Structural Modelling and *in-silico* characterization of a Novel Thermophilic *β*-amylase from *Clostridium thermosulfuregenes*

Shafiqa Nayel[1*], Mohd Shahir Shamsir[2] and Farid Ahmad Danishfar[1]

[1]University Teknology Malaysia (UTM), Herat University (HU), Malaysia.
[2]Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Original Research Article

ABSTRACT

β-amylase is a hydrolytic enzyme that is involved in breaking down starch and producing energy. Since the discovery of β-amylase, it has been applied in various applications especially in the food industry. In this study, a novel β-amylase from *Clostridium thermosuluregen*, a thermophilic anaerobic bacterium that ferments its extracellular emulsion to ethanol at 62 ℃ was modelled and studied using bioinformatics tools and compared with *B. cereus* β-amylases that functions at mesophilic conditions. The results showed that the overall structural conformations, secondary structures, and important residues involved in active and binding sites were identified in both proteins. The results revealed that the modelled β-amylase of *C. thermosulfuregen* is very similar with respect to the global conformation, location of active and binding sites. Both proteins showed identical structural domains with the thermophilic variant possessing a high percentage of hydrophobic amino acid residues, polar amino acid residues, and differences in secondary composition such as loops and beta sheets as the potential evolutionary thermal adaptations that make it stable enzyme that functions up to 70 ℃. The results suggest that the thermal stability are not dependent on one single unique mechanism and may use one or a combination of the

_____

*Corresponding author: E-mail: Shafiqa.nayel@yahoo.com;

mechanisms to sustain its structural conformation at a higher operating temperature. Overall, considering the common properties of this modelled protein with the β-amylase of *B. cereus*, it can be assumed that if the β-amylase of *C. thermosulfuregen* were expressed *in-vitro*, it would produce a stable protein that possesses the hydrolysis function for *C. thermosulfuregen* to break down the starch and sugar formation.

Keywords: β-amylase; Clostridium thermosulfuregen; characterization; Bacillus cereus.

1. INTRODUCTION

The β-amylase (EC 3.2.1.2) enzyme, or known as α-1,4-glucan maltohydrolase, is an exo-type enzyme and catalyses the β-anomeric maltose from the non-reducing end of starch to produce maltose. It belongs to the family 14 of the glycoside hydrolase (GH) [1]. β-amylase can be found in plants, fungi, and bacteria. The β-amylase enzyme has important applications in industries due to its saccharogenic activity. It is useful in the pharmaceutical industry due to its digestive activity, used for the preparation of malto-oligosaccharides, a reagent that is used for research, as nutrients in the health industry substitute for other saccharides and used in the production of the malto-oligomer mixture, an ingredient used for the preparation of chewing gum, buttercream, cakes, jellies, canned cocoa and fruit drinks [2].

Plant-based β-amylases are well documented, the breadth of the sources is limited to five plants and one bacterial origin. Therefore, it is of interest to expand the source of the enzyme to include a thermostable β-amylase that is active at higher temperatures, increasing its potential applicability in the industry. The activity of beta-amylase in the temperature range of 20 to 90 ℃ has been investigated in various organisms [3]. β-amylases from *Bacillus cereus* (a mesophilic bacterium) and soybean on raw starch granules from various botanical sources (potato, sweet potato, wheat, rice and corn) were examined and found that *Bacillus cereus* β-amylase hydrolyzed corn granules efficiently at 45 ℃ while, soybean β-amylase was 60% less active than Bacillus cereus β-amylase at the same temperature [4]. β-amylase purified from *Bacillus cereus* (mesophilic) has been reported to possess an optimum temperature of about 50 ℃ (Takasaki, 2005). While the β-amylase purified from *Clostridium thermosulphurogene*s (thermophilic) have 55% maximal activity at 50 ℃ and 65% maximal activity at 80℃, and have an optimum temperature about 75 ℃ [5]. [6] The thermostability can be enhanced by substrate

and Ca2+ addition at 80 ℃. The enzyme has a greater operating pH range of between pH 3.5 to 6.5, unlike other β-amylase that possess the optimal activity and stable only at neutral pHs. Results have also shown that the enzyme possesses an optimum activity at pH 5.5 to 6.0 [7], making it an interesting candidate for studying the sequence and structural differences that may contribute to its ability to operate at a different temperature and pH range. Barnaud *et al.* [8] cloned and sequenced the gene encoding the thermophilic β-amylase of *Clostridium thermosulfurogenes* in *Bacillus subtilis* and showed that the mature β-amylase has 519 amino acids and molecular weight (MW) of 57167 kDa. The β-amylase sequence showed 32% homology with β-amylase of soybean and barley. However, there is no 3D structure of this thermophilic enzyme available for study.

Clostridium thermosulfurogenes is gram-negativ, straight rods in shape about 0.5 x > 2 µm. It forms long filaments and motile by peritrichous flagella. Endospores are white-refractile, spherical and swollen. When *C. thermosulfurogenes* grow on thiosulphate, it produce and store yellow granule sulphur on their cells and in the medium. In the thin section, there is no outer membranous layer. *Clostridium thermosulfurogenes* have catabolites like ethanol and lactate, that are biotechnologically interest. As an ethanol or lactate-producing thermophile bacterium, it grows at lower pH values than *C. thermocellum* [6].

The characterization of β-amylase and determination of its 3D structure has been documented for a range of species; from bacteria *Bacillus cereus* [9] *Glycine max* (soybean) [10], *Ipomoea batatas* (sweet potato) [11], *Hordeum vulgare* (barley) [12] and *Triticum aestivum* (wheat) [13]. All of these β-amylase structures share common characteristics that suggest highly conserved regions especially at the active centre in the region of (α/β)8 barrel. Only the tertiary structure barrel configuration differs between that of plant and bacterial β-amylase [9]. This study is conducted to model the

structure of the thermophilic β-amylase from *Clostridium thermosulfurogenes* and compare the predicted model of enzyme thermophilic β-amylase from *Clostridium thermosulfuregenes* with *Bacillus cereus* β-amylase and making it the first modelled thermophilic structure and the first from species other than *Bacillus cereus* and of plant origin. The model will be used to examine if there are differences that can be attributed to structural adaptations that allow it to function at higher temperatures. The knowledge of these differences would be contributing to the understanding of thermophilic adaptation in β-amylase and may be used in protein modification for future use.

2. RESEARCH METHODS

For *in silico* modelling and characterisation of β-amylase from *Clostridium thermosulfuregen*, different databases, software and other sources were applied to analyse molecular structure of this protein. Fig. 1 shows the outline methodology of this study.

2.1 Primary Structure Analysis

2.1.1 Protein sequence retrieval, analysis and comparison

The EC number for enzyme β-amylase is 3.2.1.2 in BRENDA. All information about the β-amylase of *C. thermosulfuregen* such as metabolic pathway, pH, optimum temperature and reaction were retrieved from BRENDA database ("https://www.brenda-enzymes.org/"). The FASTA format was got from UniProt database ("https://www.uniprot.org/") [14] and no structure was found for this enzyme on RCSB ("www.rcsb.org"). A BLAST (Basic Local Alignment Search Tool) was performed using NCBI database ("https://www.ncbi.-nlm.nih.gov/") [15] to obtain similar sequences of β-amylase from other organisms for the purpose of structural comparison and analysis.

The protein sequence analysis of β-amylase from *C. thermosulfuregen* was conducted by applying the FASTA format of the protein in the ProtParam tool ("https://web.expasy.org/-protparam/") [16] and calculated molecular weight, theoretical pI value, amino acid composition, total number of positively and negatively charged residues, atomic composition, chemical formula, estimated half-life, aliphatic index, and hydropathicity value of the protein. After that, these primary structure information compared with β-amylase of *Bacillus cereus* that has 44.99% similarity with β-amylase of *C. thermosulfuregen*.

2.1.2 Multiple sequence alignment

Three sequence alignments will perform using the ClustalO tool ("https://www.ebi.ac.uk/Tools/msa/clustalo/") to identify the conserved regions and comparatively analyse the different sequences [17].
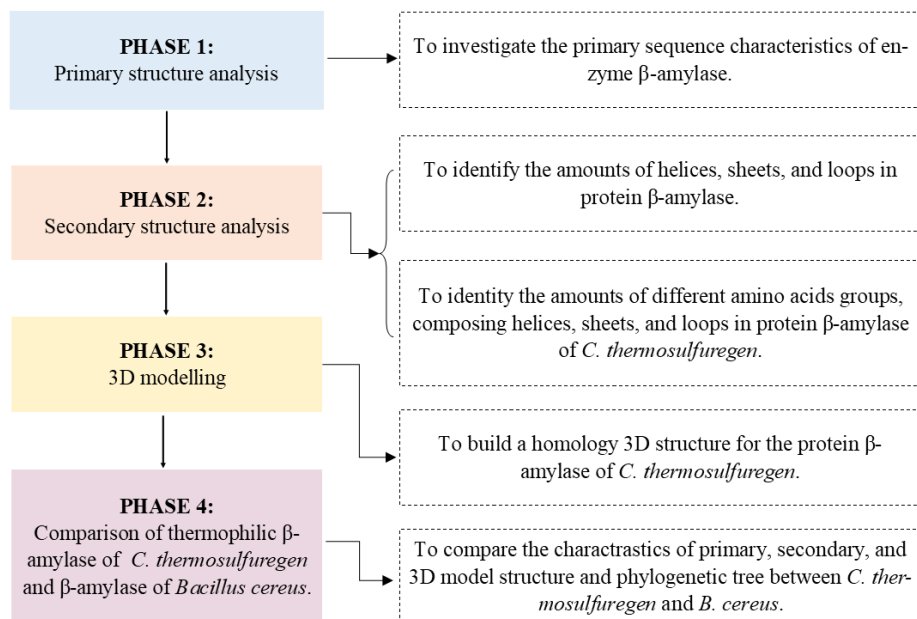

Fig. 1. Schematic outline of methodology

2.1.3 Secondary structure prediction

Secondary structure prediction of uncharacterized β-amylase protein of *Clostridium thermosulfuregen* was performed using the CFSSP Protein Sequence Analysis software ("http://www.biogem.org/tool/chou-fasman/") [18] and GOR VI ("https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4-.html") [19]. The percentages of alpha helices, β-sheets, turns, and coils were observed for further comparison and analysis. In addition, CFSSP were used to calculate the percentages of different groups of amino acid residues in an alpha helix, β-sheets, turns, and coils. Since the β-amylase of *Bacillus cereus* protein has a known structure stored as 1B90, the information about its secondary structure obtained from PDBsum database ("http://www.ebi.ac.uk/pdbsum") (Roman et al., 2017).

2.2 Tertiary Structure Analysis

2.2.1 Homology Modelling and template selection

The amino acid sequence of the uncharacterized protein (P19584) from *Clostridium thermosulfuregen* was submitted to the online modelling server SWISS-MODEL program ("https://swissmodel.expasy.org/") to obtain a recommended template for homology modelling [20]. The template was chosen based on sequence identity.

2.2.2 Homology model validation

The 3D structure of modelled protein predicted by SWISS-MODEL database was validated using the PROCHECK software. ("http://services.mbi.ucla.edu/PROCHECK/") and Ramachandran plot to investigate about the psi-phi angles and ERRAT program ("http://services.mbi.ucla.edu/ERRAT/") was used to determine the accuracy of the predicted structure. Verify 3D ("http://services.mbi.ucla.-edu/Verify_3D/") (Luthy et al., 1997) also used to assess the model of the protein 3D structure.

2.2.3 Structural comparison

The structure of β-amylase of *Clostridium thermosulfuregen* was compared with β-amylase of *Bacillus cereus*. β-amylase of *Bacillus cereus* has known structure with PDB ID (1B90) on the PDB database ("http://www.rcsb.org"). UCSF Chimera ("https://www.cgl.ucsf.edu/chimera/") [21]. was used to superimpose the modelled structure of β-amylase of *Clostridium thermosulfuregen* and the known structure of beta-amylase of *B. cereus* to compare surface structure, determining active site, binding sites, and other structural details.

For protein thermostability comparison, several factors have been compared. Amino acid composition, hydrophobicity, hydrogen bond, salt bridge and secondary structure comparison like helical content, beta-strands and shortening of loops were compared between mesophilic β-amylase *Bacillus cereus* and thermophilic β-amylase of *Clostridium thermosulfuregen*.

2.3 Phylogenetic Study

The evolutionary relationship of the P19584 protein was studied using phylogenetic tree analysis. The sequence of modelled protein was submitted to NCBI BlastP service to search for protein similarity sequence. A set of 5 proteins sequences including the template were selected and aligned using MEGA X program (version 10.1) and CastalO [17] and phylogenetic tree was constructed [22].

3. RESULT AND DISCUSSION

3.1 Protein Structure Analysis

3.1.1 Protein sequence selection and retrieval

The protein β-amylase from *Clostridium thermosulfuregen* bacterium was reached in the UniProt. The UniProt ID is P19584 (https://www.uniprot.org/uniprot/-P19584). The sequence has no known structure in UniProt and subsequent search in Protein Data Bank (PDB) (www.rcsb.org) confirmed that this sequence has no prior structure. The amino acid sequence of β-amylase from *Clostridium thermosulfuregen* bacterium have retrieved from Uniprot database (Fig. 2).

The amino acid sequences of all five β-amylases were retrieved from the Uniprot database in FASTA sequence format. β-amylases from five different species were selected from the Uniprot database as suggested templates for sequence alignment and listed in Table 1. These β-amylase sequences were used for the analysis in secondary structure determination, multiple sequence alignments, and comparison of the three-dimensional structures.

```
>sp|P19584|AMYB_THETU Thermophilic beta-amylase OS=
Thermoanaerobacterium thermosulfurigenes OX=33950 PE=1 SV=1
MIGAFKRLGQKLFLTLLTASLIFASSIVTANASIAPNFKVFVMGPLEKVTDFNAFKDQLI
TLKNNGVYGITTDIWWGYVENAGENQFDWSYYKTYADTVRAAGLKWVPIMSTHACGGNVG
DTVNIPIPSWVWTKDTQDNMQYKDEAGNWDNEAVSPWYSGLTQLYNEFYSSFASNFSSYK
DIITKIYISGGPSGELRYPSYNPSHGWTYPGRGSLQCYSKAAITSFQNAMKSKYGTIAAV
NSAWGTSLTDFSQISPPTDGDNFFTNGYKTTYGNDFLTWYQSVLTNELANIASVAHSCFD
PVFNVPIGAKIAGVHWLYNSPTMPHAAEYCAGYYNYSTLLDQFKASNLAMTFTCLEMDDS
NAYVSPYYSAPMTLVHYVANLANNKGIVHNGENALAISNNNQAYVNCANELTGYNFSGFT
LLRLSNIVNSDGSVTSEMAPFVINIVTLTPNGTIPVTFTINNATTYYGQNVYIVGSTSDL
GNWNTTYARGPASCPNYPTWTITLNLLPGEQIQFKAVKIDSSGNVTWEGGSNHTYTVPTS
GTGSVTITWQN
```

Fig. 2. FASTA format of the amino acid sequence of the β amylase enzyme from
Thermoanaerobacterium thermosulfurogenes

Table 1. Accession number, protein name, gene name and organism source for the five β-amylase

ID	Protein	Gene name	Origanism
P36924	β-amylase	Spo II	*Bacillus cereus* (mesophilic)
P10538	β-amylase	BMY1	*Glycine max* (soybean)
P16098	β-amylase	BMY1	*Hordeum vulgar* (Barley)
P10537	β-amylase	BMY1	*Ipomoea batatas* (sweet potato)
P19584	β-amylase	N/A	*Thermoanaerobacterium Thermosuifurigenes* (*Clostridium thermosulfurogenes*)

Alignment of the different β-amylase sequences showed that the *Thermoanaerobacterium thermosulfuregenes* β-amylase share many identical sequence characteristic with other β-amylases, especially to the enzyme from *Bacillus cereus* with 44.99% similarity. The active and the binding site that are present in the thermophilic enzyme are aligned to both bacterial and plant sourced enzyme (Fig. 3).

Results showed that highly similar residues that implies conservation and subsequently structure are found mostly in the regions containing beta strands. There are limited conserved residues within helices with only the helix from residue 274-300 of *Thermoanaerobacterium thermosulfurigenes* possessing the highest incidence of conserved residues (9 out of 26 amino acids). This suggests that conservation of function is located and dependent upon the beta strands. Seven amino acids that are involved in the binding site and two of the amino acids responsible for the functioning of the active sites are located on the beta strands.

3.2 Physicochemical Characterization

The result of the ProtParam analysis used for characterisation of the enzyme prior to future experimental studies [16]. The result of the analysis is presented in Table 2. The Table 2 illustrate chemical formula, molecular weight, amino acid composition, total number of positively and negatively charged residues, hydropathicity value, estimated half-life, aliphatic index and theoretical Pi value of the model protein.

3.3 Amino acid Composition

The amino acid composition between the thermophilic β-amylase of *C. thermosulfuregen* and mesophilic β-amylase from *Bacillus cereus* are presented in Table 3.

Comparison of individual amino acids between β-amylase from *C. thermosulfuregen* and *B. cereus* revealed that the changes in the number of amino acids that have been shown to be associated with higher temperature adaptations are present but not significant. In thermophilic protein, the number of Arg and Tyr residues are expected to be higher in number compared to the mesophilic homolog. Comparison between the two homologs showed that only the number of Tyr residues are higher with five more residues (36 versus 31). The number of Arg

residues are lower in the thermophilic than the mesophilic homolog (5 versus 11). This suggests that from the sequence analysis point of view, compositional change of Arg and Tyr are not an adaptation strategy. Further analysis needs to be performed to look at the 3D conformation and the location of Arg and Tyr in the protein structure.

Fig. 3. Multiple sequence alignment of β-amylase *Bacillus cereus* (mesophilic), *Glycine max* (Soybean), *Hordeum vulgare* (Barley), *Ipomoea batatas* (sweet potato) and *Thermoanaerobacterium thermosulfurigenes* (*Clostridium thermosulfurogenes*) with coloured bands representing

Table 2. Composition and characterization of sequences for the modelled proteins β amylase and template

Characteristics β-amylase	*C. thermosulfuregen*	*B. cereus*
Number of amino acids	551	546
Molecular weight	60547.58	61628.95
Theoretical PI	5.26	6.29
Total number of negatively charged residues (Asp + Glu):	37	55
Total number of positively charged residues (Arg + Lys)	27	52
Aliphatic index	73.30	75.20
Average of hydropathicity	-0.170	-0.360
Chemical formula	$C_{2740}H_{4068}N_{690}O_{833}S_{17}$	$C_{2798}H_{4237}N_{711}O_{820}S_{21}$

Table 3. Comparison of percentages of individual amino acids between β-amylase from *C. thermosulfuregen* and *B. cereus* with significant changes highlighted as grey boxes

Amino acids	β-amylase of *B. cereus*		β-amylase of *C. thermosulfuregen*	
	No	%	No	%
Ala (A)	36	7.0%	39	7.5%
Arg (R)	11	2.1%	5	1.0%
Asn (N)	36	7.0%	51	9.8%
Asp (D)	27	5.2%	22	4.2%
Cys (C)	3	0.6%	7	1.3%
Gln (Q)	20	3.9%	15	2.9%
Glu (E)	28	5.4%	15	2.9%
Gly (G)	40	7.8%	41	7.9%
His (H)	8	1.6%	8	1.5%
Ile (I)	26	5.0%	29	5.6%
Leu (L)	38	7.4%	29	5.6%
Lys (K)	40	7.8%	19	3.7%
Met (M)	14	2.7%	9	1.7%
Phe (F)	22	4.3%	22	4.2%
Pro (P)	25	4.8%	26	5.0%
Ser (S)	32	6.2%	46	8.9%
Thr (T)	35	6.8%	51	9.8%
Trp (W)	15	2.9%	16	3.1%
Tyr (Y)	31	6.0%	36	6.9%
Val (V)	29	5.6%	33	6.4%

The result showed that the expected reduction of the number Cys and Ser residues did not occur for the thermophilic amylase. The number of Cys residues increased from three to seven and Ser residues increased significantly from 32 to 46. Biggest amino acid compositional changes between mesophilic and thermophilic homolog are the increase of Asn residues from 36 to 51 (45 % increase), reduction of Glu residues from 28 to 15 (46 % reduction), reduction of Lys residues from 40 to 19 (47.5 % reduction) and increase of Thr residues from 35 to 51 (68% increase). The amino acids involved in these compositional changes have not been attributed to previous research on thermostable adaptations. However, it could be attributed to

other adaptations such as halophilic or barometric adaptations. There is no increase of Pro residues usually attributed to the increase stiffness conferred by adding Pro residues especially in the loops. The expected reduction of Gly residues often seen in thermophilic proteins are not observed between these two homologs as well.

Thermophilic β-amylase of *Clostridium thermosulfuregen* and mesophilic β-amylase of *Bacillus cereus* showed differences in the acidic and basic amino acids. The percentage of acidic and basic amino acids of modelled protein is lower than template but natural amino acids are more than template (Table 4).

3.4 The Composition of Amino Acid Groups

The result of the distribution analysis shown in percentages of amino acids for the model protein showed that as expected, there are no significant differences between hydrophobic and polar amino acid composition between the mesophilic and thermophilic homologs. There is a slight increase in frequency of small amino acids that has been attributed to thermal adaptation. Overall, comparison of both homologs show little quantitative differences between thermophiles and mesophiles (Table 5).

Table 4. Comparison of the amino acids composition in β-amylase between *C. thermosulfuregen* and *B. cereus*

Amino acid Sequence Composition	C. *thermosulfuregen*		B. *cereus*	
	No	%	No	%
Tiny (A+C+G+S+T)	184	35.453	146	28.295
Small (A+B+C+D+G+N+ P+S+T+V)	316	60.886	263	50.969
Aliphatic (A+I+L+V)	130	25.048	129	25.000
Aromatic (F+H+W+Y)	82	15.800	76	14.729
Non-polar (A+C+F+G+I+L+M+P+V+W+Y)	82	15.800	279	54.070
Polar (D+E+H+K+N+Q+R+S+T+Z)	232	44.701	237	45.930
Charged (B+D+E+H+K+R+Z)	69	13.295	114	22.093
Basic (H+K+R)	32	6.1660	59	11.434
Acidic (B+D+E+Z)	37	7.1290	55	10.659

Table 5. The composition of various amino acid groups with the groups that does not play a significant role in thermostability adaptation.

Amino-acid composition By Groups	% in β-amylase of *Bacillus cereus*	% in β-amylase of *Clostridium thermosulfuregen*
Acidic D, E	10.6%	7.1%
Aliphatic I, L, V	18%	17.6%
Aromatic H, F, W, Y	14.8%	15.7%
Basic R, H, K	11.5%	6.2%
Charged R, D, E, H, K	22.1%	13.3%
Hydrophobic A, C, F, I, L, M, V, W, Y	41.5%	42.3%
Polar R, N, D, E, Q, H, K, S, T	46%	44.7%
Big E, F, H, I, K, L, M, Q, R, W, Y	49.1%	39.1%
Small A, C, D, G, N, P, S, T, V	51%	60.8%
Tiny A, C, G, S	21.6%	25.6%

3.5 Secondary Structure Analysis

The prediction of the secondary structure of the β-amylase of *C. thermosulfuregen* carried out using the CFSSP and GOR VI. Protein Sequence Analysis software showed that the number of possible amino acid residues to form helix structure are 282 amino acid equals to 51.2%, while 397 amino acid residues form beta sheet (72.1%), 67 amino acid residue form turns (12.2%) and 314 amino acid residue form random coil (56.88%) as shown in Fig. 4. These data indicate that beta sheets are the predominant modelled protein structure. It also shows that spirals are the second largest, and coils, rings, and rotations make up the third largest material of modelled protein (Fig. 4).

About the *B. cereus*, possible amino acid residues to form helix structure are 391 amino acid which equals to 71.6%, while 249 amino acid residues form beta sheet (45.6%), 64 amino acid residue form turns (11.7%) and 259 amino acid residue form random coil (47.44%) (Fig. 4). In general, there are only slight variations in the values obtained for this aspect of comparison. However, *C. thermosulfuregen* protein residues form 51.20% helixes, 72.10% sheets, 56.99% loops, and 12.20% turns.

Studies have been reported that beta-sheets cause protein stability and increase protein strength [23], the results obtained from the GOR4 for modeled protein with 72.10% sheets, indicate the stability of the modeled protein relative to the template protein. In addition, the percentage of helixes in modeled protein is lower than template protein (51.20% for model protein) and it show the reduction of flexibility of model protein.

3.6 Tertiary Structure

3.6.1 Identification of the template

Given the importance of this protein in enzymatic action (by breaking down stable relationships) and ethanol fermentation in thermophilic bacterium, the first step in future studies is certainly to model this protein. The tertiary structure prediction of the target protein was done using Phyre2, HHPred, and SWISS Model. Among the models made by three modeller, the result of SWISS Model was selected as the best model based on the very similarity of the template protein. The results from HHPred and Swiss Model servers were evaluated.

Table 6 shows the results of two servers. The identity and e-values is within the acceptable range. The identity obtained from Swiss model database is 44.99 %, while the one from HHPred is 44% and e-values is 3.4e-65. The template and model proteins are both monomer.
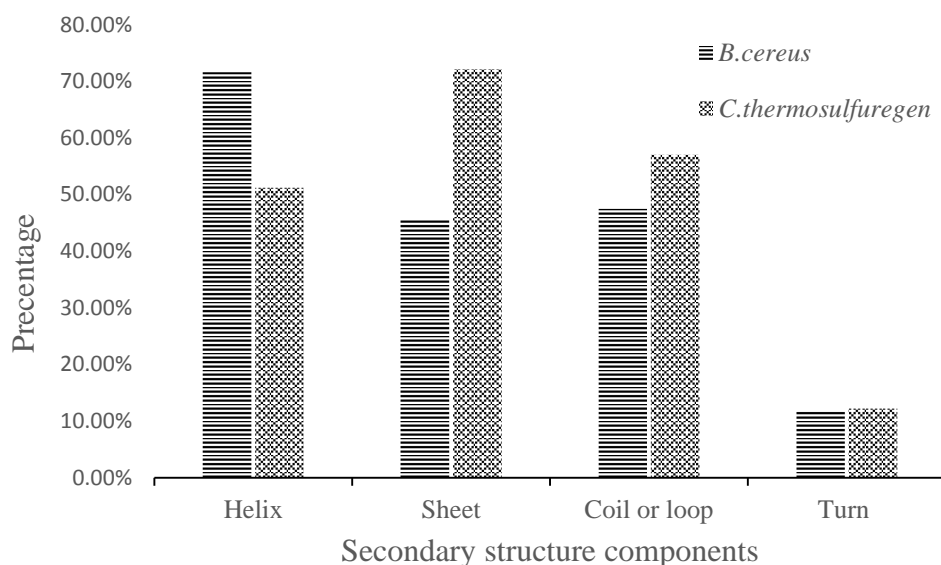

Fig. 4. Comparison of percentage of helices, sheets, loops, and turn in secondary structure of β-amylase of *C. thermosulfuregen* and β amylase of *B. cereus* using the CFSSP and GOR VI Protein Sequence Analysis software

Table 6. P19584's top four proposed templates from two server

Enzyme name	Server name	Template	Protein name	Length	Identity	E-Value
P19584	HHPred modeller	1VEM_A (BACCE)	β-amylase	516	44%	3.4e-65
		P10538 Glycine max (Soybean) (Glycine hispida)	β-amylase	495	36%	3.1e-64
		P16098,Hordeum vulgare (Barley)	β-amylase	535	37%	1.3e-64
		P10537{Ipomoea batatas (Sweet potato)}	β-amylase	498	36%	1.8e-65
		3U7V_A (Caulobacter crescentus)	Beta-galactosidase	552	14%	6.5e-11
	SWISS Model	P36924 (BACC)	β-amylase	546	44.99%	N/A
		3VOC_A (Paenibacillus polymyxa)	Beta/alpha-amylase	419	39.95%	N/A
		Q9LIR6	Beta-amylase 1, chloroplastic	575	48.2%	N/A
		O23553	Beta-amylase 3, chloroplastic	548	47.2%	N/A

3.6.2 Homology modeling

SWISS-MODEL software is a server for modeling three-dimensional comparative structure of proteins [24] that are used for prediction tertiary structure of β-amylase protein in *Clostridium thermosulfuregen*. As shown in the Fig. 5, the structure has many beta sheets, which play an important role in protein stability. The modelled structure (Fig. 5) showed high similarity to the template β-amylase from *B.* cereus (PDB ID: 1B90).

3.7 Homology Modeling validation

ERRAT2: ERRAT analyses the statistics of non-interconnected interactions between different types of atoms and breaks down the amount of error performance against the position of a non-residual sliding window in a database. Since different types of atoms are distributed randomly in proteins, the structure of the protein can be verified by distinguishing between correct and incorrect regions of the protein structure based on the interaction of the atomic characteristic. The overall quality factor of the structure must be higher than 91% in order to be considered as qualified protein structure. A high-resolution structure generally produces an overall quality factor of about 95 [25]. The overall quality factor of the predicted thermophilic β-amylase from *C. thermosulfuregen* is 913894 (Fig. 6).

Procheck: Ramachandran plot was used to validate the accuracy of the 3D structure model by visualizing the phi (Φ) and psi (ψ) dihedral angles of amino acid residues in protein structure The Ramachandran scheme shows the validity of the thermophilic β-amylase protein prediction model from the SWISS Modell as shown below in Fig. 7.

As Fig. 7 shows, there are 393 residues in the favoured region (86.8%) and the red dots on charts are less, which means that model protein is very accurate. In addition, the Residues in disallowed regions shows less than 1% and this shows the accuracy of the model (Table 7).

Verify 3D: The Verify3D program determines the compatibility of the 3D model's atomic coordinates with its amino acid sequence (1D). The three-dimensional profiles calculated from the correct protein structures match their sequences with high scores [26]. The model structure is passed from validation if at least 80% of the amino acids have scored greater or equal to 0.2 in the 3D/1D profile. Quality of the 3D structure homology of the thermophilic β-amylase protein using Verify3D is shown in Fig 8. From the results obtained, at least 80% of the amino acids have scored > 0.2 in the 3D/1D profile.
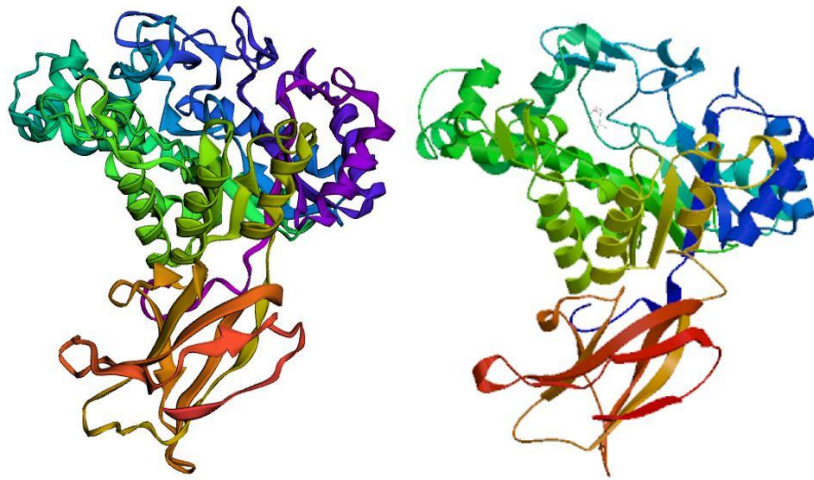
Fig. 5. (Left) Modeled structure of thermophile β-amylase from *C. thermosulfuregen* as modelled using the SWISS-MODEL software. (Right) 3D structure of β-amylase protein from *B. cereus* (pdb id: 1B90)
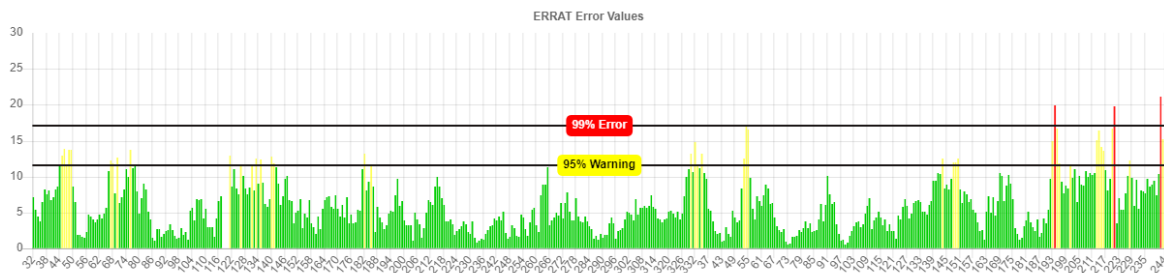

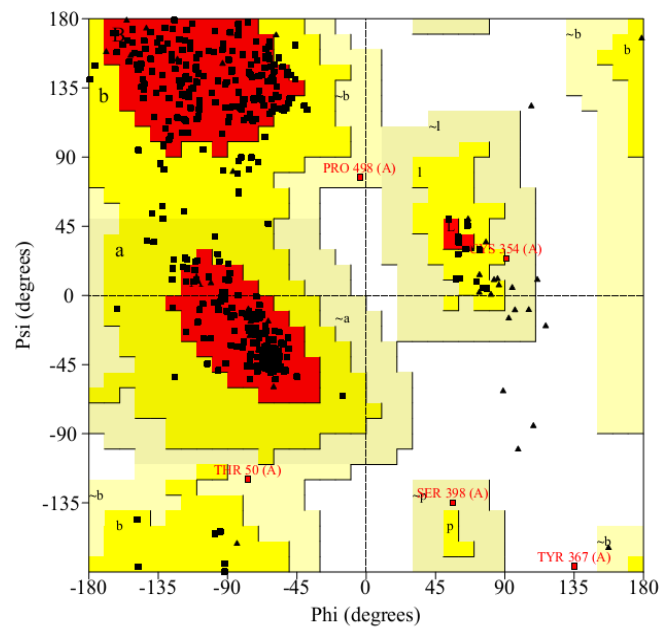**Fig. 6. Tertiary structure validation of modelled protein**

Fig. 7. Tertiary structure validation of modeled β-amylase using Ramachandran plot tool. The non-colored areas are disallowed regions that are very small

Table 7. Ramachandran plot validation percentage

Evaluation of residues	Score	%
Residues in most favoured regions [A,B,L]	393	86.8
Residues in additional allowed regions [a,b,l,p]	56	12.4
Residues in generously allowed regions [~a,~b,~l,~p]	3	0.7
Residues in disallowed regions	1	0.2
Number of non-glycine and non-proline residues	453	100.0
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	41	
Number of proline residues	26	
Total number of residues	522	
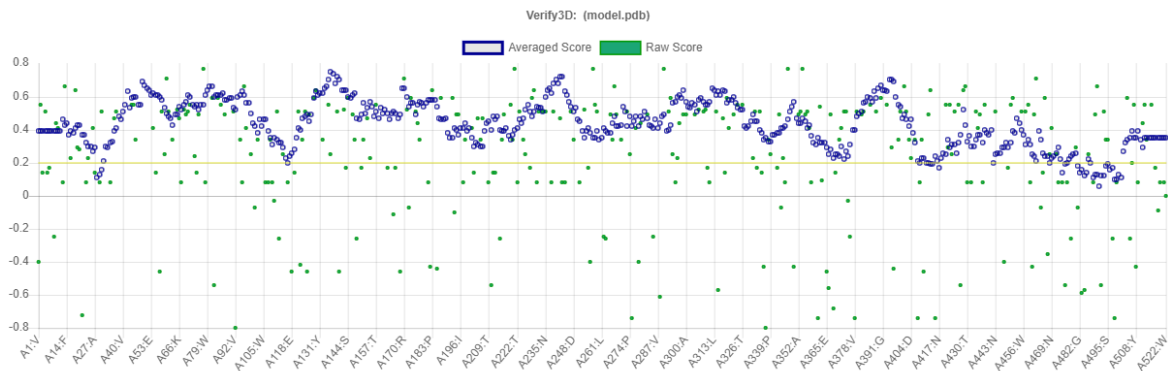
Fig. 8. Tertiary structure validation of modelled β amylase using Verify 3D program
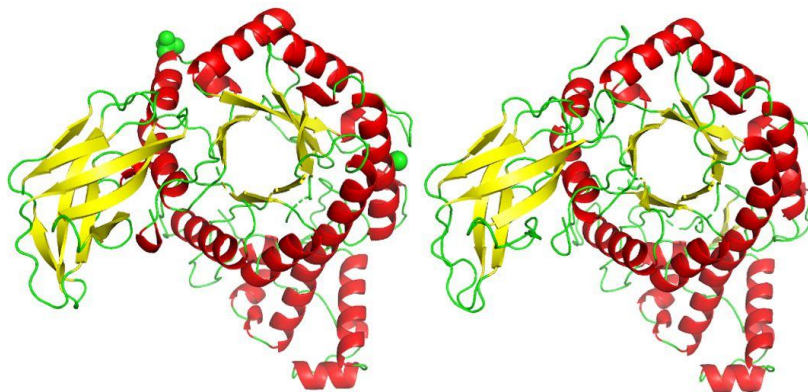

Fig. 9. Cartoon representation of the template 1B90 (right) and modelled (left) with the beta sheets in yellow and helices in red

3.8 Structural Comparison

Size and shape: The results show both β-amylase of *C. thermosulfuregen* and *B. cereus* are highly similar and have very little differences. This is expected as the β-amylase of *C. thermosulfuregen* is also monomer with only 3 amino acid difference in the length of the peptide chain. The shape of the modelled structure corresponds similarly to the shape of the template. The beta barrel located in the centre of the enzyme are highly similar in share, size and orientation, suggesting that the thermal adaptation does not involve reduction in size (Fig. 9).

Active site prediction: There is no previous report on the location of active site residues for the β-amylase of *C. thermosulfuregen*. The 1B90 structure was used as a reference to determine

the possible location of active site amino acid residues on the modelled protein. Two residues E202 and E397, have been reported as active site for the protein template in the UniProt (P36924). In the PDB database, 1-30 residues of the sequence β-amylase template (1B90) represented as a signal peptide and excluded from the structure. Therefore, the amino acids E172 and E367 are the active site for the template β-amylase. Multiple sequence alignment of the modelled protein (P19584) of *C. thermosulfuregen* with sequence of β-amylase from *B. cereus* (P36924) revealed that the active

site of the modelled β-amylase are located on the same sequence region (E195 and E392) and within highly conserved segment of the sequence (Fig. 10).

The location of the amino acid residues forming the active site are identified on the 3D model of the *C. thermosulfuregen* and compared with the template 1B90. (Fig. 11). The location and the orientation of the r-group of each amino acid are identical between the *B. cereus* and of *C. thermosulfuregen*.
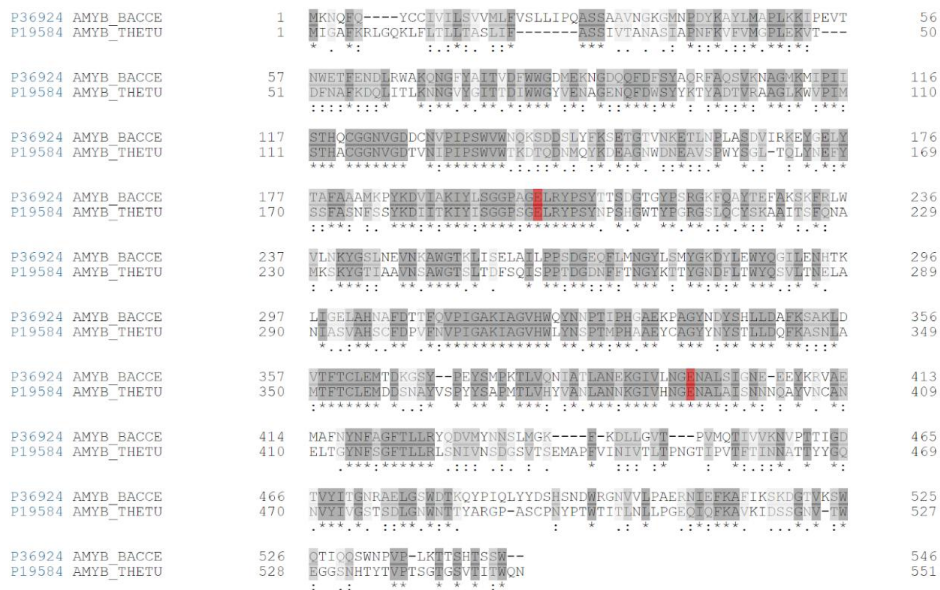

Fig. 10. Multiple sequence alignment of the sequence P19584 of *C. thermosulfuregen* (modelled protein) with the sequence of P36924 of *B. cereus* (templet protein). The active site is highlighted in red colour
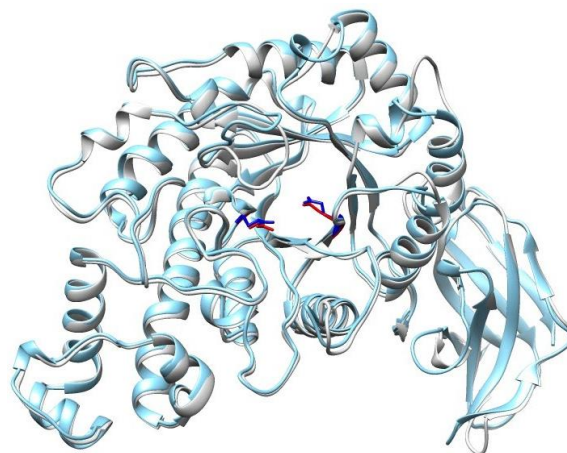

Fig. 11. Cartoon representation of the template protein 1B90 (blue color) and modelled enzyme (gray color) with the enzyme active site of 1B90 (E172 and E367; dark blue) and the model protein (E167 and E364; red color) highlighted

Binding Sites: Binding sites are small pockets of the tertiary structure in which ligands attach to it using non-covalent bonding. Protein binding sites play an important role in a wide range of applications, including molecular binding, drug design, structure identification and comparison of functional sites [27]. The seven amino acids that form the binding site for both the template protein β-amylase of *Bacillus cereus* and the *C. thermosulfuregen* model was obtained from Uniprot database and listed in Table 8.

The result shows that the amino acids involved in bindings are identical between the model and template, suggesting that the thermal adaptation did not cause changes in the binding domain of the enzyme (Fig. 12).

Measurement of the binding sites distances between amino acids in the template and model are shown in Fig. 13.

As mentioned, in the template protein, the amino acids Gly 79, Lys 119,127,360, Leu 317 and Ser 322 play a role in substrate binding site formation and in the modelled protein, the amino acids Asp 73,121, His 113,315, Lys 310,Thr 353 and Arg

423 are form the substrate binding site. In the template protein, the distance between the amino acids that make up the site is 6.48 Å and in the modelled protein the distance Asp73 with Lys310 is 6.325901 Å, Arg423 with Asp121 is 7.311912 Å, Thr353 with Arg423 is 9.846662 Å, Thr353 with Lys310 is 6.229181 Å, Asp73 with Asp121 is 6.501258 Å, Lys310 with Arg423 11.152993 Å and the distance between Asp121 and ligand and the site is modelled in the protein 9.90 Å.

Metal Binding Sites: One small difference between template and modelled protein is metal binding sites. In the template protein, the amino acids Glu86, Asp90, Gln91, Glu171, and Glu174 play a role in the formation of the metal binding site, but in the modelled protein, amino acids Glu 80 and Glu167 make the metal binding site. In template protein, the distance between the amino acids that make up the metal binding site is 12.8 Å and in modelled protein, this distance is 5.15 Å. The distance between ligands with the site, in modelled protein, is 33.01 Å and template protein, it is 12.23 Å.

Table 8. Location of amino acids that are involved with the binding of substrates in β-amylase of *Bacillus cereus* and for modeledled β-amylase of *C. thermosulfuregen*

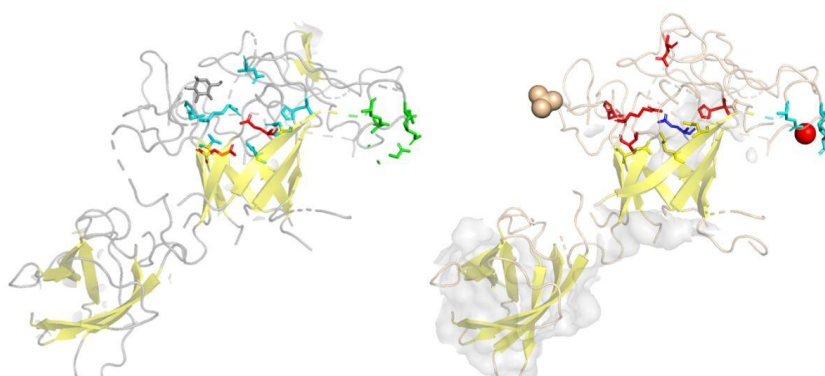| No	*B.cereus*	*C. thermosulfuregen*
1	Asp (D) 79	Asp (D) 73
2	His (H) 119	His (H) 113
3	Asp (D) 127	Asp (D) 121
4	Lys (K) 317	Lys (K) 310
5	His (H) 322	His (H) 315
6	Thr (T) 360	Thr (T) 353
7	Arg (R) 427	Arg (R) 423

Fig. 12. The cartoon representation of the amino acid involved in the binding site of β-amylase of *C. thermosulfuregen* (left): binding residues in cyan, active site in red and beta sheets in yellow and the template 1B90 (right), binding residues in red, active site in blue and beta sheets in yellow
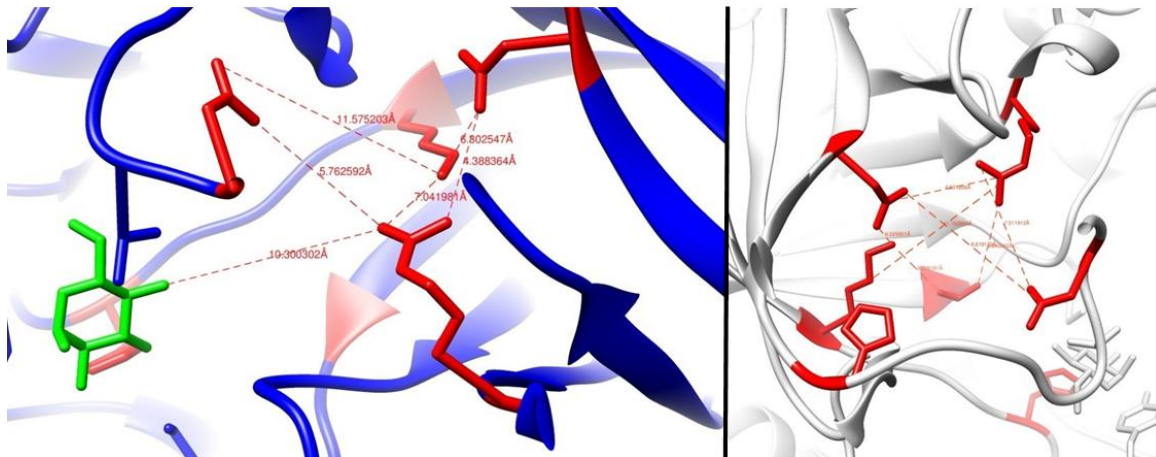
Fig. 13. Th cartoon show the distance between amino acids that form the binding sites and their relative positions in templet protein structure (1B90) and in modeled protein of *C. thermosulfuregen* (P19584)

Both amino acids Asp and Glu are polar with negative charge, so they have same feature [28]. As a result, in model protein amino acids Glu80, Glu84, Asn85, Leu164, Glu167 and in template protein amino acids Glu86, Asp90, Gln91, Glu171, and Glu174 contributing metal binding site (Table 9). Two missing information on the binding residues from the Uniprot database can now be identified using sequence and structural alignment (Asn and Leu) (Fig. 15). As the result of the sequence and structure align shows, the position of metal binding site in both model and template proteins are similar, the difference in metal binding site is 3 residues Asp90, Gln91, and Glu171of template protein (Fig. 14. Modeled (P19584) and template (P36924) proteins metal binding site (red highlights).

3.9 Comparison of Protein Thermostability

Comparison of the amino acids composition showed that there are only changes in the number of charged, aliphatic and aromatic residues. Aliphatic amino acids, which consists of Ala, Ile, Leu, and Val are involved in the formation of hydrophobic interaction [29] [30]. It is expected that the amount of Ile in the thermophilic protein is higher than mesophilic. Charged amino acids, which include Arg, Asp, His, Glu and Lys play a role in the formation of electrostatic interaction [29] [30], and Glu and Arg play a role in the formation of the salt bridge interaction in proteins. Aromatic amino acids like Phe, Trp, and Tyr have role in the protein thermostability. It is expected that the percentage of Tyr in the thermophilic protein has been

shown to be higher in thermophilic than mesophilic protein. Polar amino acids (including Asn, Gln, Ser and Thr) because of interacting with water; can reduce the stability of the protein. The amount of polar amino acids in thermophilic protein is expected to be lower than that of mesophilic protein. In addition, certain amino acids (including Cys, Gly, Met and Pro) undergo oxidation at high temperature [31] and expected to be reduced in thermophilic proteins.

The result of the Protparam database showed that not all the thermal adaptation characteristics literature reported are observed in the β-amylase of *C. thermosulfuregen*. The percentage of aliphatic and aromatic amino acids does not differ significantly, but polar and charged amino acids showing significant differences (Table 10). Polar amino acids would decrease protein thermostability by interacting with water, so it is expected that the amount of polar amino acids in thermophilic proteins is less than the mesophilic proteins. In addition, Cys and Met undergoes oxidation at high temperature and considered as thermolabile amino acids and would reduce stability of proteins [31]. Kumar *et al.* [30] have reported that Cys has lower frequency in thermophilic proteins than mesophiles. The result showed that there is a 50% reduction of Met (1.7% in mesophilic versus 2.7% in thermophilic), the opposite was observed in Cys where there is an increase 0.6% in mesophilic versus 1.3% in thermophilic. The composition of amino acids that have been observed to reduce in thermophilic protein (Gly 7.8%, and Pro 4.8%) is similar to the composition in the mesophilic beta-amylase (Gly 7.9%, and Pro 5.0%). This

suggests that the structure of β-amylase of *C. thermosulfuregen* does not exhibit all previously reported thermal adaptations.

Mapping of the locations of hydrophobic (Fig. 16), hydrophilic (Fig. 17) and polar amino acids (Fig. 18) on the surface indicates no major differences in the distribution pattern between the mesophilic and thermophilic β-amylase.

Phylogenetic Study: The full sequence was searched using the NCBI BlastP server. The alignment results have been used to construct a phylogenetic tree using the neighbour connection methods used in the MEGA X program [22]. Fig. 19 shows the phylogenetic tree of bacterial proteins. Based on the result obtained, the template protein is closest to *Clostridium thermosulfuregen* among all five species in the phylogenetic tree (Fig. 19).

Table 9. Amino acids contributing metal binding sites of both model and template proteins

β-amylase of *B. cereus*	β-amylase *C. thermosulfuregen*
Glu86, Asp90, Gln91, Glu171, and Glu174: Calcium ion binding site	Glu80, Glu84, Asn85, Leu164, and Glu167: Calcium ion binding site

```
P36924 AMYB_BACCE      1   MKNQFQ----YCCIVILSVVMLFVSLLIPQASSAAVNGKGMNPDYKAYLMAPLKKIPEVT
P19584 AMYB_THETU      1   MIGAFKRLGQKLFLTLLTASLIF-------ASSIVTANASIAPNFKVFVMGPLEKVT---
                          *  .  *:        :.:*:. ::*      *** .. .  .:  *::*.::*.**:*:

P36924 AMYB_BACCE     57   NWETFENDLRWAKQNGFYAITVDFWWGDMEKNGDQQFDFSYAQRFAQSVKNAGMKMIPII
P19584 AMYB_THETU     51   DFNAFKDQLITLKNNGVYGITTDIWWGYVENAGENQFDWSYYKTYADTVRAAGLKWVPIM
                          :::*::*:::*   *:**.*.**.*:*** :*: *::***:** : :*:*:.**:* :**:

P36924 AMYB_BACCE    117   STHQCGGNVGDDCNVPIPSWVWNQKSDDSLYFKSETGTVNKETLNPLASDVIRKEYGELY
P19584 AMYB_THETU    111   STHACGGNVGDTVNIPIPSWVWTKDTQDNMQYKDEAGNWDNEAVSPWYSGL-TQLYNEFY
                          *** ******* *:*******:.::*.: :*.*:*. ::*::.*  *.: : *.*:*

P36924 AMYB_BACCE    177   TAFAAAMKPYKDVIAKIYLSGGPAGELRYPSYTTSDGTGYPSRGKFQAYTEFAKSKFRLW
P19584 AMYB_THETU    170   SSFASNFSSYKDIITKIYISGGPSGELRYPSYNPSHGWTYPGRGSLQCYSKAAITSFQNA
                          ::**: :. ***:*:***:****:********. *.* **.**.:*.*:: * :.*:

P36924 AMYB_BACCE    237   VLNKYGSLNEVNKAWGTKLISELAILPPSDGEQFLMNGYLSMYGKDYLEWYQGILENHTK
P19584 AMYB_THETU    230   MKSKYGTIAAVNSAWGTSLTDFSQISPPTDGDNFFTNGYKTTYGNDFLTWYQSVLTNELA
                          : .***:: **.****.*.   * **:**::*: *** : **:*:* ***.:* *.

P36924 AMYB_BACCE    297   LIGELAHNAFDTTFQVPIGAKIAGVHWQYNNPTIPHGAEKPAGYNDYSHLLDAFKSAKLD
P19584 AMYB_THETU    290   NIASVAHSCFDPVFNVPIGAKIAGVHWLYNSPTMPHAAEYCAGYYNYSTLLDQFKASNLA
                          *..:**..** .*:************** **.**.** *** :** *** **:::*

P36924 AMYB_BACCE    357   VTFTCLEMTDKGSY--PEYSMPKTLVQNIATLANEKGIVLNGENALSIGNE-EEYKRVAE
P19584 AMYB_THETU    350   MTFTCLEMDDSNAYVSPYYSAPMTLVHYVANLANNKGIVHNGENALAISNNNQAYVNCAN
                          :******* *..:*  * ** * ***: :*.***:**** *****:*::* :  . .*:

P36924 AMYB_BACCE    414   MAFNYNFAGFTLLRYQDVMYNNSLMGK----F-KDLLGVT---PVMQTIVVKNVPTTIGD
P19584 AMYB_THETU    410   ELTGYNFSGFTLLRLSNIVNSDGSVTSEMAPFVINIVTLTPNGTIPVTFTINNATTYYGQ
                          . .***:****** .::: .:. :   *  *  .:.:: :*   *..:*:*. * *:

P36924 AMYB_BACCE    466   TVYITGNRAELGSWDTKQYPIQLYYDSHSNDWRGNVVLPAERNIEFKAFIKSKDGTVKSW
P19584 AMYB_THETU    470   NVYIVGSTSDLGNWNTTYARGP-ASCPNYPTWTITLNLLPGEQIQFKAVKIDSSGNV-TW
                          .***.*. ::**.*:*.       :  * .: * .:*:***. ...*.* :*

P36924 AMYB_BACCE    526   QTIQQSWNPVP-LKTTSHTSSW--
P19584 AMYB_THETU    528   EGGSNHTYTVPTSGTGSVTITWQN
                          :  .:   **   * * * :*
```

Fig. 14. Modeled (P19584) and template (P36924) proteins metal binding site (red highlights)
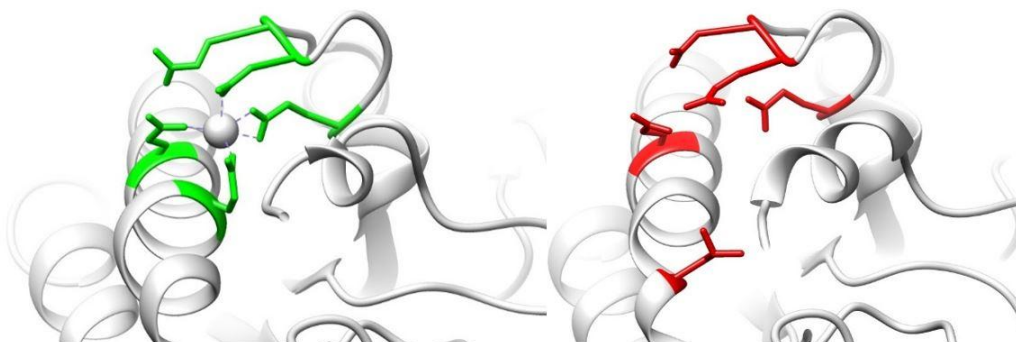

Fig. 15. Model (right; red) and template (red; green) proteins metal binding site

16

Table 10. The percentage of major amino acid groups in *C. thermosulfuregen* and *B. cereus*

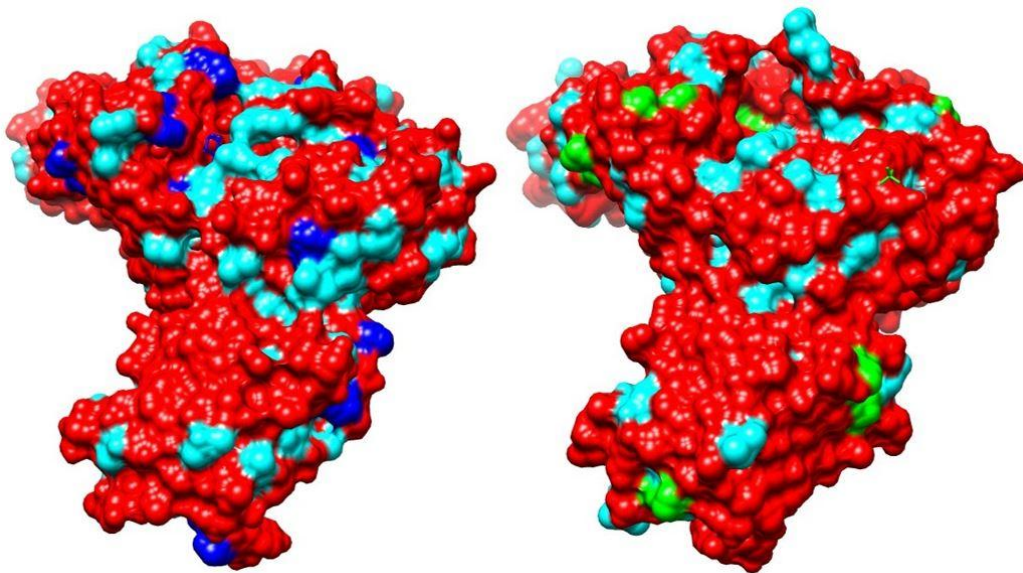| The percentage of amino acids	*C. thermosulfuregen* (%)	*B. cereus (%)*
Aliphatic amino acids	17.6	18
Charged amino acids	2.9	5.4
Aromatic amino acids	15.7	14.8
Polar amino acids	44.7	46

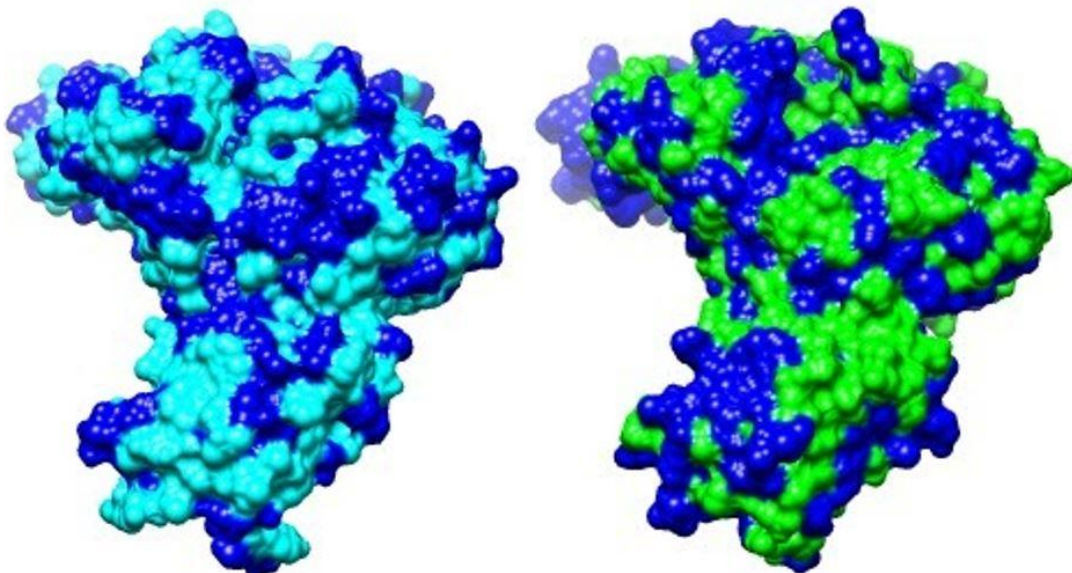**Fig. 16. Shows hydrophobic (red part) and polar (light blue) amino acid residues in both model (blue) and template (green) proteins**

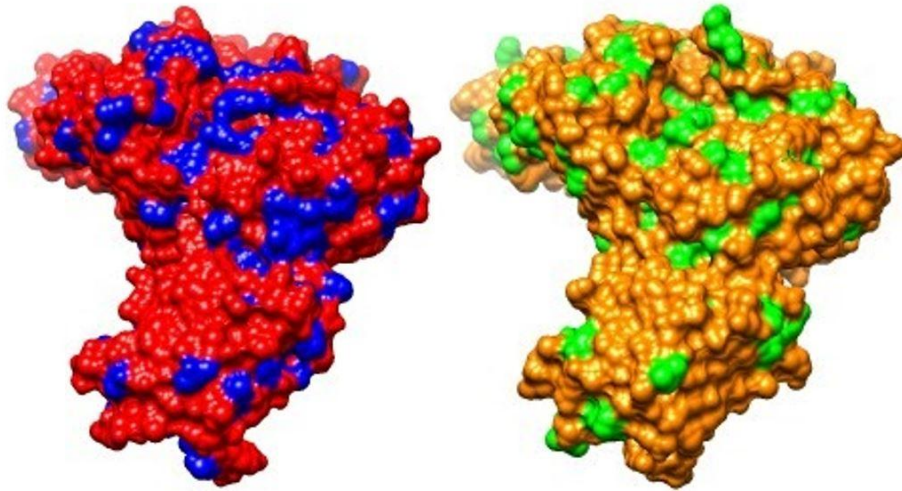**Fig. 17. Shows hydrophilic amino acid residues in both model (light blue) and template (green) proteins**

Fig. 18. Shows polar amino acid residues in both model (red part) and template (yellow part) proteins
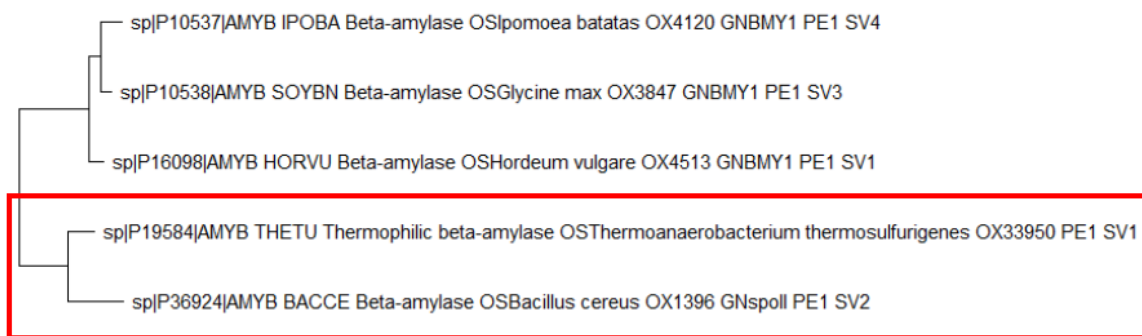

sp|P10537|AMYB IPOBA Beta-amylase OSIpomoea batatas OX4120 GNBMY1 PE1 SV4

sp|P10538|AMYB SOYBN Beta-amylase OSGlycine max OX3847 GNBMY1 PE1 SV3

sp|P16098|AMYB HORVU Beta-amylase OSHordeum vulgare OX4513 GNBMY1 PE1 SV1

sp|P19584|AMYB THETU Thermophilic beta-amylase OSThermoanaerobacterium thermosulfurigenes OX33950 PE1 SV1

sp|P36924|AMYB BACCE Beta-amylase OSBacillus cereus OX1396 GNspoII PE1 SV2

Fig. 19. Summary of Phylogenetic tree result using maximum likelihood tree neighbour-joining method from Mega X

Secondary structural comparison: The result showed that although the overall conformational structure of the protein is very similar. The comparison of helices, strands and loop properties indicates changes that may constitute part of the thermal adaptations.

Hydrogen bonds: Results of the hydrogen bonding comparison between the model and template showed that there are no significance difference. Using hydrogen bond calculating software http://cib.cf.ocha.ac.jp/bitool/HBOND/, the result showed that the model has 522 hydrogen bonds compared to 516 hydrogen bonds in the template, and that difference in the amino acid numbers did not produce a significant nett changes of hydrogen bond numbers, suggesting that additional hydrogen bond formations is not an adaptation to temperature.

Helices: Helices play an important role in the maintenance of the conformational structure, how the enzyme performs and the transfer between different conformational modes in nucleus regulation and remodeling [32]. The result showed that the percentage of helices in modelled protein is 51.20% compared to template protein that is 71.60% (Fig. 20). The TIM barrel which is a conserved protein architecture consisting of eight α-helices and eight parallel β-strands that alternate along the peptide backbone is similar between the two proteins. The helices of the mesophilic structure are shorter compared to the model.

Strands: The result showed that there is a comparable difference between the modelled and template proteins based on amount of beta-strands that forms a stable structure for protein [23]. The modelled protein possess a

significantly higher percentage of beta-sheets (72.10%) compared to the mesophilic template protein (45.60%) (Fig. 21). This increase in beta-sheets in thermophilic variants have been reported as thermal adaptation strategies [23]. However, it can be conclude that higher percentage of beta-sheets in modelled protein is one of the reason for thermal adaptation of this modelled protein.

Loop: Prior researches has shown that thermophiles can have different loop lengths between secondary structure elements. The loop is a flexible part of a polypeptide cross-linking chain that binds two secondary structures in a protein. The loop region plays an important role in protein function, including the involvement of catalytic sites in enzymes, facilitating molecular recognition and participation in ligand binding sites [33]. As the GOR VI result showed, the amount of loop in the two proteins is significantly different; in the template protein, it is 47.44% and in the [34-37] model protein, it is 56.99%. There are significant changes such as shortening in loop length (Fig. 22).
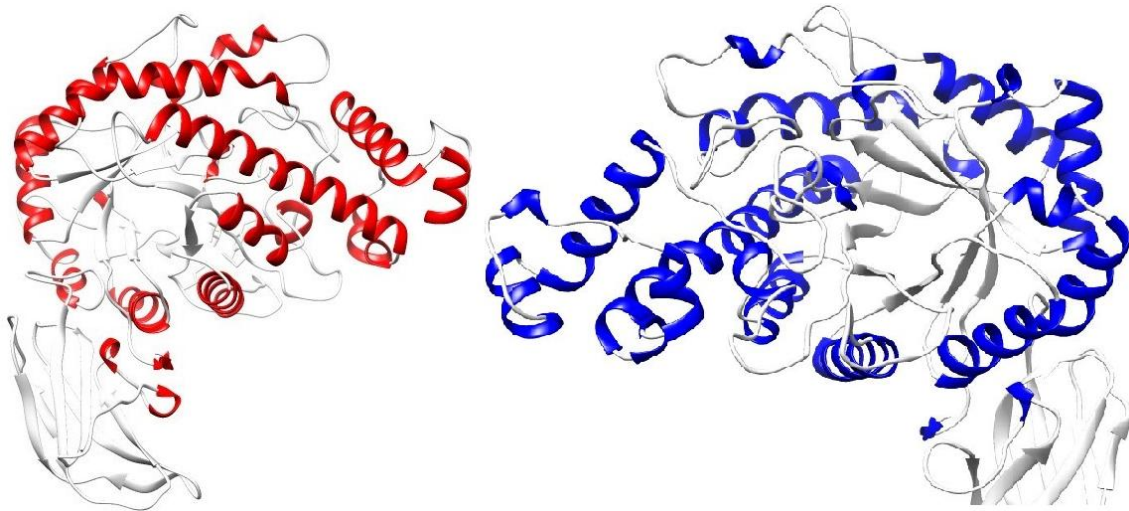

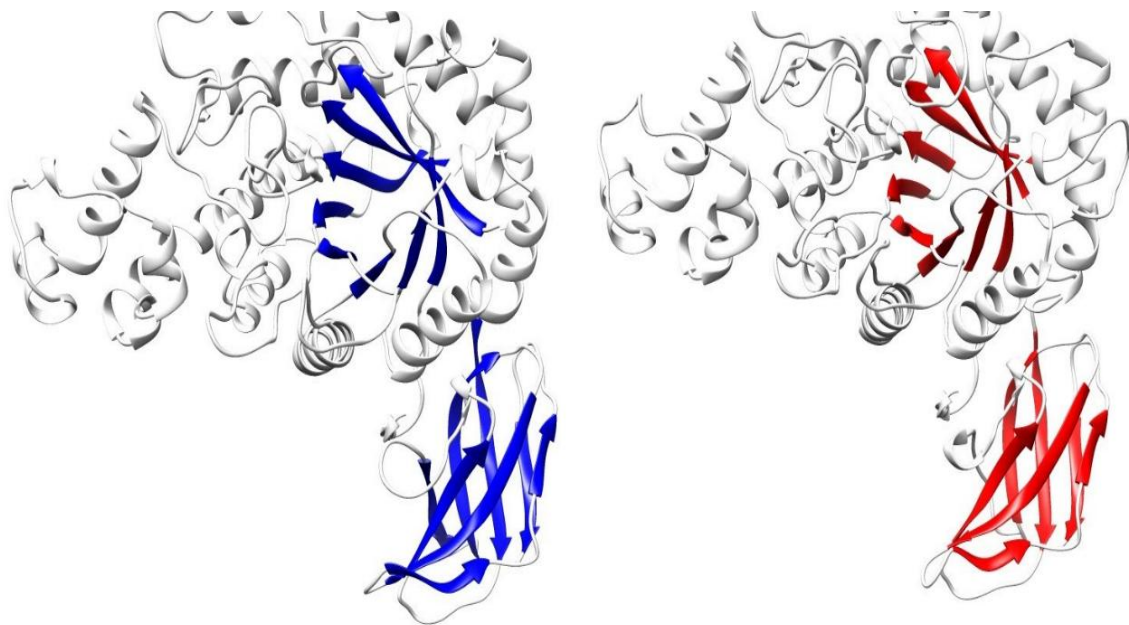**Fig. 20. Comparison of helixes structure of modelled protein (red) template protein (blue)**

Fig. 21. Comparison of the beta-strand content and orientation between the thermophilic model proteins (red) against the mesophilic template protein (blue)
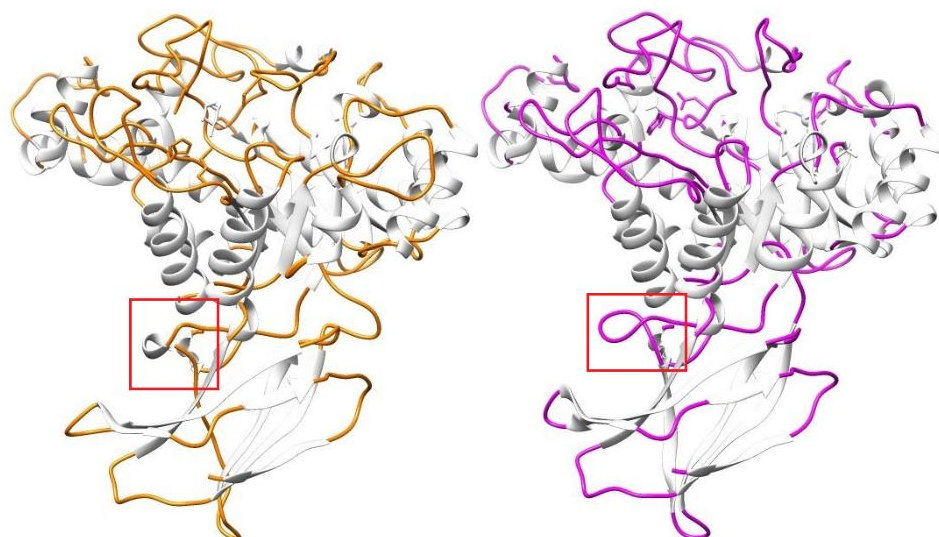
Fig. 22. Comparison of loop structure of modelled protein (yellow) against template protein (purple). The red box denotes one area highlighting significant difference in loop length where the thermophilic protein has a shorter loop length at the particular region of the protein

4. CONCLUSION

The knowledge of protein 3D (three-dimensional) structures is vitally important for protein engineering design. The 3D models of P19584 were generated by SWISS MODELL using P36924 (PDB ID: 1B90) as a template. The results revealed that the modelled β-amylase of *C. thermosulfuregen* is very similar with respect to the global conformation, location of active and binding sites. Both proteins report identical structural domains with the thermophilic variant possessing high percentage of hydrophobic amino acid residues (43.4%), polar amino acid residues (44.1%) and differences in secondary composition such as loops and beta sheets as the potential evolutionary thermal adaptations that make it stable enzyme that functions up to 70 ℃.

Overall, considering the common properties of this modelled protein with the β-amylase of *B. cereus*, it can be assumed that if the β-amylase of *C. thermosulfuregen* is expressed in-vitro, it would produce a stable protein that possess the hydrolysis function for *C. thermosulfuregen* to break down the starch and sugar formation.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Schomburg I, Chang A, Hofmann O, Ebeling C, Ehrentreich F, Schomburg D. BRENDA: A resource for enzyme data and metabolic information. Trends in Biochemical Sciences. 2002;27(1):54–56. Available: https://doi.org/10.1016/S0968-0004(01)02027-8

2. Saini R, Singh Saini H, Dahiya A, Harnek Singh Saini C. Amylases: Characteristics and industrial applications. Journal of Pharmacognosy and Phytochemistry. 2017;6(4):1865–1871.

3. Martin MF, Okpo EA, Andy IE. Microbial amylases : a review. World News of Natural Sciences. 2019;22:174–179.

4. Sarikaya E, Higasa T, Adachi M, Mikami B. Comparison of degradation abilities of α- and β-amylases on raw starch granules. Process Biochemistry. 2000;35(7):711–715.

5. Hyun HH, Zeikus JG. General biochemical characterization of thermostable extracellular β-amylase from Clostridium thermosulfurogenes. Applied and Environmental Microbiology. 1985;49(5): 1162–1167. Available:https://doi.org/10.1128/aem.49.5.1162-1167.1985

6. B, Zeikus JG. Clostridium thermosulfurogenes sp. nov., a new thermophile that produces elemental

sulphur from thiosulphate. Journal of General Microbiology. 2014;129: 1149–1158.

7. Shen GJ, Saha BC, Lee YE, Bhatnagar L, Zeikus JG. Purification and characterization of a novel thermostable β-amylase from Clostridium thermosulphurogenes. Biochemical Journal. 2004;254(3):835–840.

8. Barnaud G, Arlet G, Danglot C, Philippon A. Cloning and sequencing of the gene encoding the AmpC β-lactamase of Morganella morganii. FEMS Microbiology Letters. 1997;148(1):15–20.

9. Hirata A, Adachi M, Utsumi S, Mikami B. Engineering of the pH optimum of Bacillus cereus β-amylase: Conversion of the pH optimum from a bacterial type to a higher-plant type. Biochemistry. 2004;43(39):12523–12531.
Available:https://doi.org/10.1021/bi049173h

10. Mikami B, Adachi M, Kage T, Sarikaya E, Nanmori T, Shinke R, Utsumi S. Structure of raw starch-digesting Bacillus cereus β-amylase complexed with maltose. Biochemistry. 1999;38(22):7050–7061.
Available:https://doi.org/10.1021/bi9829377

11. Cheong CG, Eom SH, Chang C, Shin DH, Song HK, Min K, Suh, SW. (1995). Crystallization, molecular replacement solution, and refinement of tetrameric β- amylase from sweet potato. Proteins: Structure, Function, and Bioinformatics. 1995;21(2):105–117.
Available:https://doi.org/10.1002/prot.340210204

12. Rejzek M, Stevenson CE, Southard AM, Stanley D, Denyer K, Smith AM, Field RA. Chemical genetics and cereal starch metabolism: Structural basis of the non-covalent and covalent inhibition of barley β-amylase. Molecular BioSystems. 2011;7(3):718–730.
Available:https://doi.org/10.1039/c0mb00204f

13. Hofer G, Wieser S, Bogdos MK, Gattinger P, Nakamura R, Ebisawa M, Keller W. Three-dimensional structure of the wheat β-amylase Tri a 17, a clinically relevant food allergen. Allergy: European Journal of Allergy and Clinical Immunology. 2019;74(5):1009–1013.
Available: https://doi.org/10.1111/all.13696

14. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Zhang J. UniProt: The universal protein knowledgebase. Nucleic Acids Research. 2017;45(D1): D158–D169.

15. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Zaretskaya I. BLAST: a more efficient report with usability improvements. Nucleic Acids Research. 2013;41(Web Server issue):29–33.
Available:https://doi.org/10.1093/nar/gkt282

16. Gasteiger E, Wilkins MR, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. In The proteomics protocols handbook. 20055;112:571–607.
Available: https://doi.org/10.1385/1-59259-584-7:531

17. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Science. 2018;27(1):135–145.
Available: https://doi.org/10.1002/pro.3290

18. Kumar TA. CFSSP: Chou and fasman secondary structure prediction server. Wide Spectrum. 2013;1(9):15–19.

19. Garnier J, Gibrat JF, Robson B. [32] GOR method for predicting protein secondary structure from amino acid sequence. Methods in Enzymology. 1996;266(1995)540–553.
Available: https://doi.org/10.1016/s0076-6879(96)66034-0

20. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Schwede T. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research. 2014;42(W1):252–258.
Available:https://doi.org/10.1093/nar/gku340

21. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry. 2004;25(13):1605–1612.
Available: https://doi.org/10.1002/jcc.20084

22. Kumar, Sudhir, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution. 2016;33(7):1870–1874.
Available:https://doi.org/10.1093/molbev/msw054

23. Siepen JA, Radford SE, Westhead DR. β Edge strands in protein structure prediction and aggregation. Protein Science. 2009;12(10):2348–2359.

24. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Research. 2003;31(13):3381–3385. Available:https://doi.org/10.1093/nar/gkg520

25. Singh S, Jha A. Structural characterisation of 5-hydroxytryptamine2a receptor in homo sapiens by in-silico method. Asian Journal of Pharmaceutical and Clinical Research. 2018;11(Special Issue 2):81–85. Available:https://doi.org/10.22159/ajpcr.2018.v11s2.28588

26. Eisenberg D, Luthy R, Bowie JU. [20] VERIFY3D: assessment of protein models with three-dimensional profiles. Methods in Enzymology. 1997;277:396–404.

27. Roche DB, Brackenridge DA, McGuffin LJ. Proteins and their interacting partners: An introduction to protein-ligand binding site prediction methods. International Journal of Molecular Sciences. 2015;16(12): 29829–29842. Available:https://doi.org/10.3390/ijms161226202

28. Sachse S, Roeder C. A Classification Scheme of Amino Acids in the Genetic. 2013-2014;0–20.

29. Chakravarty S, Varadarajan R. Elucidation of determinants of protein stability through genome sequence analysis. FEBS Letters. 2000;470(1):65–69. Available:https://doi.org/10.1016/S0014-5793(00)01267-9

30. Kumar, Sandeep, Tsai CJ, Nussinov R. Factors enhancing protein thermostability. Protein Engineering. 2000;13(3):179–191. Available:https://doi.org/10.1093/protein/13.3.179

31. Russell RJ, Hough DW, Danson MJ, Taylor GL. The crystal structure of citrate synthase from the thermophilic Archaeon, Thermoplasma acidophilum. Structure. 1994;2(12):1157–1167. Available: https://doi.org/10.1016/S0969-2126(94)00118-9

32. Facchiano AM, Colonna G, Ragone R. Helix stabilizing factors and stabilization of thermophilic proteins: An X-ray based study. Protein Engineering. 1998;11(9):753–760. Available: https://doi.org/10.1093/protein/11.9.753

33. Li Y. Conformational Sampling in Template-Free Protein Loop Structure Modeling : An Overview Abstract : Accurately modeling protein loops is an important step to predict three-dimensional structures as well as to understand functions of many proteins . Because. Comput Struct Biotechnol J. 2013;5(6): e201302003. Available:https://doi.org/10.5936/csbj.201302003

34. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. Protein Science. 2017;27(1):129–134. Available: https://doi.org/10.1002/pro.3289

35. Luthy R, Bowei J, Einsenberg D. Verify3D: Assessment of protein models with three-dimensional profiles. Methods in Enzymology. 1997;277:396–404.

36. Takasaki Y. Purifications and enzymatic properties of β-amylase and pullulanase from Bacillus cereus var. mycoides. Agricultural and Biological Chemistry. 2005;40(8):1523–1530.

37. Vajravijayan S, Pletnev S, Mani N, Pletneva N, Nandhagopal N, Gunasekaran K. Structural insights on starch hydrolysis by plant β-amylase and its evolutionary relationship with bacterial enzymes. International Journal of Biological Macromolecules. 2017,2018;113:329–337. Available:https://doi.org/10.1016/j.ijbiomac.2018.02.138