INSIGHT INTO THE THERMOSTABILITY OF NOVEL FLUORESCENT PROTEIN ISOLATED FROM SEA ANEMONE *CRIBRINOPSIS JAPONICA: IN SILICO* STUDY

Muhammad Hariz Asraf, Razauden Mohamed Zulkifli*, Nor Azlina Ahmad, Rusna Shakira Rusli, Hilzaiti Ahmad Shaari, Mohd Shahir Shamsir Omar, Siti Noor Aishah A Rohman
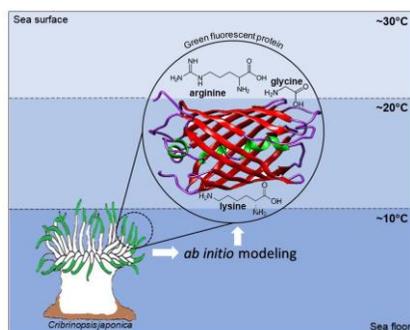
Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

Graphical abstract

Abstract

Fluorescent protein has been applied in various diagnostic and biotechnology application. However, in these applications, the temperature conditions are simulated at higher temperatures of 289K and 300K compared to natural sea anemone *Cribrinopsis japonica* fluorescent protein. This study evaluates the predictive structure and the molecular dynamic interaction of the protein with its target in different external temperature using *ab initio* similarity modeling and GROMACS/VMD respectively. Three-dimensional structure of the protein, named cjFP510 were predicted and analysed based on the highest similarity template model of *Anemonia sulcata* (2c9i) at 66.97%. The predicted model shows alternating α and β helices with longer loops and extra α-helix. cjFP510 was programmed with molecular dynamic simulation with template protein 2c9i as a reference to study its comparative adaptability in two different temperatures of 289K and 300K. cjFP510 was found to be more stable at both temperatures compared to 2c9i. Further simulation was conducted on the gyration radius to evaluate the compactness of the protein folding. Lower gyration radius of cjFP510 denotes more stable protein at 289K simulated environment than 29ci. This may be due to the presence of an extra α -helix based on the predicted model and few amino acid residues such as glycine, lysine, and arginine which contributed to the protein flexibility and thermal stability Conclusively, cjFP510 is more thermostable in the two conditional temperatures tested.

Keywords: Fluorescent protein, *ab initio*, structural prediction, molecular dynamic simulation, thermostability

Abstrak

Protein berpendarflor telah diguna pakai dalam pelbagai aplikasi diagnostik dan bioteknologi. Walaubagaimanapun, simulasi suhu dalam kajian ini jauh lebih tinggi iaitu pada 289K dan 300 K berbanding protein berpendarfluor asal daripada anemone laut, *Cribrinopsis japonica*. Kajian ini dijalankan untuk meramal struktur dan interaksi dinamik molekul protein dengan sasarannya dalam suhu luaran yang berbeza menggunakan kaedah pemodelan *ab initio* dan juga GROMACS/VMD. Struktur tiga dimensi protein, iaitu cjFP510 telah diramalkan dan dianalisis berdasarkan model persamaan tertinggi pada *Anemonia sulcata* (2c9i) iaitu sebanyak 66.97%. Model yang diramalkan menunjukkan struktur α dan β dengan gelung yang lebih panjang dan struktur tambahan α-heliks. cjFP510 telah diprogramkan dengan simulasi dinamik molekul dengan protein templat 2c9i sebagai rujukan untuk mengkaji perbezaan dalampenyesuaiannya pada dua suhu yang berbeza iaitu 289K dan

300K. cjFP510 didapati lebih stabil pada kedua-dua suhu berbanding 2c9i. Simulasi selanjutnya telah dijalankan pada jejari legaran untuk mengkaji kepadatan lipatan protein. Nilai jejari legaran yang rendah pada cjFP510 menandakan protein ini lebih stabil pada persekitaran simulasi iaitu pada suhu 289K berbanding 29ci. Keadaan ini mungkin disebabkan oleh kehadiran α-heliks tambahan pada struktur protein yang diramalkan dan juga beberapa amino asid iaitu glisina, lisina dan arginina yang menyumbang kepada kepada kelenturan protein dan kestabilan haba.Kesimpulannya, cjFP510 adalah lebih stabil haba dalam kedua-dua suhu.

Kata kunci: Protein berpendaflor, *ab initio*, ramalan struktur, simulasi dinamik molekul, kestabilan suhu

1.0 INTRODUCTION

Cribrinopsis japonica is a newly discovered species of sea anemone (Anthozoa: Actiniaria) living in the deep-sea at temperature 1.5 – 3.0 °C. In 2014, this pinkish-colored anemone was collected at a depth of about 800 m from Toyama Bay (36°58'5N, 137°22'7E) in the Sea of Japan and was further identified and characterized [1]. Further exploration done by Tsutsui, Shimada and Tsuruwaka [2] and their findings showed that *C. japonica* is photosensitive and it displays green fluorescence property emitted from the tentacles when excited by the blue light. This finding spark interest in the study of deep-sea fluorescent protein since previously, these proteins were found mostly in shallow-water cnidarians where there is good penetration of sunlight compared to deep-sea.

In industrial applications, fluorescent proteins have been applied in molecular biology and biotechnology as a fusion tagging [3], biosensors [4], cell marking [5] and epitope tag for protein purification [6]. Fluorescent proteins have also been used in drug screening [7], transgenic study [8], and act as a reporter for DNA double-strand break repair [9]. In addition, fluorescent proteins also have the potential to act as antioxidants in living animals [10].

A previous study has shown that the most similar fluorescent protein (66.97% identity) to *C. japonica's* fluorescent protein is the green fluorescent protein as(S)FP499 from another sea anemone; *Anemonia sulcata* (UniProt entry: Q9GPI6) which has been modelled. Generally, a fluorescent protein has a beta-barrel structure consisting of eleven β-strands, with an alpha helix containing the covalently bonded chromophore 4-(p-hydroxybenzylidene) imidazolidin-5-one (HBI) in the centre [11].

Moreover, this deep-sea fluorescent protein has unique characteristics, including high stability, as it needs to withstand the extreme environment of the deep sea. However, there is no known structure prediction has been made yet for *C. japonica* fluorescent protein. Therefore, a structure prediction for *C. japonica* fluorescent protein is made and compared to the existing green fluorescent protein from the cnidarian [12].

2.0 METHODOLOGY

Sequence Analysis

The length of the amino acid residue of this fluorescent protein isolated from *C. japonica*, cjFP510 (accession number A0A146FGB2) is 227 amino acids. The FASTA format of the cjFP510 sequence [13] was obtained from UniProtKB (http://www.uniprot.org/uniprot/A0A146FGB2) before proceeding further using several bioinformatics tools such as BLAST [14]. InterProScan version 5.20-59.0 [15] was used to identify the conserved domain and its specific function in the protein primary sequence to assess the idea of and provide the accurate function of the protein in cjFP510. PTMcode was used to recognize any post-translational modification that occurred within the protein sequence [16]. Multiple sequence alignment (MSA) was done using ClustalOmega comparing with the other three proteins with the highest sequence identities. ESPript (http://espript.ibcp.fr/ESPript/ESPript/) was used to perform multiple sequence alignment with color-coded conserved regions. Then, HHPRED (https://toolkit.tuebingen.mpg.de/hhpred) was executed to align the secondary structure at the residue level. In order to predict the presence and location of signal peptide cleavage sites in the amino acid sequences, SignalP version 4.1 was used. Lastly, the identification of protein glycosylation was observed using NetNGlyc 1.0 Server to know the secretion and localization of the protein. In addition to that, EXPASY's ProtParam tool was used to assess the physicochemical properties of cjFP510 and several other template proteins based on the MSA result.

Structure Prediction and Evaluation

It has been shown that the homology modeling method has been used since the WHAT IF modeling option was available back then [17]. The structural homolog search was performed in UniProt using BLAST built-in, embedded in the program database. The structural homologs in the database were

identified based on the sequence identity and the scores.

ITASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) was chosen since the programs used known protein sequences to predict and generate a model structure. Recently, the use of the *ab-initio* method alone seems to be insufficient. Other than ITASSER, PHYRE and MODWEB were used to align cjFP510 3D structure with known protein structure based on homology modeling. ITASSER combines several methods to predict the structure of the unaligned region [18]. The generated model was evaluated by its root mean square deviation (RMSD), TM score, and DOPE profile. In addition to that, VERIFY3D and ERRAT were utilized to further justify the model. Ramachandran plot was used to verify the predicted secondary structure such as the beta-strands and alpha-helices. PROCHECK was done to check the stereochemical quality of the predicted structure. All of the verifications of the models were done using online tools from UCLA (http://services.mbi.ucla.edu/SAVES/) DAS TmFilter was utilised to predict the transmembrane protein. To elucidate the predicted 3D model, UCSF Chimera version 1.14 was used to represent the structure in a ribbon form of a cartoon diagram. Additionally, the superimposition of the model structure was performed and compared with the template using UCSF Chimera version 1.14 [19].

Molecular Dynamic (MD) Simulation and Temperature Adaptation Analysis

This recently discovered fluorescent protein was simulated using MD simulation to provide a better understanding of the molecular interaction between cjFP510 and its *in silico* system or surrounding such as its thermostability. This program capable of creating an environment almost similar to the protein's surrounding nature. The simulation was performed using a high-end CPU operating with LINUX OS. GROMACS version 5.1.4 was used to discover the stability of cjFP510 at different temperatures (289K and 300K). In this study, the template structure of the fluorescent protein as(S)FP499 from *A. sulcata* PDB ID: 2c9i [20] was chosen due to its notable sequence similarity of 66.97% with our model cjFP510. The template protein was also analysed the same way with two distinct temperatures of 289K and 300K. 2c9i was used to compare how both proteins will either react or adapt in the given temperatures. The simulations were conducted simultaneously using four sets of computers. The atoms of the protein were submerged with water molecules into a cubic box created with a dimension of 10 Å using the editconf command. gmx solvate was used to solvate the protein within the boundary of the box. The protein possessed a positive charge with a value of 3.99. Therefore, the protein's charge was neutralized to pH 7 by the addition of 4 Cl^- ions. Then, the energy

minimization of the system was set at 5000 steps of the steepest descent. Energy equilibration was set up through grompp for position-restrained molecular dynamics. After that, particle-mesh Ewald method [21] was performed and the production stage at 10 ns was done at two temperature levels of 289K and 300K. The trajectories obtained were visualized as g_rms, g_rmsf, g_energy, and g_gyrate of GROMACS. xmgrace command was used to portray the trajectories into readable graphs. The simulations were run in triplicates. All the projected graphs were visualized, analysed, and edited using qtgrace and every 3D model structures were prepared using Visual Molecular Dynamics (VMD) [22].

3.0 RESULTS AND DISCUSSION

Sequence Analysis

It is worth noting that the sequence of cjFP510 is classified under green fluorescent protein (GFP)-related family (InterPro Accession No.: IPR0011584, in red colour) and homologous superfamily of GFP (InterPro Accession No.: IPR00009017, in brown colour) as reported in InterProScan server (Figure 1). The green fluorescent-like protein family consists of fluorescent proteins and non-fluorescent chromoproteins, derived from several species of Cnidarians, as well as certain diazotrophic bacteria [12]. These proteins range in their absorption wavelength maximum, and are often classified by their colour: green, yellow, red and purple-blue.
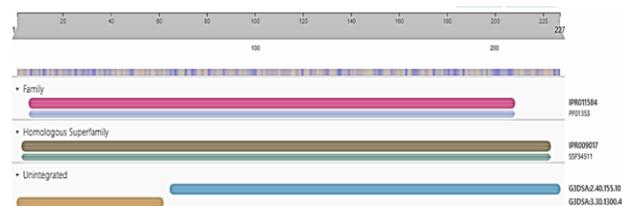
Figure 1 The output from InterProScan server showing the protein family membership of cjFP510

Moreover, cjFP510 possess no post-translational modification activity which involved in the structural stabilization and regulation of eukaryotic proteins. This information was obtained from PTMCode webserver. Besides, based on NetNGlyc 1.0 server, this sequence may not contain a signal peptide. It is reflected to the fact that proteins without signal peptides are unlikely to be exposed to the N-glycosylation machinery and thus may not be glycosylated (*in vivo*) even though they contain potential motifs. Additionally, SignalP confirms that there is no signal peptide detected which denotes the absence of cleavage site on the protein.

Block 1 (residues 1–50):

```
                                      β1              β2              β3            η1
NFCP_HETCR     ───────────────────►TT      ────────────►   ──────────────►TT      ΩΩΩΩ
GFP_AEQVI      ───────────────────►TT      ────────────►                          ΩΩΩΩ
                1       10      20      30      40      50
A0A146FGB2_9CNID ...MFPSIKESMRVKLNMEGSVNCRAFKCTGEGEGKPYEGTRSMNIEVTEGGPLPFAFDII
Q9GPI6_ANESU     ...MYPSIKETMRVQLSMEGSVNYHAFKCTGKGEGKPYEGTQSLNITITEGGPLPFAFDIL
NFCP_CONGI       ...MAGLLKESMRIKIYMEGTVNGYHFKCEGEGDGNPFEGTQNMRIRVTEGAPLPFAFDIL
NFCP_HETCR       ...MAGLLKESMRIKMYMEGTVNGHYFKCEGEGDGNPFTGTQSMRIHVTEGAPLPFAFDIL
Q7Z168_9CNID     .....MNLSKNVSVSVYMKGNVNNHEFEYDGEGGGDPYTGKYSMKMTLRGQNCLPFSYDII
GFP_AEQVI        MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTG.KLPVPWPTL
```

Block 2 (residues 60–110):

```
                      β4         α1          β5             β6
NFCP_HETCR  ΩΩΩ   TT   ──►  TT  ΩΩΩΩ   ────────────►TT  ───────────►TT
GFP_AEQVI         ΩΩΩ      ΩΩΩ  ΩΩΩΩ   ────────────►TT              TT
             60      70      80      90      100     110
A0A146FGB2_9CNID SHCFAYGQKMFIKHPKDIP..DFFKQSFPEGYTWREVTFEDGGVLAINQDTSLQNDCFIC
Q9GPI6_ANESU     SHAFQYGIKVFAKYPKEIP..DFFKQSLPGGFSWRVSTYEDGGVLSATQETSLQGDCIIC
NFCP_CONGI       SPCCAYGSKTFIKHTSGIP..DYFKQSFPEGFTWERTTIYEDGGVLTAHQDTSLEGNCLIY
NFCP_HETCR       APCCEYGSRTFVHHTAEIP..DFFKQSFPEGFTWERTTTYEDGGILTAHQDTSLEGNCLIY
Q7Z168_9CNID     TTAFQYGFRVFTKYPEGIV..DYFKDSLPDAFQWNRRIVFEDGGVLNMSSDITYKDNVLHG
GFP_AEQVI        VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVN
```
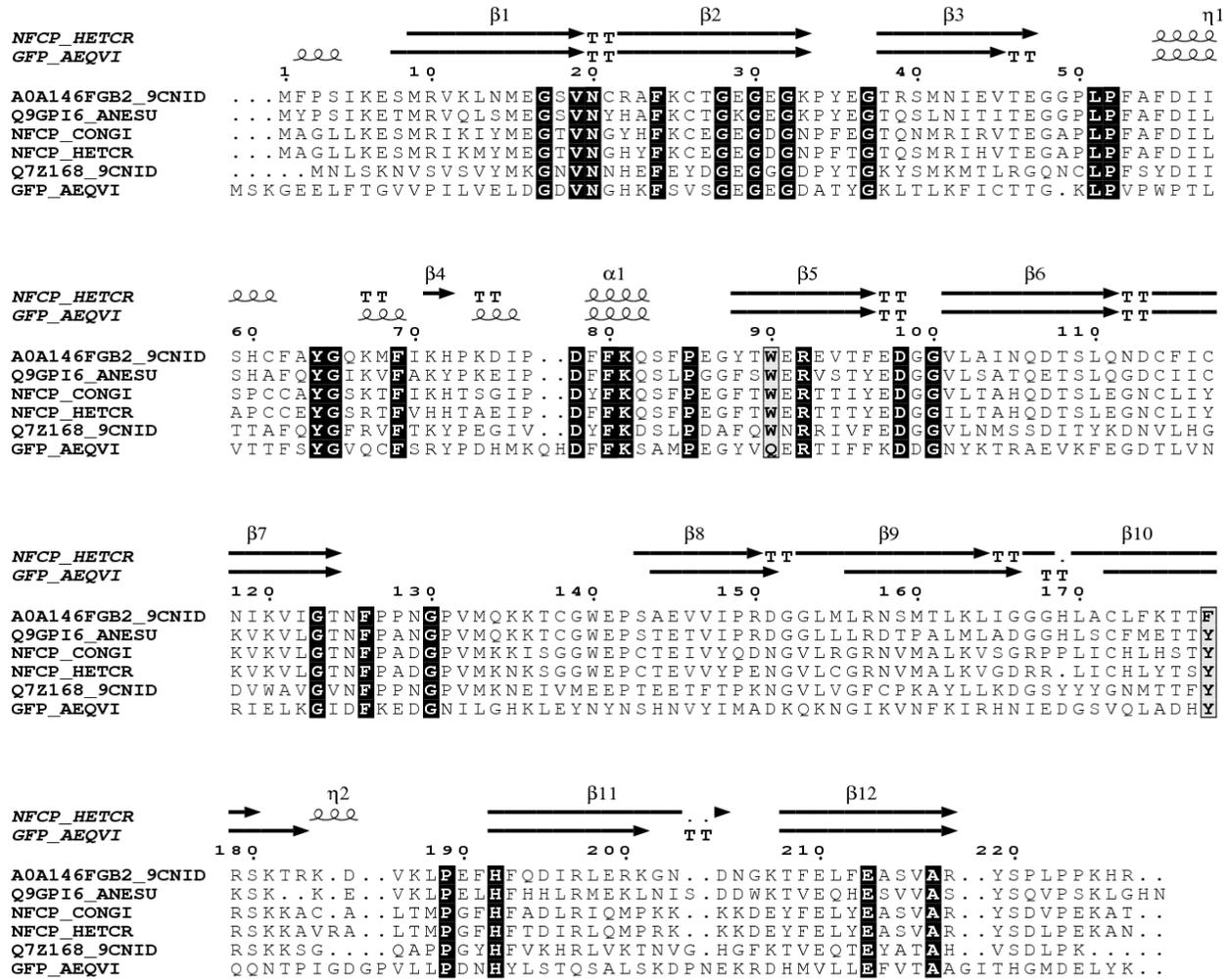
Block 3 (residues 120–170):

```
       β7                        β8           β9          β10
NFCP_HETCR ──────────►       ────────►TT  ────────►TT ►TT
GFP_AEQVI                                              TT
            120     130     140     150     160     170
A0A146FGB2_9CNID NIKVIGTNFPPNGPVMQKKTCGWEPSAEVVIPRDGGLMLRNSMTKLIGGGHLACLFKTTF
Q9GPI6_ANESU     KVKVLGTNFPANGPVMQKKTCGWEPSTETVIPRDGGLLLRDTPALMLADGGHLSCFMETTY
NFCP_CONGI       KVKVLGTNFPADGPVMKKISGGWEPCTEIVYQDNGVLRGRNVMALKVSGRPPLICHLSTY
NFCP_HETCR       KVKVLGTNFPADGPVMKNKSGGWEPCTEVVYPENGVLCGRNVMALKVGDRR.LICHLYTSY
Q7Z168_9CNID     DVWAVGVNFPPNGPVMKNEIVMEEPTEETFTPKNGVLVGFCPKAYLLKDGSYYYGNMTTFY
GFP_AEQVI        RIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHY
```

Block 4 (residues 180–220):

```
          η2          β11        β12
NFCP_HETCR ──►  ΩΩΩ       ──────►..►   ──────────►
GFP_AEQVI  ──►                    TT   ──────────►
            180     190     200     210     220
A0A146FGB2_9CNID RSKTRK.D..VKLPEFHFQDIRLERKGN..DNGKTFELFEASVAR..YSPLPPKHR..
Q9GPI6_ANESU     KSK..K.E..VKLPELHFHHLRMEKLNIS.DDWKTVEQHESVVAS..YSQVPSKLGHN
NFCP_CONGI       RSKKAC.A..LTMPGFHFADLRIQMPKK..KKDEYFELYEASVAR..YSDVPEKAT..
NFCP_HETCR       RSKKAVRA..LTMPGFHFTDIRLQMPRK..KKDEYFELYEASVAR..YSDLPEKAN..
Q7Z168_9CNID     RSKKSG....QAPPGYHFVKHRLVKTNVG.HGFKTVEQTEYATAH..VSDLPK.....
GFP_AEQVI        QQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK..
```

Figure 2 The multiple sequence alignment (MSA) of fluorescent protein consisting of green fluorescent protein of *Aequorea victoria* (Water jellyfish) (UniprotKB: P42212), orange fluorescent protein from *Cerianthus sp.* DW-2003 (UniprotKB: Q7Z168), GFP-like non-fluorescent chromoprotein of *Condylactis gigantea* (Giant Caribbean anemone) (*Condylactis Passiflora*) (UniprotKB: Q95W86), GFP-like non-fluorescent chromoprotein of *Heteractis crispa* (Leathery sea anemone) (*Radianthus macrodactylus*) (UniprotKB: Q95W85), cjFP51and 2c9i using ESPript

Table 1 Comparison of Gly, Lys and Arg compositions between the cjFP510 (model) and 2c9i (template) with other additional templates with high similarities based on MSA result obtained from ESPRIPT. M represents model structure whereas T depicts template structure

Criteria	UniProtKB accession number					
	A0A146FGB2 (M)	Q9GPI6 (T)	P42212 (T)	Q95W86 (T)	Q95W85 (T)	Q7Z168 (T)
Number of amino acid	227	228	238	227	227	222
Molecular weight (g/mol)	25654.53	25368.98	26886.32	25416.19	25637.30	25121.36
Glycine residue (%)	9.3	9.2	9.2	**9.7**	9.3	9.9
Lysine residue (%)	**8.4**	8.3	8.4	7.9	6.6	7.7
Arginine residue (%)	**5.3**	2.2	2.5	4.4	5.3	2.7

```
Q ss_pred         CccCccceEEEEEEEEECCEEEEEEEEEeeccCCccEEEEEEEEEeCCCCCCCChHHHhHhhhCChhccCCCCCCcHHHH
Q cjFP510      2   PSIKESMRVKLNMEGSVNCRAFKCTGEGEGKPYEGTRSMNIEVTEGGPLPFAFDIISHCFAYGQKMFIKHPKDIPDFFKQ    81 (226)
Q Consensus    2   ~~~~~~~m~~k~~~~~G~vNg~~F~i~G~G~g~p~~G~~~~~l~~t~g~pLP~s~~iL~~~~~yG~~~f~~Yp~gi~d~Fk~    81 (226)
                  .+|+++|+++++|+|+|||||+|+|+|+|.|+|++|+++++++|++|+||||+|+||+++|+||+++|++||++|+||||+
T Consensus    3   ~~~~~~~~~~~~~G~vNg~~F~i~G~G~g~~~~G~~~~~l~~~g~~lp~~~~il~~~~~~yG~~~f~~yp~~i~d~fk~      82 (228)
T 2c9i         3   PSIKETMRVQLSMEGSVNYHAFKCTGKGEGKPYEGTQSLNITITEGGPLPFAFDILSHAFQYGIKVFAKYPKEIPDFFKQ    82 (228)
T ss_pred          ccccccccEEEEEEEEEECCEEEEEEEEEEeeCCccceEEEEEEEEeCCCCCCCChhHHHHHHHhCCHHHhcCCCCcCcHHHH
```

```
Q ss_pred         hCCCCceEEEEEEecCCcEEEEEEEEEEEeCCEEEEEEEEEEEEeeCCCCCcccccccccccCCCeEEEEECCEEEEEEEEEEE
Q cjFP510     82   SFPEGYTWEREVTFEDGGVLAINQDTSLQNDCFICNIKVIGTNFPPNGPVMQKKTCGWEPSAEVVIPRDGGLMLRNSMTL   161 (226)
Q Consensus   82   s~p~Gyt~eR~~~fedgg~~l~~~~~i~~~~~~~~~~~~~~~~~~~~G~~f~~dgpvm~~~~~~~~~p~~e~~~p~~g~l~~~~~~~   161 (226)
                  |||+||+|+|+++|||||++++.++|+++++|+++++++.|+||++|||||++++..|+|++|+|+|+++|.|++++.++|
T Consensus   83   s~p~Gy~~eR~~~fedgg~~~~~~~~~~~~~~~~~~~~~~~~G~~f~~dgpvm~~~~~~~~~p~~e~~~p~~g~l~~~~~       162 (228)
T 2c9i        83   SLPGGFSWERVSTYEDGGVLSATQETSLQGDCIICKVKVLGTNFPANGPVMQKKTCGWEPSTETVIPRDGGLLLRDTPAL   162 (228)
T ss_pred          hCCCcEEEEEEEecCCcEEEEEEEEEEEeCCEEEEEEEEEEEEeeCCCCCcccccCccccccCCCeeEEEEcCCeEEeeEeeEE
```

```
Q ss_pred         EEcCCCEEEEEEEeEEEeCCCCCccCCCCCceEEEEEEEEEEec-CCCCEEEEEEEEEEEEEeeCCCCCCC
Q cjFP510    162   KLIGGGHLACLFKTTFRSKTRKDVKLPEFHFQDIRLERKGN-DNGKTFELFEASVARYSPLPPKH        225 (226)
Q Consensus  162   ~~~~dG~~~~~~~~~~t~y~~~~~~~~~~~~P~~Hfv~~~~~~~~-~~~~~~~~~~~E~~~A~~~~l~~~~        225 (226)
                  +++||++|+|+++|+|++++  ++.+|++|||+|++++++. ++++|+.|+|+|+++++|+++.
T Consensus  163   ~~~~dG~~~~~~~~~~t~y~~~~~--v~lP~~hfv~~~~~~~~~~~~~~~~~~~~E~~~a~~~~~~~~        225 (228)
T 2c9i       163   MLADGGHLSCFMETTYKSKK--EVKLPELHFHHLRMEKLNISDDWKTVEQHESVVASYSQVPSKL        225 (228)
T ss_pred          EeCCCCEEEEEEEeEEEeCC--CcCCCCceEEEEEEEEeecCCCCCEEEEEEEEEEEEEeCCCCccc
```

Figure 3 The alignment between cjFP510 and 2c9i in secondary structure and residue levels predicted by the HHpred server. The blue letters show the aligned beta-strands, the red letters show the aligned alpha-helices and the letter 'c' denotes the protein coils

Figure 2 and 3 shows the alignment of cjFP510 and its homologous proteins from different species and alignment of cjFP510 together with 2c9i, respectively. From Figure 3, the two protein structures possessed an identical secondary structure such as α-helices and β-strands except for extra α-helix in cjFP510 at positions 67 and 68. Moreover, the alignment shows a short gap in the template structure at positions 61 to 63, this represents a loop in the cjFP510 from position 64 to 66.

Summary of ProtParam Results

Table 1 shows the amino acid composition between cjFP510 and several other template organisms with high similarities. All of the proteins from these organisms show a significantly high percentage of glycine (Gly) residues. Gly dominates the total percentage of amino acids for all organisms compared to lysine (Lys) and arginine (Arg). Hence, this finding shows no marked difference in Gly residues between the organisms. Gly is important for cold adaptation of enzymes in alkaline phosphatase [23].

Glycine betaine is also present in most plants for cold adaptation such as mangroves [24]. Mutation of this enzyme at Gly261 and Gly262 with Alanine (Ala) resulted in lower stability with high energy activation and inactivation, respectively. In another study, a compositional bias of increased lysine-arginine ratio

was detected in cold-adapted α-amylase from *Pseudoalteramonas haloplanktis* [25]. It was stated that replacing Lysine (Lys) with homo-arginine (hR) in α-amylase will increase its stability but in return, the enzyme becomes less active. Based on the ProtParam result, cjFP510 (model) has the highest ratio of lysine-arginine residue compared to other organisms. Thus, it can be assumed that the flexibility ratio of lysine-arginine residues will affect the stability and the flexibility of the protein. Additionally, the importance of lysine itself is significant to determine the ability of the protein to adapt in a colder environment.

3D Model Prediction and Analysis of the Model

The 3D model of the cjFP510 is represented in Figure 4(A). Overall, the structural features of the cjFP510 are almost identical to other green fluorescent protein from other species. The model structure consists of α-helices and an alternating pattern of β-strands which form a β-barrel structure. The superimposition between the cjFP510 and 2c9i is represented in Figure 4(B and C). The cjFP510 is homologous to one of the chains from the 2c9i which is chain C. According to this superimposition, the cjFP510 structure has additional α-helices at the center of the protein. The flexibility and thermostability of the protein are enhanced by the additional α-helix in the protein domain of this

organism as compared to the template. Therefore, this modelled protein can adapt to extreme temperatures. The amino acid residues (Gly, Lys, Arg) involved in the adaptation might be contributing to the protein stability. The overall structure of cjFP510 has a 66.97% similarity index to that of the 2c9i. However, cjFP510 molecule has longer external loops on the surface that comprises a gap with 2c9i in four different positions of the alignment.

Evaluation of the Model

To assess the quality of the constructed 3D-model, several model evaluation tools were implemented. Firstly, PROCHECK is used as a tool to investigate the backbone conformation based on a Psi/Phi Ramachandran plot. According to the results of PROCHECK, no residues were located in the disallowed region. 77.1 % of the residues were found in the most favorable region, 18.1% in additional allowed regions and the rest of the residues were in the generously allowed regions. For the analysis using VERIFY3D, 83.70 % of the residues had an average 3D–1D score above 0.2. The quality of this constructed model is considered satisfactory as it obtains a VERIFY3D score above 80 % [26]. Another tool used is ERRAT to assess the overall quality of the model for non-bonded atomic interactions by comparing the statistics of highly refined structures. The accepted range of the ERRAT score for a good

model is above 50 %, and a higher score indicates a better quality [27]. The ERRAT score of the model was 83.40 % which is above the normal range. Table 2 shows a summary of the evaluation results. The overall scores obtained for the model using different evaluation tools are considered reasonable as it fulfilled all the normal range of scores for all three evaluation tools.

Table 2 Model structure (cjFP510) evaluation using different tools. The scores obtained reflect the quality and acceptable range of the modelled structure

Model evaluation tool	Evaluation scheme	Obtained score	Normal range of score
PROCHECK	The number of residues in an allowed region based on Psi/Phi Ramachandran plot	100.00%	>90 %
VERIFY3D	The number of residues having an average 3D-1D score above 0.2	83.70 %	>80 %
ERRAT	The overall quality for non-bonded atomic interactions	83.40 %	>50 %
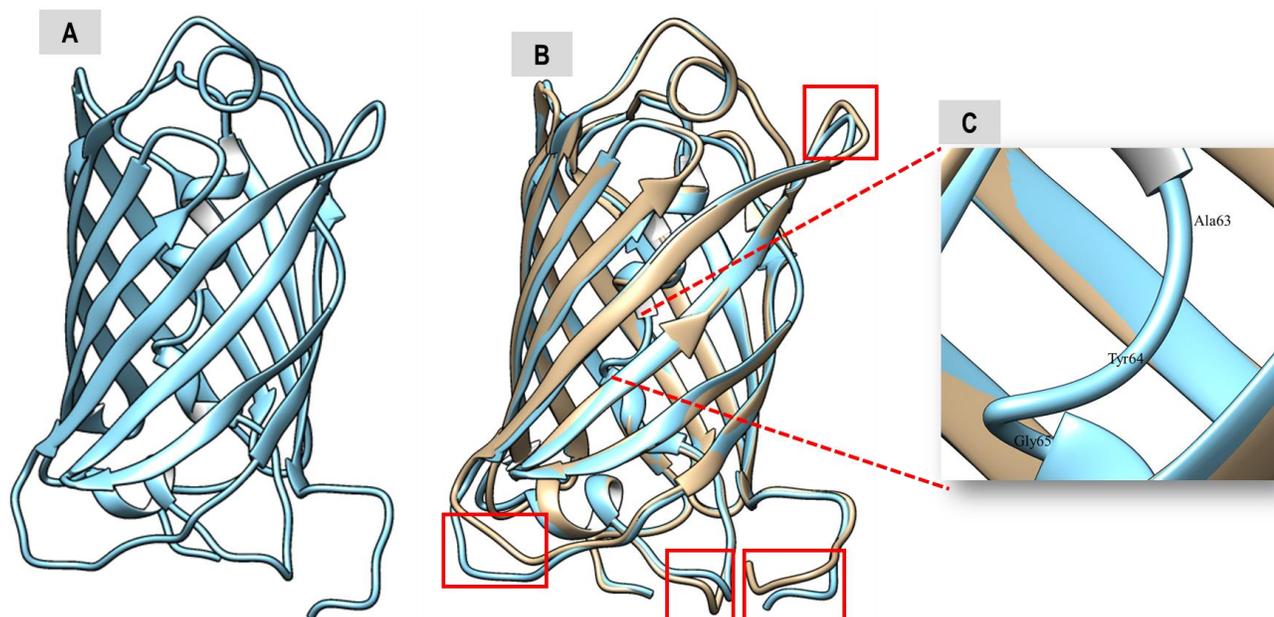
Figure 4 (A) 3D-model prediction of cjFP510 representing the secondary structure elements including α-helices, β-strands, and loops, and (B) Superimposition of cjFP510 (cyan) and its template, 2c9i (biege) in 3D visualization. Longer loops in 2c9i indicated in four regions (in red boxes) and one longer cjFP510 loop(C)

Molecular Dynamic Simulation in Different Temperatures

The stability of cjfp510 protein was evaluated at warmer (289K, equivalent to 15.85°C) and mesophilic

conditions (300K, equivalent to 26.85°C) and both model and template were subjected to 10 ns MD simulation. cjFP510 and 2c9i exhibits its fluorescence activity in the surrounding temperature of 289K and 300K, respectively [10]. The global behavior of both

cjFP510 and 2c9i were analyzed according to the root mean square deviation (RMSD) and the root mean square fluctuations (RMSF) of the protein backbone. These parameters are crucial for evaluating both stability of the structure and flexibility of individual residue in the structure during the simulation. Based on Figure 5(A), cjFP510 (in black) shows a fluctuation in the RMSD for 289K, but as it reached 10,000 ps, the RMSD graph started to stabilize. Whereas, for 2c9i (in red, the RMSD graph shows almost similar behavior as cjFP510, but as it reached 5000 ps, the RMSD graph exhibits unsteadiness as this temperature is not suitable for a mesophilic protein. Interestingly for RMSD in 300K condition, cjFP510 shows a more stable behavior starting at 2000 ps compared to 2c9i. This finding indicated that the model protein from *C. japonica* is thermostable as the protein started to adapt in a slightly warmer temperature 10].

Furthermore, Figure 5(B) illustrates the RMSF plot for cjFP510 with the highest peak at amino acid 227, and this is applied to both temperatures. It can be assumed that the amino acid arginine at amino residue number 227 might be responsible for the protein thermostable properties. Additionally, arginine (R) is located at the C-terminal of cjFP510 protein. Thus, the residue is expected to be more flexible which can lead to fluctuation. Moreover, amino acid aspartate at position 185 also shows prominent peaks for both temperatures, as it might also contribute to the unique ability of temperature adaptation for model protein from *C. japonica*. It is known that both arginine and aspartate are charged amino acids that may form salt bridges. These interactions might be crucial for the stabilization of the protein 3D structure, especially for a thermostable protein. Besides that, most thermophilic proteins have an extensive network of salt bridges on their surface that promotes thermostability thus preventing protein denaturation at elevated temperature [28]. Nevertheless, the RMSF in the psychrophilic and mesophilic environments for cjFP510 and 2c9i are almost the same, although they have a different optimum temperature. Furthermore, these RMSD and RMSF plots could be improved by increasing 10 ns up to at least 30 ns.
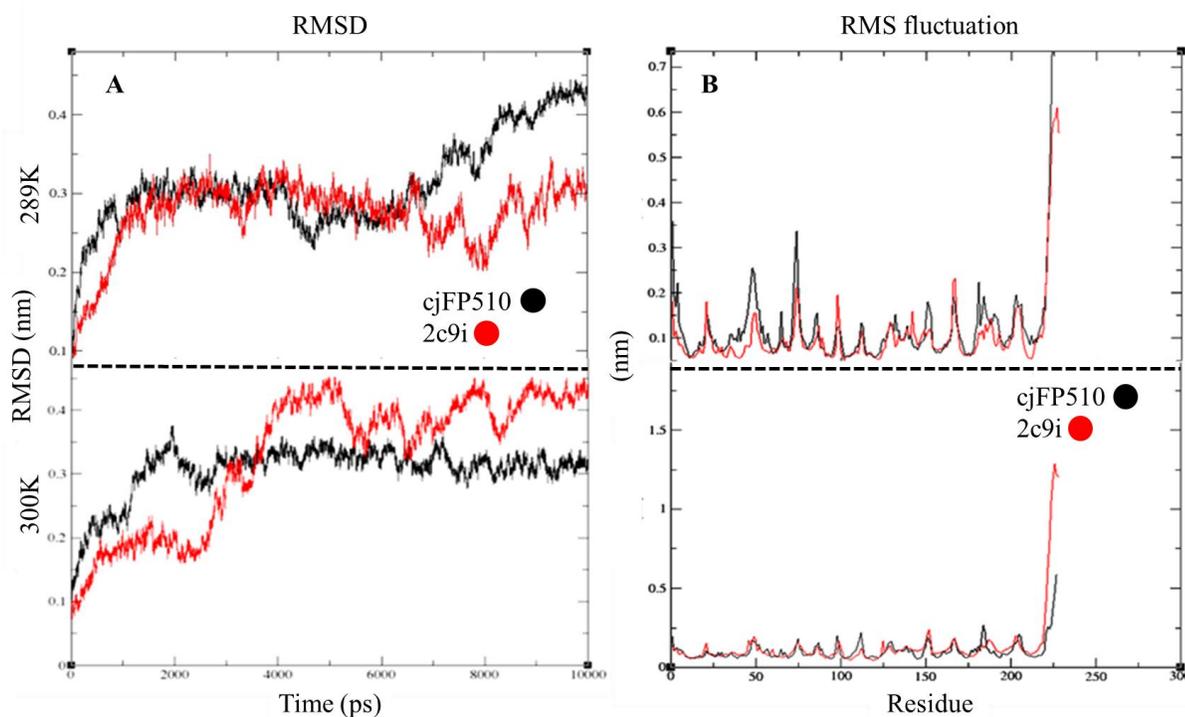

Figure 5 Dynamic changes of (A) RMSD value for cjFP510 (black) and 2c9i (red) at 289K (top) and 300K (bottom) and (B) RMSF value for cjFP510 and 2c9i at 289K (top) and 300K (bottom) during MD simulation. The RMSD plot is presented from average replicates values
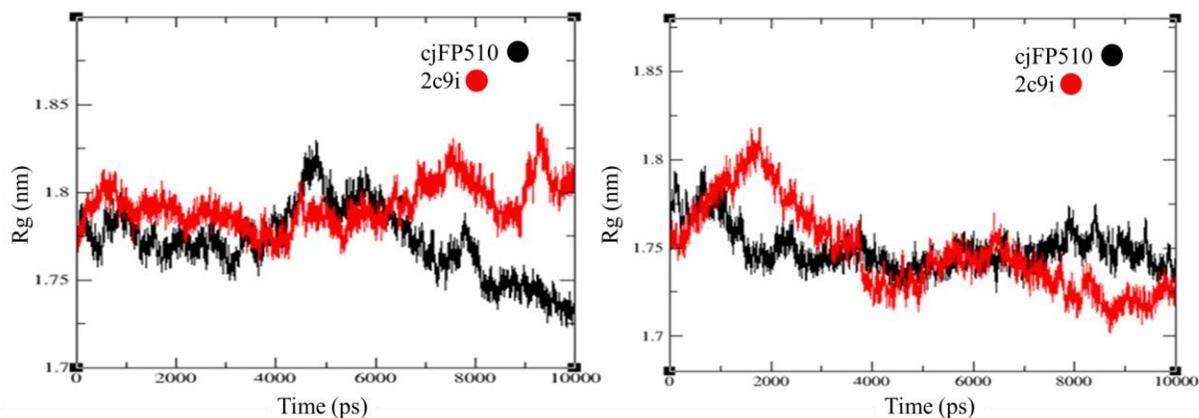
Figure 6 Comparison of the radius of gyration between model cjFP510 (black) and template 2c9i (red). The radius of gyration at 289K (left), and radius of gyration at 300K (right) for both model and template proteins. The plot is presented from average replicates values

Based on Figure 6, the Rg value of cjfp510 is less than 1.75 nm at both temperatures after the 10 ns simulation time. This scenario implies that the cjfp510 can keep its compactness at both simulated temperatures throughout the simulation. At 289K, fluctuation of the gyration radius was portrayed by the model around 5000 ps. Compared to the template in which the gyration radius increases towards the end (above 1.8 nm), the model can keep its compactness lower than 1.75 nm in the end. At 300K, the gyration radius for the model is maintained less than 1.75 nm until the end but the template showed a noticeable fluctuation at approximately 2000 ps before achieving a compact structure as the model.

These fluctuations suggested that the protein required time to adapt to various temperature changes by significantly altering its folding states. This data is supported by a fact which mentioned that a more compact protein globule possessed a slower folding rate [29]. The radius of gyration analysis leads to a conclusion that the lower the gyration radius, the protein tends to be more compact [30]. It was further discussed that the compact packing of residues contributed to the stability and folding rates. It is also deduced that cjfp510 is more stable than 2c9i based on the low Rg value at the temperature of 289K. Additionally, the trajectory of the MD simulation can be increased to at least 30 ns to obtain more reliable outputs. Hence, cold-adaptive cjFP510 from *C. japonica* can be considered as a flexible protein at warmer temperature.

4.0 CONCLUSION

The structural prediction and analysis of the 3D model of fluorescent protein isolated from *C. japonica*, cjFP510 exposed several interesting and novel characteristics of this temperature-adapted protein. The interesting features of cjFP510 compared to its template, 2c9i, have provided an insight on how the increased thermal simulation affect the model protein. The extra a-helices and amino acid residues have proven to be the important factors in order to maintain the protein thermostability. These significant findings of the novel fluorescent protein, *C. japonica* predictive structure and anlysis will create a benchmark that capable to assist future attempts in the rational *ab initio* design of the protein.

References

[1]   Tsutsui, K., Hatada, Y., Tsuruwaka, Y. 2014. A New Species of Sea Anemone (Anthozoa: Actiniaria) from the Sea of Japan: *Cribrinopsis japonica* sp. nov. *Plankt Benthos Res.* 9: 197-202. https://doi.org/10.3800/pbr.9.197.

[2]   Tsutsui, K., Shimada, E., Tsuruwaka, Y. 2015. Deep-sea Anemone (Cnidaria: Actiniaria) Exhibits a Positive Behavioural Response to Blue Light. *Mar Biol Res.* 11: 998-1003. https://doi.org/10.1080/17451000.2015.1047382.

[3]   Snapp, E. 2005. Design and Use of Fluorescent Fusion Proteins in Cell Biology. *Curr Protoc Cell Biol.* 27: 21.4.1-21.4.13. https://doi.org/10.1002/0471143030.cb2104s27.

[4]   Aguilera, R. J., Montoya, J., Primm, T. P., Varela-Ramirez, A. 2006. Green Fluorescent Protein as a Biosensor for Toxic Compounds. In: Geddes CD, Lakowicz JR (ed). Reviews in Fluorescence 2006. Springer US, Boston, MA. 463-476. https://doi.org/10.1007/0-387-33016-X_21.

[5]   De Roo, J. J. D., Vloemans, S. A., Vrolijk, H., De Haas, E. F. E., Staal, F. J. T. 2019. Development of an *In Vivo* Model to Study Clonal Lineage Relationships in Hematopoietic Cells using Brainbow2.1/Confetti Mice. *Futur Sc. OA.* 5. https://doi.org/10.2144/fsoa-2019-0083.

[6]   Luo, J., Shen, P., Chen, J. 2019. A Modular Toolset of phiC31-based Fluorescent Protein Tagging Vectors for Drosophila. *Fly.* 13: 29-41. https://doi.org/10.1080/19336934.2019.1595999.

[7] Oshiro, H., Kiyuna, T., Tome, Y., Miyake, K., Kawaguchi, K., Higuchi, T., Miyake, M., Zhang, Z., Razmjooei, S., Barangi, M., Wangsิricharoen, S., Nelson, S. D., Li, Y., Bouvet, M., Singh, S. R., Kanaya, F., Hoffman, R. M. 2019. Detection of Metastasis in a Patient-Derived Orthotopic Xenograft (PDOX) Model of Undifferentiated Pleomorphic Sarcoma with Red Fluorescent Protein. *Anticancer Res*. 39: 81-85. https://doi.org/10.21873/anticanres.13082.

[8] Bihon Asfaw, A., Assefa, A. 2019. Animal Transgenesis Technology: A Review. *Cogent Food Agric*. 5: 1-9. https://doi.org/10.1080/23311932.2019.1686802.

[9] Gassman, N. R., Holton, N. W. 2019. Targets for Repair: Detecting and Quantifying DNA Damage with Fluorescence-based Methodologies. *Curr Opin Biotechnol*. 55: 30-35. https://doi.org/10.1016/j.copbio.2018.08.001.

[10] Tsutsui, K., Shimada, E., Ogawa, T., Tsuruwaka, Y. 2016. A Novel Fluorescent Protein from the Deep-sea Anemone *Cribrinopsis japonica* (Anthozoa: Actiniaria). *Sci Rep*. 6: 1-9. https://doi.org/10.1038/srep23493.

[11] Craggs, T. D. 2009. Green Fluorescent Protein: Structure, Folding and Chromophore Maturation. *Chem Soc Rev*. 38: 2865-2875. https://doi.org/10.1039/B903641P.

[12] Labas, Y. A., Gurskaya, N. G., Yanushevich, Y. G., Fradkov, A. F., Lukyanov, K. A., Lukyanov, S. A., Matz, M. V. 2002. Diversity and Evolution of the Green Fluorescent Protein Family. *Proc Natl Acad Sci*. 99: 4256-4261. https://doi.org/10.1073/pnas.062552299.

[13] Shen, W., Le, S., Li, Y., Hu, F. 2016. SeqKit: A Caross-platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One*. 11: 1-10. https://doi.org/10.1371/journal.pone.0163962.

[14] Eric, S. D., Nicholas, T. K. D. D., Theophilus, K. A. 2014. Bioinformatics with Basic Local Alignment Search Tool (BLAST) and Fast Alignment (FASTA). *J Bioinforma Seq Anal*. 6: 1-6. https://doi.org/10.5897/ijbc2013.0086.

[15] Jones, P., Binns, D., Chang, H-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S-Y., Lopez, R., Hunter, S. 2014. InterProScan 5: Genome-scale Protein Function Classification. *Bioinformatics*. 30: 1236-1240. https://doi.org/10.1093/bioinformatics/btu031.

[16] Minguez, P., Letunic, I., Parca, L., Bork, P. 2013. PTMcode: A Database of Known and Predicted Functional Associations between Post-translational Modifications in Proteins. *Nucleic Acids Res*. 41: 306-311. https://doi.org/10.1093/nar/gks1230.

[17] Rodriguez, R., Chinea, G., Lopez, N., Pons, T., Vriend, G. 1998. Homology Modeling, Model and Software Evaluation: Three Related Resources. *Bioinformatics*. 14: 523-528. https://doi.org/10.1093/bioinformatics/14.6.523.

[18] Roy, A., Kucukural, A., Zhang, Y. 2010. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat Protoc*. 5: 725-738. https://doi.org/10.1038/nprot.2010.5.

[19] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., Ferrin, T. E. 2004. UCSF Chimera-A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem*. 25(2004): 1605-1612. https://doi.org/10.1002/jcc.20084.

[20] Nienhaus, K., Renzi, F., Vallone, B., Wiedenmann, J., Nienhaus, G. U. 2006. Chromophore-protein Interactions in the Anthozoan Green Fluorescent Protein asFP499. *Biophysical J*. 91(11): 4210-4220. https://doi.org/10.1529/biophysj.106.087411.

[21] Darden, T., Perera, L., Li, L., Lee, P. 1999. New Tricks for Modelers from the Crystallography Toolkit: The Particle Mesh Ewald Algorithm and Its Use in Nucleic Acid Simulations. *Structure*. 7: 55-60. https://doi.org/10.1016/S0969-2126(99)80033-1.

[22] Humphrey, W., Dalke, A., Schulten, K. 1996. Visual Molecular Dynamics. *J Mol Graph*. 14: 33-38.

[23] Mavromatis, K., Tsigos, I., Tzanodaskalaki, M., Kokkinidis, M., Bouriotis, V. 2002. Exploring the Role of a Glycine Cluster in Cold Adaptation of an Alkaline Phosphatase. *Eur J Biochem*. 269: 2330-2335. https://doi.org/10.1046/j.1432-1033.2002.02895.x.

[24] Hayes, M. A., Shor, A. C., Jesse, A., Miller, C., Kennedy, J. P., Feller, I. 2020. The Role of Glycine Betaine in Range Expansions: Protecting Mangroves against Extreme Freeze Events. *J Ecol*. 108: 61-69. https://doi.org/10.1111/1365-2745.13243.

[25] Siddiqui, K. S., Poljak, A., Guilhaus, M., De Francisci, D., Curmi, P. M. G., Feller, G., D'Amico, S., Gerday, C., Uversky, V. N., Cavicchioli, R. 2006. Role of Lysine Versus Arginine in Enzyme Cold-adaptation: Modifying Lysine to Homo-arginine Stabilizes the Cold-adapted Alpha-amylase from *Pseudoalteramonas haloplanktis*. *Proteins*. 64: 486-501. https://doi.org/10.1002/prot.20989.

[26] Lüthy, R., Bowie, J. U., Eisenberg, D. 1992. Assessment of Protein Models with Three-dimensional Profiles. *Nature*. 356: 83-85. https://doi.org/10.1038/356083a0.

[27] Chaitanya, M., Babajan, B., Anuradha, C. M., Naveen, M., Rajasekhar, C., Madhusudana, P., Kumar, C. S. 2010. Exploring the Molecular Basis for Selective Binding of *Mycobacterium tuberculosis* Asp kinase Toward Its Natural Substrates and Feedback Inhibitors: A Docking and Molecular Dynamics Study. *J Mol Model*. 16: 1357-1367. https://doi.org/10.1007/s00894-010-0653-4.

[28] Ban, X., Wu, J., Kaustubh, B., Lahiri, P., Dhoble, A. S., Gu, Z., Li, C., Cheng, L., Hong, Y., Tong, Y., Li, Z. 2020. Additional Salt Bridges Improve the Thermostability of 1,4-a-glucan Branching Enzyme. *Food Chem*. 316: 126348. https://doi.org/10.1016/j.foodchem.2020.126348.

[29] Gao, J., Zhang, T., Zhang, H., Shen, S., Ruan, J., Kurgan, L. 2010. Accurate Prediction of Protein Folding Rates from Sequence and Sequence-derived Residue Flexibility and Solvent Accessibility. *Proteins Struct Funct Bioinforma*. 78: 2114-2130. https://doi.org/10.1002/prot.22727.

[30] Lobanov, M. Y., Bogatyreva, N. S., Galzitskaya, O. V. 2008. Radius of Gyration as an Indicator of Protein Structure Compactness. *Mol Biol*. 42: 623-628. https://doi.org/10.1134/S0026893308040195.

.