International Journal of Pharmacy and Biological Sciences-IJPBS™ (2019) 9 (1): 1260-1266
Online ISSN: 2230-7605, Print ISSN: 2321-3272
Research Article | Biological Sciences | Open Access | MCI Approved
UGC Approved Journal

In-silico Characterization of Translated Protein Obtained from the Crab *Dromia Dehaani* (Rathbun, 1923)

Wilson Sylvester Fredrick[1]* Velayudham Sadeesh Kumar[2] and Samuthirapandian Ravichandran[3]

[1]Department of Marine Biotechnology, AMET Deemed to be University, Kanathur, Chennai- 603112.

[2]Centre of advanced study in Marine Biology, Department of Biotechnology, Faculty of Marine Sciences, Annamalai University, Parangipettai-608502.

Received: 11 Oct 2018 / Accepted: 10 Nov 2018 / Published online: 1 Jan 2019
Corresponding Author Email: sylvesterfredy@gmail.com

Abstract

Crustaceans play an important role in major fisheries. The non edible crab *Dromia dehaani* which are collected from the trash, was chosen for the study as it has the high protein estimation and molecular weight among the four crabs. The molecular characterisation was identified by 18s RNA sequence. In this present study, the nucleotide sequences of crab *D.dehaani* is translated to protein sequences through EMBOSS software in order to know the primary structure analysis, physicochemical properties and secondary structural states that the protein involved is a membrane protein with 2 transmembrane helices and it is in hydrophobic in nature. The physicochemical properties were analysed by Protparam tool. The secondary primary structure was predicted by SOPM and SOPMA. The SOSUI server performed the identification of transmembrane regions in proteins. The tool BioEdit was used to compute the Kyte and Dolittle mean hydrophobicity profile of the transmembrane regions. Multiple sequence alignment of transmembrane regions computed using MSA tool was used to generate the sequence logo of transmembrane regions. Therefore, the computational tools are helpful in understanding the essential biological processes involved in conserved sequence specific interaction between DNA/RNA and the proteins.

Keywords

Dromia dehaani, EMBOSS, Protparam, SOSUI and Transmembrane regions.

*****

INTRODUCTION

Crustaceans form a huge group of arthropods, more than 10 million tons of crustaceans are produced by fishery or farming for human consumption [1]. The crabs which are non- commercial and non-edible crabs were separated from the trash. The trash was noticed by visiting the landing centre regularly. The crab *Dromia dehaani* was predominant among the four crabs collected from the trash in the Pazhayar coast. [2]. Four trash crab species of different families which were taken for protein analysis. The maximum protein content is shown in female crab and the molecular weight of the female haemolymph of *D.dehaani*. The morphological and molecular characters were assessed to predict the cladistic

status of the crab *D. dehaani* done with 18S rRNA partial sequences and the Genbank accession number was obtained as KC130910 [3]. Therefore, the study animal was taken for further bioinformatic studies.

A large number of computational tools are available from different sources for making predictions regarding the identification and structure prediction of proteins. The statistical parameters about protein sequence, such as sequence length, the number of amino acids and the physicochemical properties of a protein such as molecular weight, atomic composition, extinction coefficient, isoelectric point, grand average hydropathy (GRAVY), aliphatic index, instability index etc. can be computed by various computational tools for the prediction and characterization of protein structure. The amino acid sequence provides most of the information required for determining and characterizing the molecule's function, physical and chemical properties.

Many researchers have reported the sequence analysis characterization of proteins using biocomputation tools [4,5,6]. The physico-chemical parameters such as theoretical isoelectric point (pI), molecular weight, extinction coefficient, half-life, instability index [7,8] aliphatic index [9] and grand average hydrophathy (GRAVY) [10,11] were described. A statistical analysis shows that the aliphatic index, which is defined as the relative volume of a protein occupied by aliphatic side chains (alanine, valine, isoleucine and leucine) of proteins of thermophilic bacteria is significantly higher than that of ordinary proteins. The index may be regarded as a positive factor for the increase of thermostabiity of globular proteins [10].

In this present study the nucleotide sequence obtained from 18S RNA sequence is translated to protein sequence in order to find out the protein structural information using computational tools such as Protparam, Peptide Mass, SOPMA, Hydropathy plot and Trans membrane region.

MATERIALS AND METHODS

Translation into protein sequences

The 18S rRNA nucleotide sequence which is derived for the molecular identification is used for translation. The nucleotide sequences were translated, and the protein sequences were retrieved from the EMBOSS software. EMBOSS is "The European Molecular Biology Open Software Suite". EMBOSS also integrates a range of currently available packages and tools for sequence analysis into a seamless whole [12].

Primary structure analysis:

Protparam tool (http://www.expasy.org/tools/protparam.html) computes various physicochemical properties that could be deduced from a protein sequence. The parameters computed by Protparam include the molecular weight, extinction coefficient, half life, instability index, theoretical pI and amino acid composition. Molecular weight and theoretical pI were computationally calculated whereas the amino acid and atomic compositions are self-explanatory. Percentages of residues were computed using the primary structural data.

Secondary structure prediction:

The secondary structure prediction was predicted by the tools SOPM, SOPMA [13] and SSCP (secondary structure content prediction) [14] server is used for the computation of percentages of α-helical, β-strand and coiled regions and secondary structure class identification.

Pepwheel draws a helical wheel diagram for a protein sequence. This displays the sequence in a helical representation as if looking down the axis of the helix. It is useful for highlighting amphipathicity and other properties of residues around a helix. By default, aliphatic residues are marked with squares; hydrophilic residues are marked with diamonds, and positively charged residues with octagons, although this can be changed.

The SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/) server performed the identification of transmembrane regions in VEGF human proteins. The tool BioEdit was used to compute the Kyte and Dolittle mean hydrophobicity profile of the transmembrane regions. Multiple sequence alignment of transmembrane regions computed using MSA tool was used to generate the sequence logo of transmembrane regions.

RESULTS

Translation into protein sequences

The 18S rRNA nucleotides sequences were translated to protein sequence and given below.

>EMBOSS_001_1
RFRFWFDLPVRNQRFEDSRMAHISYVSLDLYPLTWITVVI
HSYMLLVSDRKGRSLLLFQNRSGLGPRTLCGESELCAERT
VSAPAPHLSSVCLINFLLVMRLLGYNGRGIRVPFRRGSLRN
GYHFGRQQARKLPTPG

Primary structure prediction- PROTPARAM

ProtParam calculates the physicochemical parameters of a protein sequence such as the amino acid composition. The amino acid composition of proteins was computed given in Table 1. Physico chemical parameters of the crab *D. dehaani* were

given in Table 2. The parameter involved is 137 numbers of aminoacids with molecular weight 15794.4 and theoretical pI is 11.01.

Secondary structure analysis

The secondary structure predicted with the help of programs SOPM and SOPMA shows that the *D.dehaani* crab protein is highly random coiled are found to be mixed class of secondary structural content. The computed percentage of residues forming α-helices occupied 24.09%, extended strand at 14.60%, β-strands –nil and random coil of 61.31% are shown (Table 3).

Transmembrane region analysis

The SOSUI server classifies the crab protein as transmembrane protein the predicted region are classified as primary and secondary based on the structural residues. The various primary and secondary transmembrane regions identified by SOSUI server were given in Table 4.

The SOSUI server performed the identification of transmembrane regions. The predicted transmembrane helices were visualized and analyzed using helical wheel plots generated by the programme pepwheel included in the EMBOSS 2.7 suite is showed in Fig. 1. The identified transmembrane regions in *D.dehaani* crab proteins were found to have more hydrophobic residues and it is well documented by the Kyte and Doolittle mean hydrophobicity profile in which all the peaks are above the zero line is showed in Fig. 2.

The sequence logo of transmembrane regions generated from the multiple sequence alignment of transmembrane regions is shown in Fig. 3. The height of each letter in the sequence logo is proportional to the frequency of the amino acid at that position. The presence of more leucine amino acid in the transmembrane region is identified from sequence logo. This amino acid sequence is of a membrane protein which has two transmembrane helices.

DISCUSSION

In the present study the nucleotide sequences were translated into protein sequences by EMBOSS software and structure predictions and information was revealed with the help of computational biology. Expasy's ProtParam computes the extinction coefficient (EC) for a range of 18575, wavelength. The EC value at 280nm is favoured because proteins absorb strongly there while other substances commonly in protein solutions do not. EC of VEGF human proteins at 280nm is ranging from 7615 to 214235 M^{-1} cm^{-1} with respect to the concentration of Cysteine, Trpsine and Tyrosine. Expasy's ProtParam classifies the crab protein is unstable most of the VEGF human proteins as unstable on the basis of Instability index (II>40). The aliphatic index (AI) that is defined as the relative volume of a protein occupied by aliphatic side chains (Ala, Val, Ile and Leu) is regarded as a positive factor for increase of thermal stability of globular proteins is low (57 - 94) for all proteins and it infers that the VEGF proteins may become unstable at high temperature. Grand Average hydropathy (GRAVY) index of all the VEGF human proteins are ranging from -0.1 hence this indicates that all these proteins may interact equally and easily with water. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge 11.01 but net charge of the protein is zero. The computed isoelectric point (pI) is useful for developing buffer system purification of isoelectric focussing methods. At pI proteins are stable and compact. The computed protein concentration and extinction coefficients help in the quantitative study of protein-protein and protein-ligand interactions in solution.

Similarly [5] chosen seventeen fish antifreeze proteins mainly to study their physico-chemical properties, primary and secondary structures by using computational tools and servers. Primary structure analysis reveals the AFPs under study are hydrophobic in nature. Expasy's ProtParam computes the extinction coefficient for a range of 276, 278, 279, 280 and 282 nm wavelength, Extinction coefficient of AFPs at 280 nm is ranging from 1280 to 42990 m^{-1} cm^{-1}. Ten different silk fibroin proteins (SFs) retrieved from Swiss-Prot database are analyzed for physico chemical characterization and structure analysis was studied using *Insilico* tools. The computed pI value of Q967G5, Q8I5B3, Q9BLL9, Q8IT50, Q9BIU3, Q4F623 and A8IM76 (pI<7) indicates that these SFs are acidic and the pI of A8IM39, Q9BLL6 and Q9GUB4 (pI > 7) reveals that these are basic in character. Extinction coefficient (EC) for a range of 276,278, 279,280 and 282nm wavelength, Grand Average Hydropathy (GRAVY) index of SFs are ranging from -0.010 to 0.451[15] *Insilico* characterization reports done in decapoda are very limited novel immune related gene serine proteinase of *Fenneropeaneaus indicus* predicted by *Insilico* homology modelling studies [16].

Insilico protein analysis is a well-established technique for assessment of allergenicity and immunological cross-reactivity [17,18]. Cyclophilins constitute a family of proteins involved in many important cellular functions. They have also been identified as a pan-allergen family able to elicit IgE-mediated hypersensitivity reactions [19, 20].

Potential allergens are identified among the characterized proteins of *Periplaneta americana* using web-based and publicly available allergen prediction tools [21].

The secondary structure analysis of the crab *D. dehaani* protein is predicted by SOPMA and it is found to be highly random coiled are found to be a mixed class of secondary structure content. The computed percentage of residues forming α-helices, β-strands and coils were also noted. The transmembrane regions were analyzed by SOSUI server and states that amino acid sequence is of a membrane protein which has 2 transmembrane helices. The identified transmembrane region in the haemolyph protein of *D.dehaani* is hydrophobic in nature and documented by hydropathy plot. The transmembrane helices were also visualized by Pepwheel. Likewise [22] characterized viral envelope proteins such as P04290, P06477, P04486 which having a great importance in the keratitis disease caused by HSV and these three important proteins are studied by using a Bioinformatics tools. Secondary structure shows that some are predominant alpha helices with random coils. Transmembrane region prediction by SOSUI server predicted that P04290 and contain only one transmembrane region while P06477 soluble protein. Four transmembrane regions found in P04486 protein all predicted regions were analyzed by the helical plots.

CONCLUSION:

From the present study the nucleotide sequences of crab *D.dehaani* is translated to protein sequences in order to know the primary structure analysis, physicochemical properties and secondary structural states that the protein involved is a membrane protein with 2 transmembrane helices and it is in hydrophobic in nature. The information derived from the computational tools are reliable and the cost-effective way to know the protein information.

RESULTS

Table. 1: Aminoacid composition through Protparam tool

Alanine (A) 5	3.6%
Arginine (R) 18	13.1%
Asparagine (N) 5	3.6%
Aspartine (D) 4	2.9%
Cystine (C) 3	2.2%
Glutamine (Q) 4	2.9%
Glutamine acid (E) 4	2.9%
Glycine (G) 11	8.0%
Histidine (H) 4	2.9%
Isoleucine (I) 5	3.6%
Leucine (L) 20	14.6%
Lysine (K) 2	1.5%
Methionine (M) 3	2.2%
Phenylalanine (F) 8	5.8%
Proline (P) 8	5.8%
Serine (S) 12	8.8%
Threonine (T) 5	3.6%
Trptophan (W) 2	1.5%
Tyrosine (Y) 5	3.6%
Valine (V) 9	6.6%
Pyl (O) 0	0.0%
Sec (U) 0	0.0%

Table. 2: Physico- chemical characteristics predicted by Expasys ProtParam Tool

Length	137
Molecular weight	15794.4
Isoelectric point	11.01
Extinction coefficient at 280nm	18575
Instability Index	45.69 Unstable
Aliphatic Index	93.87
Grand Average Hydropathy	-0.137
Number of negative residues –R	-8
Number of positive residues +R	20
Length	137
Molecular weight	15794.4
Isoelectric point	11.01
Extinction coefficient at 280nm	18575
Instability Index	45.69 Unstable
Aliphatic Index	93.87
Grand Average Hydropathy	-0.137
Number of negative residues –R	-8
Number of positive residues +R	20

Table 3: SOPMA prediction result for secondary structure

Alpha Helix %	24.09
Extended Strand %	14.60
Beta Turn %	-
Random Coil	61.31
Class	mixed

Table 4: Transmembrane regions identified by SOSUI server

No	N-Terminal	Transmembrane region	C terminal	Type	Length
1.	26	VSLDLYPLTWITVVIHSYMLLVS	48	Secondary	23
2.	85	APHLSSVCLINFLLVMRLLGYNG	107	Primary	23

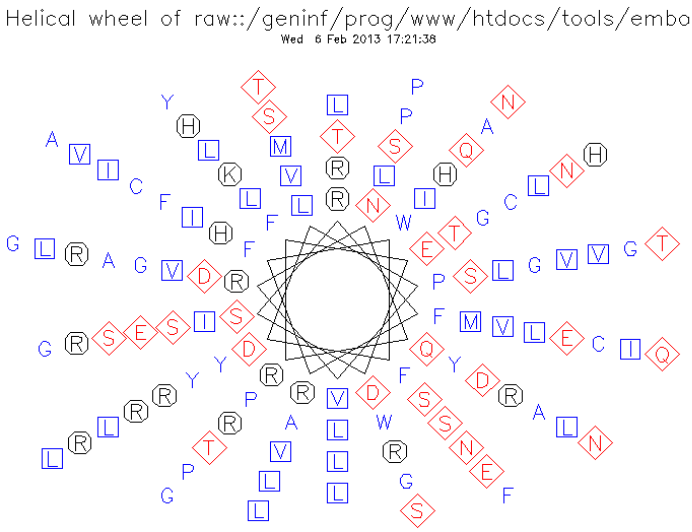

Fig. 1: Helical wheel representation of predicted helix of *D.dehaani.* Hydrophobic resides (V, L, I) are represented as blue squares and violet letters (A, G, P, Y), polar residues (E, Q, S, T) as red diamonds

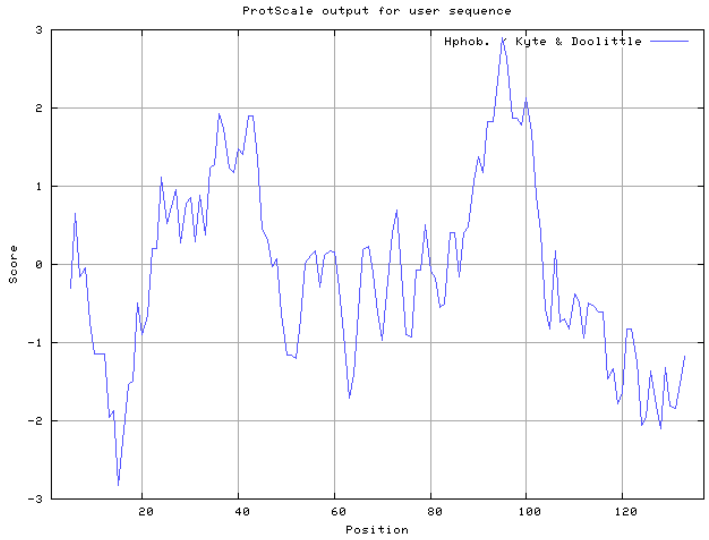

Fig. 2: Kyte and Doolittle mean hydrophobicity profile computed for the transmembrane region.

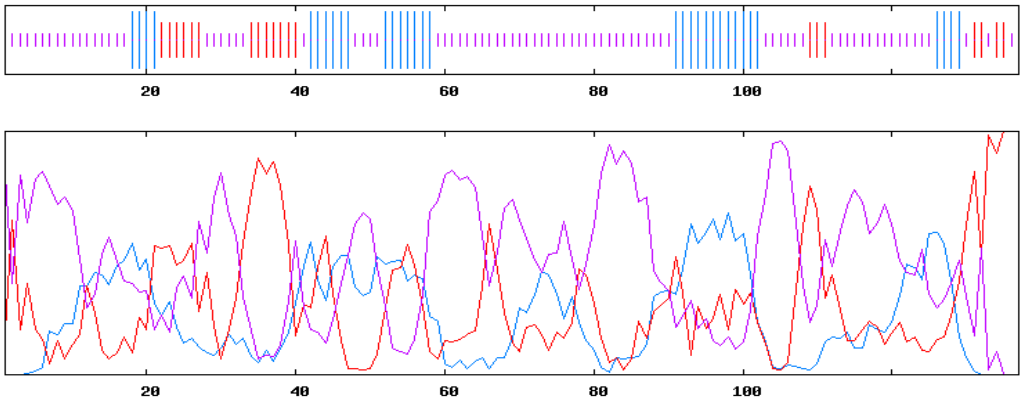

Fig. 3: The sequence logo of the transmembrane regions by multiple sequence alignment.

REFERENCES

1. Sylvester Fredrick W and S Ravichandran, Hemolymph proteins in marine crustaceans. Asian Pac J Trop Biomed 2012; 2(6): 496-502. doi:10.1016/S2221-1691(12)60084-7.
2. Sylvester Fredrick W, V. Sadeesh Kumar And S. Ravichandran. Protein analysis of the crab Haemolymph Collected from The Trash., Int J Pharm Pharm Sci, Vol 5, Issue 4, 304-308.
3. Ravichandran S, D. Karthick Rajan and W. Sylvester Fredrick., Molecular characterization of brachyuran crab Dromia dehaani (Rathbun, 1923) along the Pazhayar Coast (Southeast Coast of India), Int. J. Zool. Res., 13 (1): 1-5, 2017. DOI: 10.3923/ijzr.2017.1.5
4. Chitta, S. K., C. M. Anuradha, K. Venkata Rao and K. Venkateswara Swamy, 2005. *In silico* characterization of fatty acid synthase of Mycobacterium tuberculosis H37Rv. *Int. J. Genomics Proteomics*, 2(1): 420-429.
5. Sivakumar, K., S. Balaji and Gangaradhakrishnan, 2007. *In silico* characterization of anti freeze proteins using computational tools and servers. *J. Chem. Sci.*, 119(5): 571–579.
6. Ling, K. H., S. Loo, R. Rosli., M. N. Shamsudin, R. Mohammed and K. L. Wan, 2007. *In silico* identification and characterization of a putative phodsphatidylinositol 4-phosphate 5-Kinase (PIP5K) gene in *Eimeria Lenella*. *Silico biol*., 70011: 1002-10218.
7. Tobias, J. W., T. E. Shrader, G. Rocap and A. Varshavsky, 1991. The N-end rule in bacteria. *Science,* 254: 1374.
8. Ciechanover, A. and A.L. Schwartz, 1989. How are substrates recognized by the ubiquitin-mediated proteolytic system? *Trends Biochem. Sci*., 14(12): 483–488.
9. Guruprasad, K., B.V.B. Reddy and M.W. Pandit, 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach

for predicting in vivo stability of a protein from its primary sequence. *Prot. Eng.,* (4): 155.
10. Ikai, A. 1980. Thermostability and Aliphatic Index of Globular Proteins. *J. Biochem.,* (88): 1895.
11. Kyte, J. and R. F. Doolittle, 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.,* (157): 105.
12. Rice, P., I. Longden and A. Bleasby, 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trend. Genet.,* 16(6): 276-277.
13. Blanchet, C., C. Combet, C. Geourjon and G. Deléage, 2000. MPSA: integrated system for multiple protein sequence analysis with client/server capabilities. *TIBS,* 25(291): 147.
14. Eisenhaber, F., F. Imperiale. P. Argos and C. Froemmel, 1996. Prediction of secondary structural content of proteins from their amino acid composition alone. *Proteins Struct. Funct. Design,* 25: 157.
15. Ashokan, K. V. and M. M. Pillai, 2008. *In silico* characterization of silk fibroin protein using computational tools and servers. *Asian J. Exp. Sci.,* 22(3): 265-274.
16. Vaseeharan, B. and S. J. Valli, 2011. In silico Homology modeling of prophenoloxidase activating factor serine proteinase gene from the haemocytes of *Fenneropenaeus indicus*. *J. Proteomics Bioinform.,* 4: 53-57.
17. Binder, M. V., B. Mahler, W. R. Hayek, M. Sperr, S. Scholler and G. Prozell, 2001. Molecular and immunological characterization of arginine kinase from the Indian meal moth, *Plodia interpunctella*, a novel crossreactive invertebrate pan-allergen. *J. Immunol.*, 167: 5470-5477.
18. Flückiger, S., H. Fijten, P. Whitley, K. Blaser and R. Crameri, 2002. Cyclophilins, a new family of cross-reactive allergens. *Eur. J. Immunol.,* 32: 10-17.
19. Cadot, P. D., P. Proost, J. Van Damme, Y. Engelborghs, E. A. Stevens and J. L. Ceuppens, 2000. Purification and characterization of an 18- kd allergen of birch (*Betula verrucosa*) pollen: identification as a cyclophilin. *J. Allergy Clin. Immunol.*, 105: 286–291.
20. Fujita, C., T. Moriyama and T. Ogawa, 2001. Identification of cyclophilin as an IgE binding protein from carrots. *Int. Arch. Allergy Immunol.,* 125: 44–50.
21. Ahmed, A., K. Minhas, Namood-e-Sahar, O. Aftab and F. Sher Khan, 2010. *In Silico* identification of potential American cockroach (*Periplaneta americana*) allergens. *Iranian J. Publ. Health*, 39(3): 109-115.
22. Akhilesh, U., Bhaskar Anusha, Issar Sakshi and V. Senthamarai Selvi, 2012. *In Silico* charecterization of keratitis causing Herpes Simplex Virus (HSV 1) membrane proteins using computational tools and servers. *Res. J. Recent Sci.,* 1(11): 27-31