Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/17276
DOI URL: http://dx.doi.org/10.21474/IJAR01/17276

RESEARCH ARTICLE

3D Structure prediction and visualization of protein of the novel strain of *Rhodopseudomonas faecalis*

Aysha Sherieff[1], M. Mohibbe Azam[2] and K. Sesha Maheswaramma[3]

1. Research Scholar, JNTUA, Anantapur, Andhra Pradesh, India.
2. Indian Institute of Rice Research (ICAR-IIRR), Rajendra Nagar, Hyderabad, Telangana State, India.
3. Department of Chemistry, JNTUACEP, Pulivendula, Kadapa, Andhra Pradesh, India.

Manuscript Info

Manuscript History
Received: 20 May 2023
Final Accepted: 24 June 2023
Published: July 2023

Key words:-
Rhodopseudomonas Faecalis, Motif, Galaxy Web, Phyre 2, CATSp, RASMOL

Abstract

The secondary protein structure of *Rhodopseudomonas faecalis* was obtained from the partial protein sequence, protein analysis and domain prediction were characterized further using PSIPRED tool. Motifs were analyzed using Motif Scan-Prosite pattern tool. The 3D Structure prediction and visualization was performed using Phyre2 program, Galaxy Web and CASTp platforms. The visualization of the predicted and concrete structure of the protein was done using RASMOL. The results shows that the protein structure has sites for amidation, phosphorylation and myristylation along with 30% alpha helix and 19% beta strands, 3_{10} helix , Pi helix, turns and bends with repeated 11 different aminoacids occurrence.

Introduction:-

Purple non-sulfur phototrophic (PNSP) bacteria grow in a vast range of environmental conditions. Most of the PNSP bacteria belongs to the genus Rhodopseudomonas. These are able to grow anaerobically in the presence of light or aerobically in the dark conditions with different substrates like carbon sources and electron donors. In Bergey's Manual of Systematic Bacteriology, the genus Rhodopseudomonas has been included with seven species: *Rhodopseudomonas palustris, Rhodopseudomonas viridis, Rhodopseudomonas blastica, Rhodopseudomonas sulfoviridis, Rhodopseudomonas rutile, Rhodopseudomonas acidophila* and *Rhodopseudomonas marina* (Imhoff & Trusper, 1989). *Rhodopseudomonas faecalis* was isolated and characterized from an anaerobic reactor that digests chicken faeces for the first time by Demin Zhang et al., 2002.

A study was designed to understand and annotate a hypothetical protein of an unknown function of the bacterium *Rhodopseuodomonas capsulatus* through insilico approach. Different computational tools and wide range of bioinformatics workflow was established with 3D structure and biological function (Spencer Mark Mondol et al., 2022). PSIPRED Program was run for the protein structure prediction for a 42 kDa subunit pili protein of *Salmonella typhi* that causes typhoid fever (Sri Daramawati et al., 2022). 3D Structure of the viral S protein PBD ID: 6VYB was studied where the protein structure and docking were visualized using RasMol (Subbaiyan A, Ravichandran K, Singh SV, et al., 2020). Active site prediction was done by using CASTp for bacterial metalloprotease of family M4 (Hasan R et al.,2021). The 3D models of the NTDs of the selected bacterial Tcps were constructed using three homology modeling servers, including Phyre2 (Kelly et al.,2015).

Corresponding Author:- Aysha Sherieff
Address:- Research Scholar, JNTUA, Anantapur, Andhra Pradesh, India.

In the present work, the protein sequence from ORF's of the *Rhodopseudomonas faecalis* was obtained and used as a template for secondary structure prediction by using different tools that are authentic and highly potential methods and platforms. The wet lab method was performed by the authors previously and the 16s rRNA sequencing was done. The organism was found to be a novel one based upon the sequencing analysis. The partial nucleotide sequence was converted in to a protein and then ORF's were identified by ORF Finder tool. Later, the protein sequence was used for the study as described in the methodology section.

Methodology:-

Source of the sequence

Previously the authors happen to isolate and characterize a novel strain of *Rhodopseudomonas faecalis* bacterium. The organism's 16s r RNA partial sequence was performed and it was converted in to protein sequence. ORF's were identified and the ORF's protein sequences were used for the present study.

Secondary structure prediction

PSIPRED is a protein analysis workbench platform that provides many available analysis tools into a single web based framework (Daniel W A Buchan and David T Jones, 2019). The protein sequence was analysed using the tool. MOTIF SCAN is to find all known motifs that occur in a sequence. The purpose of this tool is to identify motifs or patterns found in the query protein sequence. Determining the motifs of the query protein will assist in the classification of a protein according to its family or domain. (Pagni M et al .,2007).

3D Structure visualization

The Galaxy WEB server predicts the protein structure from a sequence by template-based modeling and refines loop or terminus regions by ab initio modeling (Junsu Ko et al.,2012). Phyre2 is a protein structure prediction server that uses a combination of threading and comparative modeling to predict the 3D structure of a protein. Computed Atlas of Surface Topography of proteins (CASTp) provides an online resource for locating, delineating and measuring concave surface regions on three-dimensional structures of proteins. CASTp can be used to study surface features and functional regions of proteins (Binkowski TA et al.,2003). RasMol is a good viewer for showing PDB format files and it is the most widely used 3D molecular graphics (Sayle RA et al.,1995).

Results & Discussion:-

Secondary structure prediction

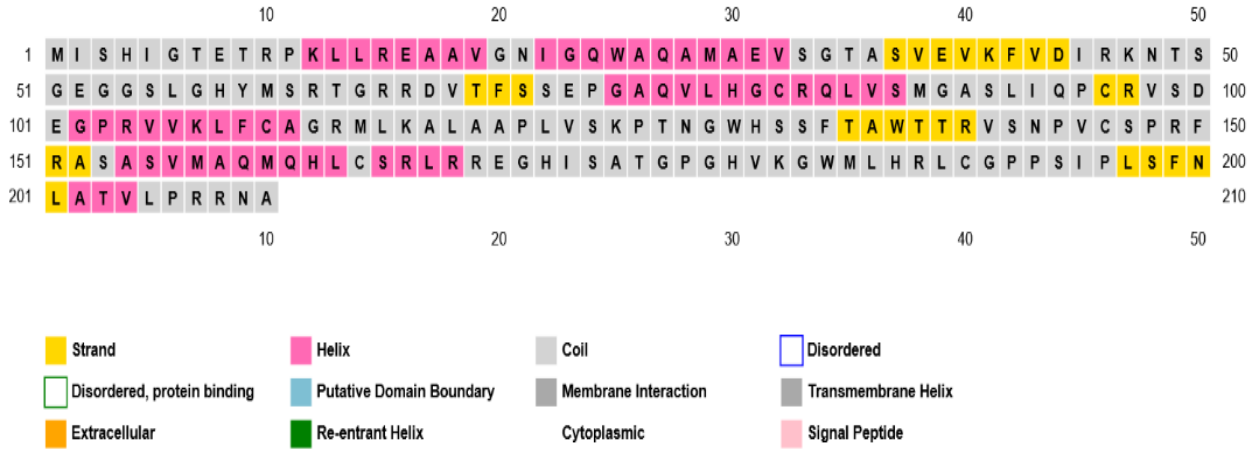


Fig 1a:- PSIPRED analysis of protein sequence of *Rhodopseudomonas faecalis*.

Figure1a shows the analysis from PSIPRED output that has helix, strand, membrane interaction and trans membrane helix regions of the secondary structure of protein of *Rhodopseudomonas faecalis*. There were about total 210

aminoacids consisting of 7 helix regions, ~6 strand regions and rest of the regions are transmembrane and membrane regions.

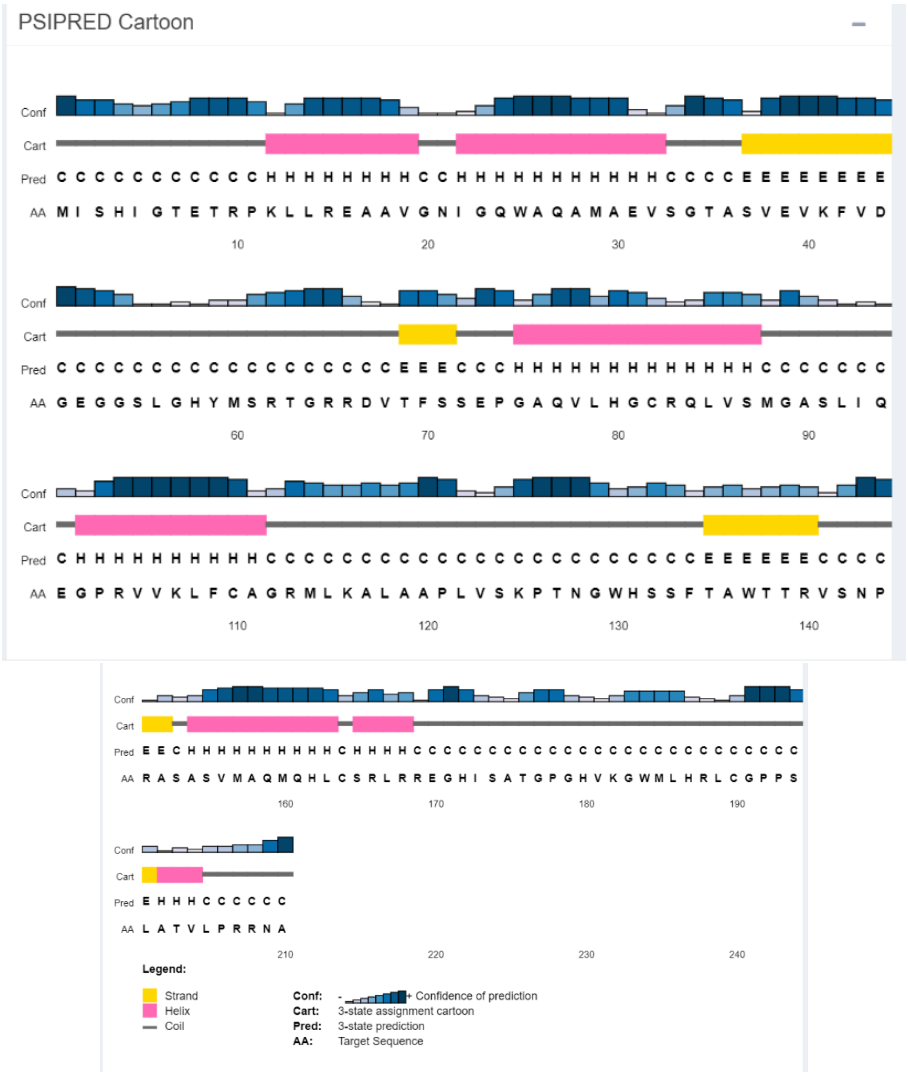


Fig 1b:- PSIPRED analysis of secondary structure of protein sequence of *Rhodopseudomonas faecalis*.

Figure 1b shows the PSIPRED secondary structure analysis of the protein of *Rhodopseudomonas faecalis* with Confidence of Prediction, 3-state assignment cartoon, 3-state prediction with respect to the target or query sequence. The protein has 5 strands (~6) , 7 helix and rest of the coil forms were present as shown in the legend.

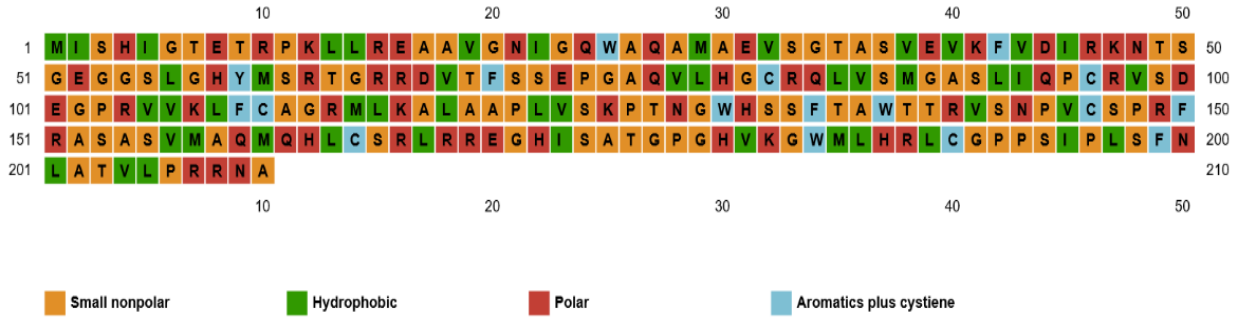


Fig 2:- PSIPRED analysis for protein sequence of *Rhodopseudomonas faecalis* showing nature of the aminoacids.

Figure 2 shows the nature of the aminoacids of the secondary structure of protein of *Rhodopseudomonas faecalis* performed through PSIPRED. There are small polar, hydrophobic, polar and aromatic plus cysteine aminoacids found in the entire secondary structure of the protein.

The protein domain prediction was done using Domain Prediction option in the PSIPRED tool and the result is obtained as follows :

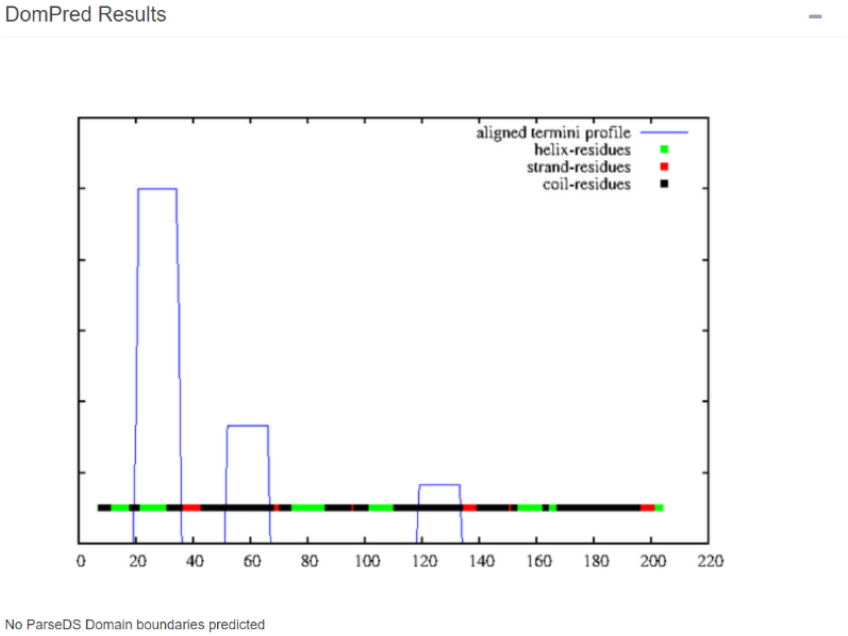


Fig 3:- Domain Prediction of the secondary protein structure of *Rhodopseudomonas faecalis*.

Figure 3 Shows the Domain Prediction of secondary protein structure of *Rhodopseudomonas faecalis*. It has helix residues, strand residues and coil residues. Out of 210 aminoacids, the aligned termini profile shows the prediction from aminoacids 20 to 38 having helix residues, 50 to 70 and 118 to 135 having coil residues whereas the strand residues are out of the aligned termini profile.

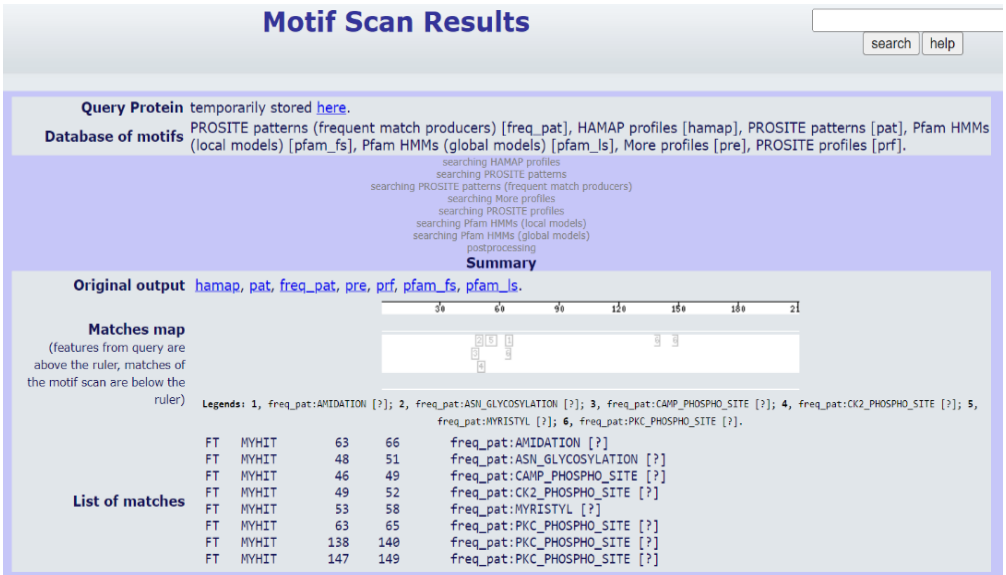


Fig 4a:- Motif Scan Results of the Protein structure of the *Rhodopseudomonas faecalis*.

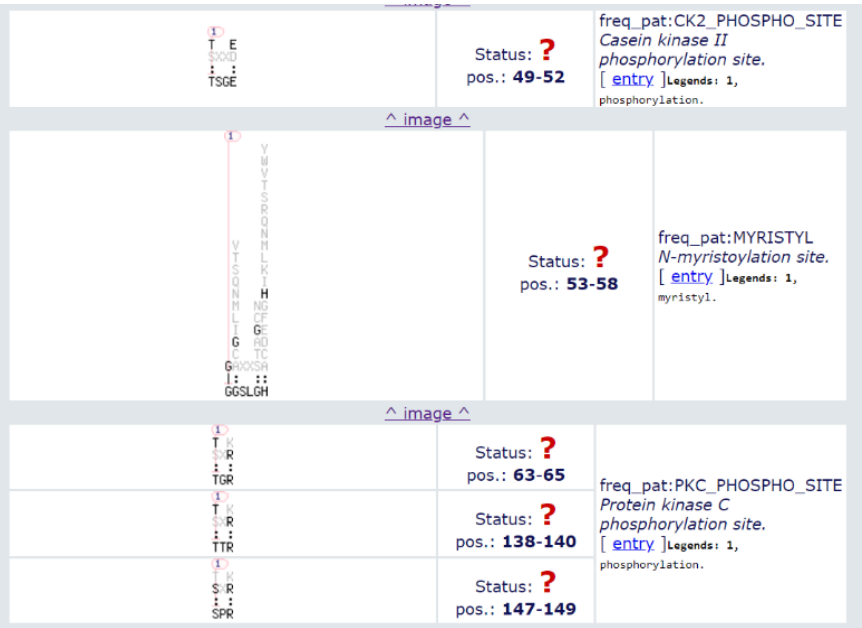


Fig 4b:- Motif Scan-Prosite patterns for the protein of *Rhodopseudomonas faecalis.*

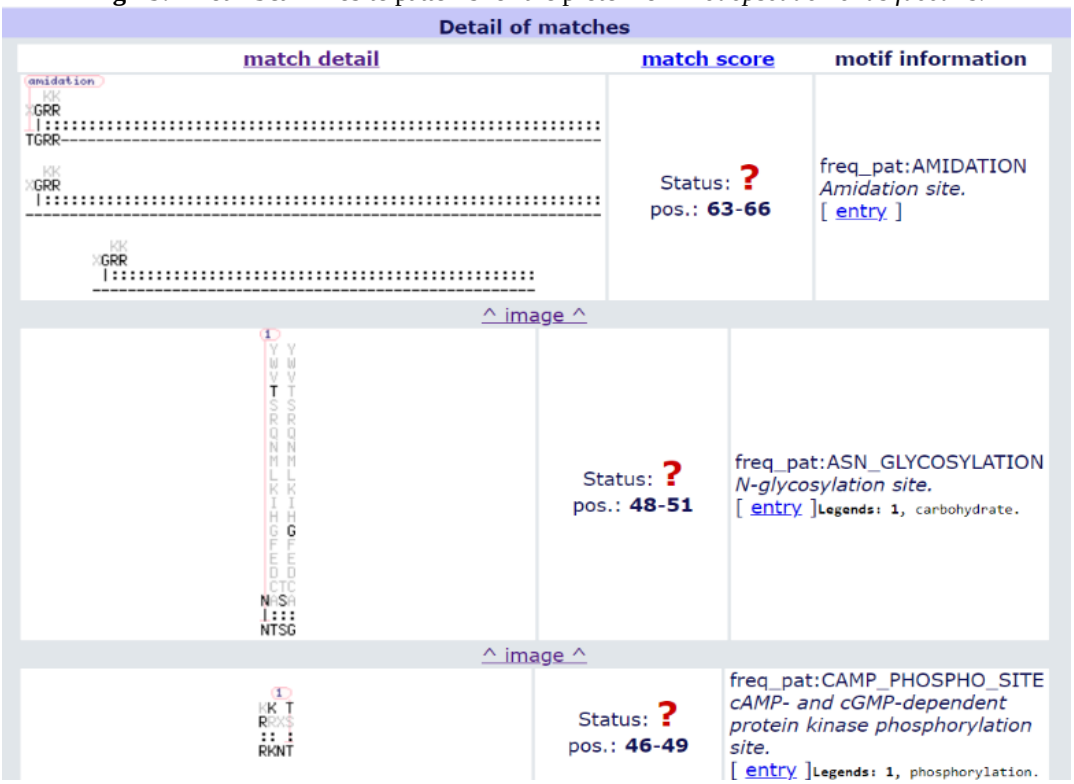
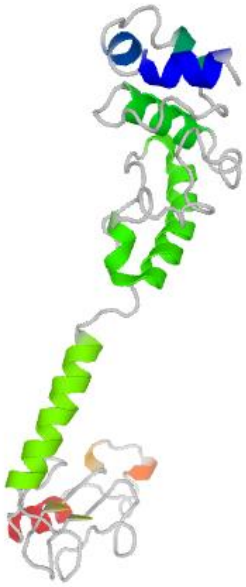


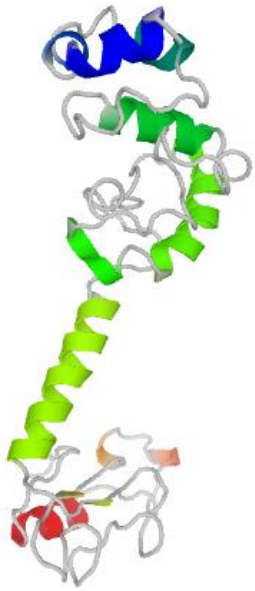
Fig 4c:- Motif Scan-Prosite patterns for the protein of *Rhodopseudomonas faecalis.*

Figure 4a, 4b and 4c shows the Motif Scan results with Prosite patterns for the protein of Rhodopseudomonas faecalis. The output shows the Amidation site, three Protein kinase phosphorylation sites (cAMP and cGMP phosphorylation, Casein Kinase II Phosphorylation site and Protein Kinase C phosphorylation site) and Myristylation site.

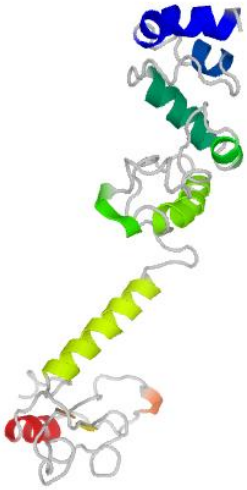
3D Structure visualization



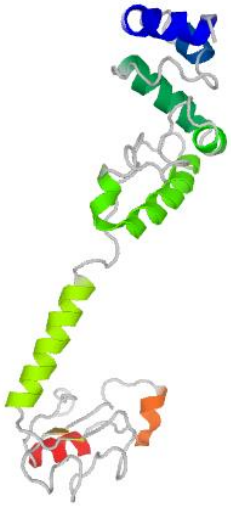
Model 1



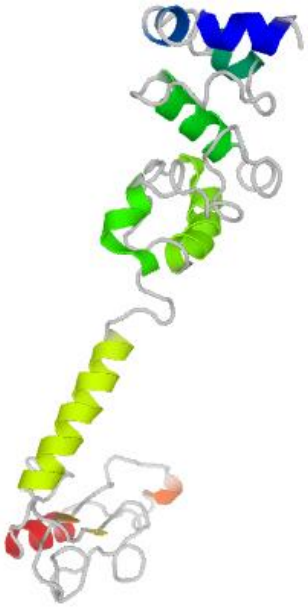
Model 2



Model 3



Model 4



Model 5

Fig 5:- Galaxy Web results of the protein 3D structure of the protein of *Rhodopseudomonas faecalis*.

The Galaxy Web protein 3D structure reveals 5 different possible models (Model 1, Model 2, Model 3, Model 4 and Model 5) for the protein of *Rhodopseudomonas faecalis* as shown in Figure 5.

3D Structure features

The protein sequence was submitted to Phyre 2 program wherein the sequence was converted in to PDB form and a 3D structure was visualized as follows. The file with PDB extension was also used for CASTp analysis.

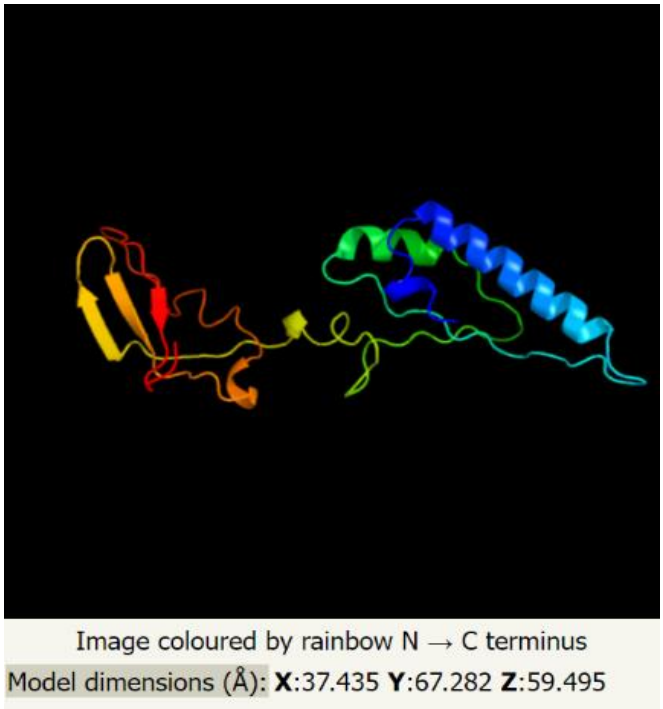


Fig 6a:- Phyre 2 result showing 3D structure of protein of *Rhodopseudomonas faecalis*.

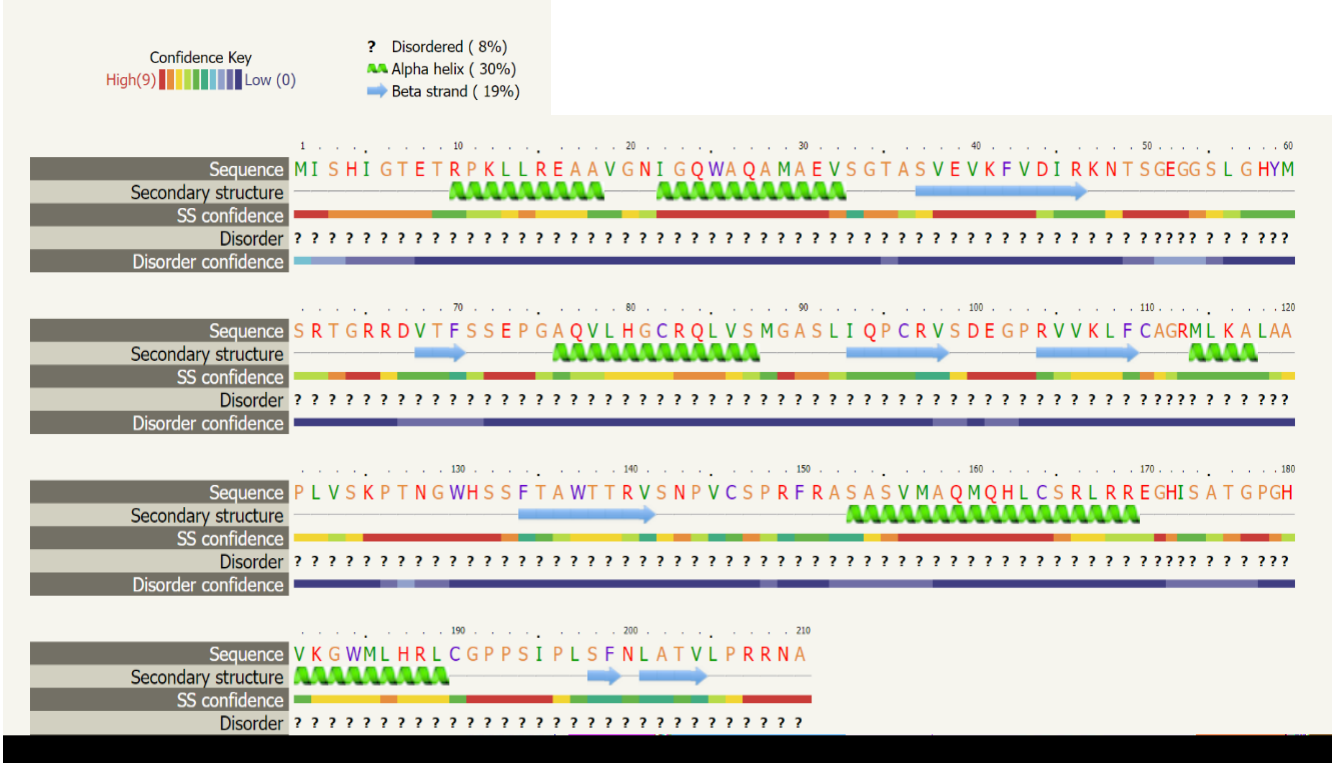


Fig 6b:- Phyre 2 result showing protein features of *Rhodopseudomonas faecalis*.

From Figure 6a and 6b, there were 124 residues (Coverage: 59% of your sequence) have been modelled with 20.2% confidence by the single highest scoring template. The secondary structure with helix, coils and strands were shown along with SS confidence and disorder confidence. About 8% disordered, 30% alpha helix and 19% beta strands were identified.

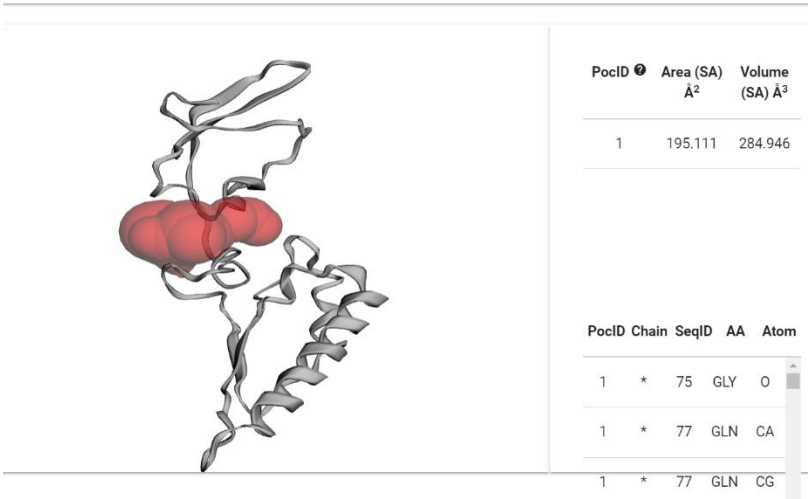


Fig 7a:- 3D structure of protein of *Rhodopseudomonas faecalis* using CASTp showing the various parameters of aminoacids.

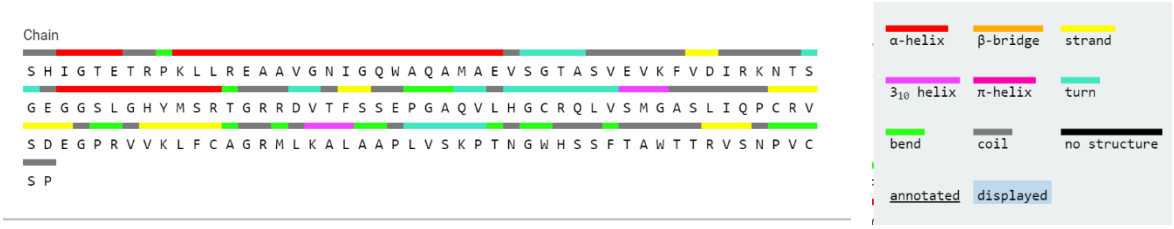


Fig 7b:- 3D structure of protein of *Rhodopseudomonas faecalis* using CASTp showing helix, beta form, coils, turns and strands.

The CASTp analysis of the protein structure of *Rhodopseudomonas faecalis* shows the 3D structure with aminoacid's characteristics and also the complete chain with presence and confirmation of alpha helix form, strands, 3_{10} helix , Pi helix, turns, bends and coils as shown in Figiure 7a and 7b.

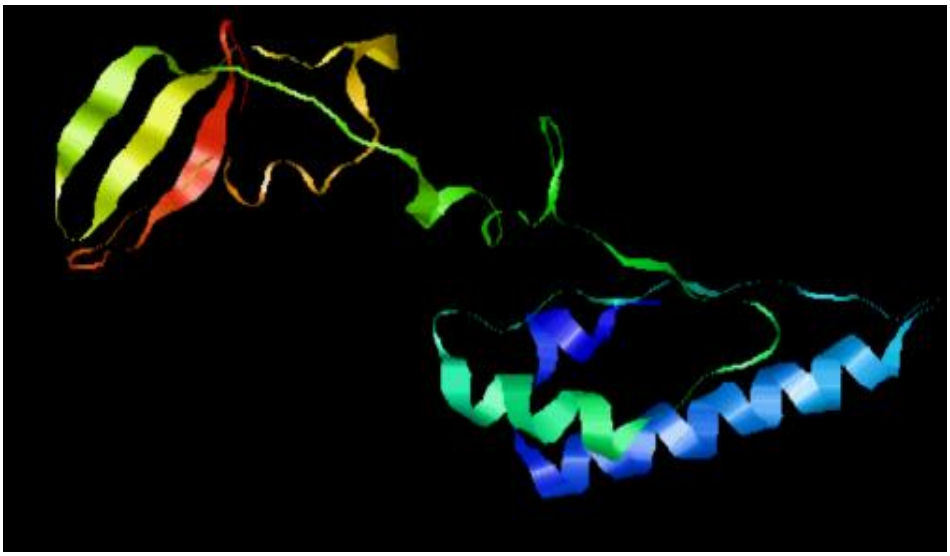
S.NO	PocID	Chain	SeqID	AA	Atom
1	1	*	75	GLY	O
2	1	*	77	GLN	CA
3	1	*	77	GLN	CG
4	1	*	77	GLN	NE2
5	1	*	78	VAL	N
6	1	*	78	VAL	CA
7	1	*	78	VAL	CG2
8	1	*	79	LEU	O
9	1	*	82	CYS	N
10	1	*	82	CYS	O
11	1	*	82	CYS	CB
12	1	*	82	CYS	SG
13	1	*	83	ARG	NE
14	1	*	83	ARG	CG
15	1	*	83	ARG	NH1
16	1	*	86	VAL	O
17	1	*	86	VAL	CB
18	1	*	86	VAL	CG1
19	1	*	86	VAL	CG2
20	1	*	87	SER	CA
21	1	*	87	SER	C
22	1	*	87	SER	O
23	1	*	87	SER	CB
24	1	*	87	SER	OG
25	1	*	88	MET	CA
26	1	*	88	MET	SD
27	1	*	90	ALA	O
28	1	*	90	ALA	CB
29	1	*	116	LYS	O
30	1	*	116	LYS	CE
31	1	*	116	LYS	CG
32	1	*	116	LYS	NZ
33	1	*	117	ALA	CA
34	1	*	117	ALA	O
35	1	*	119	ALA	O
36	1	*	120	ALA	N
37	1	*	120	ALA	CA

38	**1**	*****	**120**	**ALA**	**CB**
39	**1**	*****	**121**	**PRO**	**N**
40	**1**	*****	**121**	**PRO**	**CA**
41	**1**	*****	**121**	**PRO**	**CD**
42	**1**	*****	**122**	**LEU**	**N**
43	**1**	*****	**122**	**LEU**	**CB**
44	**1**	*****	**122**	**LEU**	**CG**
45	**1**	*****	**122**	**LEU**	**CD1**
46	**1**	*****	**123**	**VAL**	**N**
47	**1**	*****	**123**	**VAL**	**CG2**

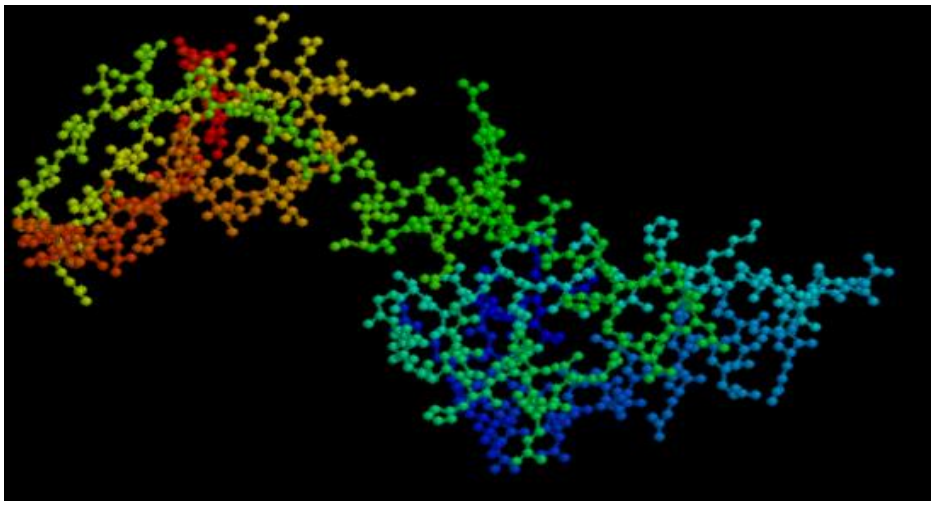
Table 1:- The complete number and type of aminoacids and atoms of the protein of *Rhodopseudomonas faecalis* using CASTp.

The Table 1 shows the 47 number of aminoacids present in the 3D structure along with their atoms and Poc ID , Seq ID and the name of the aminoacid in the protein of *Rhodopseudomonas faecalis*. Most of the aminoacids that were found are Alanine, Proline, Leucine, Valine, Lysine, Methionine, Serine, Arginine, Cysteine, Glutamine and Glycine.

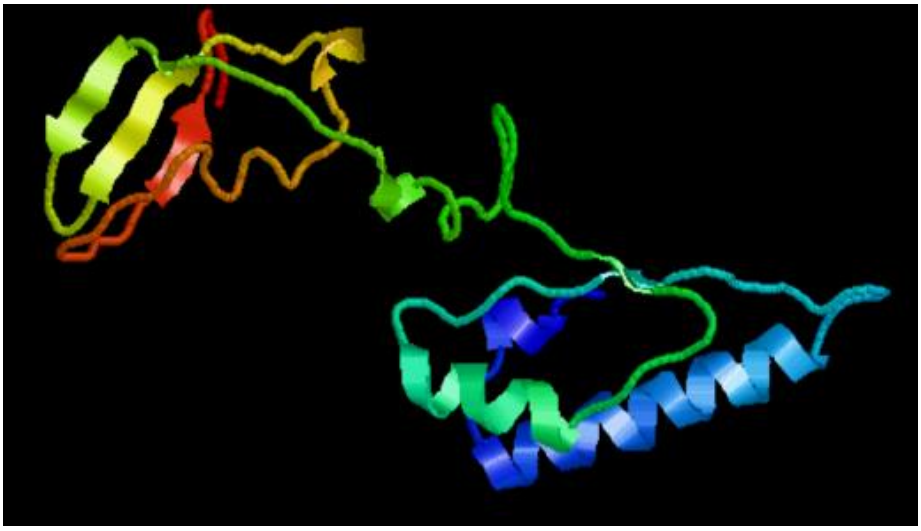
The PDB format of the protein from CASTp is obtained and the same was used as an input file to perform visualization using RASMOL.



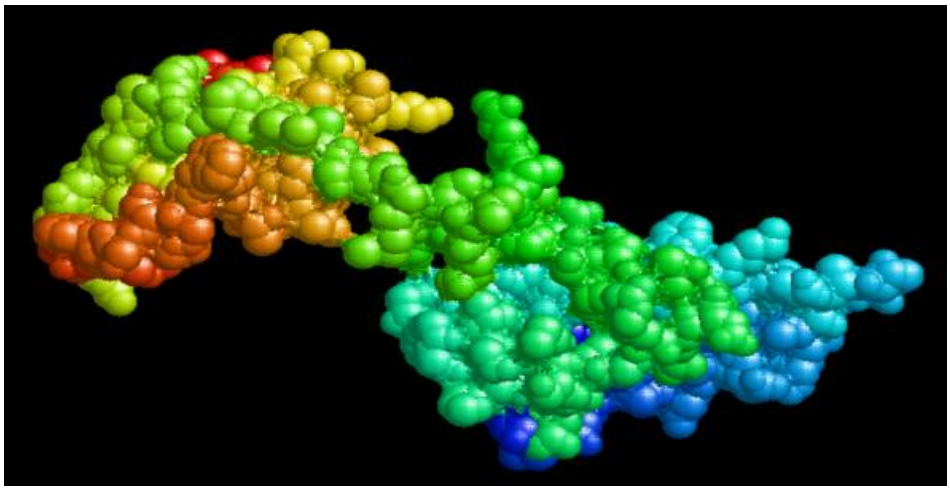
(1)



(2)



(3)



(4)

Fig 8:- 3D Structure Visualization of the protein of *Rhodopseudomonas faecalis* using RASMOL 1.Ribbon form 2.Ball and Stick 3. Cartoon form 4. Molecular surface form

The 3D structure visualization was done using RASMOL for the protein of *Rhodopseudomonas faecalis* . The Ribbon form, Ball and Stick form, Cartoon form and Molecular surface forms are shown in the Figure 8. It shows the strands, helix and coils.

Conclusion:-

The secondary structure protein analysis for *Rhodopseudomonas faecalis* by performing PSIPRED tool has revealed that the protein has 7 Helix, ~6 Strand regions out of total 210 aminoacids with small polar, hydrophobic, polar and aromatic plus cysteine aminoacids. The domain prediction shows the aligned termini profile from aminoacids 20 to 38 having helix residues, 50 to 70 and 118 to 135 having coil residues. Motif Scan results with Prosite patterns have confirmed the presence of Amidation site, Protein kinase phosphorylation sites and Myristylation sites.

The 3D Structure prediction and visualization for the protein of *Rhodopseudomonas faecalis* using Phyre 2 shows the presence of 30% alpha helix and 19% beta strands. The CASTp analysis shows additionally the presence and confirmation of alpha helix form, 3_{10} helix , Pi helix, turns and bends with repeated 11 different aminoacids occurrence ie., Alanine, Proline, Leucine, Valine, Lysine, Methionine, Serine, Arginine, Cysteine, Glutamine and Glycine. The 3D Structure with possible 5 models were generated using Galaxy Web and protein visualization was done by using RASMOL.

Acknowledgement:-

The Authors would like to thank Dr Zahoorullah S MD, Founder, Dr Z Labs and Head, Biotechnology Unit, Atal Incubation Centre ALEAP WE HUB, Supported by AIM, NITI AAYOG, GOI, Hyderabad, Telangana State, India for his insightful suggestions.

Conflict of interest statement

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. Besides there is no ethical committee approval is required as the work is related to bioinformatics.

References:-

1. Binkowski TA, Naghibzadeh S, Liang J. CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Res. 2003 Jul 1;31(13):3352-5.
2. Daniel W A Buchan , David T Jones, The PSIPRED Protein Analysis Workbench: 20 years on, Nucleic Acids Research, Volume 47, Issue W1, 02 July 2019, Pages W402–W407.
3. Demin Zhang, Huifang Yang, Zhiyong Huang, Wei Zhang and Shuang-Jiang Liu, "Rhodopseudomonas faecalis sp. nov., a phototrophic bacterium isolated from an anaerobic reactor that digests chicken faeces", International Journal of Systematic and Evolutionary Microbiology (2002), 52, 2055–2060.
4. Hasan, R., Rony, M.N.H. & Ahmed, R. In silico characterization and structural modeling of bacterial metalloprotease of family M4. J Genet Eng Biotechnol 19, 25 (2021).
5. Imhoff, J. F. & Tru$per, H. G. (1989). Purple nonsulfur bacteria. In Bergey's Manual of Systematic Bacteriology, vol. 3, pp. 1658–1682. Edited by J. T. Staley, M. P. Bryant, N. Pfennig & J. G. Holt. Baltimore: Williams & Wilkins.
6. Junsu Ko and others, GalaxyWEB server for protein structure prediction and refinement, Nucleic Acids Research, Volume 40, Issue W1, 1 July 2012, Pages W294–W297.
7. Kelley et al. (2015) Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 2015;10:845–858.
8. Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, Martin O, Kuznetsov D, Falquet L. MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W433-7.
9. Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. Trends Biochem Sci. 1995 Sep;20(9):374.
10. Spencer Mark Mondol, Depro Das, Durdana Mahin Priom, M Shaminur Rahman, M Rafiul Islam and Md Mizanur Rahaman. "In Silico Identification and Characterization of a Hypothetical Protein From Rhodobacter capsulatus Revealing S-Adenosylmethionine-Dependent Methyltransferase Activity", Bioinformatics and Biology Insights, Volume 16: 1–16, 2022.

11. Sri Darmawati, Stalis Norma Ethica, Muhammad Evy Prastiyanto, Sulaiman Ngongu Depamede, Eleventi Oktarina Putri, Mudyawati Kamaruddin, "Molecular characterization of a 42 kDa subunit pili protein of Salmonella typhi causes typhoid fever",Biodiversitas, Volume 23, Number 2, February 2022, Pages: 962-968.
12. Subbaiyan A, Ravichandran K, Singh SV, et al. In silico Molecular Docking Analysis Targeting SARS-CoV-2 Spike Protein and Selected Herbal Constituents. J Pure Appl Microbiol. 2020;14(suppl 1):989-998.