BIODIVERSITAS
Volume 25, Number 4, April 2024
Pages: 1605-1614

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d250429

Cloning and expression of Rubisco Like Protein (RLP) from halophilic bacterium *Chromohalobacter salexigens* BKL 5

IRWAN SUDARMANTO[1,2,♥], LINCE SITOHANG[2], SAIFUR ROHMAN[2,3,♥♥], WAYAN TUNAS ARTAMA[2,4]

[1]Department of Chemical, Faculty of Science, Institut Teknologi Sumatera. Sabah Balau, South Lampung 35365, Lampung, Indonesia. Tel.: +62-721-8030188, ♥email: irwan.soedarmanto@fa.itera.ac.id

[2]Graduate Program of Biotechnology, Universitas Gadjah Mada. Jl. Teknika, Sleman 55281, Yogyakarta, Indonesia

[3]Department of Agriculture Microbiology, Faculty of Agriculture, Universitas Gadjah Mada. Jl. Flora, Sleman 55281, Yogyakarta, Indonesia. Tel.: +62-274-523065, Fax.: +62-274-563062, ♥email: saifur@ugm.ac.id

[4]Faculty of Veterinary Medicine, Universitas Gadjah Mada. Jl. Fauna 2, Sleman 55281, Yogyakarta, Indonesia

Manuscript received: 31 October 2023. Revision accepted: 17 April 2024.

Abstract. *Sudarmanto I, Sitohang L, Rohman S, Artama WT. 2024. Cloning and expression of Rubisco Like Protein (RLP) from halophilic bacterium* Chromohalobacter salexigens *BKL 5. Biodiversitas 25: 1605-1614.* This research seeks to identify and determine the Rubisco-like protein characteristics from *Chromohalobacter salexigens* BKL 5 tested using a recombinant method approach. *C. salexigens* BKL 5 is a halophilic bacterium found in Bledug Kuwu Mud, Central Java. The Open Reading Frame (ORF) gene encoding Rubisco-like protein from that bacterium was successfully amplified using the polymerase chain reaction method using the primers we designed. Cloning of that gene on the pCold plasmid was carried out using the double digest method with the restriction enzymes NdeI and EcoRI. The results of plasmid transformation in *Escherichia coli* BL 21 were successful, indicating that the bacterial culture was grown in Luria Bertoni agar media containing the antibiotic ampicillin. Overexpression protein of the transformed *Escherichia coli* BL 21 shows that the effective temperature that produces maximum results is 16°C. The results of sequencing and phylogenetic analysis showed that the protein belongs to the Rubisco family, namely Rubisco-like protein with a residue of 429 amino acids and a weight of 46 kDa. Prediction of 3D structures using the Alphafold tool gives an accuracy above 90%. Observation of the 3D structure shows that glutamic acid is more distributed on the surface of the protein which is important for protein solubility. Tests with SOPMA showed that this protein was dominated by random coil structure of 41.96% which was identical to that found in bacterial species in extreme habitats. The dynamic molecular analysis using Yasara at 10 ns showed that the protein had a tendency to undergo considerable deformation but remained stable.

Keywords: *Chromohalobacter salexigens* BKL 5, cloning, halophilic, molecular dynamics, Rubisco Like Protein (RLP)

Abbreviations: AT: *Arabidopsis thaliana*, BS: *Bacillus subtilis*, CS: *Chromohalobacter salexigens* BKL 5, CT: *Chlorobium Tepidum*, GK: *Geobacillus kaustophilus*, ORF: Open Reading Frame, RLP: Rubisco Like Protein, RR: *Rhodospirillium rubrum*

INTRODUCTION

In previous study, we have isolated bacteria *Chromohalobacter salexigens* BKL 5 which grows in the Bleduk Kuwu Mud Volcano, Central Java (Rohman et al. 2012). Based on its habitat, this type of bacteria is classified as an extremophile's species, namely organisms that tolerate and thrive in the most extreme and challenging conditions of life. As a result of these extreme environmental, this species has developed a number of interesting adaptations to cellular membranes, proteins and extracellular metabolites. Optimal temperature and pH for growth of *C. salexigens* were 37°C and 7.5 (Arahal et al. 2001). *Chromohalobacter salexigens* species is a moderately halophilic gram-negative bacterium, in the form of single or paired cells. Bacterial colonies were creamy, round and 2 mm in diameter. This bacterium has previously been tested to have the ability to grow in an environment with a salt concentration of 15% with a specific growth rate of 0.1 per hour. Analysis of the protease produced by *C. salexigens* BKL5, namely DegP, showed that the protein had the characteristics of a protein derived from a group of halophilic bacteria (Fitriani et al. 2017).

One of the characteristics of halophilic proteins is the high content of negatively charged amino acid residue. The presence of negatively charged amino acids on the surface of proteins produced by halophilic bacteria can increase protein stability in environments with high salt stress, negatively charged amino acids will interact with hydrated salt ions and form salt bridges (Gunde-Cimerman et al. 2018; Sudarmanto et al. 2024). These negatively charged residues contribute to maintaining the stability of the protein structure through the interaction of intramolecular charges and the neutralization of oppositely charged ions in the medium. Another characteristic of halophilic proteins is the low hydrophobic residue. This is an adaptation of proteins to reduce hydrophobic interactions so that the stability of halophilic proteins can be maintained even though the salt concentration in the surrounding environment is high (Gunde-Cimerman et al. 2018).

Rubisco is a type of enzyme (protein) that is slow to evolve (Bouvier et al. 2021). The initial hypothesis said that the Rubisco family originated from group III Rubisco which developed into several classes through the Horizontal Gene Transfer (HGT) process. Another hypothesis says that the ancestor of the Rubisco family is group IV (RLP) which is based on the fact that all the Rubisco family have activity as enolase catalysts and RLP comes from an ancient era that was in an anoxic state where methane-producing archaea began to synthesize methane so that the atmosphere was almost saturated with CO_2 and almost free of oxygen, in such conditions methanogenic archaea synthesize N-carboxymethanofuran (carbamate) through chemical reactions rather than enzymatic reactions (Yokota et al. 2017). Based on their constituent units, during the evolution of Rubisco they were grouped into three different groups, namely groups I, II, and III. In its development, proteins that are structurally similar to Rubisco have been found and are grouped as Rubisco IV. Groups I, II and III function to catalyze the carboxylation or oxygenation of ribulose 1,5-bisphosphate (RuBP). Group IV or what is known as Rubisco Like Protein (RLP) does not have the ability to fix CO_2 because it undergoes substitution at the amino acid residue on the active site (Hanson et al. 2001). Based on BLAST analysis showed that the genome of *C. salexigens* strain DSM 3043 (1H11) had a large subunit Rubisco encoding gene. These data were used as the basis for isolating the ORF of Rubisco from *C. salexigens* BKL 5.

Enolase is a rate limiting step in the CO_2 fixation process into phosphoglyceric acid (a glucose precursor compound) (Erb et al. 2018). A good understanding of the enolation process by RLP *C. salexigens* BKL 5 from extremophiles habitat is expected to be used as a basis for improving fixation process by the Rubisco family (Flamholz et al. 2019; Carter et al. 2018). This research seeks to identify and determine the RLP characteristics of the *C. salexigens* BKL 5 tested using a recombinant method approach by carrying out the cloning process and expressing the RLP gene in the *Escherichia coli* system.

MATERIALS AND METHODS

Materials

Protein structures obtained from Research Collaboratory for Structural Bioinformatics (RCSB; https://www.rcsb.org), namely *Chlorobium Tepidum* (CT) (1YKW, RLP), *Rhodospirillium rubrum* (RR) (5RUB, RLP), *Arabidopsis thaliana* (AT) (5IUO, Rubisco), *Bacillus subtilis* (BS)(2ZVI, RLP) and *Geobacillus kaustophilus* (GK)(2OEM, RLP) which were downloaded in Fasta and PDB format.

Bacteria colony of *C. salexigens* BKL 5 (Rohman et al. 2012), The Qiagen Miniprep Kit (Qiagen, Hilden, Germany), ampisilin 100 µg/mL, primary base which were designed by our team and produced by 1st Base, Singapore through Genetica Science Co., Indonesia (forward (5'-ATATACATATGGAGTTATCCCGGCATG-3') and reverse (5'-ATATAGAATTCTCATCCTGAAACGGGCTGTG-3'), plasmid pCold IV (Takara, Japan), *Presto™ Mini Plasmid kit* (Geneaid), *FastDigest NdeI* dan *EcoRI* (Thermo Scientific), *Quick-stick ligase* (Bioline), *GenepHlow™ Gel/PCR kit* (Geneaid), KOD FX Neo *kit* (Toyobo), *Presto™ Mini Plasmid kit* (Geneaid), gel agarose 0.8% (Calbiochem Omnipur), buffer TAE, *ethidium bromide*, DNA *loading dye* (Geneaid), DNA *ladder* 1 kb (Novagen), SDS sample buffer, gel SDS-PAGE (12% resolving gel dan 6% stacking gel), 1x SDS running buffer, *Coomassie Brilliant Blue*, destaining solution.

Methods

Cloning and expression

The genomic DNA was extracted from 2 mL of overnight culture of *C. salexigens* BKL 5 (Rohman et al. 2012) by centrifugation at 16000 rpm for 5 minutes using The Qiagen Miniprep Kit (Qiagen, Hilden, Germany). The gen target (ORF RLP) of the bacteria in extracted DNA was amplified using two primers. The PCR reaction was prepared with a total volume of 50 µL which consist of 11 µL of *Nuclease Free Water* (NFW), 25 µL of 2X PCR Buffer, 10 µL of dNTPs, 1 µL of *Taq Polymerase* KOD Neo, 1 µL of *Template* DNA and 1 µL of each primer. It was then placed in a Applied Biosystems™ 2720 Thermal Cycler (Thermo Fisher Scientific, UK) for 35 cycles amplification at 94°C for 2 min (pre-denaturation), 98°C for 10 sec (denaturation), 52.3°C for 30 sec (annealing), and 68°C for 30 sec (extension), followed by 68°C for 10 min as final cycle.

Visualization of PCR results by electrophoresis on 0.8% agarose gel in 1X TAE buffer. The resulting amplification was cloned into the pCold plasmid IV. The pCold IV plasmid was first prepared as a vector, digested using restriction enzymes NdeI and EcoRI and placed in a water bath at 37°C for 15 minutes. ORF RLP and pCold vectors are combined using a ligase enzyme (T4 DNA ligase (Toyobo)) to connect two DNAs by creating a bond between the two ends of the DNA. Ligation was performed at 16°C overnight. Once the ligase is produced, the pCold IV vector carrying the ORF RLP enters the bacterial cell *Escherichia coli* BL21(DE3) by the transformation process.

Escherichia coli BL21(DE3) carrying the expression vector pCold IV and ORF RLP, was incubated on liquid LB medium containing 100 µg/mL ampicillin at 37°C until the culture OD_{600} reached 0.4-0.8, then the culture was incubated at 16°C for 30 minutes without shaking then IPTG 1 mM was added and incubation was continued overnight at 37°C. After induction, cells were harvested by centrifugation at 6000 rpm for 5 minutes. The supernatant was discarded and the cell pellet was resuspended with 2X SDS sample buffer and boiled for 5 minutes. Analysis of recombinant protein expression using 12% SDS-PAGE. Sequencing was carried out to determine the nucleotide sequences and estimate the amino acid sequences for further analysis. The 1.5 mL microtubes were labeled according to the sample code, the total volume of plasmid isolation was 30 µL each and the primary volume of primer forward base and primer reverse base was 20 µL each, then packed in a sputum column and sent to being sequenced. The amplicons and the primer pairs were sent to 1St Base

Co., Singapore through Genetica Sciences Co., Indonesia. Sequencing procedures were conducted bi-directionally using a BigDye® Terminator v.3.1 Cycles Sequencing Kit (Applied Biosystems, USA) and further read with an ABI PRISM® 377 automatic DNA sequences. Its quality was accessed using the Sequence Scanner version 2.0 Software (Applied Biosystem). Subsequently, sequences were trimmed, assembled, and manually edited using Geneious Prime version 2020 (www.geneious.com) prior to the application of BLAST (Basic Local Alignment Tools) analysis at NCBI (www.ncbi.nlm.nih.gov/).

Structure prediction and validation

Amino acid sequences were predicted in 3D structure using the AlphaFold, a protein homology method that can provide high accuracy (Jumper et al. 2021). The structure obtained then validated using PROCHECK, Verify 3D (Kaur et al. 2023; De oliviera et al. 2019) and ProSA (Shill et al. 2023).

Analysis using SOPMA tool (www.npsa-prabi.ibcp.fre) was carried out to determine the composition of the secondary structure of amino acids (Ling et al. 2022). Measurement of the physicochemical properties of proteins using Protparam (www.expasy.org) (Kaur et al. 2020).

Visualization of the homologated structure using Chimera (www.cgl.ucsf.edu) (Pettersen et al. 2021). Calculation of PCA (Principal Component Analysis) to determine the similarity of the tested proteins based on chemical-physical parameters using the R program (Hasan et al. 2021).

Amino acid alignment and phylogenetic tree

Alignment of amino acids used to determine the conserve region of *C. salexigens* BKL 5 against Rubisco family (type I, II and III) and RLP (type IV) from several other species. Alignment is also intended to determine the level of identity/similarity. The method used is a web server-based method PROMALS 3D (Arunima et al. 2019). Prediction of pocket site that catalyze enolation process using Chimera tool. Comparison of the RLP and Rubisco family structures was carried out by superimposing using Biovia Studio Discovery on the 3D structures of CT (1YKW), RR (5RUB), AT (5IUO), BS (2ZVI) and GK (2OEM) in PDB format (Haque et al. 2022). Phylogenetic analysis with MEGA 11 software were conducted to obtain an overview of the relationship with other species that express RLP and Rubisco (Tamura et al. 2021). The data used are protein sequences from various Rubisco families (Type I, II and III) and RLP (Rubisco type IV) from the NCBI portal (www.ncbi.nlm.nih.gov).

Dynamic molecular simulation

Molecular dynamic determination using the YASARA at 10 ns. The structural topology was created using AMBER Force Field 14 (González-Paz et al. 2023). SPC was chosen as a model solvent (a triclinic water box with a size of 50 × 75 × 70 Å) for protein complexes. This system is neutralized by adding sodium or chlorine ions based on the total charge. To minimize the system before MD, the steepest descent (5000 steps) algorithm was applied. MD simulations were carried out using constant temperature (300 K) and pressure (1.0 bar). The approximate number of frames per simulation is 1000. Parameter tested include Rg, RMSF and RMSD (Kamenik et al. 2020; Yamamoto et al. 2021).

RESULTS AND DISCUSSION

Cloning and expression

The ORF encoding RLP of *C. salexigens* BKL 5 was successfully amplified from DNA extracts obtained using primers we previously designed (Figure 1.A). Insertion of ORF RLP was carried out in the pCold IV plasmid (±4359 bp) which had previously been cut at 2 sides of the MCS (Multiple Cloning Site) with the restriction enzymes NdeI and EcoRI (Figure 1.B). This method is called double digest. The advantages of using multiple restriction enzymes include increased control and reduced chances of errors, and they can be used in combination with each other for more complex cuts (Wang et al. 2017). The selection of the restriction enzymes NdeI and EcoRI was based on analysis and simulations that had previously been carried out where the two enzymes did not cut ORF RLP fragment. This is a consideration for using the pCold IV plasmid which provides a cloning site suitable for NdeI and EcoRI, pColdIV plasmid also able to produce protein products with high purity in *E. coli* (Qing et al. 2004).

The cloning results were confirmed by cutting them using the restriction enzymes NdeI and EcoRI. Figure 1.C shows that cloning has been successful, indicated by the presence of 2 different types of bands on the agarose plate, indicating that ORF RLP has been inserted right into the cloning site of pCold IV plasmid.

The success of cloning was strengthened by the cultivation of transformed *E. coli* BL 21 which grew on media containing the antibiotic ampicillin 100 μg/mL (Figure 2.B). As is known, pCold IV plasmid has a gene that is resistant to ampicillin.

To clarify whether the plasmid that has been inserted with the target gene can carry out its function properly, an overexpression approach is needed which is characterized by the production of high copies of a target protein (Joshi et al. 2022). This method is carried out by inducing the transformed *E. coli* BL 21 culture using an IPTG inductor. The results of overexpression on SDS PAGE 12% showed that RLP protein was maximally expressed and measured at a weight of around 46 kDa with a total of 429 amino acids residues (Figure 2.A). This value is equivalent to the base pair composition of the ORF RLP sequencing results obtained from PCR amplification.

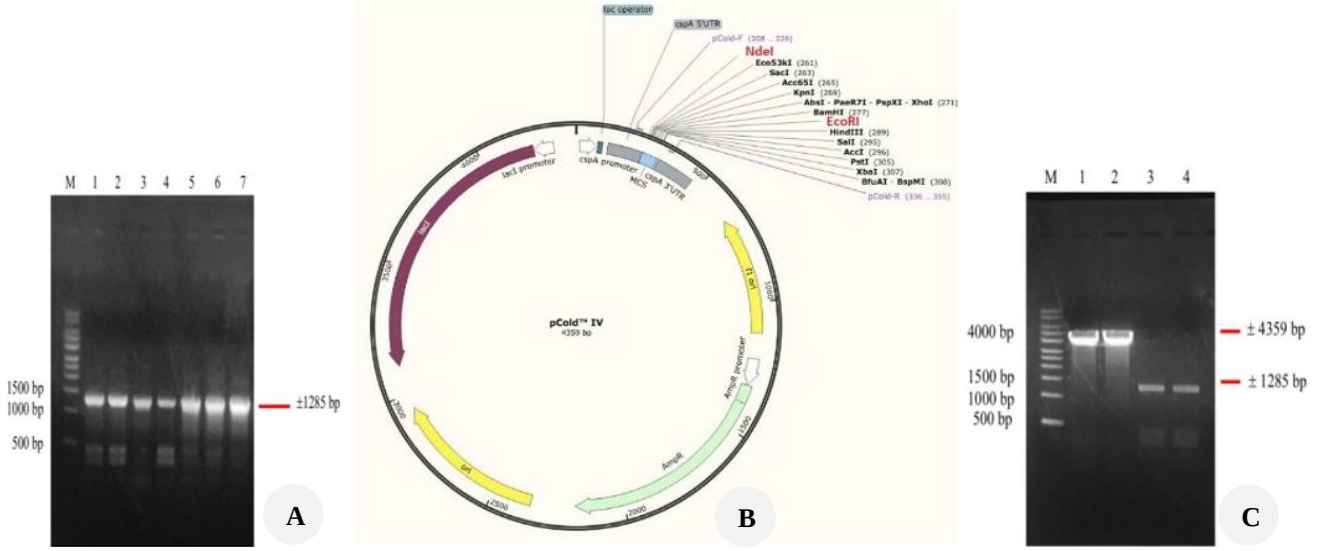

Figure 1. PCR visualization of ORF RLP. Lane M: DNA ladder (Novagen; 1 Kb), Lanes 1–7: ORF RLP (A). Target of cloning of ORF RLP on plasmid pCold IV (Takara; ±4359bp) (red letter) (B) Cutting with the restriction enzymes NdeI and EcoRI on the results of cloning ORF RLP on pCold IV plasmid. Lane M: DNA ladder (Novagen; 1 Kb), Lanes 1 and 2: pCold IV, Lanes 3 and 4: ORF RLP (C)

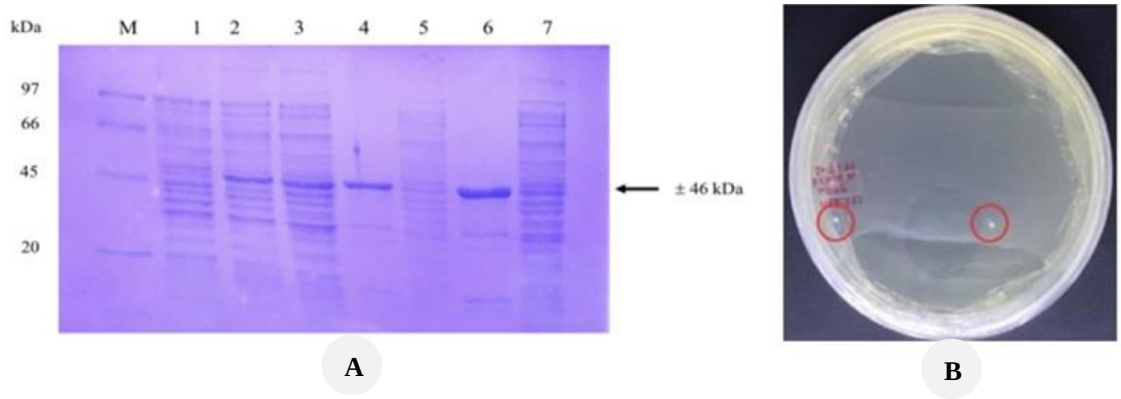

Figure 2. Expression of RLP in protein fractionation of *E. coli* BL 21 (DE3) in SDS PAGE 12%. M is a marker (A) Results of transformation of pCold IV containing ORF RLP in *E. coli* BL 21 (DE3) (Luria Bertoni agar medium consist of 100 µg/mL ampicillin). The red circle indicates a colony that has successfully grown (B)

In this study, we did not carry out a purification process. The protein sequencing process is carried out by translating the sequencing results obtained from 1st BASE which have previously been edited (Geneious Prime version 2020; www.geneious.com) using the BLASTx application provided by NCBI (www.ncbi.nlm.nih.gov/).

In BLASTx, a nucleotide sequence is used as a query, which is first translated in all six reading frames, and then each of the translated amino acid sequences is compared to the protein sequences in protein databases. The programs implement variations of the BLAST algorithm, which is a heuristic method for rapidly finding local alignments with scores sufficiently high to be statistically significant. BLAST is a commonly-used software package for comparing a query sequence to a database of known sequences (Boratyn et al. 2012)

The sequencing results have been registered in the NCBI database portal with access number MW373099.1 and are displayed as follows.

MELSRHEGQLPRITARYRIETPCDPERAAEAMAGEQ
SSGTFLKLAGETEALRERHAARVESVTCLEVGTTPSL
PGALPGETYTSADVVLSWPLENIGTSLPNLLATVFGN
LTELRELSGIRLTDLVMPACFIEALPGPQFSIEGTRRL
CGTEQGPLVGTIIKPSVGLTPEETATLVKQLIEGGIDFI
KDDELIADPPYSPLSERVPAIMQVINEHAQRTGRKPM
YAFNITGDTDEMRHRHDLVRDHGGTCVMASVNWIG
LGALAALRRHSQLAIHGHRNGWGVFYRHPQLGISYR
AYQALLRLAGTDHLHVNGLASKFAEEDASVMDSAR
ACLTPLLEGAERDDRAMPVFSSAQTVHQACPTFQQL
QSTDLIYTCGGGIMAHPDGVAAGCRSIRAAWEAASI
GVSLEDHAKQEPDLAVALRAFHSPFSG

Structural analysis and validation

Amino acid sequences were predicted in 3D structure using the AlphaFold tool built into Chimera used for further analysis. The results of 3D structural prediction with AlphaFold are shown in Figure 3.A.

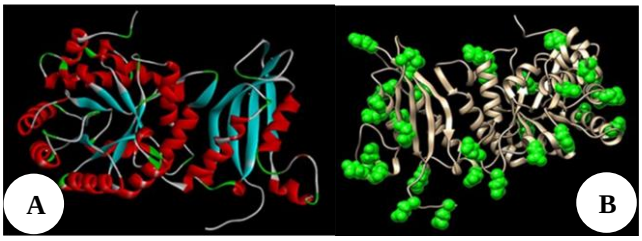

Figure 3. Structure of Rubisco Like Protein (RLP) from *C. salexigens* BKL 5 based on AlphaFold prediction (A) distribution of glutamic acid at surface of protein (green sphere) (B)

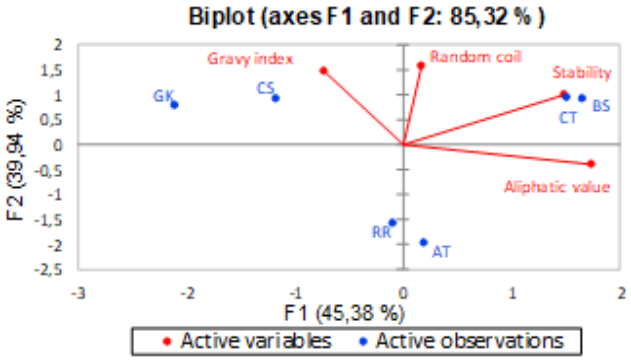

Figure 4. Principal Component Analysis (PCA) of GK, CS, CT, BS, RR and AT. The variables used are the parameters and physical chemical properties of protein

The structural validation of the homologation results gave a value of 92.54% (Verify 3D) and z- score of -6.57 (ProSA). ProSA calculates the energy needed for amino acids to fold, the more negative the value obtained, the higher the protein stability, while the z score is a value that describes the quality of the protein model being tested. The analysis of the Ramachandran plot using the PROCHECK web server shows that the structure contains amino acid residues in the allowed positions, namely the most favored region at 89.8%, the additional allowed region at 9.6% and in the generously region 0.5%. This shows that structure prediction with AlphaFold gives excellent results and can further be used as a reference in testing and analysis related to the RLP structure of *C. salexigens* BKL 5. Observation of the 3D structure of the halophilic structure from *C. salexigens* BKL 5 shows that glutamic acid is more distributed on the surface of the protein (Figure 3.B) which confirms that this amino acid will interact with water which is limited in halophilic environment so that protein solubility will be maintained (Nomoto et al. 2021). Halophilic protein surfaces have a greater negatively charged (acidic) amino acid composition than their mesophillic homologs. This allows the protein to be more competitive in interacting with positive ions which are abundant in the halophilic environment. The amino acids in question are aspartic acid and or glutamic acid. The results of the analysis show that halophilic RLP CS tends to have more glutamic acid than aspartate, this is presumably because glutamic acid has a larger size so it tends to be more exposed and accessible to interact with solvents. Glutamic acid, like aspartic acid, has a high ability to bind water and is very important for the process of protein solubilization.

Calculations with SOPMA software resulted in the following secondary structure compositions, namely random coil 41.96, alpha helix 39.63%, strand 14.69% and turn 3.73%. The high composition of random coils is identical to proteins in extreme environments, one of which is halophilic and is common in early-life organisms. The existence of this secondary structure is required for protein stabilization besides it is important for flexibility that supports protein dynamic activity (Richard 2019; Shekhar et al. 2022).

Statistical analysis using the PCA (Principal Component Analysis) method showed that proteins grown in extreme habitats had distinctive chemical-physical characteristics as well as in RLP *C. salexigens* BKL 5 (CS) (in a halophilic environment; Rohman et al. 2012) which showed close with *Geobacillus kaustophilus* (GK) (thermophilic environment; Imker et al. 2007) (Figure 4). The approach with PCA also provides the fact that Rubisco type I represented by *Arabidopsis thaliana* (AT) is statistically very different from RLP (Rubisco type IV), meaning that in order for a protein to act as Rubisco, apart from having identical binding sites, it must also have similar chemical-physical characteristics.

Amino acid alignment and phylogenetic tree

The goal of the alignment process is to determine the conserve region of *Chromohalobacter salexigens* BKL 5 against Rubisco family (type I, II and III) and RLP (type IV) from several other species and to determine the level of identity/similarity of them. Alignment of RLP *C. salexigens* BKL 5 with another RLP and Rubisco family shows an interesting stuff where there are several conserved residues, namely Lys (K), Asp (D), Asp (D) and Glu (E) which are positioned close to each other. Meanwhile, in other positions, Histidine (H) and Lysine (K) residues also indicate the existence of a conserved region that makes up the configuration a site area in the form of a pocket is called the active site of the protein (Figure 5; yellow shaded). In Rubisco type I (proteins that catalyze carboxilase/oxygenase for photosynthesis process), those residue positions (K, D, D, E) are always in the order of 201, 202, 203 and 204 (*Arabidopsis thaliana*, 5IUO; Rubisco type I) (Erb et al. 2018), while in RLP (Rubisco type IV), the positions of the residues vary in order. The active area with those amino acid configurations is important for catalyzing the enolase process. Enolation is the initial process of the carboxylation and oxygenation of RuBP by the enzyme Rubisco (a lyase enzyme) and is the main reaction of the enolase enzyme (an isomerase enzyme) like RLP (Tommasi 2021). This process catalyzes the formation of tautomeric compounds called as enediol as intermediate species of RuBP. Enolase is a rate limiting step in the CO_2 fixation process into phospho glyceric acid (a glucose precursor compound) by Rubisco family. Rubisco are slow catalysts. The average Rubisco turnover frequency is only between 1 and 10 s^{-1} (Erb et al. 2018).

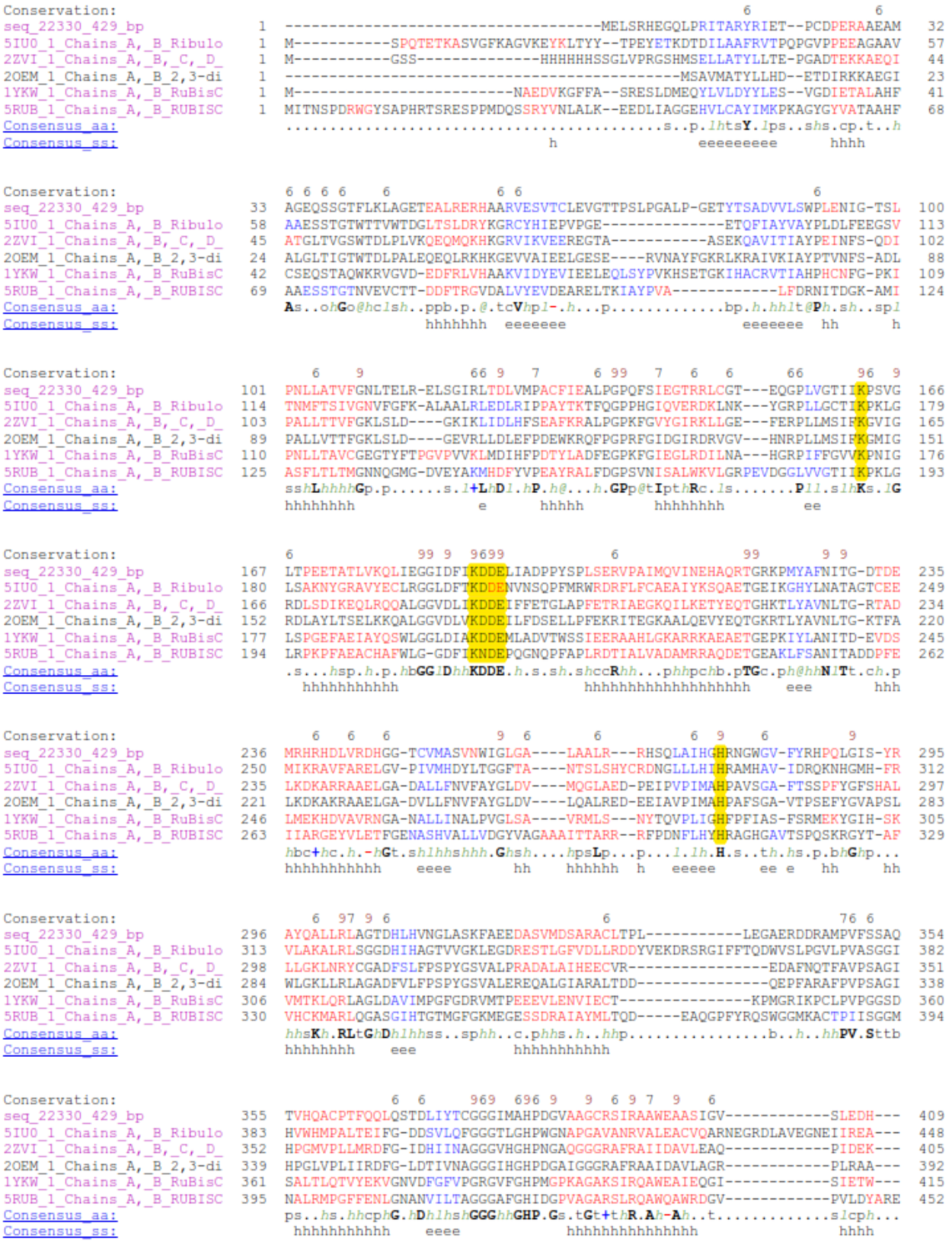

Figure 5. Alignment of GK(2OEM), CS (Seq_2230), CT(1YKW), BS(2ZVI), RR(5RUB) and AT (5IUO). The yellow shaded residue is essential for the enolation process

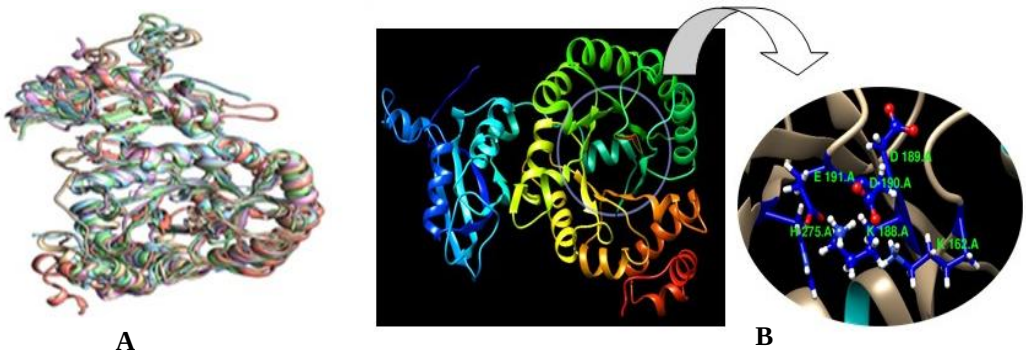

Figure 6. Comparative structure of CT (magenta), RR (red), AT (brown), BS (green) and CS (blue**)** (A) Prediction of active region of protein (B)

The prediction of the active site of the RLP CS protein which catalyzes the enolase process based on the residue arrangement resulting from the alignment is shown in Figure 6.B using the Chimera tool. This prediction is identical to the results of crystallization data of a similar RLP from *Bacillus subtilis* sp found in the RCSB database (2ZVI; www. RCSB.org) with a ligand in the form of a RuBP analog, namely DKMTP (2,3-diketo-5-methylthiopentyl 1-phosphate) (Carter et al. 2018). This strengthens the presumption that RLP CS has activity as an enolase catalyst just like other RLPs whose activities are already known.

The analysis of the similarity level of RLP based on the arrangement and configuration of amino acids using Chimera tool are *C. salexigens* BKL 5 to the structure of *Chlorobium Tepidum* (1YKW), *Rhodospirillium rubrum* (5RUB), *Arabidopsis thaliana* (5IUO*), Bacillus subtilis* (2ZVI) and *Geobacillus kaustophilus* (2OEM) were 29.1, 27.3, 28.9, 28.9 and 29.4% respectivelyz. Rubisco Like Protein (RLP) is a Rubisco family protein that has a high diversity of residues. This result analysis strengthens theory that the similarity between RLP to Rubisco and other RLP is very low (Ashida et al. 2008).

The big question from the alignment results is, even though RLP has high differences in the composition and configuration of amino acids, it has the same activity (enolase). To answer this stuff, we must remember that the activity of proteins (enzymes) solely depends on the active site (binding site) in the form of a pocket found in the protein. In this context, RLP has a low similarity around 30%, but has an identical conserved region (based on alignment results). To prove that low similarity does not affect enolase activity, we superimposed the 3D structures of several RLPs and Rubisco (type I) using BioVia Studio Discovery software. The analysis results show that although the level of similarity of the RLP arrangement and configuration to each other is low, in general the 3D structure of the RLP is similar (RMSD <2A), namely the RMSD value of CS against CT (1YKW), RR (5RUB), AT (5IUO), BS (2ZVI) and GK (2OEM) were 1.10 A; 1.23 A; 1.19A; 1.17A and 1.15A respectively (Figure 6.A). These results indicate that differences in the arrangement of amino acid residues in RLP do not affect the active side of the protein. The differences in the arrangement and configuration of amino acids in RLP are mainly an adaptation to the environment where the RLP is located (eg; *C. salexigens* BKL 5 in a halophilic environment and *Geobacillus caustophilus* in a thermophilic habitat)

Phylogenetic analysis was carried out to determine the class of RLP *C. salexigens* BKL 5 (Figure 7). The data used are protein sequences from the NCBI portal in Fasta format (www.ncbi.nlm.nih.gov). The protein used is the Rubisco family of various types. The inclusion of the Rubisco type is based on data available from NCBI. Protein sequence data in the form of amylase (CAD 20312.1) is used as a comparison for other types of proteins from the halophilic environment.

The results of phylogenetic observations show that RLP CS has great similarities with *C. salexigens* DSM which originates from sea water (Figure 7). This raises the presumption that *C. salexigens* BKL 5 which comes from fresh water has a strong relationship with *Chromohalobacter* species from sea water. Another interesting thing is that the phylogenetic tree shows that the halophilic RLP CS has similarities with other Rubisco families originating from halophilic environments (*Halomonas* species). This phenomenon shows that proteins in a halophilic environment have distinctive characteristics that are important for adapting to their environment.

Dynamic molecular simulation

Dynamic molecular simulations on proteins are used to determine the movement of each atom in a protein over time. These simulations are important for predicting the influence of the environment in the tested system on the overall protein response and conformation (Collier et al. 2020; Badar et al. 2022). Parameters measured in molecular dynamic simulations include Rg, RMSF and RMSD **(**Kandeel and Alzahrani 2020; Chen et al. 2021**).**

The simulation results show that the RMSD value and superimpose position of the RLP CS protein produces a relatively small (~2 A; Figure 8.C and 8.D) measured per unit time within 10 ns. This indicates that this protein is relatively stable in environmental conditions with high salt levels. RMSD is a parameter commonly used to determine the stability of the "protein" of a system. RMSD can describe changes in protein shape and geometry. The higher the RMSD value, the more unstable the protein. The steady state has an RMSD value of ~2 A, with ~10 distinct regions. The total energy in these regions is lower than in region (Libera et al. 2020).

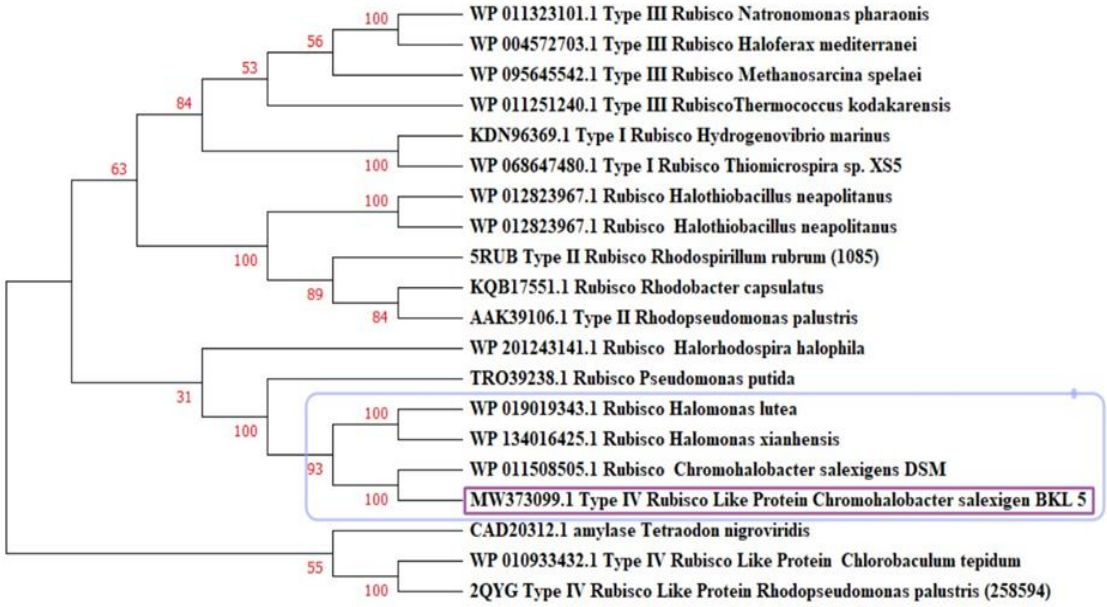

Figure 7. Phylogenetic tree of *Chromohalobacter salexigens* BKL5 (in square box) with other species expressing Rubisco. The phylogenetic tree was constructed using neighbor joining and bootstrap 1000

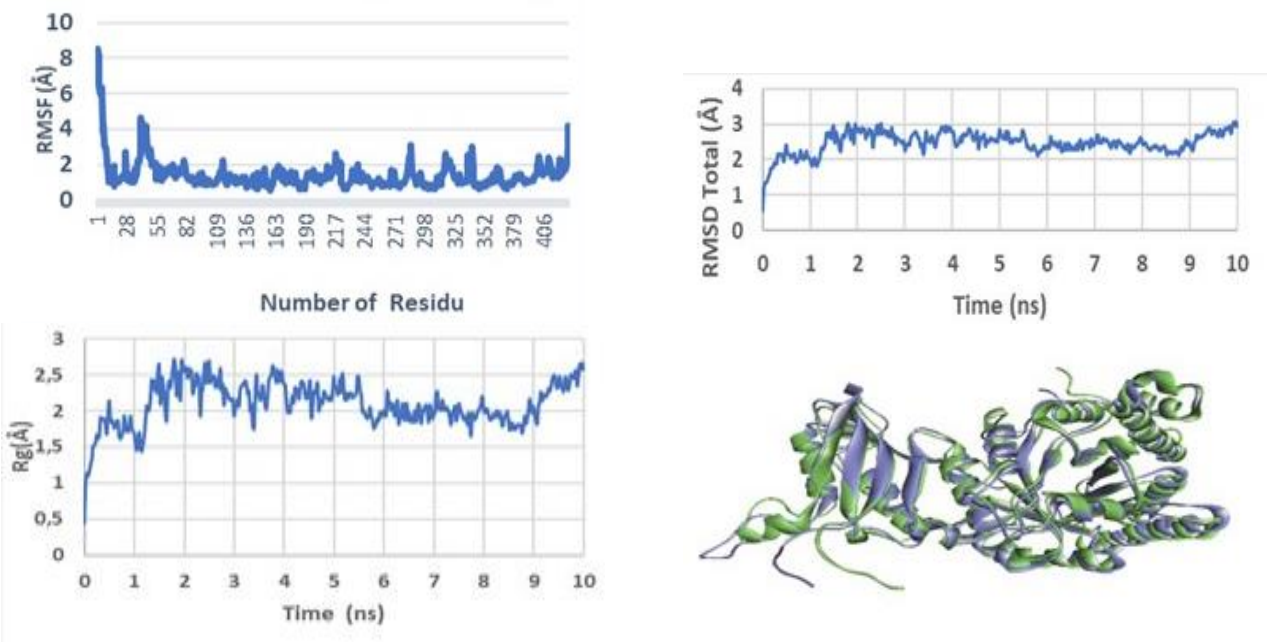

Figure 8. Result of dynamic molecular simulation test of RLP *Chromohalobacter salexigens* BKL 5 at 10 ns a) RMSF b) RMSD c) Rg and d) superimposed structure before simulation (green) and after simulation (purple)

Root Mean Square Fluctuation (RMSF) is a parameter that can describe the mobility/flexibility of individual amino acid residues (Narwani et al. 2019). RMSF measures how far atomic locations (amino acid residues) have deviated from the initial point (Rácz et al. 2022). Measurement of RMSF values shows that in general amino acid fluctuations at 10 ns intervals in RLP CS tend to be low (Figure 8.A). This indicates that this protein is relatively rigid and has low flexibility. This fact is understandable considering that in a halophilic environment the availability of water as a movement medium is very limited and this is a form of adaptation of the RLP protein in this environment.

The Rg of a protein describes the size of the protein globally. A small Rg value indicates that the protein is in a compact state while a large Rg value indicates that the protein has a smaller folding structure (Rampogu et al. 2022). Simulations at 10 ns show that the RLP CS protein has an Rg value of around 2 A, indicating that this protein has a low folding rate (Figure 8.B). This fact is in

accordance with the results of previous analysis which stated that this protein has a high random coil (unstructured protein) so that it will have a low folding rate and allows the protein to tend to open. This phenomenon is important so that the hydrophilic amino acids on the surface of the protein can be more exposed and interact with the limited amount of water.

The 10 ns simulation results show that the stability of the RLP CS protein is quite good in terms of the Rg, RMSF and RMSD values obtained (Kumar and Deshpande 2021; Kamenik et al. 2020; Yamamoto et al. 2021). This simulation suggests that RLP CS is adaptable in a hypersaline environment.

In conclusion, Rubisco like protein gene from *C. salexigens* BKL5 was successfully clonned and expressed in *Escherichia coli* BL21 (DE3) using the plasmid pCold IV. Homologation of the structure and validation processes gave very good results can further be used as a reference in testing and analysis related to the RLP structure of *C. salexigens* BKL 5. Observation of the 3D structure of the halophilic structure from *C. salexigens* BKL 5 shows that glutamic acid is more distributed on the surface of the protein. Amino acid residue sequencing and phylogenetic analysis showed that the protein was Rubisco Like Protein (RLP). Analysis with SOPMA denotes that this RLP was dominated by random coil structure which was identical to that found in species growing in extreme habitats. The 10 ns simulation results show that the stability of the RLP CS protein is quite good in hypersaline habitat.

ACKNOWLEDGEMENTS

We thank the Institut Teknologi Sumatera, Lampung, Indonesia for awarding the scholarship and *Hibah Bersaing Pasca Sarjana* UGM Year 2023 for funding this research.

REFERENCES

Arahal DR, García MT, Vargas C, Cánovas D, Nieto JJ, Ventosa A. 2001. *Chromohalobacter salexigens* sp. nov., a moderately halophilic species that includes *Halomonas elongata* DSM 3043 and ATCC 33174. Intl J Syst Evol Microbiol 51 (4): 1457-1462. DOI: 10.1099/00207713-51-4-1457.

Mishra A, Tripathi BK, Soam SS. 2019. Performance evaluation of leading protein multiple sequence alignment methods. Intl J Eng Adv Technol 9 (1): 771-776. DOI: 10.35940/ijeat. A1369.109119.

Ashida H, Saito Y, Nakano T, Tandeau DN, Sekowsk A, Danchin A, Yokota A. 2008. Rubisco like proteins as the enolase enzyme in the methionine salvage pathway: Functional and evolutionary relationships between Rubisco like proteins and photosynthetic Rubisco. J Exp Bot 59 (7): 1543-1554. DOI: 10.1093/jxb/ern104.

Badar MS, Shamsi S, Ahmed J, Alam MA. 2022. Molecular dynamics simulations: Concept, methods, and applications. In: Rezaei N (eds.). Transdisciplinarity. Integrated Science 5. DOI: 10.1007/978-3-030-94651-7_7.

Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. Biol Dir 7 (12): 1-14. DOI: 10.1186/1745-6150-7-12.

Bouvier JW, David ME, Timothy Rhodes, Jai SB, Amelia B, Alice E, Jochem RN, Anastasia U, Spencer MW, Steven K. 2021. Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off. Mol Biol Evol 38 (7): 2880-2896. DOI: 10.1093/molbev/msab079.

Carter MS, Zhang X, Huang H, Bouvier JT, Francisco BS, Vetting MW, Al-Obaidi N, Bonanno JB, Ghosh A, Zallot RG, Andersen HM, Almo SC, Gerlt JA. 2018. Functional assignment of multiple catabolic pathways for D-apiose. Nat Chem Biol 14: 696-705. DOI: 10.1038/s41589-018-0067-7.

Chen YW, Wang Y, Leung Y-C, Wong K-Y. 2021. Parameterization of large ligands for gromacs molecular dynamics simulation with LigParGen. Methods Mol Biol 2199: 277-288. DOI: 10.1007/978-1-0716-0892-0_16.

Gunde-Cimerman N, Plemenitaš A, Oren A. 2018. Strategies of adaptation of microorganisms of the three domains of life to high salt concentration. FEMS Microbiol Rev 42 (3): 353-375. DOI: 10.1093/femsre/fuy009.

Collier TA, Piggot TJ, Allison JR. 2020. Molecular dynamics simulation of proteins. Protein Nanotechnol 2073: 311-327. DOI: 10.1007/978-1-4939-9869-2_17.

De Oliveira CC, Pereira GR, De Alcantara JY, Antunes D, Caffarena ER, De Mesquita JF. 2019. In silico analysis of the V66M variant of human BDNF in psychiatric disorder: An approach to preciccion medicine. PLoS ONE 14 (4): e0215508. DOI: 10.1371/journal.pone.0215508.

Shill DK, Jahan S, Alam MM, Limon MB, Alam M, Rahman MZ, Rahman M. 2023. S-Adenosyl-l-Homocysteine Exhibits Potential Antiviral Activity Against Dengue Virus Serotype-3 (DENV-3) in Bangladesh: A Viroinformatics-Based Approach. Bioinformatics Biol Insights 17: 11779322231158249. DOI: 10.1177/11779322231158249.

Erb JT, Zarzycky J. 2018. A short history of Rubisco: The rise and fall (?) of Nature's predominant CO_2 fixing enzyme. Curr Opin Biotechnol 49: 100-107. DOI: 10.1016/j.copbio.2017.07.017.

Fitriani D, Rahman MS, Priyambada ID. 2017. Proteolytic activity of recombinan DegP from *Chromohalobacter salexigens* BKL 5. Electron J Biotechnol 29: 7-12. DOI: 10.1016/j.ejbt.2017.06.004.

Flamholz AI, Prywes N, Moran U, Davidi D, Bar-On YM, Oltrogge LM, Alves R, Savage D, Milo R. 2019. Revisiting trade-offs between Rubisco kinetic parameters. Biochemistry 58 (31): 3365-3376. DOI: 10.1021/acs.biochem.9b00237.

Hanson TE, Tabita FR. 2001. A ribulose 1,5 biphosphate carboxylase/oxygenase like protein from *Chlorobium tepidum* that involved with sulfur metabolism and the response to oxidative stress. Proc Natl Acad Sci 98 (8): 4398-4402. DOI: 10.1073/pnas.081610398.

Haque A, Baig GA, Alshawli AS, Sait KHW, Hafeez BB, Tripathi MK, Alghamdi BS, Mohammed Ali HSH, Rasool M. 2022. Interaction analysis of MRP1 with anticancer drugs used in ovarian cancer: In silico approach. Life 12 (3): 383. DOI: 10.3390/life12030383.

Imker HJ, Fedorov AA, Fedorov EV, Almo SC, Gerit JA. 2007. Mechanistic diversity in the RUBISCO superfamily: The enolase in the methionine salvage pathway in *Geobacillus kaustophilus*. Biochemistry 46 (13): 4077-4089. DOI: 10.1021/bi7000483

Joshi SH-N, Yong C, Gyorgy A. 2022. Inducible plasmid copy number control for synthetic biology in commonly used *E. coli* strains. Nat Commun 13: 6691. DOI: 10.1038/s41467-022-34390-7.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596: 583-589. DOI: 10.1038/s41586-021-03819-2.

Kamenik AS, Handle PH, Hofer F, Kahler U, Kraml J, Liedl KR. 2020. Polarizable and non-polarizable force fields: Protein folding, unfolding and misfolding. J Chem Phys 153: 185102. DOI: org/10.1063/5.0022135.

Kaur A, Pati PK, Pati AM, Nagpai AK. 2020. Physicochemical characterization and topological analysis of pathogeneis related proteins from *Arabidopsis thaliana* and *Oryza sativa* using insilico approach. PLoS One 15 (9): e0239836. DOI: 10.1371/journal.pone.0239836.

Kaur G, Prajapat M, Singh H, Sarma P, Bhadada SK, Shekhar N, Sharma S, Sinha S, Kumar S, Prakash A, Medhi B. 2023. Investigating the novel-binding site of RPA2 on Menin and predicting the effect of point mutation of Menin through protein–protein interactions. Sci Rep 13 (1): 9337. DOI: 10.1038/s41598-023-35599-2.

Kandeel M, Alzahrani A. 2020. Molecular dynamics simulation of carbonyl reductase 1 clarifies the structural switch in drug metabolism. J Taibah Univ Sci 14 (1): 1326-1334. DOI: 10.1080/16583655.2020.1821502.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular evolutionary genetics analysis version 11. Mol Biol Evol 38 (7): 3022-3027. DOI: 10.1093/molbev/msab120.

Kumar S, Deshpande PA. 2021 Structural and thermodynamic analysis of factors governing the stability and thermal folding/unfolding of SazCA. PLoS One 16 (4): e0249866. DOI: 10.1371/journal.pone.0249866.

González-Paz L, Lossada C, Hurtado-León ML, Vera-Villalobos J, Paz JL, Coll DS, Jeffreys LN, Marrero-Ponce Y, Alvarado YJ. 2023. Thermodynamics and conformational dynamics of primer nucleotide sequences associated with Monkeypox Virus (MpxV): Comparative analysis of stability and specificity with a biophysical-computational approach. Biointerface Res Appl Chem 13 (5): 435. DOI: 10.33263/BRIAC135.435.

Libera JL, Durán VF, Valdés JA, Núñez VG, Caballero J. 2020. LigRMSD: a web server for automatic structure matching and RMSD calculations among identical and similar compounds in protein-ligand docking. Bioinformatics 36 (9): 2912-2914. DOI: 10.1093/bioinformatics/btaa018.

Liang YY, Yan LQ, Tan MH, Li GH, Fang JH, Peng JY, Li KT. 2022. Isolation, characterization and genome sequencing of novel chitin deacetylase producing *Bacillus aryabhactai*. Front Microbiol 13: 999639. DOI: 10.3389/fmicb.2022.999639.

Shekhar M, Smith Z, Seeliger MA, Tiwary P. 2022. Protein flexibility and dissociation pathway differentiation can explain onset of resistance mutations in kinases. Angew Chem Intl Ed 61 (28): e202200983. DOI: 10.1101/2021.07.02.450932.

Nomoto A, Nishinami S, Shiraki K. 2021. Solubility parameters of amino acids on liquid–liquid phase separation and aggregation of proteins. Front Cell Dev Biol Sec 9: 691052 DOI: 10.3389/fcell.2021.691052.

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF Chimera X: Structure visualization for researcher, educator and developer. Protein Sci 30 (1): 70-82. DOI: 10.1002/pro.3943.

Rácz A, Mihalovits LM, Bajusz D, Héberger K, Miranda-Quintana RA. 2022. Molecular dynamics simulations and diversity selection by extended contiuous similarity indices. J Chem Inf Model 62 (14): 3415-3425. DOI: 10.1021/acs.jcim.2c00433.

Rampogu S, Lee G, Park JS, Lee KW, Kim MO. 2022. Molecular docking and molecular dynamics simulations discover curcumin analogue as a plausible dual inhibitor for SARS-CoV-2. Intl J Mol Sci 23 (3): 1771. DOI: 10.3390/ijms23031771.

Richard JP. 2019. Protein flexibility and stiffness enable efficient enzymatic catalysis. J Am Chem Soc 141 (8): 3320-3331. DOI: 10.1021/jacs.8b10836.

Rohman MS, Prijambada ID, Indriyani YA, Hendrosatriyo H. 2012. Identification of protease producing halophilic bacteria from Bledug Kuwu Mud Volcano. Indones J Biotechnol 17 (1): 35. DOI: 10.22146/ijbiotech.15995.

Hasan BM, Abdulazeez AM. 2021. A review of principal component analysis algorithm for dimensionality reduction. J Soft Comput Data Mining 2 (1): 20-30. DOI: 10.30880/jscdm.2021.02.01.003.

Sudarmanto I, Rohman MS, Artama WT. 2024. The effect of the D185 mutation on the stability and functionality Of Rubisco Like Protein (RLP) from *Chromohalobacter salexigens* BKL 5. J Chem 18: 1 DOI: 10.24843/JCHEM.2024.v18.i01.p01.

Narwani TJ, Etchebest C, Craveur P, Léonard S, Rebehmed J, Srinivasan N, Bornot A, Gelly JC, de Brevern AG. 2019. In silico prediction of protein flexibility with local structure approach. Biochimie 165: 150-155. DOI: 10.1016/j.biochi.2019.07.025.

Tommasi IC. 2021. The mechanism of Rubisco catalyzed carboxylation chemical aspect involving acid-base chemistry and functioning of the molecular machine. Catalysts 11 (7): 813. DOI: 10.3390/catal11070813.

Wang Y, Cao X, Zhao Y, Fei J, Hu X, Li N. 2017. Optimized double-digest genotyping by sequencing (ddGBS) method with high-density SNP markers and high genotyping accuracy for chickens. PLoS One 12 (6): e0179073. DOI: 10.1371/journal.pone.0179073.

Qing G, Ma LC, Khorchid A, Swapna GV, Mal TK, Takayama MM, Xia B, Phadtare S, Ke H, Acton T, Montelione GT. 2004. Cold-shock induced high-yield protein production in *Escherichia coli*. Nat Biotechnol 22 (7): 877-882. DOI: 10.1038/nbt984.

Yokota A. 2017. Revisiting Rubisco. Biosci Biotechnol Biochem 81 (11): 2039-2049. DOI: 10.1080/09168451.2017.1379350.

Yamamoto E, Akimoto T, Mitsutake A, Metzler R. 2021. Universal relation between instantaneous diffusivity and radius of gyration of proteins in aqueous solution. Phys Rev Lett 126: 128101. DOI: 10.1103/PhysRevLett.126.128101.