Comparative Analysis of Carbonic Anhydrase with Reference to *Anopheles gambiae* – A Vector of Malaria and its Homology Model Prediction

Nathan Vinod Kumar

School of Chemical and Biotechnology
SASTRA Deemed University
Thanjavur 613401, Tamil Nadu, India
E-mail: nvkibt@gmail.com

Received: July 24, 2018 **Accepted: March 21, 2020**

Published: September 30, 2020

Abstract: *Carbonic anhydrase (CA) is widely distributed enzyme which plays a crucial role in hydration of carbon dioxide. CA of Anopheles gambiae, a common vector that transmits the malaria disease in tropical countries is not well studies. CA is essential for the normal physiological function and inhibition of it can lead to lethal effect. Recently, CA was identified as a potential drug target for developing mosquito larvicidal drugs. CA of A. gambiae was characterized and compared with 12 other members of the class insecta. CA of A. gambiae comprised of 309 amino acid residues with 34.56 kDa and theoretical pI of 6.34. The Leu rich protein had more of negatively charged residues and was classified as a stable protein. Mitochondrial targeting peptide containing enzyme had about 46.28% random coils, 28.16% α-helix and 3.24% β-turns. The homology model for the CA of A. gambiae was predicted using template 4lu3.1.A and the model quality was validated by least RMSD score, Z-score and Ramachandran plot analysis. The present study is a preliminary report exploring proteomics aspects of CA of A. gambiae using in silico tools.*

Keywords: *Carbonic anhydrase, Homology model, Protein structure.*

Introduction

The highly anthropophilic mosquito, *Anopheles gambiae*, is the vector of the malarial parasite affecting 500 million people worldwide [15]. Most mosquito larvae have a remarkable capacity for ionic and osmotic regulation when compared to other organisms in the animal kingdom. Their ability to survive under extreme conditions has been attributed to strong transport processes that take place in the anal papillae as well as the midgut epithelium [7]. Due to this mechanism, there is a high challenge in the process of its eradication. There are various chemicals widely used to kill the vectors, however, their toxicity is high and they also contribute to the bioaccumulation phenomena in the environment.

Recently, there are many drugs developed based on enzyme inhibition mechanism which are crucial for normal physiology. Carbonic anhydrase (CA) is one such enzyme ubiquitously found in all organisms, also in mosquitoes, which could act as a potential drug target for its irradiation. Hence, in depth understanding of the variations among CA among different species is required. CA is necessary for mosquito larval alkalization and survival [9]. The role of CA in the alkalization mechanism of the *Aedes aegypti* larval midgut was already demonstrated. It is crucial in the maintenance of pH within the midgut of the mosquitoes [9]. Two *α*-CAs have been characterized from the mosquito *A. gambiae*. CA was detected within the epithelial cells of the mosquito larvae in both *A. gambiae* and *Aedes aegypti* using various methods [8-9, 18, 21, 23]. However, the specific roles of each CA member are unknown.

CA localization by Hansson's histochemical method revealed that the enzyme is present in different regions of the midgut according to each particular species. Even species closely related such as *Ae. aegypti* and *Ae. albopictus* displayed different distribution of CA in the midgut [9]. Since the development of methazolamide and acetazolamide, novel CA inhibitors more potent than these two drugs have been developed [20]. Perhaps by taking advantage of the alkaline pH inside the midgut as well as of these novel inhibitors it might be possible to develop formulations containing CA inhibitors that become active only at very high pH.

Since different CA isozymes are present in living organisms, an effective mosquito larvicide based solely on CA inhibition would have to selectively target the mosquito CA without affecting other species. In order to develop such larvicides, further studies are necessary to have a better understanding of the structure of the mosquito CA and the environment that surrounds it [9]. Hence this paper focuses on the physico-chemical and structural characterization of CA of *A. gambiae* using *in silico* tools and compare the features with other CA sequence of class Insecta.

Materials and methods

Sequence retrieval and phylogenetic analysis

CA sequence of *A. gambiae* was retrieved from the NCBI databank in the FASTA format. The BlastP was performed to obtain about 12 similar protein sequences of same and different Genera. The obtained sequence was subjected to Clustal W multiple sequence alignment using BioEdit5.0 [26]. Phylogenetic relations among the selected CA of *A. gambiae* and other sequences were analyzed based on the Neighbor-joining method [19] using MEGA 4.0 [25]. Midpoint root was done and the confidence level was analyzed using bootstrap of 1000 replications.

Physicochemical properties

Physicochemical parameters of the retrieved sequences were carried out using the ExPASy server tool ProtParam [12]. ProtParam software analysis was done to understand about the amino acid composition, molecular weight, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Prediction of subcellular localization was carried out using the TargetP webserver [11] (http://www.cbs.dtu.dk/services/TargetP/).

Prediction of CA secondary structure

Secondary structure of CA of *A. gambiae* was predicted using the ExPASy server [12]. The protein confirmation depends on the number of helix, sheet, and turn of amino acid sequences in the secondary structure. Self-optimized prediction method with alignment (SOPMA) was performed to understand the presence of helices, beta turns and coils in the protein structure. The presence of the helix, sheet, and turn in various CA sequence of *A. gambiae* were predicted using secondary structure prediction (PSIPRED) [6].

Homology modelling and validation

Homology model of CA of *A. gambiae* was predicted using SWISS-MODEL server [2-5, 13, 28]. The template identification was carried out to obtain a suitable template for homology modelling. The quality of the models predicted were evaluated using VMD 1. 9.1 software [14]. Root-mean-square deviation (RMSD) values were calculated using the RMSD calculator and the best homology model was predicted. Further, it was analyzed using PROSA Web server [22, 29]. Ramachandran plot for the model was developed using RAMPAGE software [17].

Results and discussion

Phylogenetic analysis using Neighbor-joining method

CA of *A. gambiae* was retrieved from NCBI databank. Using this as a query, 12 CA sequences of class Insecta were obtained by performing BlastP. Phylogenetic tree was constructed following the Multiple sequence alignment. Fig. 1 depicts the Neighbor-joining tree among the CA sequence of class insect. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Although, it is to be noted that bootstrap values provide a confidence interval that contains the phylogeny estimated from the repeated sampling through bootstrap analysis. Rather than measuring the accuracy of the data, bootstrap values represent the repeatability of the data that resulted after 1000 replications. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 13 amino acid (AA) sequences (12 CA sequences representative of class Insecta and the CA sequence of *A. gambiae*). CA sequences of *Drosophila sp*. were the most primitive whereas, CA of *Aedes sp*. were most divergent. CA sequences of *Anopheles sp.* were neighbours and formed the divergent part of the tree.

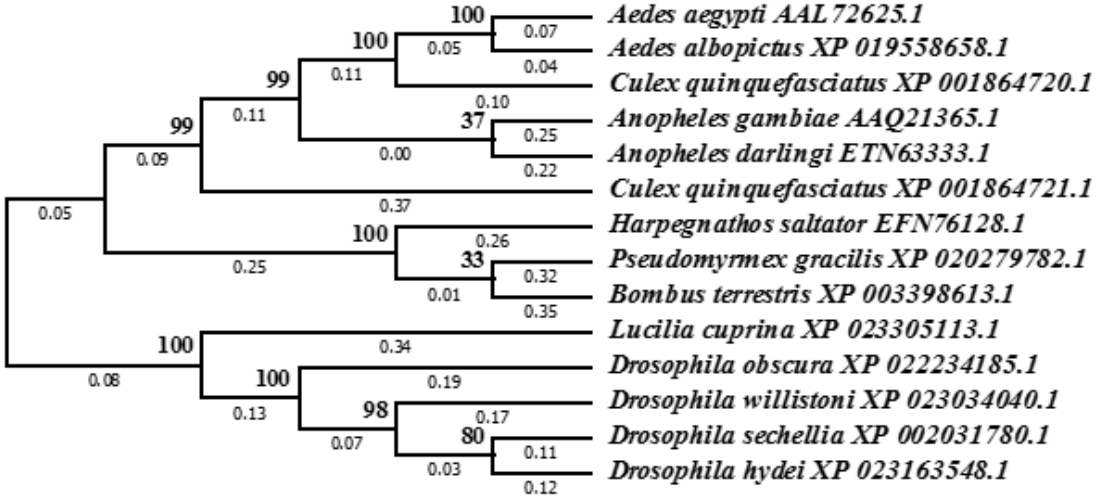

Fig. 1 Neighbor-joining tree showing evolutionary relationship among CA sequences of *A. gambiae* using MEGA 4.0.
Boot strap values are depicted at the nodes and branch lengths are also shown.

Physico-chemical parameters of CA

While comparing the physico-chemical parameters of the CA of the 13 sequences, there were many variations observed among the sequences (Table 1). The number of amino acid residues varied from 293 to 312. CA sequence of *A. gambiae* comprised of 309 amino acid residues with a molecular weight of 34.56 kDa and theoretical pI of 6.34. However, the molecular weight of all CA sequences was in the range of 32.71 to 34.87 kDa. Similarly, pI was also found in the range of 5.12 to 6.45. But, in case of the 13 CA sequences used in the study, pI was in acidic range. Total number of negatively charged residues (Asp + Glu) and positively charged residues (Arg + Lys) were calculated for the 13 sequences.

Table 1. Physicochemical characterization of CA sequences

Organism	AA	MW	pI	Total number of residues		Instability index (II)	Stability status	Aliphatic index	GRAVY	Dominant residues	
				– [a]	+ [b]					Residues	%
Anopheles gambiae	309	34.56	6.34	29	25	36.78	stable	86.44	-0.212	Leu	10.7
Anopheles darlingi	312	34.00	6.45	26	23	40.11	unstable	102.79	0.072	Leu	10.7
Aedes aegypti	298	32.71	6.14	28	22	35.74	stable	87.72	-0.167	Leu	9.7
Aedes albopictus	307	33.68	6.18	28	22	34.68	stable	93.06	-0.121	Leu	11.7
Culex quinque-fasciatus	320	34.87	6.08	30	23	34.43	stable	93.56	0.003	Leu	11.6
Drosophila obscura	304	34.07	5.32	36	23	33.17	stable	93.62	-0.176	Leu	10.9
Lucilia cuprina	305	34.72	5.12	40	27	33.48	stable	90.46	-0.275	ILe	8.2
Drosophila sechellia	304	33.91	5.22	35	21	36.80	stable	93.59	-0.121	Leu	10.2
Drosophila hydei	303	33.75	5.55	31	20	37.17	stable	94.26	-0.202	Leu	10.9
Pseudomyrmex gracilis	293	33.18	5.83	32	26	32.92	stable	99.42	-0.140	Leu	10.6
Drosophila willistoni	305	33.86	5.23	33	20	39.09	stable	94.62	-0.118	Leu	11.1
Bombus terrestris	296	33.25	6.29	26	21	29.68	stable	95.54	-0.140	Leu	11.1
Harpegnathos saltator	298	33.50	6.34	29	25	29.34	stable	89.16	-0.158	Val, Ser	10.1

[a] – negative; [b] – positive.

CA sequence of *A. gambiae* had about 29 negatively charged residues, whereas only 25 positively charged residues were present. In all the sequences, the number of negatively charged residues was high in number than the positively charged residues. Instability index (II) was calculated and all the CA sequence except that of *A. darlingi* were classified as stable protein. The CA sequences had a higher aliphatic index and ranged from 86.44 to 102.79. Moreover, the GRAVY values was negative in all species studied. It was notable that the bacterial strains had lower GRAVY values indicating the better possibilities of aqueous interaction [16]. The majority of the CA sequences were Leu rich (9.7-11.7%), however, the CA of *Lucilia cuprina* was Isoleucine rich (8.2%) and *Harpegnathos saltator* was Val and Ser rich (10.1%).

Subcellular localization prediction

Based on the TargetP v1.1 prediction, the CA of *A. gambiae* was classified as mitochondrial targeting peptide (mTP). *β*-CAs have been identified in the mitochondria of a variety of different organisms, such as plants, green algae, fungi [1, 10] and *Drosophila melanogaster* [24]. In mTPs, arginine, alanine and serine are over-represented, while negatively charged amino acid residues (Asp and Glu) are rare. Furthermore, the mTPs are believed to form an amphiphilic *α*-helix, which is important for the import of the nascent protein into the

mitochondrion [11]. Parasites rely on complex metabolic systems to satisfy their lipid needs. The present findings open a new avenue to investigate whether mitochondrial *β*-CAs are functionally involved in these processes. The single *β*-CA of *Anopheles darlingi* is the first predicted secretory *β*-CA.

Secondary structure of CA

Comparison of secondary structure components obtained from SOPMA revealed that there were variations in the values of each component. In 13 CA sequence analysed, random coils were the dominant component found in the range of 43.24-53.29% (Table 2). Similarly, the beta turns were present in lowest percentage (1.97-5.37%). In case of *Anopheles gambiae*, the CA sequence comprised of 28.16% alpha-helix, 22.33% extended strands, 3.24% beta turns and 46.28% of random coils. The graphical representation of the secondary structure of CA of *A. gambiae* is depicted in Fig. 2. This was not even comparable with the structure of *Anopheles darlingi* CA though belonging to the same genus. *Anopheles darling* had a higher percentage of beta turns and random coils and lower percentages of *α*-helix and extended strands when compared to CA of *A. gambiae.* Though studies reported that *α*-helices are most stable element in protein folding and are the first element which formed during protein folding [27], all the CA proteins studied were classified stable.

Table 2. Percentage of amino acids sequence forming secondary structure in SOPMA prediction

Organisms	***α*-helix**	**Extended strand**	**Beta turn**	**Random coils**
Anopheles gambiae	28.16	22.33	3.24	46.28
Anopheles darlingi	24.68	22.44	4.17	48.72
Aedes aegypti	24.16	24.50	4.03	47.32
Aedes albopictus	27.69	20.52	2.93	48.86
Culex quinquefasciatus	22.19	24.69	4.06	49.06
Drosophila obscura	26.64	24.34	2.96	46.05
Lucilia cuprina	25.57	22.95	1.97	49.51
Drosophila sechellia	23.03	19.08	4.61	53.29
Drosophila hydei	20.13	22.44	4.95	52.48
Pseudomyrmex gracilis	28.33	22.87	4.10	44.71
Drosophila willistoni	30.49	20.66	3.93	44.92
Bombus terrestris	25.34	27.36	4.05	43.24
Harpegnathos saltator	20.18	21.48	5.37	52.35

Homology model of CA

Homology model of CA of *A. gambiae* was predicted using SWISS-MODEL server and the model quality was evaluated. Among the 3 templates (3ryv.1.A, 4lu3.1.A and 4ygj.1.A) used, the model predicted using the template 4lu3.1.A had a least Z-score of -7.81 based on prosa web analysis (Fig. 3). This was also supported by the least RMSD score from VMD analysis. Ramachandran plot was depicted for the 3 models, and for the model obtained from the template 4lu3.1.A, 98.1% of the residues were in favoured region and 1.9% in the allowed region. No residues were found in the outlier region. While other models had only 96.5% residues within the favoured region. The models obtained were visualized using swiss-pdb viewer and depicted in fig (Fig. 3e).

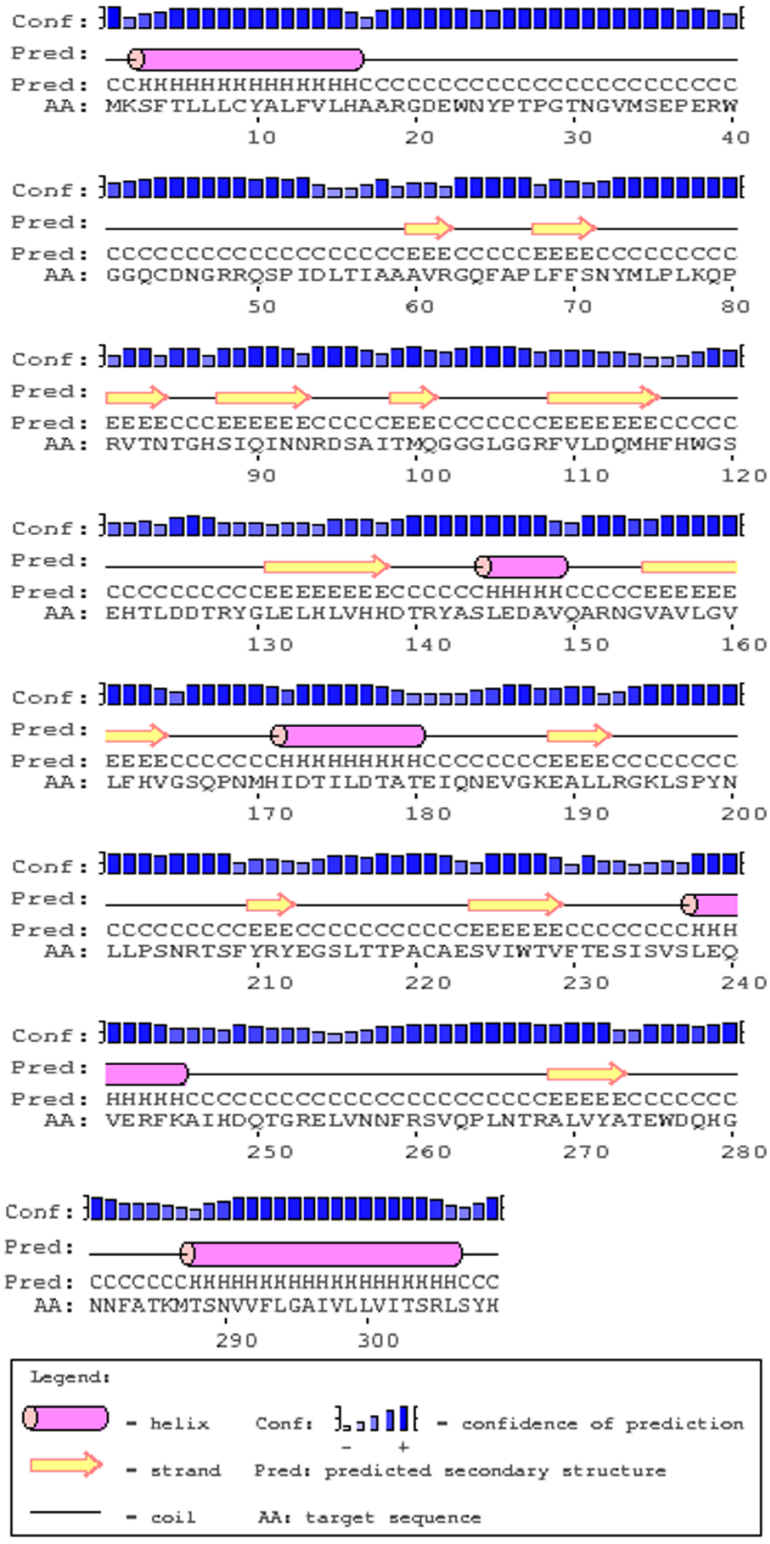

Fig. 2 Predicted secondary structure of CA sequences of *A. gambiae*

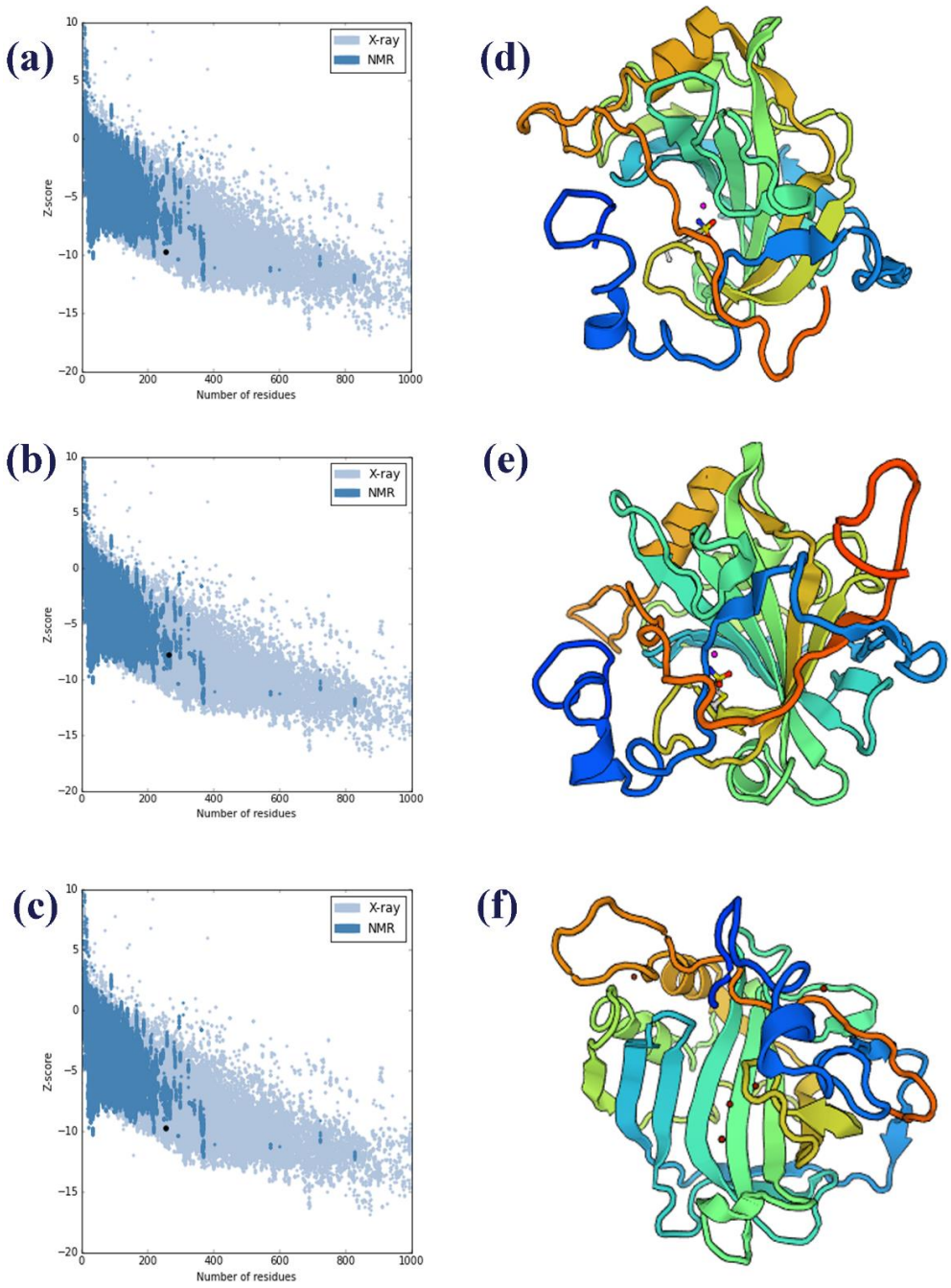

Fig. 3 Homology model of CA sequences of *A. gambiae* predicted using SWISS-MODEL and the ProSA web quality analysis of the model. Best model predicted was using the template 4lu3.1.A;
a) and d) for model generated from template 3ryv.1.A;
b) and e) for model generated from template 4lu3.1.A;
c) and f) for model generated from template 4ygj.1.A.

Conclusion

This paper elucidated the structure and physico-chemical properties of CA of *A. gambiae*, a vector that transmits the Malaria disease. The Leu rich, stable protein had dominant random coils. The activity was stable in a wide range of temperature due to higher aliphatic index. The structure predicted could help in developing a CA inhibitor based drug development as a

larvicide. Though a preliminary approach, this could lead to better larvicidal drug with less toxic effect to the environment and could replace the use of chemicals that contribute to the bio-magnification effect.

Acknowledgements

NVK is grateful to Dr. Parvathi Ammini CSIR-NIO (RC), Kochi, for her support. He also acknowledges the financial support of Science and Engineering Board (SERB), Government of India through National Post-doctoral Fellowship (PDF/2016/000438).

References

1. Bahn Y. S., F. A. Mühlschlegel (2006). CO_2 Sensing in Fungi and Beyond, Current Opinion in Microbiology, 9(6), 572-578.
2. Benkert P., M. Biasini, T. Schwede (2010). Toward the Estimation of the Absolute Quality of Individual Protein Structure Models, Bioinformatics, 27(3), 343-350.
3. Bertoni M., F. Kiefer, M. Biasini, L. Bordoli, T. Schwede (2017). Modeling Protein Quaternary Structure of Homo- and Hetero-oligomers beyond Binary Interactions by Homology, Scientific Reports, 7(1), 10480, doi: 10.1038/s41598-017-09654-8.
4. Biasini M., S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli, T. Schwede (2014). SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information, Nucleic Acids Research, 42(W1), W252-W258.
5. Bienert S., A. Waterhouse, T. A. de Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede (2016). The SWISS-MODEL Repository – New Features and Functionality, Nucleic Acids Research, 45(D1), D313-D319.
6. Buchan D. W., F. Minneci, T. C. Nugent, K. Bryson, D. T. Jones (2013). Scalable Web Services for the PSIPRED Protein Analysis Workbench, Nucleic Acids Research, 41(W1), W349-W357.
7. Clements A. N. (1992). The Biology of Mosquitoes, Chapman & Hall Publ, London, UK.
8. del Pilar Corena M., T. J. Seron, H. K. Lehman, J. D. Ochrietor, A. Kohn, C. Tu, P. J. Linser (2002). Carbonic Anhydrase in the Midgut of Larval *Aedes aegypti*: Cloning, Localization and Inhibition, Journal of Experimental Biology, 205(5), 591-602.
9. del Pilar Corena M., M. M. Fiedler, L. VanEkeris, C. Tu, D. N. Silverman, P. J. Linser (2004). Alkalization of Larval Mosquito Midgut and the Role of Carbonic Anhydrase in Different Species of Mosquitoes, Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 137(3), 207-225.
10. Elleuche S., S. Pöggeler (2010). Carbonic Anhydrases in Fungi, Microbiology, 156(1), 23-29.
11. Emanuelsson O., H. Nielsen, S. Brunak, G. Von Heijne (2000). Predicting Subcellular Localization of Proteins Based on Their N-terminal Amino Acid Sequence, Journal of Molecular Biology, 300(4), 1005-1016.
12. Gasteiger E., C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, A. Bairoch (2005). Protein Identification and Analysis Tools on the ExPASy Server, In: The Proteomics Protocols Handbook, Humana Press, 571-607.
13. Guex N., M. C. Peitsch, T. Schwede (2009). Automated Comparative Protein Structure Modeling with SWISS-MODEL and Swiss-PdbViewer: A Historical Perspective, Electrophoresis, 30(S1), S162-S173.
14. Humphrey W., A. Dalke, K. Schulten (1996). VMD: Visual Molecular Dynamics, Journal of Molecular Graphics, 14(1), 33-38.

15. Janairo J. I. B., P. I. K. Bravo, N. L. G. Morano, D. E. C. Yu (2016). *In silico* Site-directed Mutagenesis of the *Anopheles gambiae* Odorant Binding Protein 20, Nova Biotechnologica et Chimica, 15(2), 156-165.
16. Kumar N. V., M. E. Rani, R. Gunaseeli, N. D. Kannan, J. Sridhar (2012). Modeling and Structural Analysis of Cellulases Using *Clostridium thermocellum* as Template, Bioinformation, 8(22), 1105-1110.
17. Lovell S. C., I. W. Davis, W. B. Arendall, P. I. De Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, D. C. Richardson (2003). Structure Validation by Cα Geometry: ϕ, ψ and Cβ Deviation, Proteins: Structure, Function, and Bioinformatics, 50(3), 437-450.
18. Oviedo M. N., L. Vanekeris, M. D. P. Corena-Mcleod, P. J. Linser (2008). A Microarray-based Analysis of Transcriptional Compartmentalization in the Alimentary Canal of *Anopheles gambiae* (Diptera: Culicidae) Larvae, Insect Molecular Biology, 17(1), 61-72.
19. Saitou N., M. Nei (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees, Molecular Biology and Evolution, 4(4), 406-425.
20. Scozzafava A., M. D. Banciu, A. Popescu, C. T. Supuran (2000). Carbonic Anhydrase Inhibitors: Synthesis of Schiff Bases of hydroxybenzaldehydes with Aromatic Sulfonamides and Their Reactions with Arylsulfonyl Isocyanates, Journal of Enzyme Inhibition, 15(6), 533-546.
21. Seron T. J., J. Hill, P. J. Linser (2004). A GPI-linked Carbonic Anhydrase Expressed in the Larval Mosquito Midgut, Journal of Experimental Biology, 207(26), 4559-4572.
22. Sippl M. J. (1993). Recognition of Errors in Three-dimensional Structures of Proteins, Proteins: Structure, Function, and Bioinformatics, 17(4), 355-362.
23. Smith K. E., L. A. VanEkeris, P. J. Linser (2007). Cloning and Characterization of AgCA9, a Novel α-carbonic Anhydrase from *Anopheles gambiae* Giles Sensu Stricto (Diptera: Culicidae) Larvae, Journal of Experimental Biology, 210(22), 3919-3930.
24. Syrjänen L., M. Tolvanen, M. Hilvo, A. Olatubosun, A. Innocenti, A. Scozzafava, J. Leppiniemi, B. Niederhauser, V. P. Hytönen, T. A. Gorr, S. Parkkila (2010). Characterization of the First Beta-class Carbonic Anhydrase from an Arthropod (*Drosophila melanogaster*) and Phylogenetic Analysis of Beta-class Carbonic Anhydrases in Invertebrates, BMC Biochemistry, 11, 28, doi: 10.1186/1471-2091-11-28.
25. Tamura K., J. Dudley, M. Nei, S. Kumar (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0, Molecular Biology and Evolution, 24(8), 1596-1599.
26. Thompson J. D., T. J. Gibson, D. G. Higgins (2002). Multiple Sequence Alignment Using ClustalW and ClustalX, Current Protocols in Bioinformatics, Chapter 2, Unit 2.3, doi: 10.1002/0471250953.bi0203s00.
27. Verma A., V. K. Singh, S. Gaur (2016). Computational Based Functional Analysis of *Bacillus phytases*, Computational Biology and Chemistry, 60, 53-58.
28. Waterhouse A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede (2018). SWISS-MODEL: Homology Modelling of Protein Structures and Complexes, Nucleic Acids Research, 46(W1), W296-W303.
29. Wiederstein M., M. J. Sippl (2007). ProSA-web: Interactive Web Service for the Recognition of Errors in Three-dimensional Structures of Proteins, Nucleic Acids Research, 35(Web Server Issue), W407-W410.

Nathan Vinod Kumar, Ph.D.
E-mail: nvkibt@gmail.com

Vinod Kumar Nathan is a Post Doctorate researcher working at National Institute of Oceanography – CSIR, Kochi, India. He was conferred his Ph.D. in Biotechnology from Madurai Kamaraj University, India. His field of interest includes microbial biotechnology and enzymology. He is an expert in proteomics and bioinformatics research.