Tertiary structure prediction of bromelain from *Ananas Comosus* using comparative modelling method

Fatahiya Mohamed Tap [1], Fadzilah Adibah Abd Majid [2, *], Nurul Bahiyah Ahmad Khairudin [1]

[1]Chemical Energy Conversion and Applications Group, Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
[2]Institute of Marine Biotechnology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia

A R T I C L E I N F O

A B S T R A C T

Bromelain is a general name for a family of sulfhydryl which can be found in the proteolytic enzyme group from pineapples (*Ananas Comosus*). This study focuses on the prediction of three dimensional structures (3D) of stem bromelain using the method of comparative modelling. The amino acid sequence of bromelain was obtained from the NCBI database was used as a tool to search for proteins with known 3D structures related to the target sequence. Suitable template was chosen based on >30% sequence similarity and lowest e-value. Based on these criteria, 1YAL was selected as the best template with 55% of sequence similarity and 9×10^{-52} of e-value.

1. Introduction

Bromelain can be derived from stem and fruit of pineapples. Stem bromelain (EC.3.4.22.32), ananain (EC.3.4.22.31) and comosain were isolated from the stem juice of pineapples by precipitation and centrifugation methods (Gautam et al., 2010). Bromelain also is classified as a protease because it refers to the group of enzyme which belongs to the catalytic function that hydrolyses the peptide bonds of proteins. In the presence of oxidizing agents the sulfhydryl group of cysteine will be oxidized and lead to the formation of disulphide bond. The effects of the oxidation process of the cysteine group will alter the important properties in substrate binding, structure stabilization, and thus cause changes in active site which results in the loss of its catalytic activity. Besides that, bromelain draws high attention in industrial applications such as meat tenderization, baking industry, anti-browning agent, protein hydrolysate, alcohol production, textile industry and cosmetic industry because of its unique properties and rich with proteolytic activity (Arshad et al., 2014). Bromelain has also widely used in therapeutic applications such as platelet aggregation, fibrinolysis, anti-inflammatory activity, modulation of cell adhesion and antibodies (Pavan et al., 2012). Since bromelain has great commercial values in both industrial and pharmacology area, therefore it is essential to maintain and improve bromelain properties in conformation and catalytic activity in order to preserve its industrial and therapeutic values. Even though researchers have discovered many methods to enhance the extraction, purification, optimum condition and the specific activity of bromelain for different substrates, the mechanism of this enzyme is still in question because of the unknown three dimensional structures. Therefore, this study was carried out to predict three-dimensional (3D) structure of Bromelain from stem pineapple by comparative modelling.

2. Materials and methods

2.1. Bromelain sequence retrieval

The amino acid sequence of stem bromelain was retrieved from the NCBI's protein database [Gene Bank Accession Numbers: ADY68475] (Pruitt et al., 2002). This target sequence has 291 amino acid residues.

2.2. Template selection, 3D model development, binding site prediction, and model assessments

Protein Basic Local Alignment Search Tool (BLASTp) (Altschul et al., 1997) was used to search for possible templates and the best template was selected based on the highest sequence identity. Once the suitable template was identified, pair-wise sequence alignment was carried out between the

target and the template sequences using LALIGN (Huang and Miller, 1991). Next, the 3D model was generated using the comparative modelling software MODELLER (Šali and Blundell, 1993). The model then was evaluated using Ramachandran plot (Morris et al., 1992) Verified 3D (Luthy et al., 1992), and ERRAT (Colovos and Yeates, 1993). The best model was selected and subjected for the analysis of the active sites and binding pockets using COFACTOR (Roy and Zhang, 2012). All the graphics presentations were prepared using UCSF Chimera (Pettersen et al., 2004) and VMD (Humphrey et al., 1996).

3. Results and discussions

3.1. Model development

Bromelain sequence required a template of known crystal structure derived from experimental method. The template was identified from BLASTp program against the protein data bank (PDB) (Berman et al., 2000) and Swiss-Prot (Apweiler et al., 2004) databases. From the BLASTp results, Chymopapain (PDB ID: 1YAL) (Maes et al., 1996) was chosen as the best template because it has the highest sequence identity which is 55%. Fig. 1 shows pair wise sequence alignment between 1YAL and target bromelain sequence. From the figure, most of the residues were conserved to each other from Ala136 –Asp291. The target residues Met1-Ser78, Leu81-Ala97, Tyr99 – Ile103, Arg105 – Asp119, Gln127 – Gly135, Cys149 and Asn201 were found to have no template residues. These missing template residues may lead to improper model development.

From the single template modelling, 100 models were predicted using MODELLER. Model 72 ($Model_{72}$) was chosen as the best template and used for further evaluation.
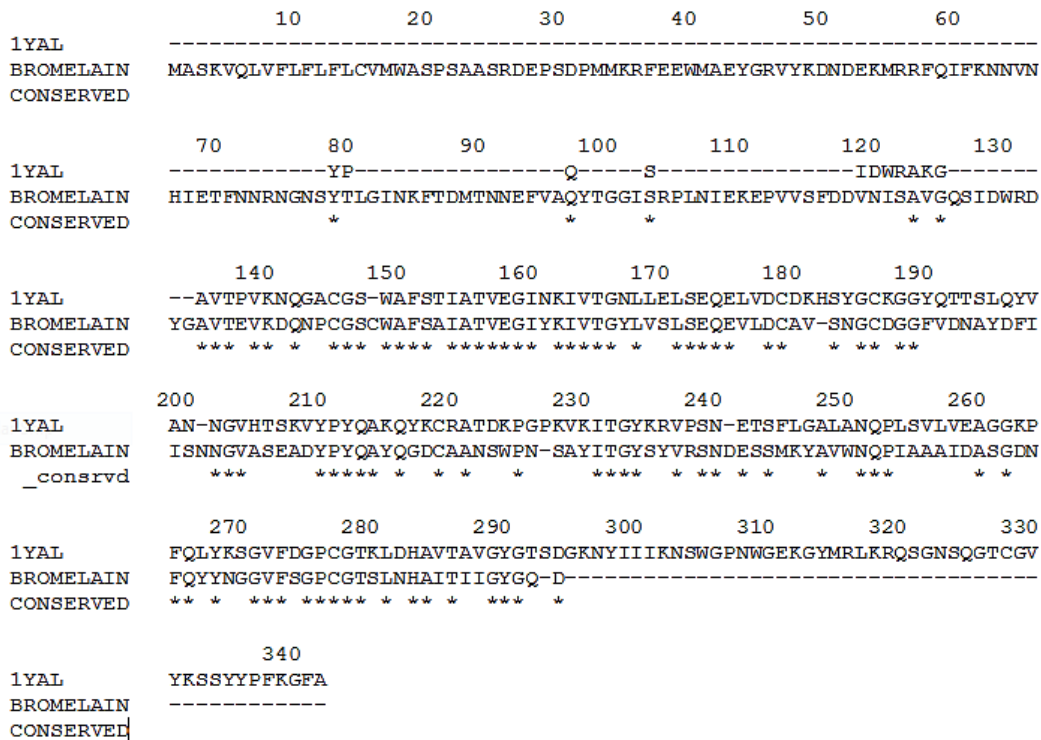
```
                    10        20        30        40        50        60
1YAL        ------------------------------------------------------------
BROMELAIN   MASKVQLVFLFLFLCVMWASPSAASRDEPSDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVN
CONSERVED

                    70        80        90       100       110       120       130
1YAL        -----------YP---------------Q-----S---------------IDWRAKG-------
BROMELAIN   HIETFNNRNGNSYTLGINKFTDMTNNEFVAQYTGGISRPLNIEKEPVVSFDDVNISAVGQSIDWRD
CONSERVED              *                 *     *                         * *

                    140       150       160       170       180       190
1YAL        --AVTPVKNQGACGS-WAFSTIATVEGINKIVTGNLLELSEQELVDCDKHSYGCKGGYQTTSLQYV
BROMELAIN   YGAVTEVKDQNPCGSCWAFSAIATVEGIYKIVTGYLVSLSEQEVLDCAV-SNGCDGGFVDNAYDFI
CONSERVED     *** ** *   *** **** ******* ***** *   *****  **   * ** **

            200       210       220       230       240       250       260
1YAL        AN-NGVHTSKVYPYQAKQYKCRATDKPGPKVKITGYKRVPSN-ETSFLGALANQPLSVLVEAGGKP
BROMELAIN   ISNNGVASEADYPYQAYQGDCAANSWPN-SAYITGYSYVRSNDESSMKYAVWNQPIAAAIDASGDN
 _consrvd    ***      ***** * * *     *     ****  * ** *     *   ***      *  *

                    270       280       290       300       310       320       330
1YAL        FQLYKSGVFDGPCGTKLDHAVTAVGYGTSDGKNYIIIKNSWGPNWGEKGYMRLKRQSGNSQGTCGV
BROMELAIN   FQYYNGGVFSGPCGTSLNHAITIIGYGQ-D------------------------------------
CONSERVED   ** *   *** ***** * ** *   *** *

                    340
1YAL        YKSSYYPFKGFA
BROMELAIN   ------------
CONSERVED
```

Fig. 1: Pair wise sequence alignment between bromelain and 1YAL (the alignment showed 55% sequence identity and 59% similarity)

The $Model_{72}$ has seven helices and two strands as shown in Fig. 2. Helices were the major part consists in the model. All the helices were abundance in the middle, while two strands were located at the C-terminal of the model. The N-terminal model (Met1 to Ala135) was not properly modelled since it doesn't have template residues.

$Model_{72}$ was selected because it had the lowest Discrete Optimized Protein Energy (DOPE) (Shen and Sali, 2006) and Modeller objective function (molpdf) which were -20019.08 and 1915.35, respectively. The DOPE energy was represented in graph as shown in Fig. 3. The model had more stable energy at residues Val136 to Gly291 because it had template residues.

3.2. Model assessments

The $Model_{72}$ was superimposed with the 1YAL template as shown in Fig. 4. Met1 – Val119 was the loop of $Model_{72}$. This loop is independent because it has no template residues as mention in Fig. 1. I120 – D291 residues from $Model_{72}$ (purple) were correctly superimposed with the template structure (blue) and 53% of the residues at this superimposed were conserved to each other.

$Model_{72}$ was validated using SAVES (The Structure Analysis and Verification Server) which utilizes three different validations programs; Ramachandran Plot, ERRAT, and Verify-3D. From the Ramachandran plot (Fig. 5) analysis seven residues

were found to be located in disallowed region. These residues were Gly45, Gly76, Gly134, Gly184, Asn225, Ser242, and Gly287. Based on the ERRAT and Verify-3D analyses, was found that Model$_{72}$ scored 41.28% and 53.61% respectively.
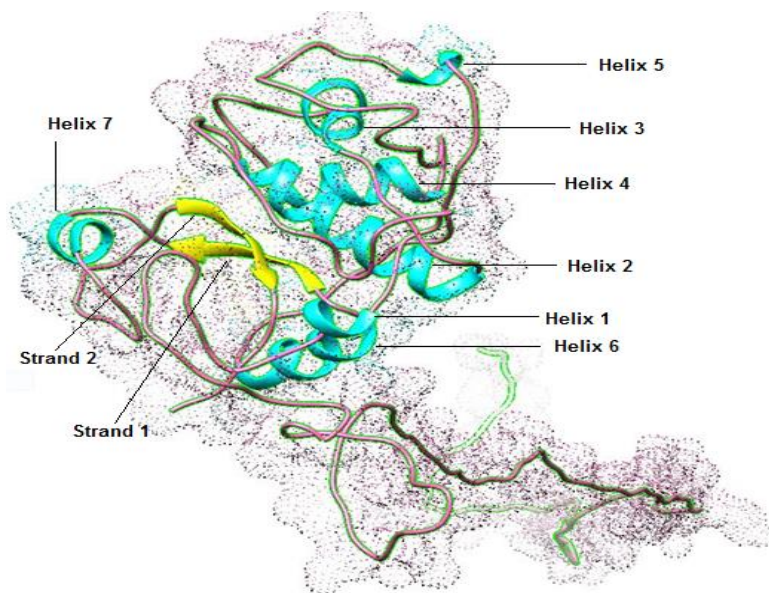

Fig. 2: The predicted tertiary structure of bromelain (Ribbons and dots represented the secondary structure and surface of the Model$_{72}$)
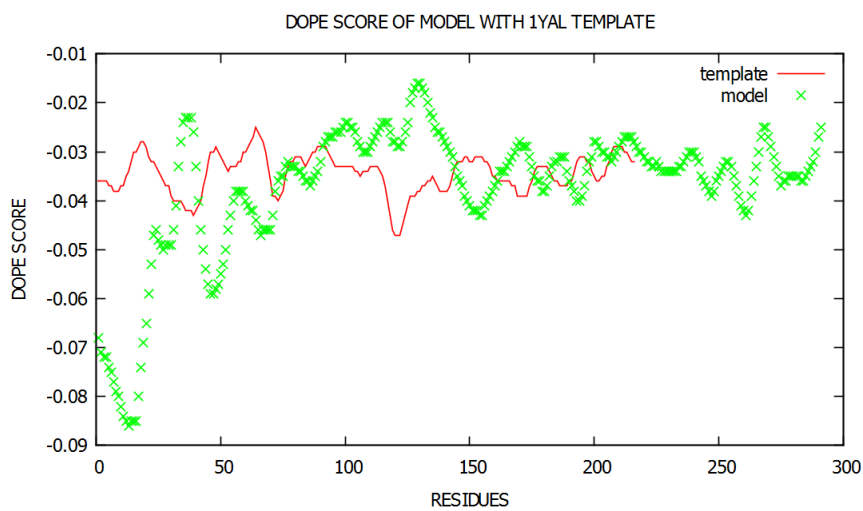

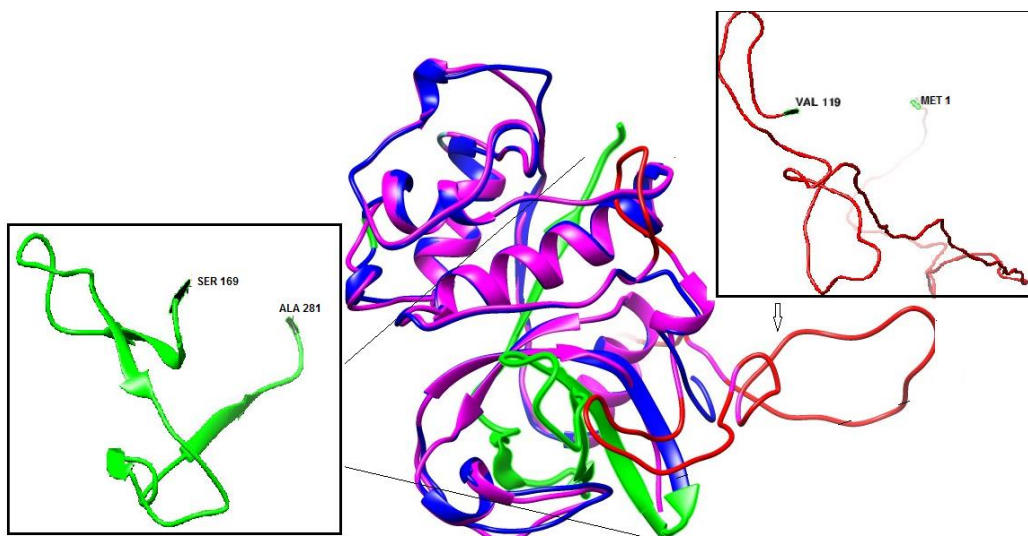**Fig. 3:** DOPE energy score between template (red) and model (green)

Fig. 4: The superimposed model between Model$_{72}$ and 1YAL template (Red and green regions represent the region with no template structure)
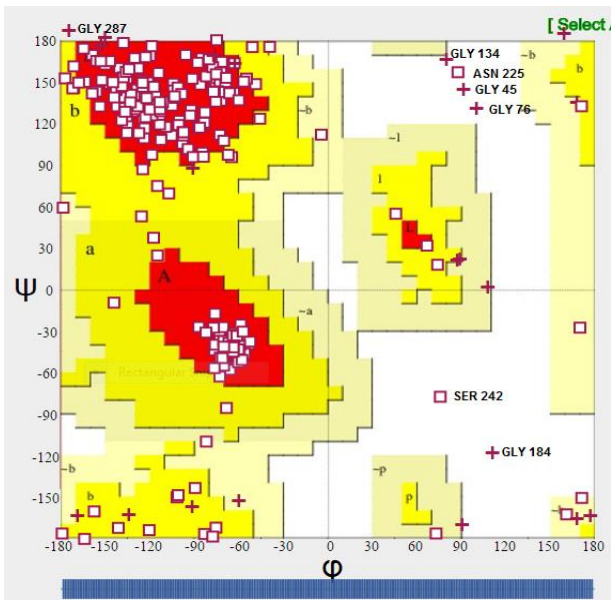
Fig. 5: Ramachandran plot of the model from multiple templates modelling of bromelain (Red, yellow, light yellow and white region represent the most favoured region, allowed region, generously allowed region, and disallowed region, respectively)

3.3. Binding site prediction of model₇₂

Model₇₂ was selected for next analysis, which is binding site identification using the COFACTOR program. This program generates binding site prediction using two comparative methods which are binding-specific substructure and sequence similarity comparisons. Results showed the best binding pocket was pocket 1 which has confidence score 0.70 and sequence alignment 51%. The binding site consists of Gln142-

Trp149, Asp187-Gly189, Ala255 and Thr279-Leu281 as shown in Fig. 6. The best binding pocket was selected based on its confidence score ranging from 0 to 1 in which the highest score indicates the most reliable binding pockets and sequence alignment is more than 30%.

It was also found that the three catalytic residues (Gln142, Cys148, and His281) located in the binding pocket.

4. Conclusion

In this study, 3D structure of bromelain was predicted using single template modelling. 1YAL was chosen as the template structure because it has the highest sequence identity. Based on the assessments of Model₇₂, it was showed that this model needed further refinement at the loop region consists of Met1 – Val119. Subsequent studies on the refinement of this model are highly recommended in order to better understand the functional properties of bromelain.
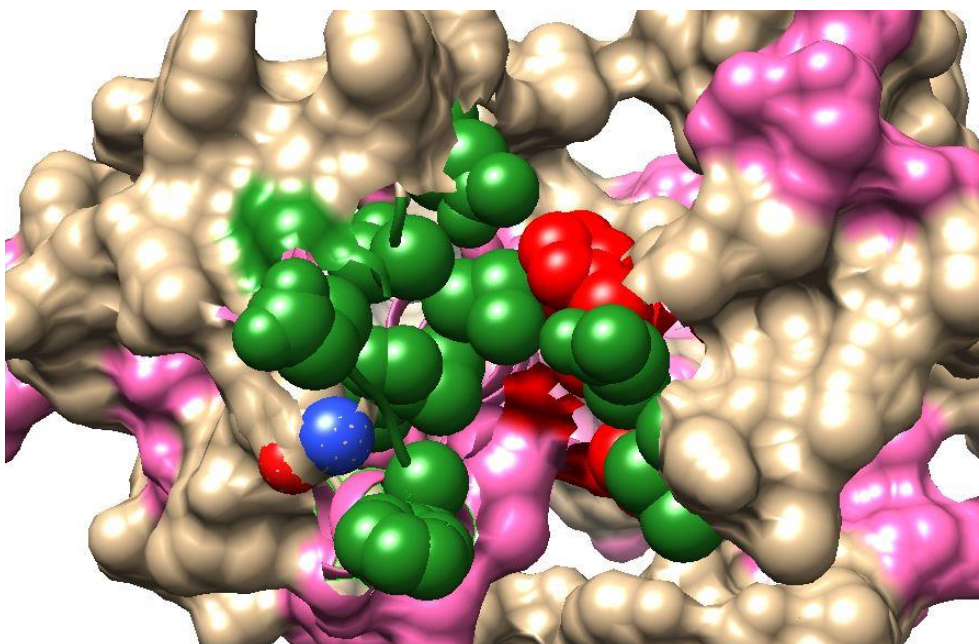

Fig. 6: Surface representation of binding pockets, purple indicates the surface of helices structure (the sphere model represents the location of binding pocket in the model)

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25(17): 3389-3402.

Apweiler R, Bairoch A, and Wu CH (2004). Protein sequence databases. Current Opinion in Chemical Biology, 8(1): 76-80.

Arshad ZIM, Amid A, Yusof F, Jaswir I, Ahmad K, and Loke SP (2014). Bromelain: An overview of industrial application and purification strategies. Applied Microbiology and Biotechnology, 98(17): 7283-7297.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, and Bourne PE (2000). The protein data bank. Nucleic Acids Research, 28(1): 235-242.

Colovos C and TO Yeates (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Science, 2(9): 1511-1519.

Gautam SS, Mishra SK, Dash V, Goyal AK, and Rath G (2010). Comparative study of extraction, purification and estimation of bromelain from stem and fruit of pineapple plant. Thai Journal of Pharmaceutical Sciences, 34(2): 67-76.

Huang X and Miller W (1991). A time-efficient, linear-space local similarity algorithm. Advances in Applied Mathematics, 12(3): 337-357.

Humphrey W, Dalk A, and Schulten K (1996) VMD: Visual molecular dynamics. Journal of Molecular Graphics, 14(1): 33-38.

Luthy R, Bowie JU, and Eisenberg D (1992). Assessment of protein models with three-dimensional profiles. Nature, 356(6364): 83-85.

Maes D, Bouckaert J, Poortmans F, Wyns L, and Looze Y (1996). Structure of chymopapain at 1.7 A resolution. Biochemistry, 35(50): 16292-16298.

Morris AL, MacArthur MW, Hutchinson EG, and Thornton JM (1992). Stereochemical quality of protein structure coordinates. Proteins: Structure, Function and Bioinformatics, 12(4): 345-364.

Pavan R, Jain S, Shraddha, and Kumar A (2012). Properties and therapeutic application of bromelain: A review. Biotechnology Research International, 2012: Article ID 976203, 6 pages. https://doi.org/10.1155/2012/976203

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25(13): 1605-1612.

Pruitt K, Brown G, Tatusova T, and Maglott D (2002). The reference sequence (RefSeq) database. In: McEntyre J and Ostell J (Eds.), The NCBI handbook, National Center for Biotechnology Information: 1-24. Bethesda Softworks, Rockville, USA.

Roy A and Zhang Y (2012). Recognizing protein-ligand binding sites by global structural alignment and local geometry refinement. Strcucture, 20(6): 987-997.

Šali A and Blundell TL (1993). Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology, 234(3): 779-815.

Shen MY and Sali A (2006). Statistical potential for assessment and prediction of protein structures. Protein Science, 15(11): 2507–2524.