International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 11 (2019), pp. 1935-1940
© International Research Publication House. http://www.irphouse.com

Computational Prediction of *Corynebacterium matruchotii* Protein's 3D Structure Reveals its Capacity to bind to DNA domain site in the Malaria Vector, Anopheles

Marion Olubunmi Adebiyi[1, 2], Damilare Falola[3], Okesola Olatunji[4], Roseline Oluwaseun Ogundokun[2] and Emmanuel A. Adeniyi[2]

[1.]Department of Computer and Information Sciences, Covenant University. Ota. Nigeria.

[2.]Department of Computer Science, Landmark University Omu Aran, Nigeria.

[3.]Friedrich Schiller Universitat, Jena.

[4.]Department of Computer and Information Sciences, First Technical University, Ibadan, Nigeria.

Corresponding authors emails: marion.adebiyi@covenantuniversity.edu.ng and ogundokun.roseline@lmu.edu.ng

Abstract

Background: The production of proteins in our body system is controlled by the presence of transcriptional regulators. Transcriptional regulation is a mechanism that allows the translation of selected proteins to stop when they are not required or surplus in the body system. Corynebacterium matruchotii is a redox-responsive transcriptional regulator. Corynebacterium matruchotii is a protein inferred from homology and has a very short sequence of length 86. This protein was implicated in the work of Adebiyi, 2014 where it was predicted to be an insecticide target (after estimating it's likely success in crystallization) i.e. estimating the protein's propensity to produce the diffraction quality crystal which was the measure used in selecting best target.

Objective: objective of the study is to computationally predict a best, quality 3D structure of Corynebaterium matruchotii protein out of many structures predicted, perform a functional analysis of the predicted structure to elucidate additional probable function of the protein in the Anopheles.

Materials and Method: C0E150 protein sequence was obtained from Uniprot databases and selection of template protein was concluded. As part of the method for this study, the comparative modeling tool MODELLER program and SWISS Model Server were deployed to generate the 3D structure, JMOL and VMD were used for predicted model surface visualization, while PROCHECK finally was employed to expunge the best model of all models and to perform the functional analysis.

Results: A total of three 3D protein structure of target protein were obtained from the comparative modeling analysis. Ramachandran plot was generated for each of the structures obtained. This exposed the detailed outline of the theoretically favoured regions, generously allows regions and the amino acid disallows regions. This was used to determine the significance of the best of the three structures by clearly revealing the structure from modeler with multiple alignment templates which has the best residue in the favourable region. It also unveiled the surface protein of the structure with distinct active site gathered with cavity, such active site was hypothesized as the binding site for DNA domain, which reveals its function as a protein that binds to DNA, thus modifying the DNA and regulates the gene expression in the malaria vector, Anopheles.

Conclusion: It was concluded that the study protein has a capacity to bind to DNA domain and this makes the protein very significant in the study's organism of interest.

Keywords: Corynebacterium matruchotti, Protein, Swiss Model, Transcriptional regulator, Homology, C0E150, Blast Analysis, Modeller, Procheck.

INTRODUCTION

The number of protein sequences has enlarged speedily in the past years. In spite of the happening that the X-ray crystallography is the principal method for the determination of protein structure, it is consuming time and succeeds only if appropriate circumstances for growing crystals are possible[1]. In this concern, three main methods of computational prediction for protein structure are used to establish three-dimensional (3-D) structure of a protein from its sequence which are homology, modeling, threading and de novo methods[1]. Functional characterization of a protein sequence is one of the most numerous problems in biology. This undertaking is usually facilitated by an exact three-dimensional (3D) structure of the studied protein[1].

Three-dimensional (3D) protein structures delivers valued insights into the molecular basis of protein function, permitting an effective design of experiments[2]. In the absence of an experimentally determined structure for a particular protein, comparative or homology modeling often provides a useful 3-D model for a protein that is associated to at least one recognized protein structure[3,4]. It is possible to have both experimental and computational structure for a protein and in-fact it is possible to have more than one of both experimental and computational structure for a protein as long as the structures give high quality in it structures. The PROCHECK

programs are advantageous for measuring the quality not only of protein structures in the process of being solved but also of prevailing structures and of those being modelled on identified structures[5].

Proteins are biochemical molecules comprising of polypeptides amalgamated by peptide bonds among the amino and carboxyl groups of amino acid residues. Corynebaterium matruchotii is a protein gathered from homology and has a very short sequence of length 86. Early investigations of Corynebaterium matruchoti were focused on its classification and its 72 abilities to calcify and co-aggregate with other oral bacteria[6]. Corynebacterium matruchoti is the sole Gram-positive genera found in the population of Anopheles mosquitoes[1,3,21]. This protein was implicated in the work of as predicted as an insecticide target, after estimating is likely success in crystallization i.e. estimating the protein's propensity to produce the diffraction quality crystal which was the measure used in selecting best target. In that work, the biochemical metabolic network of the Anopheles was computationally analyzed using the choke point and reaction without deviation concept algorithm to extract the essential reactions. 61 essential genes were implicated in the list of possible insecticidal targets elucidated by that analysis.

The motivation came from a previous work which involve the development of Pathway genome database (PGDB) for Anopheles gambiae (https://biocyc.org/organism-summary?object=ANO) where a graph based model was deployed to analyze the topology of metabolic network of this organism to determine essential enzymatic reactions in the network. A refined list of 61 potential insecticide candidates target was revealed. Interesting, the target protein of this study was one of these 61-protein insecticide target, which require further functional analysis to reveal the part it plays in the malaria vector.

Objective of the study is to computationally predict a quality 3D structure of Corynebaterium matruchotii protein, perform a functionally analysis of the predicted structure and reveal additional function of the protein in vectors.

This study deals with the computational prediction of Corynebaterium matruchotii Protein's 3D structure. The paper contains other parts such as materials and methods, results and discussion, conclusion, acknowledgement and references.

MATERIALS AND METHODS

This research work spans between March, 2016 and January, 2018 because of the HPC server which is not always available. The target protein (protein of interest) was first identified in March 2016, the sequence of the protein was obtained from UNIPROT, BLAST analysis was performed, then MODELLER program and SWISS MODEL server were deployed to build the model of the target protein.

The short abstract of this initial analysis was presented at the 3rd Covenant University International Conference on African Development Issues (CUICADI) in May, 2016. In January, 2017 the research was extended and began with the use of JMOL and VDM to do the surface visualization of the resulting structure. PROCHECK and Ramachandran plot analysis was deployed in June, 2017 for structure assessment and quality validation and functional analysis was completed at Covenant University Center for Research, Innovation and Development by authors in January, 2018.

Computational modeling of 3D structure of the Anophele's COE150 Protein:

The straight/standard protocol for homology modeling is the computational prediction of the tertiary or 3D structure of the protein of interest, which must have been sequenced. The prediction is based on using an experimentally determined homologous structure (templates protein) with high percentage of sequence identity when aligned with the target protein sequence. Modeling the target protein (protein of Interest) gives opportunity to explore methods that predicts how other molecules (such as drugs or other compounds) can interact with the protein thus elucidating more functional details of such protein.

There are established tools with complete protocols such as the SWISS-MODEL server[2,3,5,8,9,11]. Method in [7] was adapted and hybridize with other tool.

The Corynebacterium matruchotii protein sequence was taken from UNIPROT databases (ID code C0E150)[10]. The Corynebacterium matruchotii has a very short sequence, after a proper search on the database, only a few template sequences with high sequence identity were available. For surety and accuracy of results, four template sequences of average sequence identity was used in building the model. 1KQJ, 3N5N, 1KG2, 1KG7 (PDB code) were selected after blasting the query sequence[9,10,12]. The sequences of these four proteins together with the query sequence were passed into the MODELLER[3] program to build atom models of *Corynebacterium matruchotii*, C0E150 protein according to the methods described for comparative protein modeling. 1KQJ which has a sequence identity of 56% and resolution of 1.7A was chosen by MODELLER as the template sequence, and the model was built with it. Another model was also built using both the best template sequence (1KQJ) and the next best (1KG2) to get more accurate results and validate.

The sequence of C0E150 protein was loaded into SWISS MODEL SERVER[6], using its automated mode for building models.

Finally, JMOL and VMD were used as our model visualization tool. PROCHECK[5] was used to finally select the best model and to verify the stereo-chemical quality of the models.

RESULTS AND DISCUSSION

In-silico prediction result elucidates the target protein binding to DNA domain for gene expression and transcription regulation in the Anopheles.

The modeled structure of the target protein, C0E150 obtained by homology modeling has alpha helices, beta sheets and loops.

Ramachandran plots were generated to further elucidate the data points and distribution encompassed in each of the structures. Details of results obtained from PROCHECK for target protein C0E150 from the three methods are as discussed:

Predicted Structure and Ramachandran Plot from Modeler using Author's Template.

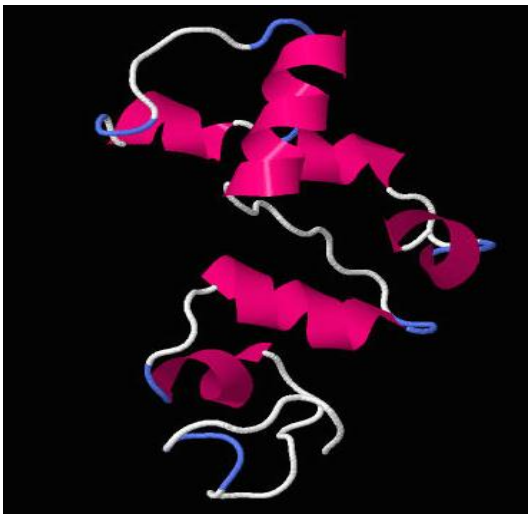

Figure 1(a): Structure from modeler using Author's template.

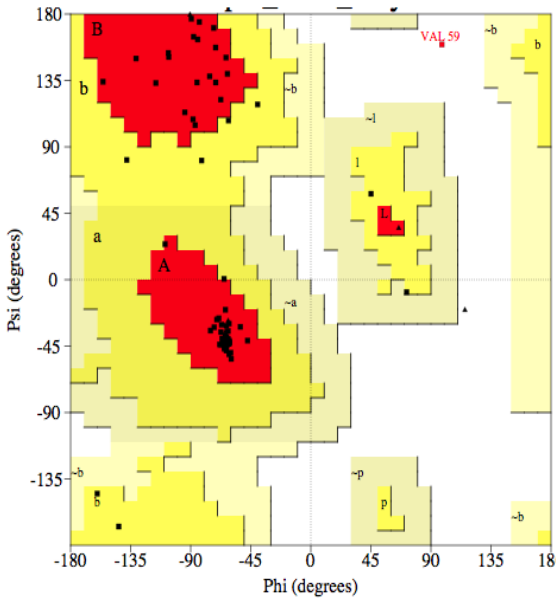

Figure 1(b): Ramachandran plot for Structure from modeler with Author's template.

The first structure from Modeler (Figure 1(a)) on Ramachandran plot, (Figure 1(b)) shows the amino acids in favorable position as 85.9%, and amino acid in disallowed region as 1.4%.

Predicted Structure and Ramachandran Plot from Modeler Program with Multiple Alignments Template.

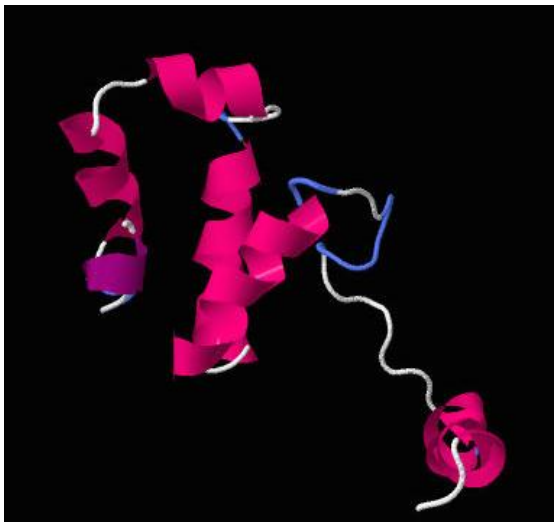

Figure 2(a): Structure from modeler program with multiple alignments template.

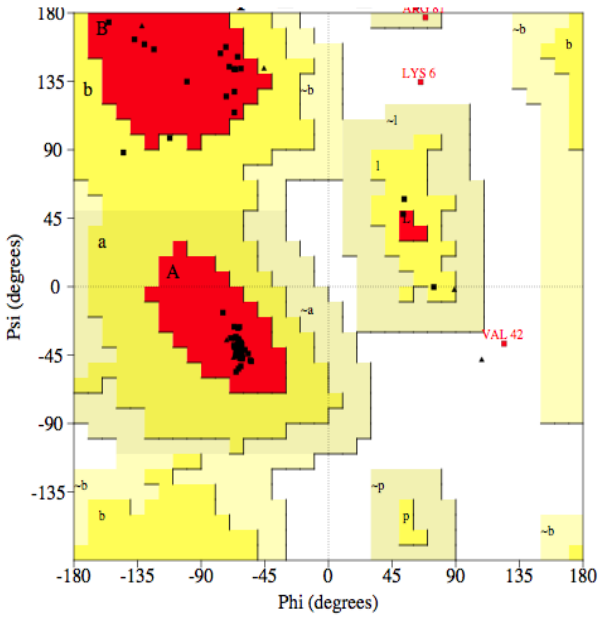

Figure 2(b): Ramachandran Plot for Structure from Modeler with Multiple Alignments template.

The second model structure (figure 2(a)) from Modeler with multiple alignments template, and its Ramachandran plot, figure 2(b) showed 90.3% for residues allowed in favorable region and 2.8% in disallowed region. This is significant as it distinctly identified Structure from Modeler with multiple alignments template, (Figure 2(a)) as the best predicted structure and structure with the highest quality because of the high residues deposited in the favorable region.

Predicted Structure and Ramachandran Plot from SWISS Model Server with SWISS-Selected Template.

Figure 3(a): Structure from SWISS MODEL server with the Swiss-selected template.

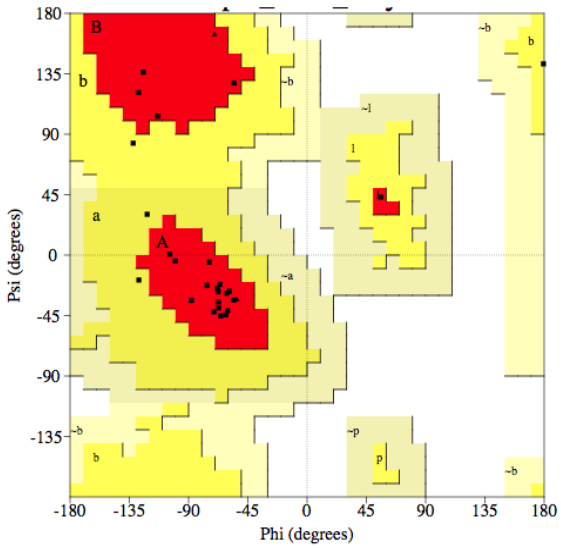

Figure 3(b): Ramachandran Plot for Structure from Swiss Model Server with the Swiss-selected template.

The model from Swiss server with SWISS template, Figure 3(a) was used to generate the Ramachandran plot in Figure 3(b) which shows residues in most favored region at 84% and 0% residues in disallowed region.

The generated Ramachandran plots (Figures 1b, 2b and 3b) revealed the backbone of the amino acid residue in each of the three structures. It also expands the empirical distribution of data points embedded in each of the single structure. JMOL and VDM visualized the backbone of the structure, a part that reflects the outline in the theoretically favored regions. Red, brown, and yellow regions represent the favored, disallowed, and generously allowed regions respectively as defined by ProCheck tool.

Surface Protein of the Structure with the Best Quality

Figure 2(a) is predicted to be the structure with the highest quality because its show an amino acid residue deposition in allowed and favorable region of up to 90.3% with surface of protein with cavity represented as the red edges while the active site is revealed with the green colored spottings.

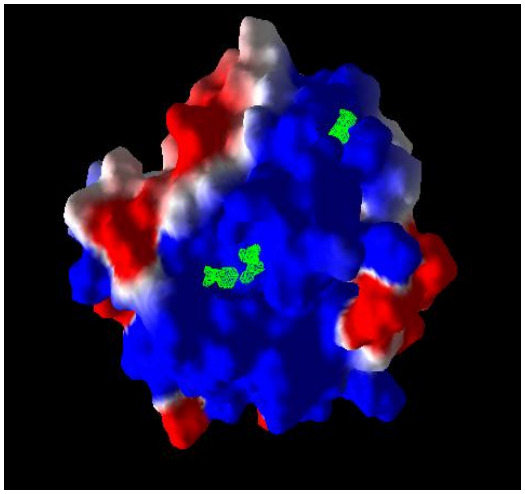

Figure 4: Surface of protein with cavity

Figure 4 depicts the surface protein of the predicted structure for target protein C0E150, which is the best structure selected and represented in Figure 2(a). This active site is hypothesized to be the predicted probable biding site for DNA as confirmed in other literature and UNIPROT server functions[13].

Justification of Result

The following are the literature review that supports the findings in this study.

The protein was revealed to be involved in enzymatic function that catalyzes multi-stage reactions[14].

These proteins are part of normal flora in vertebrates; they are contaminant in clinical specimen[15] which shows their activity in DNA contamination or repairs, this is an activity that happens when they are down regulated or up regulated[16,17].

This protein's involve in hydrolase activation[17]. This has to do with gene expression in the DNA domain. It could bind to DNA as redox-responsive transcriptional regulator and/ or DNA repairs. It is characterized by its "whip handle" appearance on gram staining an evidence of a characteristic of catalytic activity [16,17].

Future Recommendation

Finally, since the protein has been established an essential catalytic activity enzyme[16,18]. Further work may involve subjecting the encoding sequence of this protein to a position specific iteration BLAST (PSI-BLASTp) against the human genome database to eliminate target, in case it is homologous to non-target organism.

Significant Statement

This study discovered the best quality 3D structure in-silico for the Anopheles' corynebacterium maturchotii protein. It also elucidates the surface of protein with cavity and active site binding to DNA domain. This can be beneficial for gene regulation in the malarial vector, Anopheles. This study will help researchers to uncover the critical aspect of the gene function and important insecticidal target in the Anopheles that many researchers have not paid attention to. This suggests if there is the possibility of eliminating these gene in the mosquito, attention will be paid to predicting chemical compounds as insecticide, that is capable of eliminating the gene in Anopheles without destructing the same gene in the Human host.

CONCLUSION

This protein (*Corynebactarium matruchotii*) was implicated with the capacity to bind to DNA domain by its surface cavity; this makes the protein very significant in the organism's (Anopheles) metabolic analysis. This is essential for biochemical compounds combination and recommendation for insecticidal target development and design.

ACKNOWLEDGEMENT

We sincerely appreciate Landmark University for the funding of this article publication and Covenant University for the platform and resources provided by this great citadel, which made this project a possibility.

CONFLICT OF INTEREST

The authors declared that the received funding does not lead to any conflict of interest.

REFERENCES

[1] Shokry A. M., Ismail M. A., Yassin H. M., Mostafa S. A., Salama M. S. & Shahin M. A., 2012. Bioinformatics Analysis using Homology Modeling to Predict the Three-Dimensional Structure of Spodoptera Littoralis (Lepidoptera: Noctuidae) Aminopeptidase N Receptor. Egypt. J. Genet. Cytol., 41: 269-284.

[2] Schwede T., Kopp J., Guex N., & Peitsch M. C. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research, Vol. 31, No. 13, 3381–3385. DOI: 10.1093/nar/gkg520

[3] Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen M., Pieper, U., & Sali A., 2007. Comparative Protein Structure Modeling UNIT 2.9 Using MODELLER. Current Protocols in Protein Science, Unit 2.9

[4] Eswar, N., M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper and A. Sali, 2006. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Supplement 15: 5.6.1-5.6.30.

[5] Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton, 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl.Cryst., 26: 283-291.

[6] Luong T. T., Tirgar R., Reardon –Robinson M., Joachimiak A., Osipiuk J., & Ton –That H., 2018. Structural basis of a thiol -disulfide oxidoredu ctase in the hedgehog –forming actinobacterium Corynebacterium matruchotii. Journal of Bacteriology, doi:10.1128/JB.00783-17

[7] Singh, S., kumar A, Patel A., Tripathi A., kumar, D., and Verma M., 2010. In Silico 3d Structure Prediction and Comparison of Nucleocapsid Protein of H1N1. Journal of Modelling and Simulation of Systems, 1(2): 108-111

[8] Fulton L., Clifton, S., Chinwalla, A. T., Mitreva, M., Sodergren, E., Weinstock, G., Clifton, S., Dooling, D. J. Fulton, B., Minx, P., Pepin, K. H. Johnson, M., Bhonagiri, V., Nash, W. E., Mardis, E. R. and Wilson, R. K. Submitted (Sept. 2018) to the EMBL/GenBank/DDBJ databases

[9] Guex, N., 1996. Swiss-PdbViewer: A new fast and easy to use PDB viewer for the PC. Experientia, 52: A26.

[10] Kiefer, F., K. Arnold, M. Künzli, L. Bordoli, and T.Schwede, 2009. The SWISS-MODEL Repositoryand associated resources. Nucleic Acids Res. 37: D387-D392.

[11] http://molegro-virtual-docker-2012.software.informer.com

[12] PDB; http://www.rcsb.org/pdb/

[13] Kanehisa, M. and Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28: 27-30

[14] https://www.uniprot.org/keywords/KW-0378*:* GO:0016787

[15] Denise F. Bratcher, 2012., *Principles and Practice of Pediatric Infectious Diseases (Fourth Edition)*

[16] https://www.ebi.ac.uk/QuickGO/term/GO:0003674.

[17] Rose Kim, Annette C. Reboli, 2015. in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition)*

[18] Adebiyi M. O., 2014. Computational Analysis of Anopheles gambiae metabolism to facilitate insecticidal target and complex resistance mechanism discovery. Covenant University repository.

[19] http://www.uniprot.org/uniprot/C0E15

[20] Kanehisa, M., S. Goto, Y. Sato, M. Furumichi, and M. Tanabe, 2012. KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Res. 40: D109-D114.

[21] Slyvie M, Chung T.N, Krajan T, Waraporn J, Theeraphap C, Anne-Laure M and Estelle J., 2013. Bacrerial Biodivesity in Midguts of Anopheles Mosquitoes, Malaria Vectors in Southeast Asia. http://dx.doi.org/10.5772/55610