspartic proteinases. Acta. Crystallogr. **58**, 250–259 (2002). https://doi.org/10.1107/S0907444901018327
63. Tanaka, T.; Teo, K.S.L.; Lamb, K.M.; Harris, L.J.; Yada, R.Y.: Effect of replacement of the conserved Tyr75 on the catalytic properties of porcine pepsin A. Protein Pept. Lett. **5**, 19–26 (1998)

64. Kufareva, I.; Abagyan, R.: Methods of protein structure comparison. Methods Mol. Biol. **857**, 231–257 (2012). https://doi.org/10.1007/978-1-61779-588-6_10
65. Rueda, M.; Orozco, M.; Totrov, M.; Abagyan, R.: A web tool for the superimposition of biomolecules and assemblies with rotational symmetry. BMC Struct. Biol. **13**, 32 (2013). https://doi.org/10.1186/1472-6807-13-32
66. Rao, G.N.; Rao, A.A.; Rao, P.S.; Muppalaneni, N.B.: A tool for the post data analysis of screened compounds derived from computer-aided docking scores. Bioinformatics **9**(4), 207–209 (2013)
67. Umezawa, H.; Aoyagi, T.; Morishima, H.; Matsuzaki, M.; Hamada, M.: Pepstatin, a new pepsin inhibitor produced by *Actinomycetes*. J. Antibiot. (Tokyo) **23**, 259–262 (1970). https://doi.org/10.7164/antibiotics.23.259
68. Dostál, J.; Brynda, J.; Vaňková, L.; Zia, R.S.; Pichová, I.; Heidingsfeld, O.; Lepšík, M.: Structural determinants for subnanomolar inhibition of the secreted aspartic protease Sapp1p from *Candida parapsilosis*. J. Enzyme Inhib. Med. Chem. **36**(1), 914–921 (2021). https://doi.org/10.1080/14756366.2021.1906664
69. Dunn, B.M.; Gustchina, A.; Wlodawer, A.; Kay, J.: Subsite preferences of retroviral proteinases. Methods Enzymol. **241**, 254–278 (1994). https://doi.org/10.1016/0076-6879(94)41068-2
70. Dunn, M.B.; Goodenow, M.M.; Gustchina, A.; Wlodawer, A.: Retroviral proteases. Genome Biol. (2002). https://doi.org/10.1186/gb-2002-3-4-reviews3006