

Publication Date: 23 April 2025

Archs Sci. (2025) Volume 75, Issue 1 Pages 8-10, Paper ID 2025202.

<https://doi.org/10.62227/as/75202>

A Comprehensive Guide to Cardiolipin Binding Residues of Different Organisms Using Python

Stecy Antony Selvakumar^{1,*} and Arul Mugilan²

¹Research Scholar, PG Research Department of Physics, Kamarajar Government Arts College, Surandai-627859, Tamil Nadu (India), Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu (India)

²Assistant Professor, PG Research Department of Physics, Kamarajar Government Arts College, Surandai-627859, Tamil Nadu (India), Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu (India)

Corresponding authors: (e-mail: astecy2705@gmail.com).

Abstract Cardiolipin, a type of phospholipid, is naturally produced within the mitochondrial membrane. It was initially discovered in animal hearts and is predominantly located on bacterial membranes. Several research has been made to understand the structure and function and their role inside the mitochondrial membrane. In this study, we present accurate details regarding the CL binding residues for various organisms namely Bos Taurus, Mus Musculus, Cereibacter sphaeroides and Methanosarcina acetivorans. To achieve this, we developed a Python program specifically designed for CL binding. We extracted data from the Protein data bank. The results show that polar aliphatic positively charged amino acids are highly preferred in CL. The structure of the binding residues with lipid was visualized with the help of the AVAGADRO tool.

Index Terms cardiolipin, python, PDB, binding residue, avagadro Tool

I. Introduction

Cardiolipin, also referred to as cardiolipid, is a specialized phospholipid abundantly present in the cell membrane. A distinctive characteristic of cardiolipin is its occurrence primarily within the mitochondria, the powerhouse of the cells [1], [2]. The constitution of CL consists of approximately 20 percent lipids. The study of CL is essential because its structural and chemical properties [3]–[5]. CL is a dimerized lipid composed of four fatty acid chains linked to a glycerol backbone with two phosphate groups having glycerol at its third position [2], [6], [7]. This unique composition makes cardiolipin to form double layered structure within the inner mitochondrial membrane [8], [9]. The presence of cardiolipin plays a crucial role in preserving the structural integrity and ensuring the optimal functioning of mitochondria [10], [11]. It plays a crucial role in facilitating the assembly of the electron transport chain and provides stability to the respiratory complexes involved in oxidative phosphorylation [12].

Cardiolipin, a vital molecule, actively participates in cellular signalling by effectively interacting with a wide range of proteins and enzymes [13]. Cardiolipin regulates mitochondrial fusion and fission through protein binding [14]. Cardiolipin is a vital cofactor for numerous enzymes responsible for lipid metabolism and signalling pathways. Dysfunction of cardiolipin can result in major disturbances in the electron transport chain, which plays a crucial role in the effective generation of energy in the mitochondria. This disturbance ultimately results in reduced energy production and a rise in oxidative stress. The consequences of these effects can

manifest as cardiac dysfunction, impacting the overall health and functioning of the heart [1]. In addition to its function as a structural component cardiolipin also plays a crucial role in various biochemical processes occurring within mitochondria [11]. When the metabolism of Cardiolipin is disrupted, it leads to harm the functional process of neurons which in turn contributes to neuronal injury [15], [16]. Cardiolipin dysfunction leads to neurological illnesses such as Alzheimer's disease and Parkinson's disease [13], [17].

To properly understand the mechanism of action and annotation of CL binding proteins, we must first identify the amino acid residues that interact with CL. For proteins that are still unknown, homology and sequence similarity-based approaches have shown to be ineffective. We created an CL binder technique that predicts CL binding proteins and interacting residues based solely on protein sequence, without requiring structural information.

II. Materials and Methods

The protein sequences needed for this study were obtained from the Protein database [18]–[20]. The selected sequences were found to be non-homologous and their structures were accurately determined at a high resolution. Our analysis focused on investigating the interactions between cardiolipin residues both interacting and non-interacting in order to gain a deeper understanding of lipid interactions [21], [22]. Residues are considered to be interacting if their atoms fall within a range of 3Å. The contact between residues and CL was accurately calculated using a Python program [23]. In this study

we had taken protein sequences with similarities between 30% and 90% and their differences were taken into an account. The data set with similar protein sequences was removed and the remaining sequences were retained.

After selecting the protein, a Python programme is written with a range less than 13 Å to find the bond length [24], [25]. It will display the outcomes of amino acids rapidly reacting with the Phospholipid CL. After separating the amino acids from the results, the data was statistically evaluated [26]. The probability of amino acid distribution was found out with the help of frequency of occurrence. The frequency of occurrence of amino acid for 458 proteins of both singlet and triplet groups was calculated using the formula (1),

$$P(x) = \frac{(\text{No of counts in particular AA})}{(\text{Total No of aminoacid in proteins})} \quad (1)$$

We used the values of the frequency of occurrence of singlet and triplet as expected count and observed count in this technique. The Deviation parameter for AA singlet and triplet in a particular structural element is computed using the formula (2),

$$DP = \frac{(\text{Observed count}-\text{Expected count})}{(\text{Expected count})} \quad (2)$$

This formula is used to determine the residue that preferentially binds to the CL as well as the amino acid that is in the "non-preferential" zone. A few proteins were chosen at random, and the procedure was carried out in the same way for all other proteins. The visualization shows the shortest distances between the amino acids that are interacting with the Phospholipid CL.

III. Result and Discussion

A. Graphical representation of the Chi-square test of various protein IDs with Phospholipid CL

The output of the program was extracted and presented graphically.

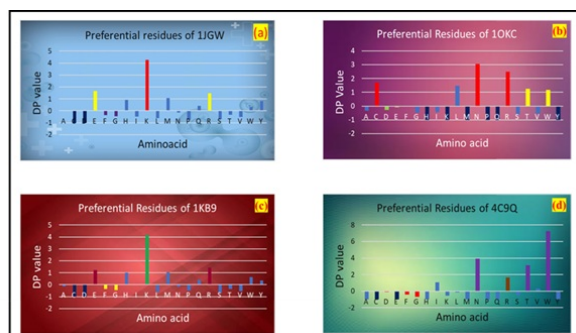


Figure 1: (a, b, c, d) represents the interactions of amino acids with CL in the preferential and non-preferential regions of Cereibacter sphaeroides, Bos Taurus, Mus Musculus, Methanosarcina acetivorans

The above findings were evaluated using the Deviation parameter, which aids in determining which amino acids are

favoured to bind with the ligand. We classified the distribution of amino acid into four groups namely Highly preferential, Low preferential, tends to prefer and out of range. For that we categorize the deviation parameter values into positive side and negative side. DP values in the +0.5 range are assumed to be in the preferential zone, whereas +1 or greater than 1 designates the extremely favourable sector. Similarly, DP values in the -0.5 range are considered low favoured zones, while values in the -1 range are considered out of range.

As shown in Fig for organisms like Cereibacter sphaeroides and Mus Musculus the residues such as K is the highly preferred one followed by W,H,N. Residues like E and R tends to be in the preferred zone where as residues phenylalanine and glycine have minimal probability to interact with CL and residues Cysteine and Aspartic acid are in the non-CL interacting site. For Bos Taurus residue N is highly preferred followed by R and C for organism Methanosarcina acetivorans W and N are highly preferred. Residue K being positively charged easily binds with the ligand CL. The presence of polar aliphatic positively charged amino acids was more prevalent in the CL-binding sites.

B. Interactions and structure of aminoacid with ligand using Avagadro tool

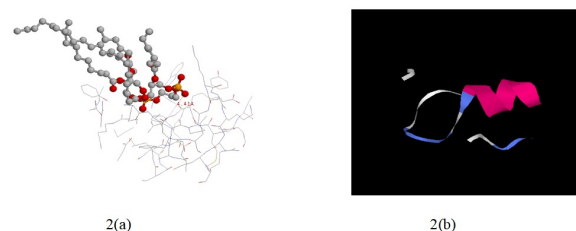


Figure 2: Shortest distance (4.41 Å) between Hetatm CL ligand and surrounding residues, with CL-bound proteins shown within 13 Å

The figure above illustrates the shortest distance between the Hetatm CL and the amino acid. In this representation, the ligand is depicted in ball and stick form, while the other residues are shown in stick form and the nearest distance of the ligand and residue is around 4.41 Å. The next picture displays the structure of CL binded proteins around 13 Å.

IV. Conclusion

In order to understand how CL works, we first need to know which amino acid residues bind to CL. We have created a Python program to find the binding residue. We chose Python because it is widely used and has a wide range of uses in biosciences. This method is easy and time-consuming, as it only involves a protein sequence. It allows us to perform and analyse multiple sequences at the same time, without any prior knowledge of structural data. The main finding is that the residues in the binding site of CL are highly conserved, and CL favours mixed combination of alpha helix and beta turn

structure and the residues like K,W,H,N has larger propensity to bind with CL. The structural analysis shows that the nearest distance bounded by Cardiolipin is around 4.41 Å.

Funding

This study was conducted with no funding.

Competing Interests

The authors declare that they have no conflict of interest.

Data availability statements

The data that support the findings of this study are available in Protein Data Bank at <https://doi.org/10.2210/pdbXXXX/pdb>

References

- [1] Claypool, S. M., & Koehler, C. M. (2012). The complexity of cardiolipin in health and disease. *Trends in Biochemical Sciences*, 37(1), 32-41.
- [2] LeCocq, J., & Ballou, C. E. (1964). On the structure of cardiolipin. *Biochemistry*, 3(7), 976-980.
- [3] Lewis, R. N., & McElhaney, R. N. (2009). The physicochemical properties of cardiolipin bilayers and cardiolipin-containing lipid membranes. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1788(10), 2069-2079.
- [4] Boyd, K. J., Alder, N. N., & May, E. R. (2018). Molecular dynamics analysis of cardiolipin and monolysocardiolipin on bilayer properties. *Biophysical Journal*, 114(9), 2116-2127.
- [5] Li, X. X., Tsoi, B., Li, Y. F., Kurihara, H., & He, R. R. (2015). Cardiolipin and its different properties in mitophagy and apoptosis. *Journal of Histochemistry & Cytochemistry*, 63(5), 301-311.
- [6] Schlame, M., Rua, D., & Greenberg, M. L. (2000). The biosynthesis and functional role of cardiolipin. *Progress in Lipid Research*, 39(3), 257-288.
- [7] Macfarlane, M. G. (1958). Structure of cardiolipin. *Nature*, 182(4640), 946-946.
- [8] Paradies, G., Paradies, V., De Benedictis, V., Ruggiero, F. M., & Petrosillo, G. (2014). Functional role of cardiolipin in mitochondrial bioenergetics. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1837(4), 408-417.
- [9] Unsay, J. D., Cosentino, K., Subburaj, Y., & García-Sáez, A. J. (2013). Cardiolipin effects on membrane structure and dynamics. *Langmuir*, 29(51), 15878-15887.
- [10] Paradies, G., Ruggiero, F. M., Petrosillo, G., & Quagliariello, E. (1997). Age-dependent decline in the cytochrome c oxidase activity in rat heart mitochondria: role of cardiolipin. *FEBS Letters*, 406(1-2), 136-138.
- [11] Chicco, A. J., & Sparagna, G. C. (2007). Role of cardiolipin alterations in mitochondrial dysfunction and disease. *American Journal of Physiology-Cell Physiology*, 292(1), C33-C44.
- [12] Gonzalez, F., & Gottlieb, E. (2007). Cardiolipin: setting the beat of apoptosis. *Apoptosis*, 12, 877-885.
- [13] Ren, M., Phoon, C. K., & Schlame, M. (2014). Metabolism and function of mitochondrial cardiolipin. *Progress in Lipid Research*, 55, 1-16.
- [14] Ikon, N., & Ryan, R. O. (2017). Cardiolipin and mitochondrial cristae organization. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1859(6), 1156-1163.
- [15] Pointer, C. B., & Klegeris, A. (2017). Cardiolipin in central nervous system physiology and pathology. *Cellular and Molecular Neurobiology*, 37, 1161-1172.
- [16] Kirkland, R. A., Adibhatla, R. M., Hatcher, J. F., & Franklin, J. L. (2002). Loss of cardiolipin and mitochondria during programmed neuronal death: evidence of a role for lipid peroxidation and autophagy. *Neuroscience*, 115(2), 587-602.
- [17] Chu, C. T., Bayır, H., & Kagan, V. E. (2014). LC3 binds externalized cardiolipin on injured mitochondria to signal mitophagy in neurons: implications for Parkinson disease. *Autophagy*, 10(2), 376-378.
- [18] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28(1), 235-242.
- [19] Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer Jr, E. F., Brice, M. D., Rodgers, J. R., ... & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112(3), 535-542.
- [20] Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., & Velankar, S. (2017). Protein Data Bank (PDB): the single global macromolecular structure archive. *Protein Crystallography: Methods and Protocols*, 627-641.
- [21] Chauhan, J. S., Mishra, N. K., & Raghava, G. P. (2010). Prediction of GTP interacting residues, dipeptides and tripeptides in a protein from its evolutionary information. *BMC Bioinformatics*, 11, 1-9.
- [22] Ansari, H. R., & Raghava, G. P. (2010). Identification of NAD interacting residues in proteins. *BMC Bioinformatics*, 11, 1-8.
- [23] Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., ... & De Hoon, M. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422.
- [24] Ekmekci, B., McAnany, C. E., & Mura, C. (2016). An introduction to programming for bioscientists: a Python-based primer. *PLoS Computational Biology*, 12(6), e1004867.
- [25] Bassi, S. (2007). A primer on python for life science researchers. *Plos Computational Biology*, 3(11), e199.
- [26] Aloy, P., Cedano, J., Oliva, B., Aviles, F. X., & Querol, E. (1997). 'Trans-Mem': a neural network implemented in Excel spreadsheets for predicting transmembrane domains of proteins. *Bioinformatics*, 13(3), 231-234.

...