



Physicochemical Characterization and Comparative Analysis of Ribulose 1,5 Biphosphate Carboxylase–Oxygenase Like Proteins (RLP) from Halophilic *Chromohalobacter salexigens* BKL 5 and Non-Halophilic Counterparts Using in Silico Approaches

Irwan Sudarmanto ^{1,2}, Wayan Tunas Artama ^{2,4}, Muhammad Saifur Rohman ^{2,3,*}

¹ Chemical Study Program, Faculty of Science, Institute Technology of Sumatera, Jl Terusan Ryacudu, WayHui, Lampung Selatan, 35365, Lampung, Indonesia

² Post Graduate Program of Biotechnology, Post Graduate Program, Universitas Gadjah Mada, Jl Teknik Utara, Catur Tunggal, Sleman, 55281, Yogyakarta, Indonesia

³ Department of Agriculture Microbiology, Faculty of Agriculture, Universitas Gadjah Mada, Karang malang, Catur Tunggal, Sleman, 55281, Yogyakarta, Indonesia

⁴ Faculty of Veterinary Medicine, Universitas Gadjah Mada, Jl Fauna No 2, Karang gayam, Catur tunggal, Sleman, 55281, Yogyakarta, Indonesia

* Corresponding author: saifur@ugm.ac.id

<https://doi.org/10.14710/jksa.27.4.151-159>

Article Info

Article history:

Received: 02nd November 2023

Revised: 04th March 2024

Accepted: 01st April 2024

Online: 30th April 2024

Keywords:

halophilic; metaboanalyst 5.0; *Chromohalobacter salexigens* BKL5; RuBisCo-like Protein; Partial Least Square

Abstract

This study is intended to analyze and compare halophilic RuBisCo-like proteins (RLP) with non-halophilic homologs using computational analysis approaches. The data used were protein sequences from NCBI and protein sequences from cloned and expressed RuBisCo-like proteins *Chromohalobacter salexigens* BKL 5 from our previous study. The analysis was in the form of Principal Component Analysis (PCA) and Partial Least Square (PLS). The tools used were Origin Lab Full Version 2019 and Metaboanalyst 5.0. The parameters tested were isoelectric point (pI), negative charge (acidic amino acids), aliphatic index, and GRAVY index using the ProtParam tool from ExPASy. Calculations analysis showed that the amino acid residues in halophilic RuBisCo-like proteins that differ significantly from their homologs were glutamic acid and alanine, while the variables that could be observed as differentiating were negative charge and aliphatic index. Analysis shows that some parameters obtained from the program can be used as discriminants to differentiate halophilic and non-halophilic RuBisCo-like proteins.

1. Introduction

Halophilic bacteria are a type of bacteria capable of thriving in environments with high salt concentrations. These organisms can be classified based on their optimal salt requirements for growth. Specifically, weakly halophilic bacteria require 0.2-0.85 M (1-5%) sodium chloride (NaCl), moderately halophilic bacteria require 0.85-3.4 M (5-20%) NaCl, and extremely halophilic bacteria require 3.4-4.5 M (20-30%) NaCl [1]. Based on several studies conducted, the adaptation of halophilic microorganisms mainly occurs at the protein level [2], where there are different characteristics between halophilic and non-halophilic proteins [2, 3, 4].

One of the characteristics of halophilic proteins is the high content of negatively charged amino acid residues, especially those from extreme halophilic bacteria, which can reach 20% of the total residue [5]. 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 their low content of hydrophobic residues [6]. This is an adaptation of proteins to reduce hydrophobic interactions, thereby maintaining the stability of halophilic proteins in high-salt environments [7].

Researchers have conducted several explorations to evaluate the relationship between the structure of halophilic proteins and their stability. They conducted several studies at all structural levels, from primary and secondary to tertiary, between halophilic and non-halophilic proteins. The results showed that the residue composition differs between halophilic and non-halophilic proteins on the surface and interior areas [7]. Electrostatic interactions and unique amino acid composition are the unique characteristics that distinguish halophilic and non-halophilic proteins.

Computational analysis has become a new approach to accelerate research in the protein field [8]. The essence of computational analysis is finding patterns and distribution in existing data and subsequently making predictions for new datasets. These data typically comprise data points characterized by various features or descriptors, such as protein sequences, their secondary and tertiary structures, and the physicochemical properties of amino acids. The number of features in such datasets usually varies from dozens to thousands, rendering these problems high-dimensional. In terms of halophilic proteins, research based on computational analysis has been carried out by Zhang and Ge [4] and Asy'ari *et al.* [9].

We have isolated the RuBisCo-like proteins (RLP) gene from *Chromohalobacter salexigens* BKL 5, sourced from the hypersaline environment of the Bledug Kuwu Mud Volcano in Central Java, where the salt content is approximately 10% [10]. This gene was cloned onto the pCold plasmid and subsequently transformed and expressed in the *Escherichia coli* BL 21 system. These findings remain unpublished. Based on its habitat, this type of bacteria is classified as an extremophile species, organisms that tolerate and thrive in the most extreme and challenging conditions of life. As a result of these extreme environmental insults, extremophiles have developed several interesting adaptations to cellular membranes, proteins, and extracellular metabolites [11].

These environments are often so unique that the organisms within them exhibit highly specialized adaptations at the protein level. For instance, they possess enzymes capable of functioning effectively in extreme conditions without denaturing. Such proteins can operate under conditions where mesophilic proteins might fail. Some variables can lead to an environment being considered extreme, such as pH, temperature, relative salinity, or the presence of other external factors, such as heavy metals or radiation [12]. In this case, *Chromohalobacter salexigens* BKL 5 are categorized as halophiles, organisms that can survive the ionic stresses placed upon them by saline environments.

Chromohalobacter salexigens BKL 5 species is a moderately halophilic gram-negative bacterium in the form of single or paired cells. The colonies of this bacterium are creamy, round, and 2 mm in diameter. These bacteria grow optimally at a salt concentration of 7.5 to 10%. In a freshwater environment, CS was found

and isolated from mud in the Bledug Kuwu Mud Volcano, Central Java [10].

RuBisCo was grouped into four groups: I, II, III, and IV. Groups I, II, and III function to catalyze the carboxylation or oxygenation of ribulose 1,5-bisphosphate (RuBP). Group IV, known as RuBisCo-like protein (RLP), is unable to fix CO₂ due to substitution at the amino acid residue in its active site [13]. RLP plays a role in the methionine metabolic pathway and catalyzes the enolization reaction of the RuBP analog substrate, specifically the 2,3-keto-5-methylthiopentyl-1-P analog [13, 14].

This study represents the first instance of using computational methods for a comparative analysis of halophilic and non-halophilic RuBisCo-like proteins (RLPs). The objective was to identify and compare the differences between these proteins based on their amino acid compositions. The analysis utilized protein sequences sourced from the National Center for Biotechnology Information (NCBI) at www.ncbi.nlm.nih.gov. Parameters related to amino acids, such as isoelectric point (pI), aliphatic index, GRAVY index, and negative charge, were key factors in this comparative study.

2. Experimental

2.1. Materials

Fifteen RLP sequences were obtained from NCBI (www.ncbi.nlm.nih.gov) in FASTA format (Table 1). The literature search results showed that six of these RLP types belonged to halophilic bacteria, while the remaining nine were derived from non-halophilic bacteria. The dataset comes from various bacterial classes, with the halophilic group specifically originating from bacteria classified as moderately halophilic.

2.2. Extractions

2.2.1. Prediction of 3D Protein Structure, Validation, and Visualization

In this study, the 3D RLP protein structure used as an illustration to describe the position of amino acids is the RLP structure of *Chromohalobacter salexigens* BKL 5 resulting from homology obtained using the validated AlphaFold. Amino acid sequences were predicted in 3D structure using the AlphaFold ([www.github.com/google-deepmind/alphafold](https://github.com/google-deepmind/alphafold)), a protein homology method that can provide high accuracy [15]. The structure obtained was then validated using PROCHECK [<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>] [16, 17] and ProSA (<https://prosa.services.came.sbg.ac.at/prosa.php>) [18].

2.2.2. Analysis of Amino Acid Composition

RLP amino acid sequences were analyzed using the ProtParam tool available on ExPASy (www.expasy.org). The results obtained were the composition of amino acid residues and several parameters, such as pI (isoelectric point), aliphatic index, and GRAVY index [19, 20].

Table 1. Fifteen RLP protein sequences from different organisms (bacteria) (www.ncbi.nlm.nih.gov)

Species	Accession number	Halophilic	Number of amino acids	Reference
<i>Achromobacter marplatensis</i>	EJO30418	no	430	Trimble <i>et al.</i> [21]
<i>Acidovorax</i> sp	AFU45989	no	425	Ohtsubo <i>et al.</i> [22]
<i>Actinobaculum suis</i>	KMY22626	no	416	Arnold <i>et al.</i> [23]
<i>Advenella kashmirensis</i>	AFK61550	no	424	Ghosh <i>et al.</i> [24]
<i>Bordetella bronchiseptica</i>	Q7WVK2	no	423	Parkhill <i>et al.</i> [25]
<i>Chlorobaculum thiosulfatophilum</i>	AAK14332	yes	434	Imhoff [26]
<i>Chromohalobacter salexigens</i> BKL 5	MW373099	yes	429	Rohman <i>et al.</i> [10]
<i>Allochromatium vinosum</i> DSM 180	Q93UZ0	yes	457	Weissgerber <i>et al.</i> [27]
<i>Flammeovirgaceae</i> bacterium	AHM61158	yes	415	Yoon <i>et al.</i> [28]
<i>Halomonas</i> sp	EHA15577	yes	424	Lin <i>et al.</i> [29]
<i>Microcystis aeruginosa</i>	CAJ43366	no	386	Frangeul <i>et al.</i> [30]
<i>Paenibacillus</i> sp	ETT62068	no	423	Moreno Switt <i>et al.</i> [31]
<i>Rhizobium grahamii</i>	EPE94755	no	417	Althabegoiti <i>et al.</i> [32]
<i>Rhodospseudomonas palustris</i>	ACF00976	no	366	Li <i>et al.</i> [33]
<i>Roseivivax marinus</i>	ETW11401	yes	420	Dai <i>et al.</i> [34]

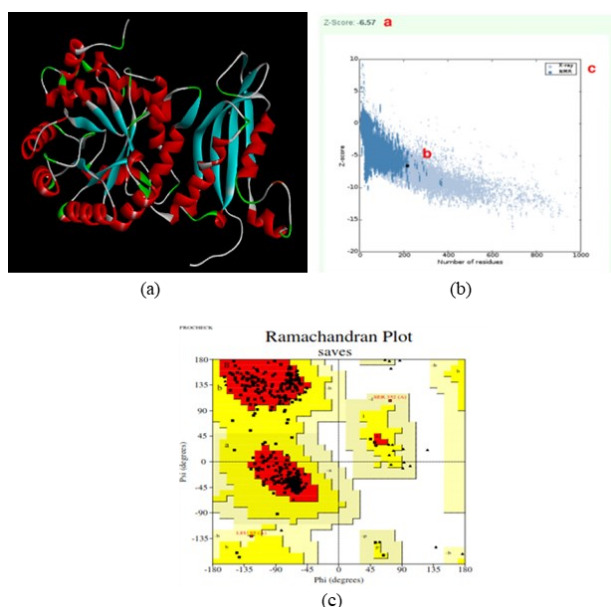


Figure 1. Structural analysis of RLP from *Chromohalobacter salexigens* BKL 5: (a) Predicted structure using AlphaFold, (b) Z-score validation performed with ProSA, and (c) Ramachandran plot assessment via PROCHECK

2.2.3. Calculation of PLS Regression

Partial Least Squares (PLS) regression is a method that reduces several variables into a smaller set of predictors [35]. These predictors were utilized to determine regressions and correlations. VIP (Variable important in Projection) analysis was performed to analyze the most significant amino acid residues and parameters that differentiate between halophilic and non-halophilic RLPs. This analysis was conducted using the Metaboanalyst 5.0 software [36]. The 3D structures of the RLP proteins and the positions of the amino acids were visualized using the ChimeraX tool, following the guidelines provided in the Chimera user manual (www.cgl.ucsf.edu) [37].

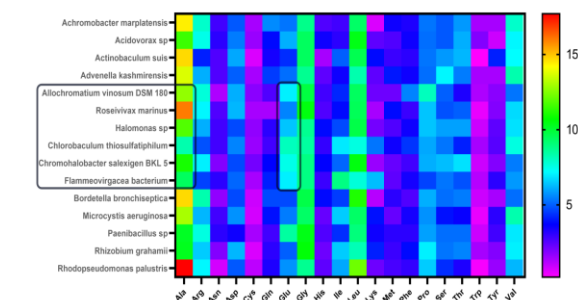


Figure 2. Heatmap of amino acid composition in several RLP. The area in the square box is the halophilic RLP. The scale on the right column describes the level of amino acid composition; the greater the value, the higher the composition

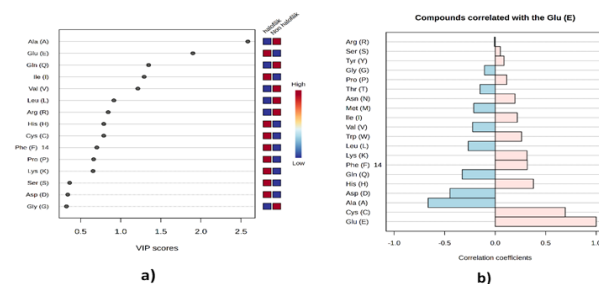


Figure 3. Results of PLS analysis of the amino acid composition of (a) VIP score and (b) correlation coefficient

2.2.4. PCA Analysis

Principal component analysis (PCA) is a dimensionality reduction technique frequently employed to reduce large datasets by transforming extensive sets of variables into smaller ones that still contain the bulk of the information from the original sets [38]. Origin Lab Full Version 2019 was utilized to analyze the heatmap and PCA. The parameters used as variables in this analysis included the isoelectric point (pI), negative charge (representing acidic amino acids), aliphatic index, and GRAVY index.

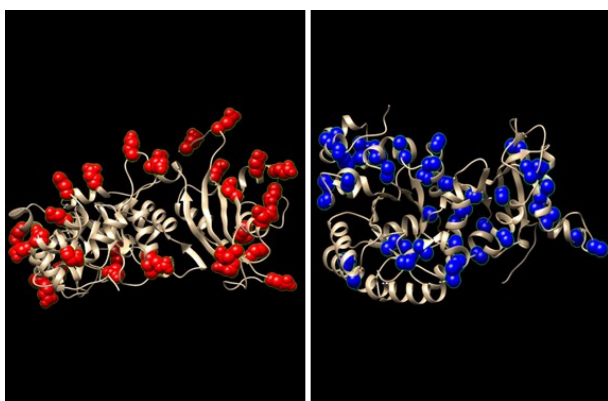


Figure 4. Distribution of glutamic acid (red) and alanine (blue) in the 3D RLP structure of *Chromohalobacter salexigens* BKL 5 visualized using Chimera

3. Results and Discussion

3.1. Prediction of 3D Protein Structure, Validation, and Visualization

The results of 3D structural prediction with AlphaFold are shown in Figure 1a. The structural validation of the homology results gave a value of 92.54% and a z-score of -6.57 (Figure 1b) (ProSA). ProSA is a tool used to calculate the energy required for amino acids to fold into their functional configurations. A more negative value indicates greater protein stability. Meanwhile, the z-score is a value describing the quality of the tested protein model. The Ramachandran plot analysis conducted through the PROCHECK web server indicates the protein structure contains amino acid residues located in the allowed positions. Specifically, 89.8% of the residues are located in the most favored regions, 9.6% in the additionally allowed regions, and 0.5% in the generous regions, as shown in Figure 1c. This shows that structure prediction using AlphaFold yields excellent results and can be further utilized as a reference for testing and analyzing the RLP structure of *Chromohalobacter salexigens* BKL 5.

3.2. Analysis of Amino Acid Composition

Amino acids are the building blocks of proteins whose composition and configuration determine function and activity. The function of a protein is strongly dependent on its structure [39]. Alterations in protein structures from organisms evolved to extreme environments vary in their mechanisms for maintaining optimal activity [40].

RLP functions as an enolase catalyst and exhibits significant structural variation. Consequently, these proteins display very low similarity to each other, with about only 30% resemblance in their structures [41, 42]. This phenomenon occurs because the configuration of amino acids in RLP tends to vary, even though the composition of the amino acids in these proteins generally remains relatively consistent. Analysis of amino acid composition using ProtParam from ExPasy, visualized with a heatmap, shows that RLP from halophilic and non-halophilic bacteria have similar patterns as depicted in Figure 2.

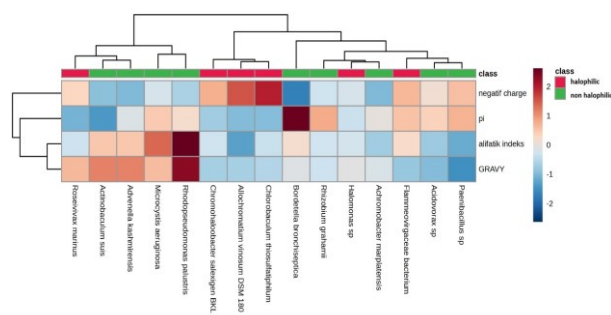


Figure 5. Halophilic and non-halophilic RLP heatmap variables

3.3. Calculation of PLS Regression

Further analysis was carried out to validate the heatmap results with a chemometric approach based on machine learning. The PLS method was used to determine the most significant variable as a discriminant for halophilic and non-halophilic classes. The results show that the two amino acids with high VIP scores are alanine and glutamic acid (Figure 3a). In the halophilic RLP group, the amount of glutamic acid was greater than in the non-halophilic class. Conversely, the alanine composition in halophilic RLP was less than in non-halophilic RLP. Analysis of the correlation coefficient on glutamic acid between halophilic and non-halophilic RLP strengthened the PLS results. Specifically, it revealed an inverse correlation between the levels of glutamic acid and alanine (Figure 3b), where an increase in the amount of glutamic acid would be followed by a decrease in the amount of alanine.

Water is less available to protein at high salt concentrations (higher than 0.1 M) because most water surrounds salt in an ionic lattice [43]. The lower availability of water can cause hydrophobic amino acids in a protein to lose hydration and tend to aggregate. Subsequently, high salt concentrations strengthen hydrophobic interactions in a protein. Salt also interferes with the electrostatic interactions between charged amino acids. Non-halophilic proteins cannot function in high salt concentrations because the hydrophobic and electrostatic interactions they normally rely on for proper folding and maintaining stability are greatly altered. This can even lead to destabilization of the protein, potentially causing global unfolding and aggregation, ultimately leading to precipitation [44].

One of the most notable adaptations of halophilic proteins is the large increase in acidic residues on the protein's surface, like glutamic and aspartic acid. This is almost ubiquitous with halophilic proteins and can distinguish between halophilic and non-halophilic protein sequences [45]. PLS analysis of the RLP protein shows that the halophilic acidic amino acid significantly different from its non-halophilic homolog is glutamic acid (Figure 3a). Glutamic acid in halophilic RLP is more abundant than its non-halophilic counterpart. Figure 4 shows that the glutamic acid position in the RLP protein is distributed more on the protein surface. This indicates that the adaptation of RLP to a hypersaline environment is that it has a higher glutamic acid composition and is distributed on its surface.

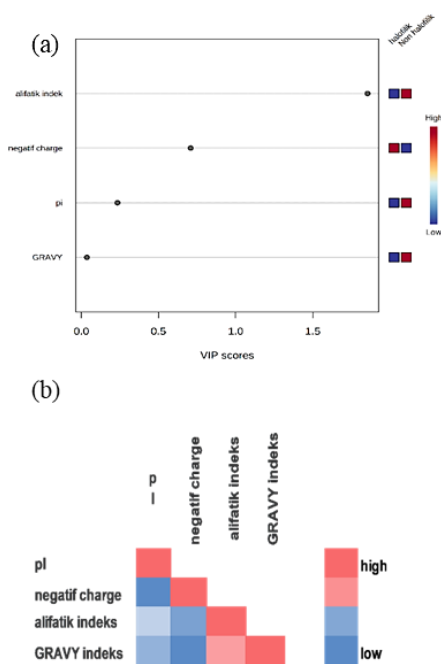


Figure 6. Results of PLS analysis of (a) several VIP score parameters and (b) correlation matrix

Glutamic acid, like aspartic acid, has a high ability and is very important for protein solubilization. There are several possible roles for these acidic residues. It is thought that the increased negative charge on the protein's surface allows the protein to compete with positively charged ions from salt and facilitates excess protein hydration [46, 47, 48].

The description of the distribution of alanine in halophilic RLP strengthens the notion that the smaller amount of alanine (10% total average on halophilic RLP compared to more than 12% average total on non-halophilic RLP) allows the RLP protein to be more flexible. Another adaptation made to maintain protein stability is increasing interaction with an environment where there is minimal availability of water as a medium by reducing the number of hydrophobic amino acids. In this halophilic RLP, the amino acid in question is alanine (Figure 3). Reduction of these amino acids is essential to decrease hydrophobic interactions, thereby potentially enhancing protein flexibility under high salt stress conditions [49].

The availability of free water molecules in a halophilic environment is deficient because water is surrounded by positive ions from salts [50]. Halophilic proteins carry out several general adaptations, including a decreased presence of hydrophobic residues and an increased abundance of acidic amino acid residues [51]. Negative charges on the protein surface will interact with hydrated ions, while low hydrophobic residues will reduce aggregation [52]. Amino acid residues with polar characteristics can enhance stability by engaging in interactions with water within the cytoplasm. Another adaptation involves abundant residues with relatively polar side chains, such as threonine or serine, which are hydrophobic residues [53].

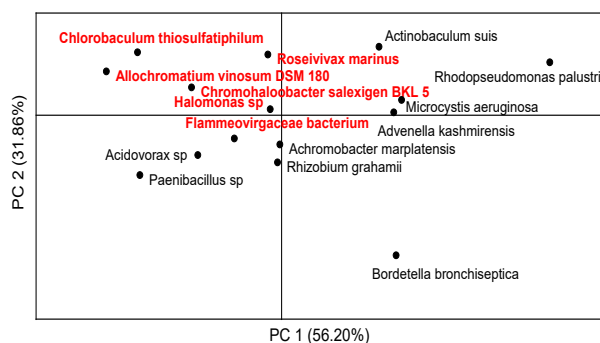


Figure 7. PCA of RLP from several bacteria. Red writing indicates halophilic, while black writing is non-halophilic

Analysis of several variables obtained from ProtParam was used to determine which factors were most used to distinguish the halophilic class. The heatmap results show that halophiles tend to have a high negative charge and a low aliphatic index (Figure 5). These results were later confirmed by conducting a PLS analysis, and it showed consistent results where the VIP with the highest score was the negative charge and aliphatic index (Figure 6).

The aliphatic index is the relative volume of a protein occupied by aliphatic/hydrophobic side chains of specific amino acids (alanine, valine, leucine, and isoleucine) [54, 55]. The VIP analysis results show that the aliphatic index of halophilic RLP is relatively lower than that of its non-halophilic homologs (Figure 6). This reduction is primarily due to a lower presence of alanine residues, which contribute to the aliphatic content, by approximately 2% (calculations not shown) [48]. The low aliphatic index value is also related to the folding rate of a protein, where the lower the index aliphatic value, the smaller the folds. The low aliphatic index of halophilic RLP allows the RLP protein to tend to be more flexible and allows the protein to be more exposed to solvents [56].

Siglioccolo *et al.* [57] determined that the hydrophobic interaction within the core of halophilic proteins is consistently less than that observed in non-halophilic proteins. They propose that the lower hydrophobic contact in the core may counterbalance the increased strength of hydrophobic interactions in high salt concentrations. Weaker hydrophobic interactions, due to smaller hydrophobic residues, can enhance the flexibility of proteins in high-salt environments. This increased flexibility helps prevent the protein's hydrophobic core from becoming excessively rigid [43].

Low hydrophobic interactions in general, particularly at the level of core and conserved hydrophobic contacts, may help the structure prevent loss of function in hypersaline environments. Shrinking of hydrophobic contacts must be even more critical for the early stages of folding when intramolecular hydrophobic nuclei must correctly form to guide the polypeptide through the folding funnel to the native state [57, 58].

3.4. PCA Analysis

Principal Component Analysis (PCA) of several variables obtained from ProtParam shows that these parameters can be effectively utilized to determine whether a protein, specifically RLP, is halophilic (Figure 7). The variables used were PC1 (negative charge; aliphatic index) and PC2 (GRAVY index; isoelectric point (pI)), which represent 88.06% of the number of variants and are considered quite representative. PCA is a method that can reduce many variables into unrelated variables and is referred to as the principal component that can be used to conclude. By utilizing parameters such as the isoelectric point (pI), negative charge (from acidic amino acids), aliphatic index, and GRAVY index in combination with computational analysis, it is possible to distinguish halophilic RLP (red writing) from its non-halophilic counterpart (black writing). This method proves extremely useful for the rapid identification of proteins, particularly when the origin or habitat of a protein with a specific sequence is unknown.

4. Conclusion

Calculations through PLS showed that glutamic acid and alanine are the amino acids with the most significant differences, as indicated by the highest VIP values. The two amino acids are unique in that they differentiate amino acids as an adaptation of RLP to a halophilic environment. Glutamic acid, an acidic amino acid, played a crucial role in maintaining the solubility of RLP under high salt stress conditions and was essential for the formation of salt bridges that help sustain protein structure. As one of the hydrophobic amino acids, alanine was present in lower amounts in halophilic RLP compared to its mesophilic counterparts. This reduction leads to decreased hydrophobic interactions within the protein, thereby increasing its exposure to solvents. PCA analysis of parameters such as the isoelectric point (pI), negative charge (from acidic amino acids), aliphatic index, and GRAVY index yielded positive results (halophilic RLP was completely separated from its mesophilic homolog); therefore, these parameters can be effectively utilized as variables in chemometric procedures to determine whether an RLP protein is halophilic or non-halophilic.

Acknowledgment

We thank the Institute of Technology of Sumatera for awarding the scholarship and funding this study.

References

- [1] Gabriel Zamith Leal Dalmaso, Davis Ferreira, Alane Beatriz Vermelho, Marine Extremophiles: A Source of Hydrolases for Biotechnological Applications, *Marine Drugs*, 13, 4, (2015), 1925-1965 <https://doi.org/10.3390/md13041925>
- [2] Bassam Oudh Aljohny, Halophilic Bacterium - A Review of New Studies, *Biosciences Biotechnology Research Asia*, 12, 3, (2015), 2061-2069 <http://dx.doi.org/10.13005/bbra/1874>
- [3] Bhramar Dutta, Rajib Bandopadhyay, Biotechnological potentials of halophilic microorganisms and their impact on mankind, *Beni-Suef University Journal of Basic and Applied Sciences*, 11, (2022), 75 <https://doi.org/10.1186/s43088-022-00252-w>
- [4] Guangya Zhang, Huihua Ge, Protein Hypersaline Adaptation: Insight from Amino Acids with Machine Learning Algorithms, *The Protein Journal*, 32, (2013), 239-245 <https://doi.org/10.1007/s10930-013-9484-3>
- [5] Yan-Huey Chen, Chia-Wei Lu, Yuan-Tay Shyu, Shih-Shun Lin, Revealing the Saline Adaptation Strategies of the Halophilic Bacterium *Halomonas beimenensis* through High-throughput Omics and Transposon Mutagenesis Approaches, *Scientific Reports*, 7, (2017), 13037 <https://doi.org/10.1038/s41598-017-13450-9>
- [6] Aharon Oren, Halophilic microbial communities and their environments, *Current Opinion in Biotechnology*, 33, (2015), 119-124 <https://doi.org/10.1016/j.copbio.2015.02.005>
- [7] Nina Gunde-Cimerman, Ana Plemenitaš, Aharon Oren, Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations, *FEMS Microbiology Reviews*, 42, 3, (2018), 353-375 <https://doi.org/10.1093/femsre/fuy009>
- [8] Qin Wang, Yanchao Wang, Xiaoming Jiang, Lei Ma, Zhaojie Li, Yaoguang Chang, Yuming Wang, Changhu Xue, Amino Acid Profiling with Chemometric Analysis as a Feasible Tool for the Discrimination of Marine-Derived Peptide Powders, *Foods*, 10, 6, (2021), 1294 <https://doi.org/10.3390/foods10061294>
- [9] Mukhammad Asy'ari, Pingkan Aditiawati, Akhmaloka, Rukman Hertadi, Structural Analysis of Halophilic Lipases Isolated from Bledug Kuwu Mud Crater, Purwodadi-Grobogan, Indonesia: A Structural Bioinformatics Study, *Procedia Chemistry*, 16, (2015), 392-399 <https://doi.org/10.1016/j.proche.2015.12.069>
- [10] Muhammad Saifur Rohman, Irfan Dwidya Prijambada, Yohanna Anisa Indriyani, Heri Hendrosatriyo, Identification of Protease Producing Halophilic Bacteria from Bledug Kuwu Mud Volcano, *Indonesian Journal of Biotechnology*, 17, 1, (2012), 35-41 <https://doi.org/10.22146/ijbiotech.15995>
- [11] Pabulo Henrique Rampelotto, Extremophiles and Extreme Environments, *Life*, 3, 3, (2013), 482-485 <https://doi.org/10.3390/life3030482>
- [12] James Charlesworth, Brendan P. Burns, Extremophilic adaptations and biotechnological applications in diverse environments, *AIMS Microbiology*, 2, 3, (2016), 251-261 <https://doi.org/10.3934/microbiol.2016.3.251>
- [13] Thomas E. Hanson, F. Robert Tabita, A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress, *Proceedings of the National Academy of Sciences*, 98, 8, (2001), 4397-4402 <https://doi.org/10.1073/pnas.081610398>
- [14] Suk Min Kim, Hyun Seung Lim, Sun Bok Lee, Discovery of a RuBisCO-like Protein that Functions as an Oxygenase in the Novel D-Hamamelose Pathway, *Biotechnology and Bioprocess Engineering*, 23, (2018), 490-499 <https://doi.org/10.1007/s12257-018-0305-6>

- [15] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstern, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, Demis Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature*, 596, 7873, (2021), 583-589 <https://doi.org/10.1038/s41586-021-03819-2>
- [16] Gurjeet Kaur, Manisha Prajapat, Harvinder Singh, Phulen sarma, Sanjay kumar Bhadada, Nishant Shekhar, Saurabh Sharma, Shweta Sinha, Subodh kumar, Ajay Prakash, Bikash Medhi, Investigating the novel-binding site of RPA2 on Menin and predicting the effect of point mutation of Menin through protein-protein interactions, *Scientific Reports*, 13, (2023), 9337 <https://doi.org/10.1038/s41598-023-35599-2>
- [17] Clara Carolina Silva De Oliveira, Gabriel Rodrigues Coutinho Pereira, Jamile Yvis Santos De Alcantara, Deborah Antunes, Ernesto Raul Caffarena, Joelma Freire De Mesquita, In silico analysis of the V66M variant of human BDNF in psychiatric disorders: An approach to precision medicine, *PLoS ONE*, 14, 4, (2019), e0215508 <https://doi.org/10.1371/journal.pone.0215508>
- [18] Dipok Kumer Shill, Shafina Jahan, Mohammad Mamun Alam, Md Belayet Hasan Limon, Muntasir Alam, Mohammed Ziaur Rahman, Mustafizur Rahman, S-Adenosyl-l-Homocysteine Exhibits Potential Antiviral Activity Against Dengue Virus Serotype-3 (DENV-3) in Bangladesh: A Viroinformatics-Based Approach, *Bioinformatics and Biology Insights*, 17, (2023), <https://doi.org/10.1177/11779322231158249>
- [19] Elisabeth Gasteiger, Alexandre Gattiker, Christine Hoogland, Ivan Ivanyi, Ron D. Appel, Amos Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Research*, 31, 13, (2003), 3784-3788 <https://doi.org/10.1093/nar/gkg563>
- [20] Amritpreet Kaur, Pratap Kumar Pati, Aparna Maitra Pati, Avinash Kaur Nagpal, Physico-chemical characterization and topological analysis of pathogenesis-related proteins from *Arabidopsis thaliana* and *Oryza sativa* using *in-silico* approaches, *PLoS ONE*, 15, 9, (2020), e0239836 <https://doi.org/10.1371/journal.pone.0239836>
- [21] William L. Trimble, Le T. Phung, Folker Meyer, Simon Silver, Jack A. Gilbert, Draft Genome Sequence of *Achromobacter piechaudii* Strain HLE, *Journal of Bacteriology*, 194, 22, (2012), 6355-6355 <https://doi.org/10.1128/jb.01660-12>
- [22] Yoshiyuki Ohtsubo, Fumito Maruyama, Hisayuki Mitsui, Yuji Nagata, Masataka Tsuda, Complete Genome Sequence of *Acidovorax* sp. Strain KKS102, a Polychlorinated-Biphenyl Degrader, *Journal of Bacteriology*, 194, 24, (2012), 6970-6971 <https://doi.org/10.1128/jb.01848-12>
- [23] Mirjam Arnold, Olivia Richard, Corinne Gurtner, Heiko Nathues, Alexander Grahofer, A case report: *Actinobaculum suis* infection associated with formation of pyogranuloma, epididymitis and azoospermia in a boar, *BMC Veterinary Research*, 17, 1, (2021), 6 <https://doi.org/10.1186/s12917-020-02680-1>
- [24] Wriddhiman Ghosh, Ashish George, Atima Agarwal, Praveen Raj, Masrura Alam, Prosenjit Pyne, Sujoy Kumar Das Gupta, Whole-Genome Shotgun Sequencing of the Sulfur-Oxidizing Chemoautotroph *Tetrathiodibacter kashmirensis*, *Journal of Bacteriology*, 193, 19, (2011), 5553-5554 <https://doi.org/10.1128/jb.05781-11>
- [25] Julian Parkhill, Mohammed Sebaihia, Andrew Preston, Lee D. Murphy, Nicholas Thomson, David E. Harris, Matthew T. G. Holden, Carol M. Churcher, Stephen D. Bentley, Karen L. Mungall, Ana M. Cerdeño-Tárraga, Louise Temple, Keith James, Barbara Harris, Michael A. Quail, Mark Achtman, Rebecca Atkin, Steven Baker, David Basham, Nathalie Bason, Inna Cherevach, Tracey Chillingworth, Matthew Collins, Anne Cronin, Paul Davis, Jonathan Doggett, Theresa Felwell, Arlette Goble, Nancy Hamlin, Heidi Hauser, Simon Holroyd, Kay Jagels, Sampa Leather, Sharon Moule, Halina Norberczak, Susan O'Neil, Doug Ormond, Claire Price, Ester Rabinowitsch, Simon Rutter, Mandy Sanders, David Saunders, Katherine Seeger, Sarah Sharp, Mark Simmonds, Jason Skelton, Robert Squares, Steven Squares, Kim Stevens, Louise Unwin, Sally Whitehead, Bart G. Barrell, Duncan J. Maskell, Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*, *Nature Genetics*, 35, 1, (2003), 32-40 <https://doi.org/10.1038/ng1227>
- [26] Johannes F. Imhoff, Phylogenetic taxonomy of the family *Chlorobiaceae* on the basis of 16S rRNA and *fmo* (Fenna-Matthews-Olson protein) gene sequences, *International Journal of Systematic and Evolutionary Microbiology*, 53, 4, (2003), 941-951 <https://doi.org/10.1099/ijs.0.02403-0>
- [27] Thomas Weissgerber, Renate Zigann, David Bruce, Yun-juan Chang, John C. Detter, Cliff Han, Loren Hauser, Cynthia D. Jeffries, Miriam Land, A. Christine Munk, Roxanne Tapia, Christiane Dahl, Complete genome sequence of *Allochrocatium vinosum* DSM 180T, *Standards in Genomic Sciences*, 5, 3, (2011), 311-330 <https://doi.org/10.4056/sigs.2335270>
- [28] Jaewoo Yoon, Kyoko Adachi, Sanghwa Park, Hiroaki Kasai, Akira Yokota, *Aureibacter tunicatorum* gen. nov., sp. nov., a marine bacterium isolated from a coral reef sea squirt, and description of *Flammeovirgaceae* fam. nov, *International Journal of Systematic and Evolutionary Microbiology*, 61, 10, (2011), 2342-2347 <https://doi.org/10.1099/ijs.0.027573-0>
- [29] Yanbing Lin, Haixin Fan, Xiuli Hao, Laurel Johnstone, Yao Hu, Gehong Wei, Heng A. Alwathnani, Gejiao Wang, Christopher Rensing, Draft Genome Sequence of *Halomonas* sp. Strain HAL1, a Moderately Halophilic Arsenite-Oxidizing Bacterium Isolated from Gold-Mine Soil, *Journal of Bacteriology*, 194, 1, (2012), 199-200 <https://doi.org/10.1128/jb.06359-11>
- [30] Lionel Frangeul, Philippe Quillardet, Anne-Marie Castets, Jean-François Humbert, Hans C. P.

- Matthijs, Diego Cortez, Andrew Tolonen, Cheng-Cai Zhang, Simonetta Gribaldo, Jan-Christoph Kehr, Yvonne Zilliges, Nadine Ziemert, Sven Becker, Emmanuel Talla, Amel Latifi, Alain Billault, Anthony Lepelletier, Elke Dittmann, Christiane Bouchier, Nicole Tandeau de Marsac, Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium, *BMC Genomics*, 9, 1, (2008), 274 <https://doi.org/10.1186/1471-2164-9-274>
- [31] Andrea I. Moreno Switt, Alexis D. Andrus, Matthew L. Ranieri, Renato H. Orsi, Reid Ivy, Henk C. den Bakker, Nicole H. Martin, Martin Wiedmann, Kathryn J. Boor, Genomic comparison of sporeforming bacilli isolated from milk, *BMC Genomics*, 15, 1, (2014), 26 <https://doi.org/10.1186/1471-2164-15-26>
- [32] M. Julia Althabegoiti, Luis Lozano, Gonzalo Torres-Tejerizo, Ernesto Ormeño-Orrillo, Marco A. Rogel, Víctor González, Esperanza Martínez-Romero, Genome Sequence of *Rhizobium grahamii* CCGE502, a Broad-Host-Range Symbiont with Low Nodulation Competitiveness in *Phaseolus vulgaris*, *Journal of Bacteriology*, 194, 23, (2012), 6651-6652 <https://doi.org/10.1128/jb.01785-12>
- [33] Meijie Li, Tongtong Zhu, Rumeng Yang, Zhaobao Wang, Min Liu, Jianming Yang, Carotenoids synthesis affects the salt tolerance mechanism of *Rhodospseudomonas palustris*, *Frontiers in Microbiology*, 14, (2023), 1292937 <https://doi.org/10.3389/fmicb.2023.1292937>
- [34] Xiaofeng Dai, Xiaochong Shi, Xin Gao, Jingli Liu, Xiao-Hua Zhang, *Roseivivaxmarinus* sp. nov., isolated from deep water, *International Journal of Systematic and Evolutionary Microbiology*, 64, Pt_8, (2014), 2540-2544 <https://doi.org/10.1099/ijs.0.062760-0>
- [35] Joseph Hair, Abdullah Alamer, Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example, *Research Methods in Applied Linguistics*, 1, 3, (2022), 100027 <https://doi.org/10.1016/j.rmal.2022.100027>
- [36] Jasmine Chong, David S. Wishart, Jianguo Xia, Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis, *Current Protocols in Bioinformatics*, 68, 1, (2019), e86 <https://doi.org/10.1002/cpbi.86>
- [37] Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Eric F. Pettersen, Gregory S. Couch, John H. Morris, Thomas E. Ferrin, UCSF ChimeraX: Meeting modern challenges in visualization and analysis, *Protein Science*, 27, 1, (2018), 14-25 <https://doi.org/10.1002/pro.3235>
- [38] Óscar Álvarez, Juan Luis Fernández-Martínez, Celia Fernández-Brillet, Ana Cernea, Zulima Fernández-Muñiz, Andrzej Kloczkowski, Principal component analysis in protein tertiary structure prediction, *Journal of Bioinformatics and Computational Biology*, 16, 02, (2018), 1850005 <https://doi.org/10.1142/S0219720018500051>
- [39] Cristina Sotomayor-Vivas, Enrique Hernández-Lemus, Rodrigo Dorantes-Gilardi, Linking protein structural and functional change to mutation using amino acid networks, *PLoS ONE*, 17, 1, (2022), e0261829 <https://doi.org/10.1371/journal.pone.0261829>
- [40] Anindya S. Panja, Smarajit Maiti, Bidyut Bandyopadhyay, Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges, *Scientific Reports*, 10, (2020), 1822 <https://doi.org/10.1038/s41598-020-58825-7>
- [41] Tobias J. Erb, Jan Zarzycki, A short history of RubisCO: the rise and fall (?) of Nature's predominant CO₂ fixing enzyme, *Current Opinion in Biotechnology*, 49, (2018), 100-107 <https://doi.org/10.1016/j.copbio.2017.07.017>
- [42] Hiroki Ashida, Yohtarō Saito, Toshihiro Nakano, Nicole Tandeau de Marsac, Agnieszka Sekowska, Antoine Danchin, Akiho Yokota, RuBisCO-like proteins as the enolase enzyme in the methionine salvage pathway: functional and evolutionary relationships between RuBisCO-like proteins and photosynthetic RuBisCO, *Journal of Experimental Botany*, 59, 7, (2008), 1543-1554 <https://doi.org/10.1093/jxb/ern104>
- [43] Moshe Mevarech, Felix Frolow, Lisa M. Gloss, Halophilic enzymes: proteins with a grain of salt, *Biophysical Chemistry*, 86, 2, (2000), 155-164 [https://doi.org/10.1016/S0301-4622\(00\)00126-5](https://doi.org/10.1016/S0301-4622(00)00126-5)
- [44] Christopher J. Reed, Hunter Lewis, Eric Trejo, Vern Winston, Caryn Evilia, Protein Adaptations in Archaeal Extremophiles, *Archaea*, 2013, (2013), 373275 <https://doi.org/10.1155/2013/373275>
- [45] Guangya Zhang, Ge huihua, Lin Yi, Stability of halophilic proteins: From dipeptide attributes to discrimination classifier, *International Journal of Biological Macromolecules*, 53, (2013), 1-6 <https://doi.org/10.1016/j.ijbiomac.2012.10.031>
- [46] Ram Karan, Melinda D. Capes, Shiladitya DasSarma, Function and biotechnology of extremophilic enzymes in low water activity, *Aquatic Biosystems*, 8, (2012), 4 <https://doi.org/10.1186/2046-9063-8-4>
- [47] K. Linda Britton, Patrick J. Baker, Martin Fisher, Sergey Ruzheinikov, D. James Gilmour, María-José Bonete, Juan Ferrer, Carmen Pire, Julia Esclapez, David W. Rice, Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloferax mediterranei*, *Proceedings of the National Academy of Sciences*, 103, 13, (2006), 4846-4851 <https://doi.org/10.1073/pnas.0508854103>
- [48] Akira Nomoto, Suguru Nishinami, Kentaro Shiraki, Solubility Parameters of Amino Acids on Liquid-Liquid Phase Separation and Aggregation of Proteins, *Frontiers in Cell and Developmental Biology*, 9, (2021), <https://doi.org/10.3389/fcell.2021.691052>
- [49] Renganathan Senthil, Singaravelu Usha, Konda Mani Saravanan, Importance of fluctuating amino acid residues in folding and binding of proteins, *Avicenna Journal of Medical Biotechnology*, 11, 4, (2019), 339-343
- [50] Lea Weinisch, Steffen Kühner, Robin Roth, Maria Grimm, Tamara Roth, Daili J. A. Netz, Antonio J. Pierik, Sabine Filker, Identification of osmoadaptive strategies in the halophile, heterotrophic ciliate *Schmidingerothrix salinarum*, *PLoS Biology*, 16, 1, (2018), e2003892 <https://doi.org/10.1371/journal.pbio.2003892>

- [51] C. Brininger, S. Spradlin, L. Cobani, C. Evilia, The more adaptive to change, the more likely you are to survive: Protein adaptation in extremophiles, *Seminars in Cell & Developmental Biology*, 84, (2018), 158-169
<https://doi.org/10.1016/j.semcdb.2017.12.016>
- [52] Dixita Chettri, Ashwani Kumar Verma, Lija Sarkar, Anil Kumar Verma, Role of extremophiles and their extremozymes in biorefinery process of lignocellulose degradation, *Extremophiles*, 25, (2021), 203-219 <https://doi.org/10.1007/s00792-021-01225-0>
- [53] Pu Ge, Gen Luo, Yi Luo, Wei Huang, Hongbin Xie, Jingwen Chen, Jingping Qu, Molecular understanding of the interaction of amino acids with sulfuric acid in the presence of water and the atmospheric implication, *Chemosphere*, 210, (2018), 215-223
<https://doi.org/10.1016/j.chemosphere.2018.07.014>
- [54] Atsushi Ikai, Thermostability and Aliphatic Index of Globular Proteins, *The Journal of Biochemistry*, 88, 6, (1980), 1895-1898
<https://doi.org/10.1093/oxfordjournals.jbchem.a133168>
- [55] Qingzhen Hou, Raphaël Bourgeas, Fabrizio Pucci, Marianne Rooman, Computational analysis of the amino acid interactions that promote or decrease protein solubility, *Scientific Reports*, 8, (2018), 14661
<https://doi.org/10.1038/s41598-018-32988-w>
- [56] Xiangyi Jiang, Ji Yu, Zhongxia Zhou, Jacob Kongsted, Yuning Song, Christophe Pannecouque, Erik De Clercq, Dongwei Kang, Vasanthanathan Poongavanam, Xinyong Liu, Peng Zhan, Molecular design opportunities presented by solvent-exposed regions of target proteins, *Medicinal Research Reviews*, 39, 6, (2019), 2194-2238
<https://doi.org/10.1002/med.21581>
- [57] Alessandro Siglioccolo, Alessandro Paiardini, Maria Piscitelli, Stefano Pascarella, Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface, *BMC Structural Biology*, 11, (2011), 50
<https://doi.org/10.1186/1472-6807-11-50>
- [58] David March, Valentino Bianco, Giancarlo Franzese, Protein Unfolding and Aggregation near a Hydrophobic Interface, *Polymers*, 13, 1, (2021), 156
<https://doi.org/10.3390/polym13010156>