Pak. J. Bot., 55(2): 489-500, 2023. DOI: http://dx.doi.org/10.30848/PJB2023-2(11)

MOLECULAR DISSECTION AND AN *IN-SILICO* APPROACH OF A NOVEL *GIBBERELLIN 20-OXIDASE* GENE OF *HIBISCUS CANNABINUS* L.

NOR AINI AB SHUKOR[1,2*], YUSUF CHONG YU LOK[3*], SURES M. KUMAR[1], RAMBOD ABIRI[1] AND MOHD PUAD ABDULLAH[4]

[1]*Department of Forestry Science and Biodiversity, Faculty of Forestry and Environment,*
Universiti Putra Malaysia, Serdang, Selangor DE 43400 UPM, Malaysia
[2]*Laboratory of Bioresource Management, Institute of Tropical Forestry and Forest Products (INTROP),*
University Putra Malaysia, 43400 UPM Serdang, Selangor Malaysia
[3]*Faculty of Plantation and Agrotechnology, Universiti Teknologi MARA, Kampus Jasin, 77300 Merlimau, Melaka, Malaysia*
[4]*Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences,*
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia
[*]*Corresponding author's email: dnoraini@upm.edu.my; yusufchong@uitm.edu.my*

Abstract

Kenaf (*Hibiscus cannabinus* L) is widely grown for its fibre. The *Gibberellin 20-oxidase* (*HcGA20ox*) gene is responsible for the improvement in quality of kenaf fibre. The two transcripts designated as *HcGA20ox1* (KY399834) and *HcGA20ox1* (KY399835) were isolated from the shoot tissue of three-week-old seedlings using the Rapid Amplification of cDNA Ends PCR (RACE), and PCR walking approaches, which encoded the polypeptides of 430 and 345 amino acids, respectively. Therefore, *HcGA20ox1b* was presumed as a pseudogene and only *HcGA20ox1* was focused in downstream works. A high level of similarity has been observed between amino acid sequence of HcGA20ox1 and the GA20ox of other species. In addition, substrate binding site "LPWKET" and two histidine residues which serve as Fe^{2+} binding domains were found in the sequence. The BLAST results showed that *HcGA20ox1* shared high identity with the GA20ox from other species including *Gossypium hirsutum, Citrus sinensis, Populus trichocarpa*, *Jatropha curcas, etc.* The highest identity was attributed to the GA20ox of *Gossypium hirsutum* (accession no.: XP_016668427) at 83% identity. Using the Conserved Domain Database (CDD) available in NCBI, the non-heme dioxygenase superfamily, 2OG-Fe (II) oxygenase superfamily and gibberellin 20-oxidase domains were detected in HcGA20ox1. The molecular weight of the protein was 122224.45 with 4.97 theoretical pI.

Key words: *HcGA20ox*, RNA extraction, ProtParam, PROCHECK, I-TASSER, and Protein structure.

Introduction

Hibiscus is quite a large genus of native sub-tropical, tropical, and warm-temperate flowering herbs (Ebenezer *et al*., 2019). Hibiscus are belonging to the family of mallow or Malvaceae (cotton family), which includes several hundred species, distributed in the high humidity regions throughout the globe (Bismarck *et al*., 2005). *Hibiscus cannabinus* (kenaf) is an annual commercial fast-growing plant, which is harvested for fibres in many parts of the world. The fibres derived from bast fibres (the outer fibrous bar), have cellulose content of 35% to 60%, which makes them a reliable cellulose source (Rowell, 2000; Kargarzadeh *et al*., 2012). Kenaf is widely cultivated in Malaysia, India, China, Bangladesh, and Thailand for its fibres due to the favourable weather conditions, and the governments have shown some willingness in supporting the development of new usages. Industrially, the fibre of kenaf is used as a filter composite material. Since it is biodegradable, it is not abrasive during processing, has specific mechanical properties and a low density (Edeerozey *et al*., 2007).

Up until now, acetylation, acrylonitrile grafting, permanganate, alkaline, benzoylation, isocyanate, silane, peroxide, maleated coupling agents and other chemicals (sodium chlorite, triazine, stearic acid) have been applied to improve the quality material and affect on the strength of bonding between polymer matrix and natural fibre surface in which they chemically and/ or physically adjust the fibre surface and raises its strength (Thiruchitrambalam *et al*., 2012). The stalk of kenaf contains an outer and inner fibre in which the outer fibre is called "bast". However, the inner part of kenaf fibre, also known as core, contains approximately 39% and 58% of the dry weight of stalk, respectively (Jamaludin, 2008). Moreover, bast fibres as a cortical extraxylary sclerenchymatous structure can provide mechanical network to support the phloem and conduct elements of the phloem which naturally exists in bundles held together by lignin and pectins. The fibre crops cell wall is composed of pectin, hemicelluloses, cellulose, and lignin (Pejic, 2008). So far, various varieties of kenaf have been modified and marketed to meet the needs for high-fibre quality, yield, seed quality, and disease-/ insect- and drought- tolerance (Bitzer, 2000). Although the novel kenaf varieties have played significant roles in kenaf studies as the genetic resource, kenaf identification systems using agronomical and morphological traits are always problematic (Siepe, 1997). The kenaf elongation and increase of fibre quality can be obtained using genetic modification and environmental manipulation. Isolation and characterisation of the genes related to the fibre quality pathway have been carried out in some plants, such as cotton (*Gossypium arboreum*) (Lee *et al*., 2007).

Gibberellin (GA), which is a group of tetracyclic diterpenoid carboxylic acid, is a plant growth regulator in many plants (Hirano *et al*., 2007; Takehara & Ueguchi-Tanaka, 2018). GA regulates various pathways in the plant germination, growth, maturation, development as well as reproduction (Fleet & Sun, 2005). Reportedly, dwarfism during plant development phase is critically affected by GA-insensitive or -deficient mutants in all angiosperms (both monocot as well as dicot plant species) including, the rice GA receptor mutant gid1 and *Arabidopsis* biosynthesis mutant ga1-3. *Arabidopsis* contains five GA20-oxidase (*GA20ox1* to *5*) and studies results demonstrated that *GA20ox1, -2*, and *-3* are the dominant paralogs of *GA20* gene (Xiao *et al.*, 2006; Plackett *et al.*, 2012). In another experiment, Nelissen *et al* (2018) demonstrated that high levels of GA in maize overexpressing GA20-Oxidase. In coconut, *CnGA20ox1* is responsible for the height, whereas this characteristic varies from other crops due to lack of mutation in *CnGA20ox1* of dwarf type coconut (Boonkaew *et al*., 2018). Interestingly, *FveGA20ox4* controls runnering-flowering and improves the pathway of strawberry productivity by regulating the trade-off between vegetative propagation and sexual reproduction (Tenreira *et al*., 2017).

To date, there is no report of *GA20ox* gene in kenaf and the genome sequence of kenaf is not available. Therefore, the study of *HcGA20ox* gene is constrained and the effort to isolate the *HcGA20ox* gene becomes a prerequisite for this study. Currently, the *HcGA20ox* gene fragment has been isolated through PCR and cloned in this study. Identity of the gene fragment was verified by sequencing and BLAST search in NCBI. Subsequently, Rapid Amplification of cDNA End (RACE) method was used to obtain the full-length of *HcGA20ox* cDNA sequence. A putative *HcGA20ox* cDNA and another pseudogene were obtained in this study. In the current study, the 3D model along with the primary and secondary structures of the HcGA20ox protein, were predicted and the best 3D protein model were proposed for future studies.

Materials and Methods

Cultivation of kenaf plants: Seeds of kenaf (variety V36) were obtained from INTROP, UPM. The seeds were sown in peat moss and watered daily. Three-week-old seedlings were transferred to a polybag and grown on mixed soil in the green house. A small amount of NPK green fertiliser was applied to the plants every two weeks. To extract the RNA, young shoots of the three-week-old seedlings were collected and stored at -80ºC until used for isolation.

Total RNA extraction from kenaf tissues: Several attempts were performed using different approaches to obtain high quality total RNA from kenaf. For total RNA isolation, three various methods such as easy-BLUE Total RNA Extraction Kit (iNtRON Biotechnology, Korea), TransZol Up kit (TransGen Biotech, China) and TRIzol methods were applied according to the instructions of companies.

Synthesis of cDNA and isolation of *HcGA20ox* gene fragment: To synthesis the cDNA from RNA (5 µg), RevertAid H minus first strand cDNA synthesis kit (Thermo Scientific, USA) was applied based on the instructions of manufacturer. Except that Oligo d(T)-anchor primer (custom made) was used as the RNA binding primer. Subsequently, the cDNA was diluted with equal volumes of sterile water and stored at -80 ˚C until used. *HcGA20ox* gene fragment was amplified using the PCR anchor primer (5’-GAC CAC GCG TAT CGA TGT CGA C-3’) and a degenerate primer (5’-TCC AAK YTG CCW TGG AAR GAR AC-3’) designed according to the conserved region of *GA20ox* genes from other closely related species. The RT-PCR was performed using the DreamTaq Green DNA Polymerase (Thermo Scientific, USA) and the cDNA prepared early as the template.

Cloning of *HcGA20ox* gene fragment: The products of PCR were resolved on agarose gel (1.5%). Target band of approximately 750 bp was excised from the gel and purified using the MEGA-spin Agarose Gel Extraction Kit (iNtRON Biotechnology, Korea). Next, the purified products were cloned using the pGEM®-T Easy Vector Systems (Promega, USA). DNA-spin™ Plasmid DNA Purification Kit (iNtRON Biotechnology, Korea) was applied to isolate plasmid DNA and the results sequence by First Base Labolatory Sdn. Bhd. (Selangor, Malaysia).

Isolation of *HcGA20ox* gene full-length cDNA: The full-length cDNA sequence of *HcGA20ox* gene was isolated using SMARTer RACE 5’/3’ Kit (Clontech, Japan). The 5’ RACE cDNA template was synthesised from 1 µg of total RNA. The target sequence was amplified using the universal primer mix A (provided by the kit) and GA20RACE-R primer (5’-GGG GGA GAG AAA GAA ACA AGA GAA-3’), which was designed to bind at the 3’ UTR of *HcGA20ox* gene. The PCR was performed using DreamTaq Green DNA Polymerase (Thermo Scientific, USA). The PCR product was cloned, and five colonies were sequenced as described in the previous section.

Amplification of the open reading frame of HcGA20ox1: The open reading frame (ORF) of *HcGA20ox1* was amplified by PCR using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific, USA). The primers used in the PCR were GA20 OE (F) (5’-TGA TGA TCT AGA ATG AGA ACG GCC TTG GCG TT-3’) and GA20 OE (R) (5’-GGA GCT CTC AGG TGT TTC TGT GTT GAA CCC A-3’). The *Xba* I and *Sac* I restriction sites were incorporated at the 5’ ends of the forward and reverse primers, respectively (sequence in bold). The PCR product was purified using the MEGA-spin Agarose Gel Extraction Kit (iNtRON Biotechnology, Korea).

In silico analysis

Evolutionary tree of *HcGA20ox1* gene: The extracted *HcGA20ox1* gene samples were sent for sequencing and the blast results showed 100% similarity with the targeted gene in tomato. The sequence of the gene was submitted to the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). The phylogenetic tree was designed using MEGA Molecular Evolutionary Genetic Analysis Version 6. RNA Analyser was applied to estimate the RNA structure and functions of both genes.

The complexity of HcGA20ox protein structures: Different structures of HcGA20ox1 protein were analysed to measure various factors, including estimated half-life, amino acid composition, molecular weight, extinction coefficient, instability index, grand average of hydropathicity (GRAVY), aliphatic index, and theoretical pI. Evaluating and analysing the HcGA20ox protein was done using ProtParam (http://web.expasy.org/protparam/), which is available on the database of ExPASy server. The secondary structure of HcGA20ox1 protein was predicted using Self-Optimised Prediction Method with Alignment (SOPMA) and protein prediction structure server PSIPRED.R5 (2000; Buchan *et al*., 2013; Geourjon & Deleage, 1995). The 3D structures of the HcGA20ox proteins were predicted using Iterative Threading ASSEmbly Refinement I-TASSER, server (http://zhanglab.ccmb.med.umich. edu/I-TASSER (Yang *et al*., 2013). MEMSAT-SV (Nugent & Jones, 2009), Cell-Ploc (Chou & Shen, 2010), and PSORT II (Nakai & Horton, 1993) are other prediction tools that were used in this study to monitor the subcellular localisation of the HcGA20ox protein. PROCHECK (Laskowski *et al*., 1993) tools were also applied to validate I-TASSER results, as evaluated by Ramachandran plot statistics.

Results

Extraction of total RNA: Analysis of the total RNA sample using the spectrophotometer revealed the ratios of A_{260}/A_{280} was 1.98. Additionally, the ration of A_{260}/A_{230} was 2.01. Agarose gel electrophoresis (1.5%) showed the RNA sample was intact as the 28S rRNA and 18S rRNA bands were not degraded. Hence, it is appropriate to be used for cDNA synthesis (Fig. 1A).

Isolation of *HcGA20ox* gene fragments and full-length cDNA: A distinct single band of 750 bp was generated by PCR using gradient annealing temperatures ranging from 56-65˚C (Fig. 1B). The size of gene fragments generated using PCR anchor primer and degenerate primer fell within the range of targeted size, which is 700-800 bp. This indicates that the PCR product obtained corresponds to the target gene. The 5’ RACE-PCR yielded an approximately 1.4 kb product. Sequencing result showed two cDNAs corresponded to the two *HcGA20ox* gene fragments isolated were obtained through RACE. Hence, the two cDNAs were designated as *HcGA20ox1* (Accession number: KY399834) and *HcGA20ox1b* (Accession number: KY399835). The cDNAs of *HcGA20ox1* and *HcGA20ox1b* were 1439 bp and 1520 bp, respectively (Fig. 1C).

***HcGA20ox* gene sequence analysis:** Sequencing analysis revealed the cDNA sequences of *HcGA20ox1* and *HcGA20ox1b* were highly similar. However, the latter contained additional nucleotides within the coding sequence compared to the former. *HcGA20ox1* encoded for a protein of 430 amino acids while the sequence of *HcGA20ox1b* was translated into 345 amino acids due to the presence of a premature stop codon within the additional nucleotide region. Sequencing results revealed that the gene fragment isolated consisted of two different genes, which shared high similarity in their coding sequences (CDS) and 3’ untranslated region (UTR). BLAST search results further confirmed that the gene fragments isolated corresponded to *GA20ox* gene as high sequence identity (79-89%) and was shown between the query and subjects. NCBI was used to align and retrieve both *HcGA20ox1* and *HcGA20ox1b* genes nucleotide with other genes, predict intron/exon positions of both genes and find the UTRs location.

The RNA Analyser program showed one Exon with the length of 179- 1294. RNA Analyser also predicted that the gene was flanked by 5′UTR started and ended codon 1-178; and 3′UTR started and ended codon 1295 and 1445. The poly-A- signal of 3′UTR was also flanked between start codon 1440 and stop codon 1445. A SnRNP-motif was observed in codon 347 (gaucuugg) and a Put. Sm-site was observed in codon 766 (gauuuuag). The predicted results also revealed that 45.99% of the gene sequence was G+C bps (Fig. 2A). The general information of *HcGA200x1b* gene predicted by RNA analyser showed that the Exon codon was between 179-1039 bp. The gene was flanked between 5′UTR-started codon 1 and ended codon 178, and 3′UTR started in 1040 and ended in 1533. The PlyA-Sgl was flanked between 1528- 1533 and a snRNP-motif was recorded in 347 (gaucuugg) and a Put. Sm-site was also seen in 766 (gauuuuuag). The predicted results also revealed that 45.99% of the gene sequence was G+C bps (Fig. 2B).

Multiple sequence alignment (MSA) of *HcGA20ox1b* and some other HcGAs enzymes were performed with *Arabidopsis thaliana*, *Eucalyptus grandis*, *Gossypium raimodii*, *Jatropha curcas*, *Nicotina tabacum*, *Populus tomentosa*, and *Hibiscus cannabinus*. Therefore, *HcGA20ox1b* was presumed as a pseudogene and only *HcGA20ox1* was focused on downstream works. The highest similarity of *HcGA20ox1* with other *HcGA* genes was observed in different species as follows: *Gossypium raimondii* (89%), *Gossypium hirsutum* (89%) and *Theeobrom cacao* (89%). The sequence of HcGA20ox1 was similar with the GA20ox from other species. In addition, substrate binding site “LPWKET” and two histidine residues which serve as Fe^{2+} binding domains were found in the sequence. LPWKET were in positions 214- 219 and both Fe^{2+} were reported in positions 323 and 364 (Fig. 3).

The Phylogenetic tree was built using *HsGA20ox1* and 40 other homologue sequences retrieved from NCBI. The phylogenetic tree of *HcGA20o1* produced five main clusters. Cluster 1 consisted of 5 operational taxonomic units (OTUs). Cluster 2 consisted of two sub-clades and cluster 3 was sub-divided into 3 classes. Cluster 4 was sub classed into 2 ranges and Cluster 5 consisted of 4 sub classes. To this end, the phylogenetic tree clusters were created using *HsGA*, and all clusters comprised of some plant families such as Malvaceae, Casuariaceae, Sapindaceae, Theaceae, Anacardiaceae, Salicaceae and Ericaceae. Interestingly, phylogenetic analysis showed that, *HcGA20o1* was genetically more similar to different *Gossypium* sp. (*G. hirsutum*, *G. raimondii* and *G arboreumand*) (Fig. 4).

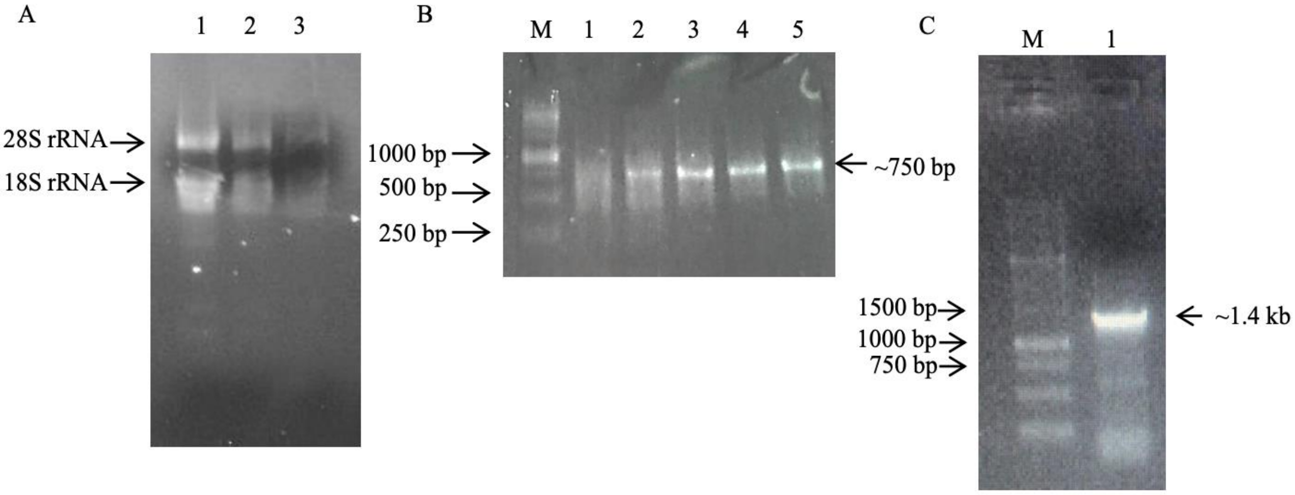


Fig. 1A. Total RNA sample extracted from kenaf. Lane 1: 2µg, lane 2: 1µg, lane 3: 0.5µ. 1B. PCR product generated at various annealing temperatures. Lane 1: 55.0˚C, Lane 2: 56.9˚C, Lane 3: 61.1˚C, Lane 4: 63.0 ˚C and Lane 5: 65.0˚C. 1C. PCR product produced by 5' RACE of *HcGA20ox*. Lane 1: *HcGA20ox* PCR product. Lane M: 1kb DNA ladder.

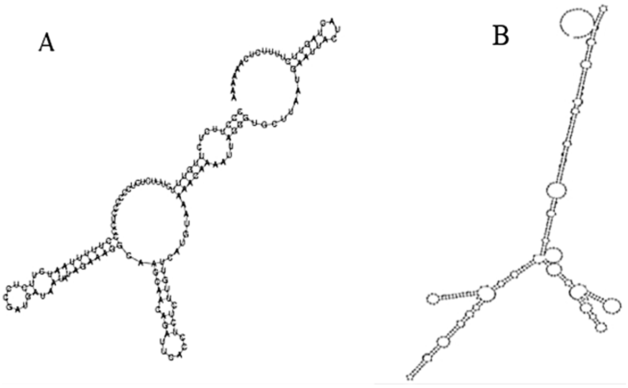


Fig. 2. RNA Analyser predicted a general model of *HcGA20ox1* (A) and *HcGA20ox1b* (B). Each circle shows a functional genomic of sequences.

HcGA20ox protein sequence analysis: ProtParam computes different physico-chemical parameters of protein that can be deduced from HcGA20ox sequence. Moreover, the analysis result of amino acid by PortParm showed total number of negatively charged residues (Asp + Glu) was 24, whereas total number of positively charged residues (Arg + Lys) was 63. The molecular weight of protein was 122224.45 with 4.97 theoretical pI. The formula of HcGA20o1 protein was C_{2291} H_{3607} N_{681} O_{626} S_{21} and the total number of atoms was 7226. Ext. coefficient of the HcGA20ox protein was 66765 with the absorbance of 0.1% (=1 g/l) 1.299 with respect to all pairs of Cys residues form cystines. In the Ext. coefficient of 65890 the Abs 0.1% (=1 g/l) was 1.282, assuming all Cys residues were reduced. The *in-silico* study of HcGA20o1 half-life protein showed that the N-terminal of the sequence measured was N-Terminal asparagine (Asn), whereas the predicted half-life of the protein was 3 min for yeast *In vivo*, 1.4 hours for mammalian reticulocytes *In vitro*, and >10 hours for *Escherichia coli In vivo*. The HcGA20ox protein is categorised as unstable protein (59.05) because the instability index (II) was higher than 40. Additionally, aliphatic index of HcGA20ox protein was 76.85 and Grand average of hydropathicity (GRAVY), which is unite of hydrophobicity measurement of a peptide, was -0.396 (Fig. 5).

The secondary structure analysis of HcGA20o1 protein consisted of 17 helixes, 8 sheets, and 9 coils (Fig. 6). Figure 6 demonstrates the result of PSIPRED diagrammatic output, in which the diagram explains the query sequence with, and confidence value (a series of blue bar graphs) and secondary structure cartoons at each position in the alignment. Understanding helix, sheet and coil of protein helps to estimate the thermodynamic, mechanism and HcGA20o1 protein structure.

Analysis of HcGA20ox protein by Hphob & Doolittle demonstrated 20 amino acids levels as follow: Met: 1.900, His: -2.200, Tyr: -1.300, Ala: 1.800, Trp: -0.900, Phe: 2.800, Arg: -2.500, Asn: -2.500, Ser: -0.800, Asp: -2.500, Cys: 2.500, Gln: -1.500, Glu: -1.500, Gly: -0.400, Ile: 2.500, Lys: -2.900, Leu: 2.800, Thr: -0.700, Pro: -1.600, Val: 2.200: -2.500, -2.500 and -0.490 (Fig. 7A). PortScale results showed a range between -2.5 to + 2.5 which this higher hydropathy confirmed higher hydrophobicity. With the output width fixed as 70, this result released 70 amino acids and associating expected structures in each line. The result of SOPMA also demonstrated that the sequence length was 457 amino acids. The ratio of each assembly was Alpha helix (Hh): 159 (34.79%), Extended strand (Ee): 62 (13.57%), Beta turn (Tt): 40 (8.75%) and Random coil (Cc): 196 (42.89%) (Fig. 7B).

Prediction of top five final models of HcGA20o1 protein using I-TASSER: The local accuracy of the protein was predicted using the I-TASSER on-line software with the distance deviation defined in Angstrom between residue positions in the model and native structure. According to the results achieved from the estimation of the 3D structure for the HcGA20o1 protein encoded by gene, the protein Homology Analogy Recognition Engine software (Phyre2) server and I-TASSER programs were applied. The output of the APhyre2 model was generated based on the template galactose-binding domain-like. The I-TASSER software predicted five models with local structure error profiles with *C*-score of -3.02, -3.04, -3.04, -3.76 and -4.83 for each model, respectively. The results also showed that the Estimated TM-score was 0.37±0.13 and Estimated RMSD was 14.6±3.7Å for the first model (Fig. 8). The results of these tests demonstrated local accuracy prediction were more suitable for the residues with higher threading alignment coverage or/ and these accurate residues should be placed at the α-helix and β-strand regions or/ and were buried (at 25% threshold).

HcGA20o1 protein structurally close to the target in the PDB: Next to simulation of assembly structure, the TM-align structural alignment was predicted by I-TASSER model to all structures in the PDB library. In this step, the top ten predicted PDB models were reported based on the closest structural similarity such as the highest TM-score. The structure similarity demonstrated the most similar protein to the target model. The confirmation of these data was done using 'Predicted function using COACH', since 'Predicted function using COACH' has been trained to derive biological function from multi-source of structure and sequence features which has on average a higher accuracy than the function annotations derived only from the global structure comparison. The PDB Hit of the best model was 2h4tB with TM-score 0.817, RMSD[a] = 3.31, IDEN[a] = 0.117, and Cov= 0.908.

Protein HcGA20o1 predicted function using COACH and COFACTOR: The predicted role of ligand binding sites reported biological results of the objective model protein by COACH and COFACTOR according to the prediction of HcGA20o1 structure by I-TASSER. Although COFACTOR deduces HcGA20o1 protein functions using protein-protein networks and structure comparison, COACH is a meta-server reveals that combines multiple function annotation results (on ligand-binding sites) from the S-SITE, COFACTOR, TM-SITE programs. The PDB Hit of the best model was 1s5oA with Lig Name 152 and different binding site residues (Fig. 9).

```
                               10        20        30        40        50        60        70        80        90       100
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Acacia mangium         -----------------------------------------------------------MVVECITNNLPSMPHPAKHGTKED---KEDEPLVFDASLLR
Arabidopsis thaliana   -----------------------------------------------------------MAVSFVT-TSPEEEDKPKLGLG-----NIQTPLIFNPSMLN
Eucalyptus grandis     -----------------------------------------------------------MAVDCLT-SKTSPAMPPQHKDEAR---EDKKHLVFDASVIR
Gossypium raimondii    ---MLPLPSPSSFSITPYSFLCLNFISYP---LFKYSQNHSYR--YPSLLGLSIFTLHAMAIDCIS-NIASMTHHPKDEKK-----DEQKKLVFDASVLK
Jatropha curcas        -----------------------------------------------------------MAVDCIK-TMP-----QHHQED-----DQNKPLVFDASVLR
Nicotiana tabacum      -----------------------------------------------------------MAIDCMI-TNVN--SPMLRILE-----DDKKPLIFDASQMK
Populus tomentosa      -----------------------------------------------------------MAIDCIK-TMPSITTPQHHPKDQDQCKDDGKSFVFDAQVLR
Hibiscus cannabinus    MRTALAFLSSHLSNLCSHTALSLFFLLTPYSLLCLTSLNHSYIYIYTIFFGRPQYTIHAMAIDCIS-NIPSMPHEPK-------------KLVFDASVLR

                              110       120       130       140       150       160       170       180       190       200
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Acacia mangium         YEHNLPKQFIWPDEEKPCLNPPELVVPLVDLRGFLSGDPVAAMETSKLVSEACRKHGFFLVVNHRIDSRLISHAHRFMDDFFELPLSQKQRAQRKAGEHC
Arabidopsis thaliana   LQANIPNQFIWPDDEKPSINVLELDVPLIDLQNLLS-DPSSTLDASRLISEACKKHGFFLVVNHGISEELISDAHEYTSRFFDMPLSEKQRVLRKSGESV
Eucalyptus grandis     HQPDIPKQFIWPDEEKPCANAPDLAVPLIDLDGFLSKDPSASEEASRLVGDACQKHGFFLVVNHGVDAGLISDAHKYMDKFFGLPLSEKQRAQRKLGEHC
Gossypium raimondii    FESQIPKEFIWPDDEKPSANAPELQVPLIDLGGFLSGDPVATMEASRFISEACQQHGFFLVVNHGVDAKLLADAHKYMDNFFLLPLRQKQRAQRKLGEHC
Jatropha curcas        YKSNVPQQFIWPDHEKPTANAPELPVPHIDLGGFLSGDPVAAMEASKLVGEACKKHGFFLVVNHGVNQKLIQDAHRYMDSFFELPLCEKQKAQRKIGEHC
Nicotiana tabacum      REYNIPTQFIWPDDEKPRAVARELPVPLIDLGGFLSGDPVAAQQASRLVGEACRNHGFFLVVNHGVNANLISNAHRYMDMFFDLPLSEKQKAQRKLEEHC
Populus tomentosa      HQTNIPQQFVWPDHEKPNINAPELQVPLVDLGDFLSGNPVAAVEASRLVGEACKKHGFFLVVNHGVDKTLIAHAHNYVDTFFKLPLSEKQKAQRKIGESC
Hibiscus cannabinus    FQSHIPKEFIWPDDEKPSANPPELQVPLIDLGGFLSGDPVATMEASRLVREACRQHGFFLVANHGVDAKLVSDAHNYMDNFFELPLCEKQRAQRKFGEDC

                              210       220       230       240       250       260       270       280       290       300
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Acacia mangium         GYASSFTSRFSSKLPWKETLSFQFSADKNSQNLVKDYLCEKVGHEFEQFGMVYQEYCEAMSNLSLVIMELLGMSLGVGRTCFREFFDKNNSIMRLNYYPP
Arabidopsis thaliana   GYASSFTGRFSTKLPWKETLSFRFCDDMSRSKSVQDYFCDALGHGFQPFGKVYQEYCEAMSSLSLKIMELLGLSLGVKRDYFREFFEENDSIMRLNYYPP
Eucalyptus grandis     GYASSFTGRFSSKLPWKETLSFGYSAEKSSANVVEDYFKNTMGEEFEQSGRVYQDYCEAMSRLSLGIMELLGMSLGIGRDHFREFFESNDSIMRLNYYPP
Gossypium raimondii    GYASSFTGRFSTKLPWKETLSFRYSAENNSSKMVEDYLVNKMGNELRQLGRVYQDYCEAMSKLSLGIMELLAISLGVGRAHFREFFDKNDSIMRLNYYPP
Jatropha curcas        GYASSFTGRFSSKLPWKETVSVRYSADNNSPKLVQHYLRNTMGETFSEFGRVYQEYCEAMSTLSLGIMELLGMSLGVSREHFREFFEENDSIMRLNYYPQ
Nicotiana tabacum      GYASSFTGRFSSKLPWKETLSFRYSAEEDSSHIVEEYFQNTMGESFSHLGNVYQEYCNSMSTLSLGIMELLGMSLGVGREHFKEFFEENESIMRLNYYPP
Populus tomentosa      GYASSFTGRFSSKLPWKETLSFRYTAEENSSKHIEEYFHNRMGEDFAEFGTVYQDYCEAMSTLSLGIMELLGMSLGVSREHFREFFNENDSIMRLNYYPP
Hibiscus cannabinus    GYASSFTGRFSTKLPWKETLSLRYSAKNSESKMVEDYVVNKLGDELRQFGRVYQDYCEAMSKLSLGIMEILAISLGVNRAHFKEFFEKNESIMRLNYYPP

                              310       320       330       340       350       360       370       380       390       400
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Acacia mangium         CQKPDLTLGTGPHCDPTSLTILHQDKVGGLQVYVDNEWHSITPNFNAFVVNIGDTFMALSNGRYKSCLHRAVVNSQTTRKSLAFFLCPGDDKVVTPPSEL
Arabidopsis thaliana   CIKPDLTLGTGPHCDPTSLTILHQDHVNGLQVFVENQWRSIRPNPKAFVVNIGDTFMALSNDRYKSCLHRAVVNSESERKSLAFFLCPKKDRVVTPPREL
Eucalyptus grandis     CQKPDLTLGTGPHCDPTSLTILHQDQVGGLQVFVDNEWRSISPNFNAFVVNIGDTFMALSNGLYKSCLHRAVVNSRTPRKSLAFFLCPRSDKVVRPPSEL
Gossypium raimondii    CQKPDLTLGTGPHCDPTSLTILHQDRVGGLQVFVDNEWHSISPNFEAFVVNIGDTFMALSNGRYKSCLHRAVVNSHKPRKSLAFFLCPEGDKVVTPPAEL
Jatropha curcas        CQKPDLTLGTGPHCDPTSLTILHQDQVGGLQVFVDNQWRSISPNFQAFVVNIGDTFMALSNGRYKSCLHRAVVNSKKPRKSLAFFLCPKSDRIVSPPSEL
Nicotiana tabacum      CQKPDLTLGTGPHCDPTSLTILHQDSVGGLQVFVDNEWRSVSPNFNAFVVNIGDTFMALSNGRYKSCLHRAVVNNKTPRKSLAFFLVPKKDKVVSPPNEL
Populus tomentosa      CQKPDLTLGTGPHCDPTSLTILHQDQVGGLQVFVDNEWRSINPNFDAFVVNIGDTFMALSNGIYKSCLHRAVVNSQTPRKSLAFFLCPKNDKMVTPPHEL
Hibiscus cannabinus    CQKPDLTLGTGPHCDPTSLTILHQDRVGGLQVFVDNEWRSVSPNSEAFVVNIGDTFMALSNGRFKSCMHRAVVNSQETRKSLAFFLSPAGDKLVAPPAEL

                              410       420       430       440
                       ....|....|....|....|....|....|....|....|....|...
Acacia mangium         VDHVSPRIYPDFTWPMLLEFTQKHYRADMNTLEQFANWVQRNKS----
Arabidopsis thaliana   LDSITSRRYPDFTWSMFLEFTQKHYRADMNTLQAFSDWLTKPI-----
Eucalyptus grandis     VAMSCPRAYPDFTWPVLLEFTQKHYRADMNTLRAFTNWLQQRTSEPVR
Gossypium raimondii    VSQNSPRVYPDFTWPMLLEFTQKHYRADMNTLQEFSNWVQQRNS----
Jatropha curcas        VDDSNPRIYPDFTWPMLLEFTQKHYRADMKTLEMFTNWVQQLQQKN--
Nicotiana tabacum      VDTNNPRIYPDFTWPTLLEFTQKHYRADMNTLQTFSNWLKQKTAQV--
Populus tomentosa      VDTCNPRIYPDFTWPMLLEFTQKHYRADMKTLEVFTNWLHQRSFT---
Hibiscus cannabinus    VNQNSPRVYPDFTWPMLHEFVQKHYRADMNTLQVFSNWVQHRNT----
```

Fig. 3. Multiple sequences alignment of GA20ox amino acids from kenaf (*Hibiscus cannabinus*) with *Acacia mangium*, *Arabidopsis thaliana*, *Eucalyptus grandies*, *Gossypim raimondii*, *Jatropha curcas*, *Nicotina tabacum*, *Populus tomentosa* and *Hibsicus cannabinus*. The substrate binding site “LPWKET” was boxed, while the Fe^{2+} binding domains were indicated with red arrow.

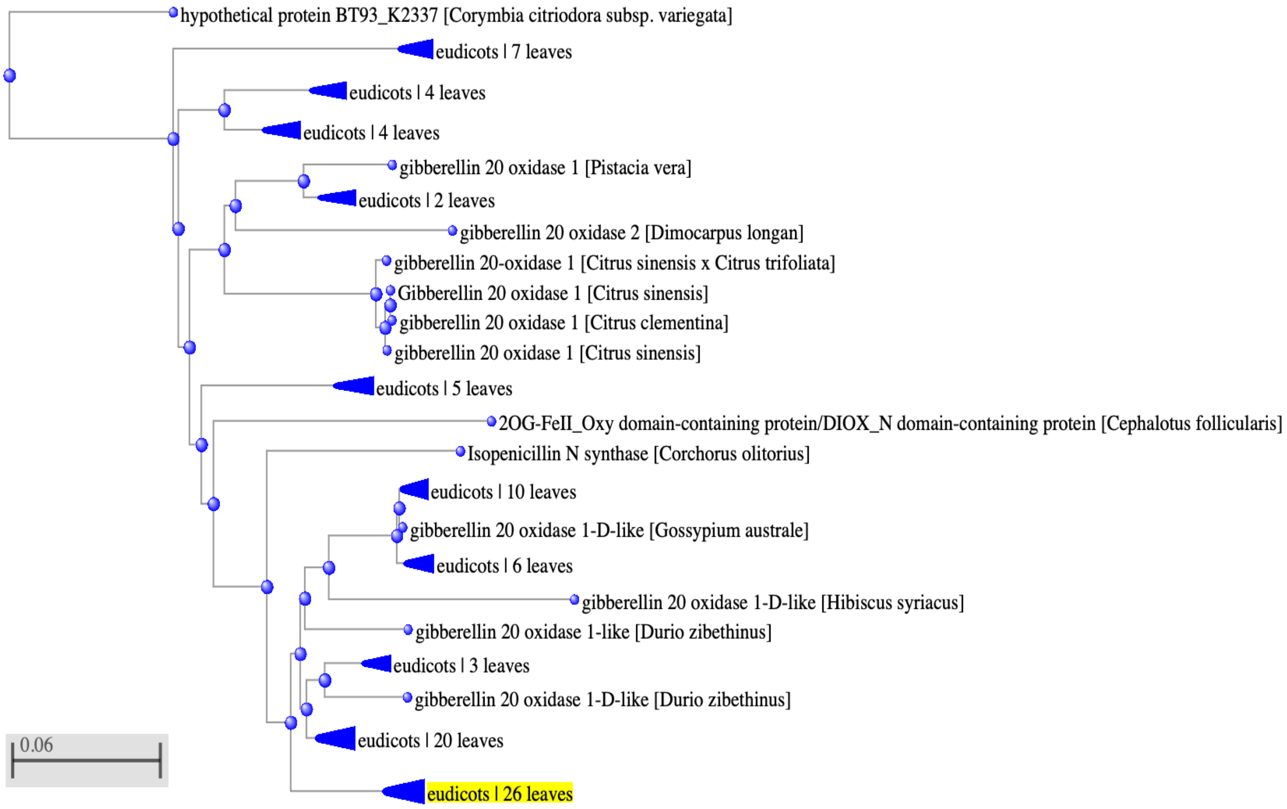

hypothetical protein BT93_K2337 [Corymbia citriodora subsp. variegata]
eudicots | 7 leaves
eudicots | 4 leaves
eudicots | 4 leaves
gibberellin 20 oxidase 1 [Pistacia vera]
eudicots | 2 leaves
gibberellin 20 oxidase 2 [Dimocarpus longan]
gibberellin 20-oxidase 1 [Citrus sinensis x Citrus trifoliata]
Gibberellin 20 oxidase 1 [Citrus sinensis]
gibberellin 20 oxidase 1 [Citrus clementina]
gibberellin 20 oxidase 1 [Citrus sinensis]
eudicots | 5 leaves
2OG-FeII_Oxy domain-containing protein/DIOX_N domain-containing protein [Cephalotus follicularis]
Isopenicillin N synthase [Corchorus olitorius]
eudicots | 10 leaves
gibberellin 20 oxidase 1-D-like [Gossypium australe]
eudicots | 6 leaves
gibberellin 20 oxidase 1-D-like [Hibiscus syriacus]
gibberellin 20 oxidase 1-like [Durio zibethinus]
eudicots | 3 leaves
gibberellin 20 oxidase 1-D-like [Durio zibethinus]
eudicots | 20 leaves
eudicots | 26 leaves
0.06

Fig. 4. Evolutionary relationship of taxa.

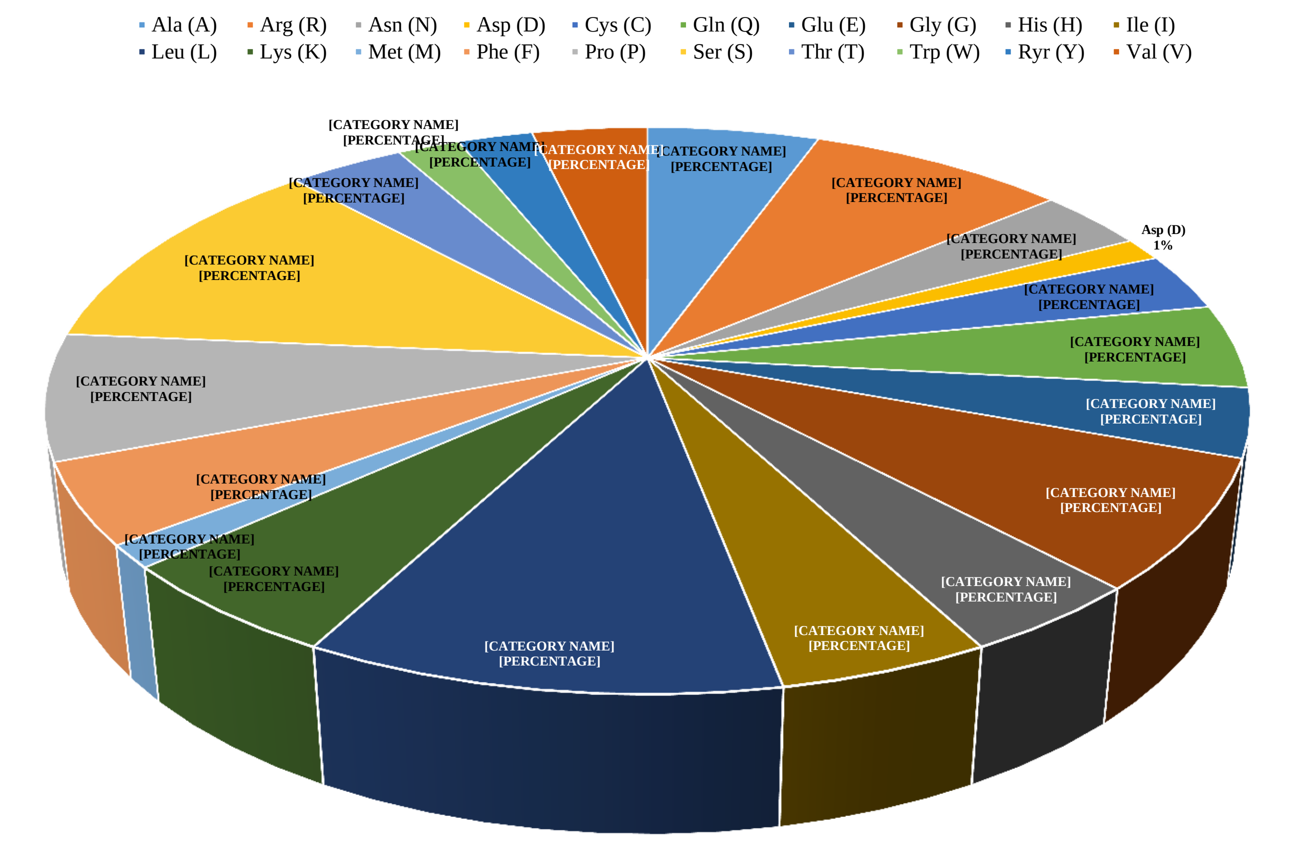
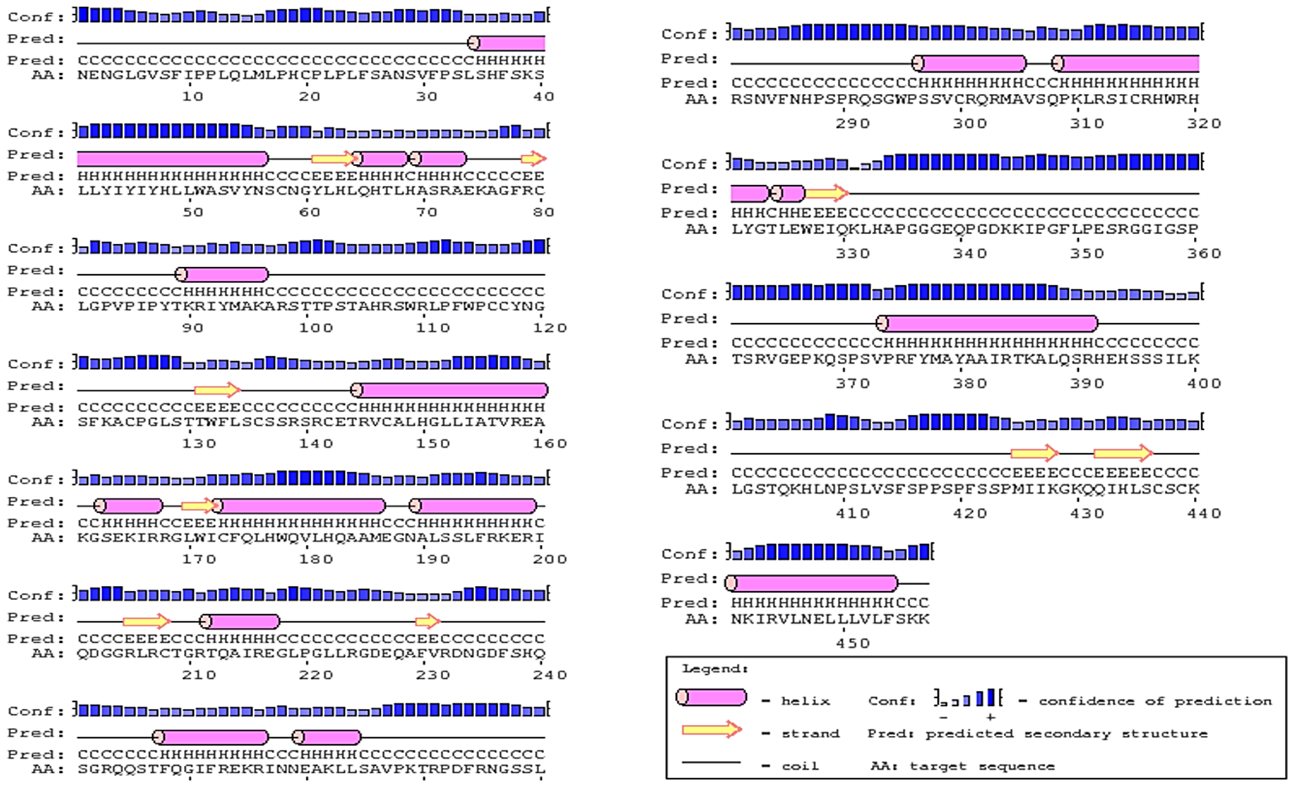


Fig. 5. The amino acid composition analysis of the protein encoded by *HcGA20o1*; predicted by Protparm (http://web.expasy.org/protscale/) in the toolkit of ExPASy.



```
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHH
  AA: NENGLGVSFIPPLQLMLPHCPLPLFSANSVFPSLSHFSKS
                10        20        30        40

Pred: HHHHHHHHHHHHHHHHCCCCEEEEHHHHCHHHHCCCCCEE
  AA: LLYIYIYHLLWASVYNSCNGYLHLQHTLHASRAEKAGFRC
                50        60        70        80

Pred: CCCCCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCC
  AA: LGPVPIPYTKRIYMAKARSTTPSTAHRSWRLPFWPCCYNG
                90       100       110       120

Pred: CCCCCCCCCCEEEECCCCCCCCCCHHHHHHHHHHHHHHHH
  AA: SFKACPGLSTTWFLSCSSRSRCETRVCALHGLLIATVREA
               130       140       150       160

Pred: CCHHHHHCCEEEHHHHHHHHHHHHHHCCHHHHHHHHHHHC
  AA: KGSEKIRRGLWICFQLHWQVLHQAAMEGNALSSLFRKERI
               170       180       190       200

Pred: CCCCEEEECCCHHHHHHCCCCCCCCCCCEECCCCCCCCCC
  AA: QDGGRLRCTGRTQAIREGLPGLLRGDEQAFVRDNGDFSHQ
               210       220       230       240

Pred: CCCCCCCHHHHHHHHHCCHHHHHCCCCCCCCCCCCCCCCC
  AA: SGRQQSTFQGIFREKRINNEAKLLSAVPKTRPDFRNGSSL
               250       260       270       280

Pred: CCCCCCCCCCCCCCCHHHHHHHHHCCCHHHHHHHHHHHHH
  AA: RSNVFNHPSPRQSGWPSSVCRQRMAVSQPKLRSICRHWRH
               290       300       310       320

Pred: HHHCHHEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: LYGTLEWEIQKLHAPGGGEQPGDKKIPGFLPESRGGIGSP
               330       340       350       360

Pred: CCCCCCCCCCCCHHHHHHHHHHHHHHHHHHCCCCCCCCCC
  AA: TSRVGEPKQSPSVPRFYMAYAAIRTKALQSRHEHSSSILK
               370       380       390       400

Pred: CCCCCCCCCCCCCCCCCCCCCCCCEEEEECCCEEEEECCCC
  AA: LGSTQKHLNPSLVSFSPPSPFSSPMIIKGKQQIHLSCSCK
               410       420       430       440

Pred: HHHHHHHHHHHHHHHCCC
  AA: NKIRVLNELLLVLFSKK
               450
```

Conf:
Pred:
Legend:
- helix
Conf: - confidence of prediction
- +
- strand
Pred: predicted secondary structure
- coil
AA: target sequence

Fig. 6. The secondary structure of *HcGA20o1* predicted by PsiPred. The bioinformatics analysis revealed that HcGA20o1 contained 17 helixes, 8 sheets, and 9 coils. The blue bar diagram shows confidence value.

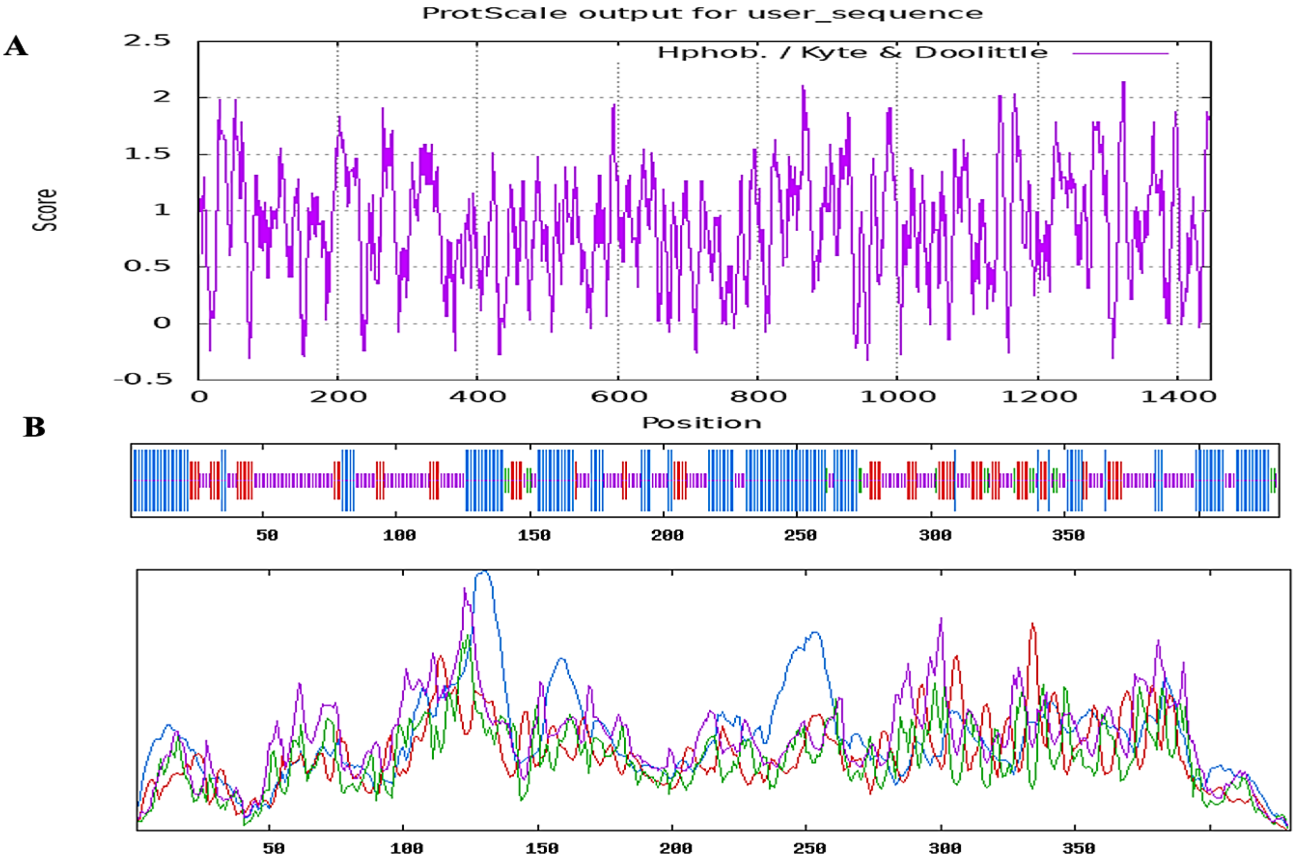


Fig. 7A. Analysis of hydrophilicity as well as hydrophobicity for the encoded protein *HcGA20o1* using the ExPASy ProtScale results of hydrophobicity as well as hydrophilicity by Kyte–Doolittle scale to evaluate the interaction degree of polar solvents such as water with specific amino acids. Fig. 7B. Shows the visualised and second contains score curve value for all predicted states of HcGA20o1 protein.

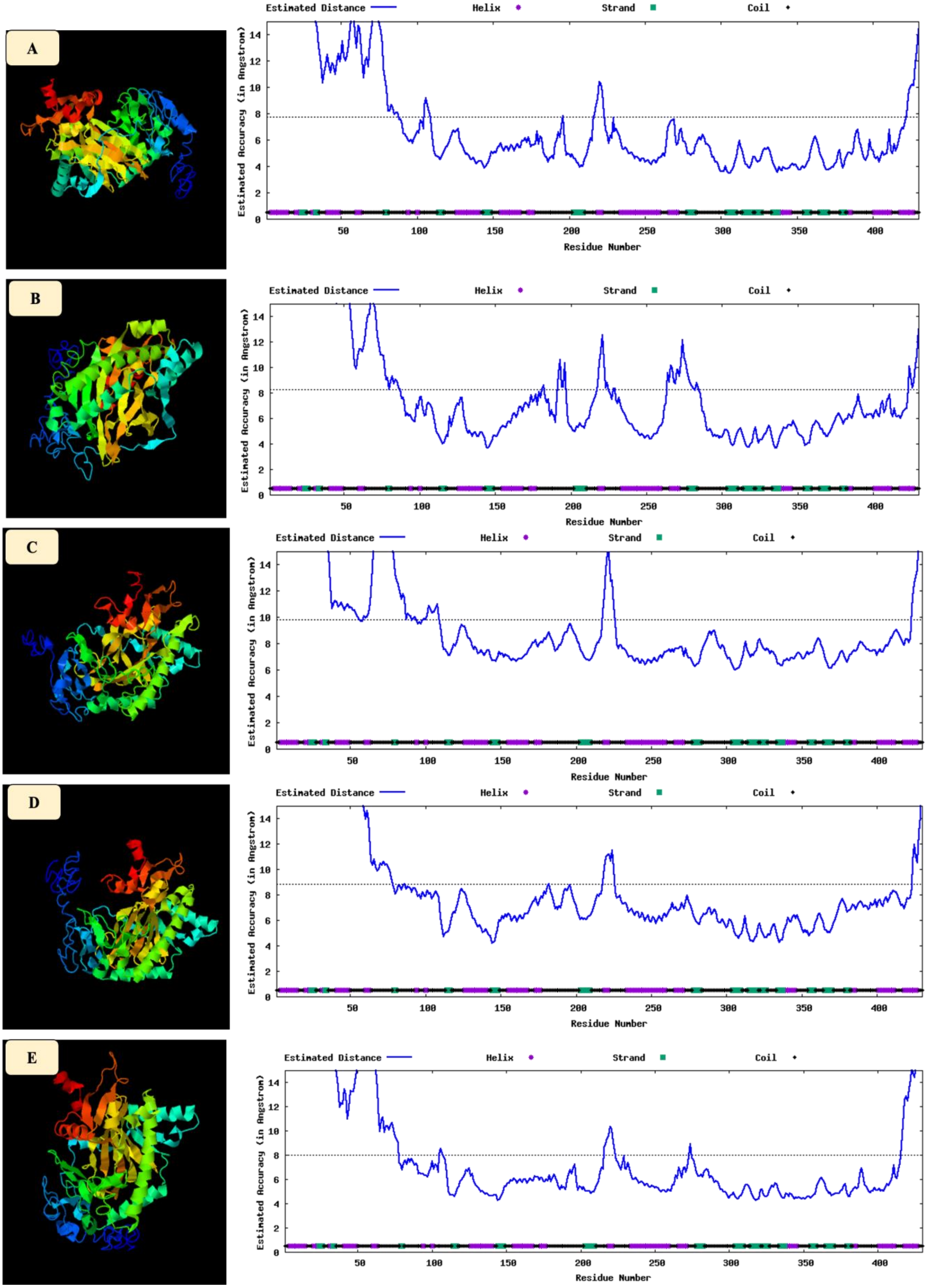


Fig. 8. The SPICKER program was used to cluster the decoys in I-TASSER. Decoys were clustered based on the pair- wise similarity structure. Five largest models were predicted and correspondent based on their structure. For each model, the confidence was predicted quantitatively using a factor named *C*-score.

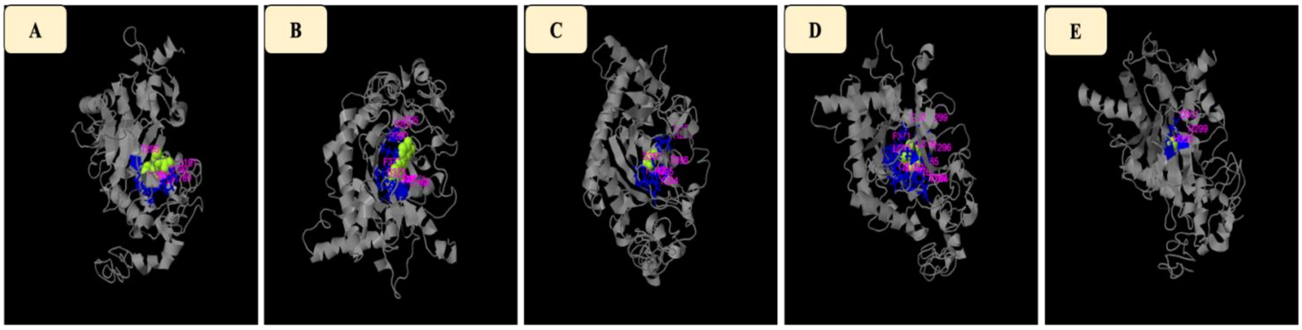


Fig. 9. Ligand binding sites of HcGA20o1 protein.

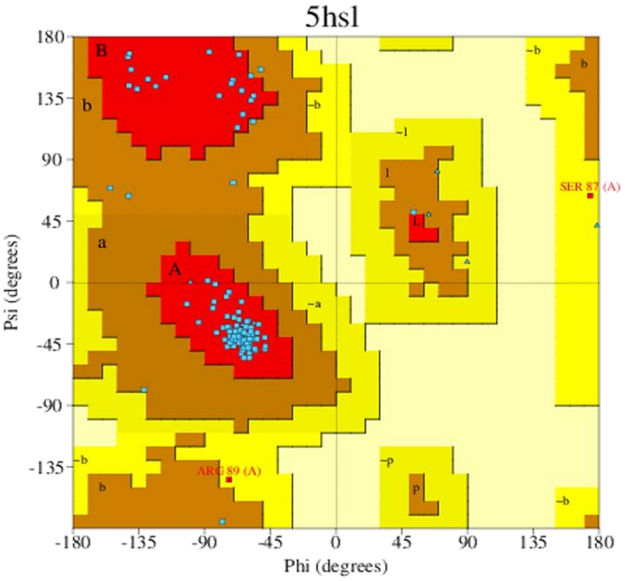


Fig. 10. Ramachandran plot of protein modelled structure validated by PROCHECK program.

Ligand binding sites of HcGA20o1 protein: The ranking of each model was predicted according to the total number of templates in cluster (cluster size). The prediction results demonstrated that the higher confidence rate was observed in the first model with (*C*- score 0.24) and cluster size 9. The *C*- score of other predicted four protein models were 0.4, 0.4, 0.2 and 0.2. PDB Hit of models 2, 3, 4 and 5 were predicted as follows: 2rcuB, 4gscB, 2o01A and 3k7Ta, respectively.

Enzyme commission (EC) numbers and active sites of HcGA20o1 protein: Enzyme commission (EC) numbers and active site results demonstrated five predicted models. The predicted *C*-score ^{EC} of five models were 0.145, 0.130, 0.129, 0.129, and 0.129. The PDB Hit of five predicted models were 2debA, 1x17B, 1x17A, 1t7nA and 1q6xA. TM-score and RMSD^a of all five models were as follows: model 1= TM-score= 0.815 and RMSD^a = 3.27, model 2= TM-score= 0.726 and RMSD^a = 3.97, model 3= TM-score= 0.719 and RMSD^a = 4.07, model 4= TM-score= 0.728 and RMSD^a = 3.98 and model 5= TM-score= 0.733 and RMSD^a = 4.18. The analysis of *HcGA20o1* ontology showed the following features of: Molecular Function=GO: 0016746, GO-Score= 0.51, Biological Process= GO: 0042221, GO: 0006810, GO: 0006631, GO-Score= 0.48, 0.35, 0.35, Cellular Component= GO: 0019866, GO: 0031966, GO: 0000267, GO: 0042579 and GO-Score= 0.51, 0.49, 0.48 and 0.48.

Ramachandran plot of HcGA20o1 protein: The predicted HcGA20o1 protein secondary structure using PSIPRED demonstrated a worthy confidence of prediction. Furthermore, quality assessment of tertiary model was evaluated applying the nearest PDB ID to the protein sequence. The PROCHECK showed that the nearest PDI ID was 5hs1 and represented by the Ramachandran plot. Accordingly, 92.0% of residue were presented in three most favourable regions (A, B, and L), 6.2% were presented in a, b, 1, and p, whereas 1.8% were presented in ~a, ~b, ~1 and ~p. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20.0, a good quality model would be expected to have over 90% in the most favoured regions [A, B, L]. G- Factors analysis showed Phi-psi distribution score was 0.19 and chil-chil distribution was 0.16. The result also demonstrated that Main- chain bond length was 0.70 and Main- chain bond angle was 0.64 and overall average was 0.37. Additionally, G-factors suggest a measure of how unusual, or out-of-the-ordinary, a property is (Fig. 10).

Discussion

Starting-up of various biological study is the extraction of high quality -RNA or/ and -DNA, while low quality may negatively affect lots of molecular biology experiments such as DNA microarray analysis, EST (expressed sequence tags) analysis, SAGE (serial analysis of gene expression) technology, subtractive hybridisation, cDNA library construction, RT-PCR, and cDNA synthesis (To, 2000; To, 2004; Shahid *et al*., 2022). High quality RNA is dependent on different factors such as plant type, age, tissue, and components. In cotton, for example, high RNA- and DNA-quality isolations have been reported using polyvinylpyrrolidone (PVP), alkaline pH, and higher buffering capacity with high concentrations of tannins and phenolic terpenoids (John, 1992). The extraction of RNA from some plants such as kenaf is difficult due to the large concentrations of polyphenol and phenolic compound, polysaccharides, and other secondary metabolites. So far, a variety of RNA extraction protocols or kits have been stated in kenaf (Footitt *et al*., 2018). Moreover, Lee *et al*., (2007) have extracted RNA from various kenaf explants such as root, stem, leaf, and inflorescence using different RNA extraction methods including TRIzol, Sodium dodecyl Sulphate (SDS), Cetyltrimethylammonium bromide (CTAB), Biotake corporation kit and hot- borate. High RNA integrity was obtained in the RNA extraction protocol using TRIzol reagent for various kenaf explants. In this regard, TRIzol is a well reported RNA extraction method, and this protocol have been introduced as a suitable liquid to liquid process for RNA extraction, but the suitability of this method is still a matter of concern for other plant types and age (Box & Tiao, 2011).

Multiple sequence alignment of *HsGA20ox1* with other *GA20ox*s revealed that the conserved and characteristic motifs of the amino acids are essential to produce gibberellin enzyme. So far, the role of a wide variety of GA biosynthesis genes has been described in different plants, especially fruits, *Arabidopsis*, and *Pyrus communis* (Plackett *et al*., 2011; Plackett *et al*., 2012). Manipulation of GA20ox gene in/ from different plants species showed the regulatory function of this gene in the metabolism and biosynthesis of GA enzyme (Plackett *et al*., 2011; Plackett *et al*., 2012; Rieu *et al*., 2008). The evaluation of *GA20ox* function in transgenic *Arabidopsis* showed a positive association in the gibberellin biochemical composition (Rieu *et al*., 2008). The pathway analysis of gibberellin mechanism demonstrated that adding soluble 2-oxo-glutarate-dependent dioxygenase (GA-oxidase) catalyses the GA biosynthesis (Plackett *et al*., 2011; Plackett *et al*., 2012). The positive regulatory function and bioactive GAs production role of GA20 have been reported in different kenaf phases such as seed germination, fruit and flower development, stem elongation as well as plant development (Yamaguchi, 2008; Jia *et al*., 2009). The identification of GA pathway in plants has been a subject of various investigations. However, a complete enzyme pathway has been reported in both *Arabidopsis* and rice. The general mechanism of GA pathway from trans-geranylgeranyl diphosphate (GGDP) is categorised into three main steps as follows: First, two different types of diterpene cyclases namely ent-kaurene synthase KS and copalyl diphosphate synthase (CPS) make ent-copalyl diphosphate (CDP) *via* transferring GGDP into the tetracyclic hydrocarbon ent-kaurene. Next, ent-kaurene is converted into GA12 using two membrane-associated P450 monooxygenases ent-kaurenoic acid oxidase (KAO) and entkaurene oxidase (KO) in the endoplasmic reticulum. Finally, some C19-GAs such as GA20 and GA9 are made using soluble 2-oxoglutarate-dependent dioxygenase (2ODDs). The main catalyser is GA 3-oxidase in the cytosol which forms C19GAs in *Arabidopsis* and rice. It has been reported that GA 2-oxidase (GA2ox) which is a 2b-hydroxylation enzyme hydroxylates C-2 of active GAs in the degradation pathway. Gene identification in rice, spinach and *Arabidopsis* shows that small genes encoded GA2oxs, and GA2oxs are classified into two main subgroups named C19GA2oxs and C20GA2oxs (Hedden & Kamiya, 1997; Lange, 1998; Hedden &Phillips, 2000).

Using PortParm helps to evaluate different factors such as the grand average of hydropathicity (GRAVY), aliphatic index, instability index, estimated half-life, extinction coefficient, atomic composition, amino acid composition, theoretical pI, and molecular weight. Protein or peptide GRAVY was calculated by dividing the hydropathy values sum to residues number of sequence of all amino acids. The aliphatic index of a protein was defined as the relative volume engaged by different aliphatic side chains like leucine, isoleucine, valine, and alanine. It has been reported that those proteins smaller than 40 instability value were stable, whereas predicted protein value more than 40 showed unstability. The extinction coefficient factor demonstrated the absorbance of light by protein at a certain wavelength, and this factor is useful to predict coefficient of purified protein by a spectrophotometer (Gasteiger *et al*., 2005). Reportedly, protein with an instability index higher than 40 possesses *In vivo* half-life less than 5 h, though 16 h and more *In vivo* half-life have been reported in the protein with instability index below 40. Since the instability index is protein factor to fin instability level, *In vivo* half-life of a protein can be measured by applying this index (Guruprasad *et al*., 1990).

In this investigation, the two key genes involved in the fibre quality and gibberellin biosynthesis of kenaf, *HcGA20ox1* (KY399834) and *HcGA20ox1* (KY399835) were partially isolated for the first time. First, RNA of three-week-old shoot was extracted using TRIzol protocol, Rapid Amplification of cDNA Ends PCR (RACE), and PCR walking approaches, which encoded the polypeptides of 430 and 345 amino acids, respectively. The RNA Analyser program showed one exon with the length of 179- 1294 in *HcGA20ox1* whereas; *HcGA200x1b* gene was flanked between the exon codon in the range of 179- 1039 bp. The highest similarity of *HcGA20ox1* with other *HcGA* genes was observed in different species as follows: *Gossypium raimondii* (89%), *Gossypium hirsutum* (89%) and *Theeobrom cacao* (89%). High similarity of HcGA20ox1 with other species of GA20ox was observed using amino acid sequencing. In addition, substrate binding site “LPWKET” and two histidine residues which serve as Fe^{2+} binding domains were found in the sequence. LPWKET were in positions 214-219 and both Fe^{2+} were reported in positions 323 and 364. The Phylogenetic tree was built using *HsGA20ox1* and 40 other homologue sequences retrieved from NCBI. The phylogenetic tree of *HcGA20o1* produced five main clusters. Multiple sequence alignment of *HsGA20ox1* with other *GA20ox*s revealed that the conserved and characteristic motif of the amino acids is essential to produce gibberellin enzyme. The molecular weight of protein was 122224.45 with 4.97 theoretical pI. The atomic compositions of protein were Carbon (C)= 2291, Hydrogen (H)= 3607, Nitrogen (N)= 681, Oxygen (O)= 626 and Sulphur (S)= 21. The instability index (II) was computed to be 59.05 and this classified the protein as unstable. Aliphatic index of protein was 76.85 and Grand average of hydropathicity (GRAVY) was -0.396. The secondary structure analysis of the protein encoded by HcGA20o1 revealed that it consists of 17 helixes, 8 sheets, and 9 coils. With the output width fixed as 70, this result released 70 amino acids and associating expected structures in each line. The result of SOPMA also demonstrated that sequence length was 457 amino acids. The ratio of each assembly was Alpha helix (Hh): 159 (34.79%), Extended strand (Ee): 62 (13.57%), Beta turn (Tt): 40 (8.75%) and Random coil (Cc): 196 (42.89%). The I-TASSER software predicted five models with local structure error profiles with the *C*-score of -3.02, -3.04, -3.04, -3.76 and -4.83 for each model, respectively. The results also demonstrated that the Estimated TM-score was 0.37±0.13 and Estimated RMSD was 14.6±3.7Å for the first model. HcGA20o1 protein structurally close to the target in the PDB showed that the PDB Hit of the best model was 2h4tB with TM-score 0.817, RMSDa = 3.31, IDENa = 0.117, and Cov= 0.908. Predicted function of HcGA20o1 protein using COFACTOR and COACH also demonstrated that the PDB Hit of the best model was 1s5oA with Lig Name 152 and different binding site residues. The prediction results of Ligand binding sites of HcGA20o1 protein demonstrated that the higher confidence rate was observed in the first model with (*C*-score 0.24) and cluster size 9. The *C*- score of other predicted four protein models were 0.4, 0.4, 0.2, and 0.2. PDB Hit of models 2, 3, 4, and 5 were predicted as follows: 2rcuB, 4gscB, 2o01A and 3k7Ta, respectively. The recommended secondary structure of HcGA20o1 protein designed by PSIPRED demonstrated a worthy confidence of prediction.

Briefly, this investigation provided the first evolutionary analysis of *Gibberellin 20-oxidase* (*HcGA20ox*) extraction and characterisation from kenaf. The results of this test may help us to identify *GA20ox* isoform(s), which are responsible for improving fibre quality and synthesis in the plant breeding programs.

Acknowledgement

The authors would like to acknowledge the Ministry of Education (MOE) Malaysia for financially supporting this research through FRGS grant (project no. 5524677). Part of this research was also funded by HICoE research grant (project no. 6369102).

References

Bismarck, A., S. Mishra and T. Lampke. 2005. Plant fibers as reinforcement for green composites. *J. Nat. Fibers,* 2: 37-108.

Bitzer, M.J, C.G. Cook and B.S. Baldwin. 2000. The development of kenaf varieties in the United States. In: Proceeding of the International Kenaf Symposium. *Hiroshima, Japan*, pp. 91-94.

Boonkaew, T.C. Mongkolsiriwatana, A. Vongvanrungruang, K. Srikulnath and S. Peyachoknagul. 2018. Characterization of *GA20ox* genes in tall and dwarf types coconut (*Cocos nucifera* L.). *Genes & Genom*., 40(7): 735-745.

Box, G.E. and G.C. Tiao. 2011. Bayesian inference in statistical analysis (Vol. 40). John Wiley & Sons.

Buchan, D.W.A., F. Minneci, T.C.O. Nugent, K. Bryson and D.T. Jones .2013. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucl. Acids Res*., 41: W349-W357.

Chou, K.C. and H.B. Shen. 2010. Cell-PLoc 2.0: An improved package of web-servers for predicting subcellular localization of proteins in various organisms. *Nat. Sci*., 2(10): 1090.

Ebenezer, A.A., A.O. David, T.A. Koyinsola, O.T. Olayemi and O.O. Olutoyi. 2019. Some effects of crude aqueous extracts of *Hibiscus sabdariffa* leaves on the testes and sperm parameters of adult male wistar rats (*Rattus norelegicus*). *J. Adv. Med.*, 1-8.

Edeerozey, A.M., H.K. Akil, A.B. Azhar and M.Z. Ariffin. 2007. Chemical modification of kenaf fibers. *Mater. Lett.*, 61(10): 2023-2025.

Fleet, C.M. and T.P. Sun. 2005. A DELLAcate balance: the role of gibberellin in plant morphogenesis. *Curr. Opin. Plant. Biol.*, 8(1): 77-85.

Footitt, S., S. Awan and W.E. Finch-Savage. 2018. An improved method for the rapid isolation of RNA from *Arabidopsis* and seeds of other species high in polyphenols and polysaccharides. *Seed Sci. Res*., 4(3): 360-364.

Gasteiger, E., C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel and A. Bairoch. 2005. Protein identification and analysis tools on the ExPASy server. *In The Proteomics Protocols Handbook*, pp. 571-560.

Geourjon, C. and G. Deleage. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, 11(6): 681-684.

Guruprasad, K., B.B. Reddy and M.W. Pandit. 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *In vivo* stability of a protein from its primary sequence. *Protein Eng. Des. Sel.*, 4(2): 155-161.

Hedden, P. and A.L. Phillips. 2000. Gibberellin metabolism: new insights revealed by the genes. *Trends Plant. Sci.*, 5(12): 523-530.

Hedden, P. and Y. Kamiya. 1997. Gibberellin biosynthesis: enzymes, genes and their regulation. *Ann. Rev. Plant Biol*., 48(1): 431-460.

Hirano, K., M. Nakajima, K. Asano, T. Nishiyama, H. Sakakibara, M. Kojima and J.A. Banks. 2007. The GID1-mediated gibberellin perception mechanism is conserved in the lycophyte Selaginella moellendorffii but not in the bryophyte Physcomitrella patens. *The Plant Cell*, 19(10): 3058-3079.

Jamaludin, J.B. 2008. Effects of fiber size modification on the mechanical properties of kenaf fiber reinforced polyester composite (Doctoral dissertation, B. Sc. Thesis submitted to the Universiti Teknikal Malaysia Melaka).

Jia, Q., J. Zhang, S. Westcott, Z.Q. Zhang, M. Bellgard, R. Lance and C. Li. 2009. GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. *Fun. Integr. Genom*., 9(2): 255-262.

John, M.E. 1992. An efficient method for isolation of RNA and DNA from plants containing polyphenolics. *Nucl. Acids. Res.*, 20(9): 2381.

Kargarzadeh, H., I. Ahmad, I. Abdullah, A. Dufresne, S.Y. Zainudin and R.M. Sheltami. 2012. Effects of hydrolysis conditions on the morphology, crystallinity, and thermal stability of cellulose nanocrystals extracted from kenaf bast fibers. *Cellulose,* 19(3): 855-866.

Lange, T. 1998. Molecular biology of gibberellin synthesis. *Planta*, 204(4): 409-419.

Laskowski, R.A., M.W. MacArthur, D.S. Moss and J.M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog*., 26(2): 283-291.

Lee, J.J., A.W. Woodward and Z.J Chen. 2007. Gene expression changes and early events in cotton fibre development. *Ann. Bot*., 100(7): 1391-1401.

Nakai, K. and P. Horton. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.,* 24(1): 34-35.

Nelissen, H., X.H. Sun, B. Rymen, Y. Jikumaru, M. Kojima, Y. Takebayashi and J. De-Block. 2018. The reduction in maize leaf growth under mild drought affects the transition between cell division and cell expansion and cannot be restored by elevated gibberellic acid levels. *Plant Biotechnol. J*., 16(2): 615-627.

Nugent, T. and D.T. Jones. 2009. Transmembrane protein topology prediction using support vector machines. *BMC Bioinform*., 10(1): 159.

Pejic, B.M., M.M. Kostic, P.D. Skundric and J.Z. Praskalo. 2008. The effects of hemicelluloses and lignin removal on water uptake behavior of hemp fibers. *Bioresour*. *Technol*., 99(15): 7152-7159.

Plackett, A.R., S.G. Thomas, Z.A. Wilson and P. Hedden. 2011. Gibberellin control of stamen development: a fertile field. *Trends Plant Sci*., 16(10): 568-578.

Plackett, A.R., S.J. Powers, N. Fernandez-Garcia, T. Urbanova, Y. Takebayashi, M. Seo and A.L. Phillips. 2012. Analysis of the developmental roles of the *Arabidopsis* gibberellin 20-oxidases demonstrates that GA20ox1,-2, and-3 are the dominant paralogs. *The Plant Cell*, 24(3): 941-960.

Rieu, I., O. Ruiz-Rivero, N. Fernandez-Garcia, J. Griffiths, S.J. Powers, F. Gong and A.L. Phillips. 2008. The gibberellin biosynthetic genes *AtGA20ox1* and *AtGA20ox2* act, partially redundantly, to promote growth and development throughout the Arabidopsis life cycle. *Plant J*., 53(3): 488-504.

Rowell, R.M. 2000. Characterization and factors effecting fiber properties. *Natural polymers and agrofibers based composites*. *Frollini*, 115-133

Shahid, M.N., S. Rasheed, M.S. Iqbal, A. Jamal, S. Khalid and Z. Shamim. 2022. *In silico* prediction of potential mirnas to target zymv in cucumis melo. *Pak. J. Bot*., 54(4): 1319-1325.

Siepe, T., D. Ventrella and E. Lapenta. 1997. Evaluation of genetic variability in a collection of *Hibiscus cannabinus* (L.) and *Hibiscus* spp. (L.). *Ind*. *Crop Prod*., 6: 343-352.

Takehara, S. and M. Ueguchi-Tanaka. 2018. Gibberellin. In: *Plant Structural Biology: Hormonal Regulations.* pp. 83-95. Springer, Cham.

Tenreira, T., M.J.P. Lange, T. Lange, C. Bres, M. Labadie, A. Monfort and B. Denoyes. 2017. A specific gibberellin 20-oxidase dictates the flowering-runnering decision in diploid strawberry. *The Plant Cell*, 29(9): 2168-2182.

Thiruchitrambalam, M., A. Alavudeen and N. Venkateshwaran. 2012. Review on kenaf fiber composites. *Rev. Adv. Mater. Sci*., 32(2): 106-111.

To, K.Y. 2000. Identification of differential gene expression by high throughput analysis. *Comb*. *Chem*. *High T. Scr*., 3: 235-241.

To, K.Y. 2004. Overview of differential gene expression by high throughput analysis. In: R. Rapley and S. Harbron (Eds.), Molecular Analysis and Genome Discovery. *John Wiley & Sons*, *Chichester*, England, pp. 167-190.

Xiao, J., H. Li, J. Zhang, R. Chen, Y. Zhang, B. Ouyang and Z. Ye. 2006. Dissection of GA 20-oxidase members affecting tomato morphology by RNAi-mediated silencing. *Plant Growth Regul*., 50(2-3): 179-189.

Yamaguchi, S. 2008. Gibberellin metabolism and its regulation. *Ann. Rev. Plant Biol*., 59: 225-251.

Yang, J., A. Roy and Y. Zhang. 2013. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, 29(20): 2588-2595.

(Received for publication 9 May 2020)