In Silico Analysis of RAG1 Protein of Labeo Calbasu

Rumpi Ghosh, A.D. Upadhayay* and A.K. Roy
Bioinformatics Centre, College of Fisheries,
CAU (I), Lembucherra, Tripura-799210
Email:*ad_up@rediffmail.com

Abstract:- **55294.5 kda Recombination activating gene 1 has been characterized in Labeo calbasu. The primary structure prediction of 483 amino acid residues RAG protein shows that the maximum number of amino acid found to be glutamine (91%) and least is Trypsin (1.4%). Total numbers of positive and negative charged residues are 76 and 66. The calculated pI is 5.64.The computed value is 7, and high aliphatic index (73.66) and instability index 50.59 and the (GRAVY) value is low -0.555. The secondary structure prediction of RAG 1 protein reveals extended strand and alpha helix. The secondary structure prediction is found to be c oil 31.26% followed by extended strand (14.08%) and alpha helix found to be frequent 49.48% and beta turn 5.18%. The tertiary structure was modelled by Swiss model workspace by using the templates from Swiss model template library and PDB sum the modelled structure showed that Tertiary structure or 3D structure prediction is done using Swiss model Automated mode for homology modeling. Swiss model server searched for the solved templates with similar sequences, the result of top seven best templates are given value in Fig: 3.Best templates were aligned with best E-value, percentage similarities and maximum number of query sequence covered were selected for homology modelling.**

Keywords: RAG1, Homology modelling, Procheck, PDB sum, Swiss model, Template library.

I. INTRODUCTION

Recombination activating gene 1 also known as RAG-1, which is a protein that in humans is encoded by the RAG1 gene. The protein encoded by this gene is involved in activation of immunoglobulin V-D-J recombination. The encoded protein is involved in recognition of the DNA substrate, but stable binding and cleavage activity also requires RAG2. Defects in this gene can be the cause of several diseases .RAG1 (Recombination Activating Gene 1) is a Protein Coding gene. Diseases associated with RAG1 include combined cellular and humoral immune defects with granulomas and alpha/beta t-cell lymphopenia with gamma/delta t-cell expansion, severe cytomegalovirus infection, and autoimmunity. Among its related pathways are signalling by GPCR and Interleukin-3, 5 and GM-CSF signalling. GO annotations related to this gene include *protein homodimerization activity* and *ubiquitin-protein transferees' activity*. The **recombination-activating genes** (RAGs) encode enzymes that play an important role in the rearrangement and recombination of the genes of immunoglobulin and T cell receptor molecules. There are two recombination-activating gene products known as RAG-1 and RAG-2, whose cellular expression is restricted to lymphocytes during their developmental stages. RAG-1 and RAG-2 are essential to the generation of mature B and T lymphocytes, two cell types that are crucial components of the adaptive immune system. RAG1 mediates the DNA-binding to the conserved recombination signal sequences (RSS) and catalyzes the DNA cleavage activities by introducing a double-strand break between the RSS and the adjacent coding segment. Catalytic component of the RAG complex, a multiprotein complex that mediates the DNA cleavage phase during V (D) J recombination. V(D)J recombination assembles a diverse repertoire of immunoglobulin and T-cell receptor genes in developing B and T-lymphocytes through rearrangement of different V (variable), in some cases D (diversity), and J (joining) gene segments. In the RAG complex, RAG1 mediates the DNA-binding to the conserved recombination signal sequences (RSS) and catalyzes the DNA cleavage activities by introducing a double-strand break between the RSS and the adjacent coding segment. RAG2 is not a catalytic component but is required for all known catalytic activities. The RAG complex also plays a role in pre-B cell allelic exclusion, a process leading to expression of a single immunoglobulin heavy chain allele to enforce clonality and monospecific recognition by the B-cell antigen receptor (BCR) expressed on individual B-lymphocytes. RAG1 also acts as a E3 ubiquitin-protein ligase that mediates mono ubiquitination of histone H3. Histone H3 monoubiquitination is required for the joining step of V (D) J recombination.

II. METHODOLOGY

Retrieval of target Sequence: - The Protein Sequence of the Recombination activating protein (RAG1) of *Labeo calbasu (orange-fin labeo)* is obtained from the Uniprot-KB having accession no. -D2JK26-1 .The

sequence length reported to be 483 amino acids residues (3). A template selection search was performed using PDB sum [3] against Swiss model template library .Simultaneously "Template library" at Swiss model tool provided by Swiss Institute of Bioinformatics was also utilised for template selection [4]. The template Library output shows the 7 significant hits having e-value [3]. Out of these which are the best template 3jbx was selected sharing about 95.86% of similarity with the query sequence. Finally the structure was predicted by using Swiss model Automated mode.

Physico-chemical characterization: - The values of theoretical isoelectric point (pI), molecular weight, and total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY) were computed. For Physico-chemical characterization Expasy's protparam server was used**.** The results are shown in Table No.1

Secondary structure prediction method: - GOR IV and SOPMA were employed for calculating the secondary structure of the protein sequence considered for this study. The secondary structure of RAG1 protein is predict by using GOR IV and SOPMA and the result showed that the sequence is mainly composed of Alpha helix and beta sheets . The comparative analysis by both GOR IV and SOPMA [2] is given table No. 2 .From the result, it is predict that the protein is chiefly composed of alpha helix (49.48%)and followed by random coil (47.00%)and extended strand(18.22%). The result is graphically represented Fig: 1.

Tertiary structure prediction:-

The tertiary structure was modelled by Swiss model workspace by using the templates from Swiss model template library and PDB sum the modelled structure showed that Tertiary structure or 3D structure prediction is done using Swiss model Automated mode for homology modeling [2]. Swiss model server searched for the solved templates with similar sequences, the result of top seven best templates are given value in Fig: 3.Best templates were aligned with best E-value, percentage similarities and maximum number of query sequence covered were selected for homology modelling.

III. RESULT S AND DISCUSSION

Primary structure prediction: - In this study primary structure of Recombination activating protein (RAG1) were predicted using Expasy's prot param server (http://expasy.org/cgi-bin/protparam) using the protein sequence and results are shown in Table-1. Results showed that Recombination activating protein had 483 amino acid residues, and estimated Molecular weight: 55294.5 kda. The maximum number of amino acid present in the sequence was found to be glutamine (9.1%) and the least was Trypsin (1.4%).The total number of positively charged residues (Asp+Glu): were 76 and the total number of negatively charged residues (Arg+ Lys) were 67.The calculated isoelectric point (pI) is useful for at pI the solubility is least and the mobility in an electric field is Zero. The Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of protein is zero. The calculated isoelectric point (pI) was computed to be 5.64. The computed value is less than 7 indicate that the protein is acidic. The high aliphatic index (73.66) indicates that this protein is stable for a wide range of temperature range .While the instability index (50.59) provides the estimated of instability of protein in a test tube .The grand average hydropathicity (GRAVY) value is low -0.555,indicates better interaction of protein with water.

Secondary structure prediction:-

The secondary structure of RAG1 protein prediction is composed of alpha helix and beta sheets and the secondary structure is predicted using GOR IV and SOPMA. (**Table 2**) represents the comparative analysis of GOR IV and SOPMA from which it is clear that random coil is predominantly present when the structure was predicted both by SOPMA and GOR IV, followed by extended strand and alpha helix [5] .The secondary structure prediction was done and random coil was found to be frequent 151 is 31.26% followed by extended strand (68 is 14.08%) and alpha helix was found to be frequent 239 is 49.48% and beta turn 25 is 5.18%.This is graphically represented in fig.1&2.

Tertiary structure prediction: - Swiss model is an online tool for 3D structure prediction based on homology modeling. Tertiary structure or 3D structure prediction is done using Swiss model workspace for homology modeling. Swiss model server searched for the solved templates with similar sequences, the result of top seven best templates are given value in Fig: 3. Best templates were aligned with best E-value, percentage similarities and maximum number of query sequence covered were selected for homology modelling. A total of 7 templates were found. The templates with the highest quality have been selected for model building. Models

are built based on the target –template alignment. Model quality was estimated by assessing the QMEAN score and Z-score. Finally the 3D structure prediction is done by using Swiss model automated mode.
Template 3jbx.1.A is select for homology modelling. Generated model was subjected to evaluation programs procheck and verify 3D. The procheck result shows that out of 483 amino acid sequences [3], 0 amino acid was fallen in disallowed region, representing the degree of accuracy in the predicted structure (refer figure 3,table 3),residues in generously allowed regions 5.

3. Table

Table No. 1: parameters computed using Expasy's prot param tool

Sl. No.	**Property**	**Value**
1.	Number of amino acids	483
2.	Molecular weight	55294.5
3.	Theoretical isoelectric point	5.64
4.	Total number of negatively charged residues(Asp+Glu)	76
5.	Total No. of positively charged residues (Arg+Lys)	67
7.	Instability index	50.59
8	Grand average of hydropathicity	-0.555
9.	Total No. of atoms	7689

Table 2: Secondary structure prediction of RAG1 protein by GOR IV and SOPMA

Secondary structure	GOR IV	SOPMA
Alpha helix (Hh)	34.78%	49.48%
3_{10} helix (Gg)	0.00%	0.00%
Pi helix(li)	0.00%	0.00%
Beta bridge	0.00%	0.00%
Extended strand (Ee)	18.22%	14.08%
Beta turn (Tt)	0.00%	5.18%
Bend region (Ss)	0.00%	0.00%
Random coil(Cc)	47.00%	31.26%
Ambiguous states (?)	0.00%	0.00%
Other states	0.00%	0.00%
Sequence length	488bp	483bp

Table-3: Ramachandran plot calculation by procheck. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R factor no greater than 20 a good quality model would be expected to have over 90% in the most favoured regions

Sl. No.	Residues in most favoured regions		Residues in additional allowed regions		Residues in generously allowed regions		Residues in Disallowed regions		Total no. of Residues
1.	**number**	**%**	**number**	**%**	**number**	**%**	**number**	**%**	**number**
2.	1437	(90.9%)	138	(8.7%)	**5**	**0.3%**	**0**	**0.0%**	1918

4. Figure

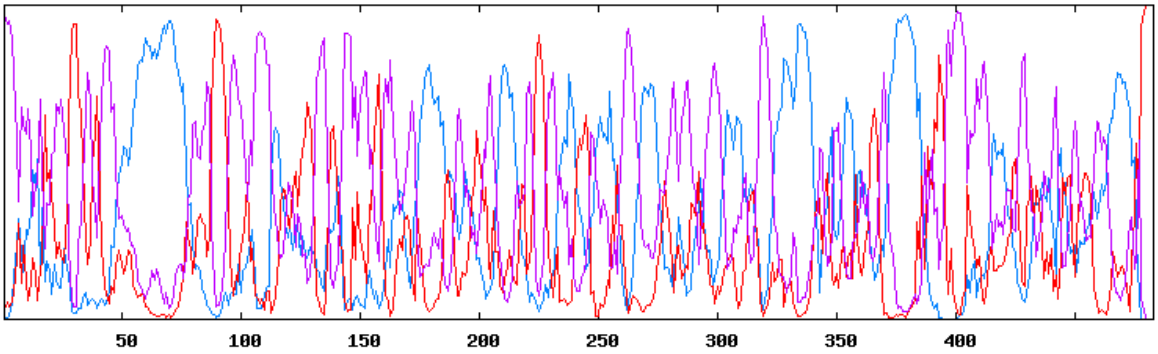


Figure 1: Graphical representation of secondary elements in RAG1 protein by GOR IV

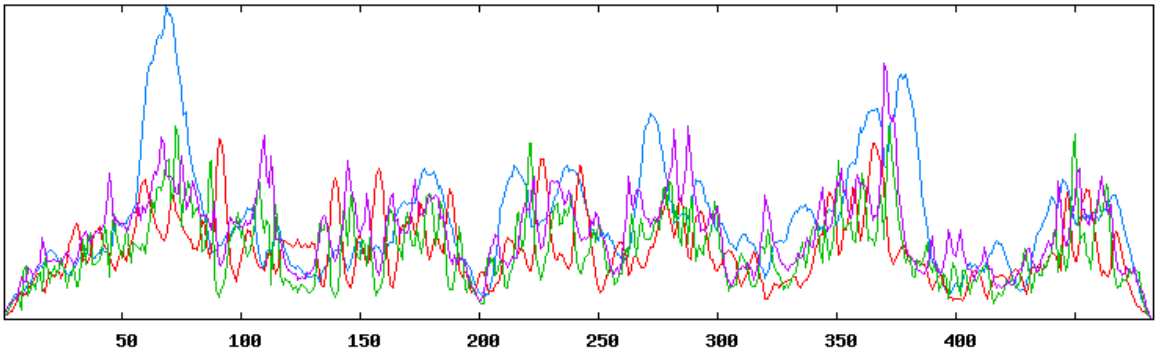


Figure 2: Graphical representation of secondary elements in RAG1 protein by SOPMA

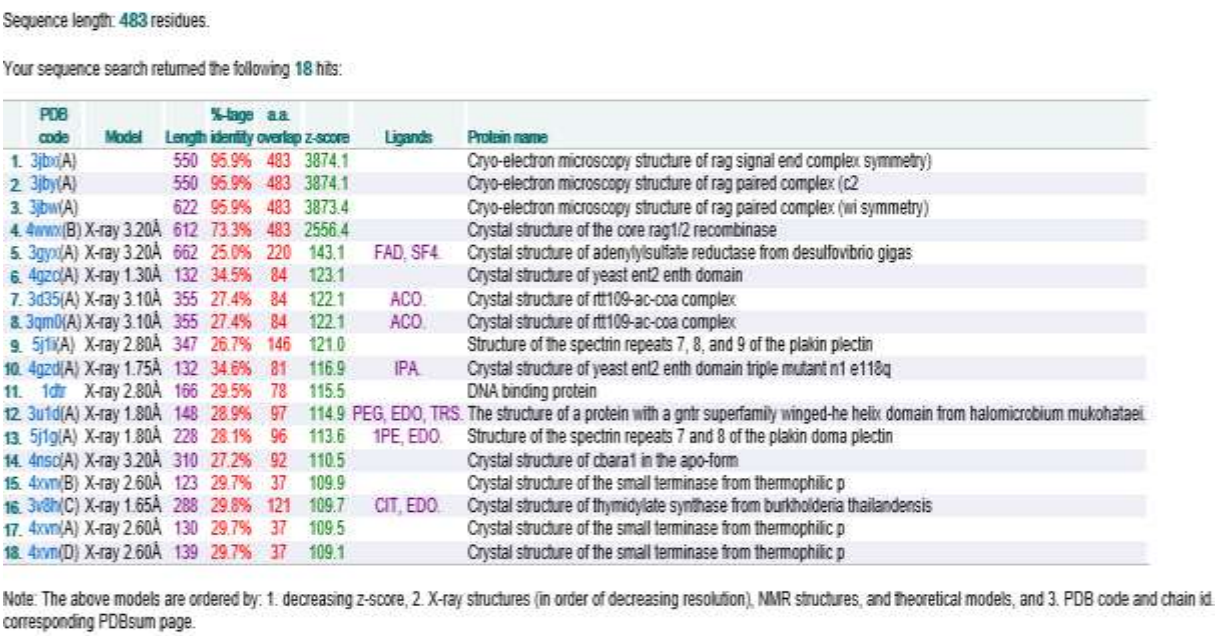
Template Results

Templates | Sequence Similarity | Alignment of Selected Templates | More

Name	Title	Coverage	Identity	Method	Oligo State	Ligands
3jbx.1.A	V(D)J recombination-activating protein 1		95.86	EM, 3.4Å	hetero-oligomer	2 x ZN , 4 x MG
3jbw.1.A	V(D)J recombination-activating protein 1		95.86	EM, 4.6Å	hetero-oligomer	2 x ZN
3jbw.1.C	V(D)J recombination-activating protein 1		95.86	EM, 4.6Å	hetero-oligomer	2 x ZN
4wwx.1.A	V(D)J recombination-activating protein 1		75.00	X-ray, 3.2Å	hetero-oligomer	1 x ZN
4wwx.2.A	V(D)J recombination-activating protein 1		75.00	X-ray, 3.2Å	hetero-oligomer	1 x ZN
4wwx.1.A	V(D)J recombination-activating protein 1		74.79	X-ray, 3.2Å	hetero-oligomer	1 x ZN
4wwx.2.A	V(D)J recombination-activating protein 1		74.79	X-ray, 3.2Å	hetero-oligomer	1 x ZN

The full list of templates matching your target sequence includes the following 15 templates which are not in the list above. The full template list is available in text or html format.

Figure: 3.Template representation of Swiss model template library



Sequence length: 483 residues.

Your sequence search returned the following 18 hits:

	PDB code	Model	Length	%-age identity	a.a. overlap	z-score	Ligands	Protein name
1.	3jbx(A)		550	95.9%	483	3874.1		Cryo-electron microscopy structure of rag signal end complex symmetry)
2.	3jby(A)		550	95.9%	483	3874.1		Cryo-electron microscopy structure of rag paired complex (c2
3.	3jbw(A)		622	95.9%	483	3873.4		Cryo-electron microscopy structure of rag paired complex (wi symmetry)
4.	4wwx(B)	X-ray 3.20Å	612	73.3%	483	2556.4		Crystal structure of the core rag1/2 recombinase
5.	3gyx(A)	X-ray 3.20Å	662	25.0%	220	143.1	FAD, SF4	Crystal structure of adenylylsulfate reductase from desulfovibrio gigas
6.	4gzc(A)	X-ray 1.30Å	132	34.5%	84	123.1		Crystal structure of yeast ent2 enth domain
7.	3d35(A)	X-ray 3.10Å	355	27.4%	84	122.1	ACO	Crystal structure of rtt109-ac-coa complex
8.	3qm0(A)	X-ray 3.10Å	355	27.4%	84	122.1	ACO	Crystal structure of rtt109-ac-coa complex
9.	5j1i(A)	X-ray 2.80Å	347	26.7%	146	121.0		Structure of the spectrin repeats 7, 8, and 9 of the plakin plectin
10.	4gzd(A)	X-ray 1.75Å	132	34.6%	81	116.9	IPA	Crystal structure of yeast ent2 enth domain triple mutant n1 e118q
11.	1dtr	X-ray 2.80Å	166	29.5%	78	115.5		DNA binding protein
12.	3u1d(A)	X-ray 1.80Å	148	28.9%	97	114.9	PEG, EDO, TRS	The structure of a protein with a gntr superfamily winged-he helix domain from halomicrobium mukohataei
13.	5j1g(A)	X-ray 1.80Å	228	28.1%	96	113.6	1PE, EDO	Structure of the spectrin repeats 7 and 8 of the plakin doma plectin
14.	4nsc(A)	X-ray 3.20Å	310	27.2%	92	110.5		Crystal structure of cbara1 in the apo-form
15.	4xvn(B)	X-ray 2.60Å	123	29.7%	37	109.9		Crystal structure of the small terminase from thermophilic p
16.	3v8h(C)	X-ray 1.65Å	288	29.8%	121	109.7	CIT, EDO	Crystal structure of thymidylate synthase from burkholderia thailandensis
17.	4xvn(A)	X-ray 2.60Å	130	29.7%	37	109.5		Crystal structure of the small terminase from thermophilic p
18.	4xvn(D)	X-ray 2.60Å	139	29.7%	37	109.1		Crystal structure of the small terminase from thermophilic p

Note: The above models are ordered by: 1. decreasing z-score, 2. X-ray structures (in order of decreasing resolution), NMR structures, and theoretical models, and 3. PDB code and chain id. corresponding PDBsum page.

Figure: 4-Representation of pdb sum template.

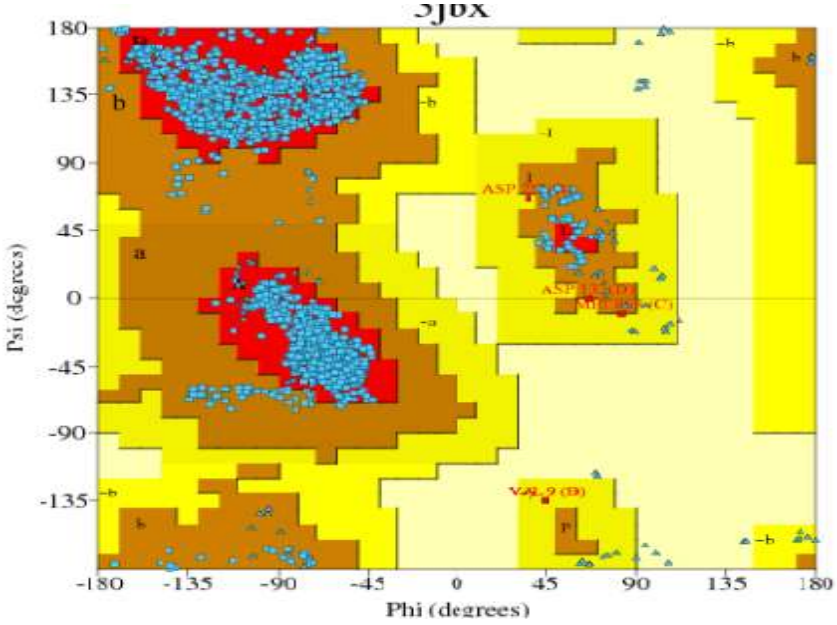


Figure 5:- Graphical representation of Ramachandran plot by Procheck.

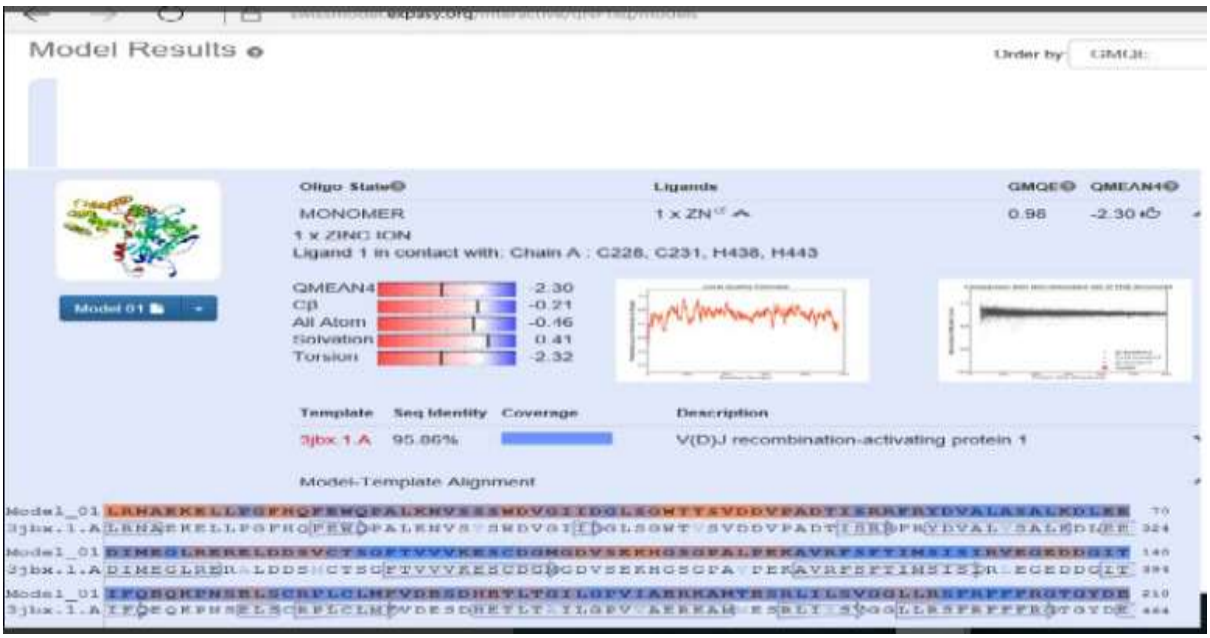


Figure5:-representation of model of RAG1 protein by Swiss model Work space.

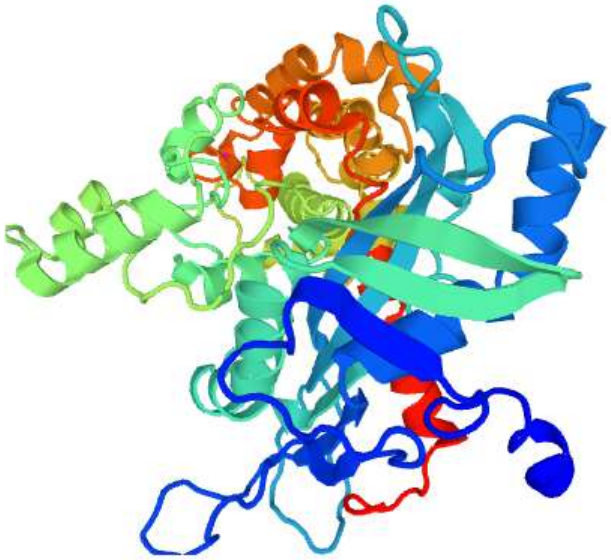


Figure6: - 3D structure of RAG 1 protein by Swiss model.

IV. Conclusion

Present study deals with the insilico sequence and structure analysis of RAG1 protein of *Labeo calbasu* [2]. It is most necessary to know the 3D structure of a protein, which may be achieved by means of X-ray crystallography or NMR spectroscopy .The experimental technique are more time consuming and tedious method also may not be always succeed in determining the structure of all the proteins. Since, large amount of data is generating day by day therefore creating a gap between available sequences and experimental solved structure. Computational biology which deals with the computational approach to analyze biological data in silico, help in this regard to reduce the gap [3]. The method like homology modeling is one of them useful to resolve the structure [4].Present study deals with the Insilco sequence analysis and structure analysis of RAG1 protein of *Labeo calbasu* by various tools and software's. On the basis of the various structural and physicochemical parameters assessment, it can be concluded that the predicted three dimensional structure of RAG1 is stable. Based on the finding, it could concluded that the protein is non polar, hydrophilic in nature. The prediction of protein structure can be used to know more about interaction of RAG1 protein domains with glucocorticoids or ligands and their role in fat content of fish. Based on the findings it could be important for evaluating how the regulation of this gene is related in the complications connected to various disease of fishes. The predicted structure of RAG1 protein may be further used in characterizing the protein and understanding the mechanism in Labeo fishes.

Acknowledgement

Authors are thankful to the Dean, COF (CAU), Lembucherra for encouragement and support. Financial assistant of DBT, India for the Bioinformatics Centre, COF CAU Tripura, under which this work has been carried out, is duly acknowledged.

REFERENCES

[1] M. Bandari, S. Kumar RS, J. kodavaty, and K. Reddy CV, "Insilico analysis and 3D structure prediction of human gastrin protein", International journal of advanced research in IT and engineering, ISSN: 2278-6244.Vol.2-no.4.april 2013.

[2] A. S. Bhalla and R. Lal, "Insilico sequence analysis, structure prediction and evaluation of fatty acid Synthase protein (FAS) of Gallus gallus (chicken) by homology modelling", IJSR, ISSN: 2319-7064, 2013.

[3] P.G. Pohekar and N.G. Chikhale, "Insilico analysis and prediction of 3D structure of soybean Trypsin inhibitor (SBTI) by homology modelling", IJIRSET, ISSN: 2319-8753.vol.2, 11, 2013.

[4] E. Kriger, S.B Nabuurs, and G .Vriend, "Homology modelling, in structural Bioinformatics", volume 44(Eds P. E. Bourne and H. Weissig), John Wiley & sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0471721204.ch25,2005.

[5] Vidhya VG, A. Bhaskar, P. purushotaman, "In silico Analysis and 3D modelling of SORD Protein in Diabetic Retinopathy", Scholars Research Library,ISSN:2231-3176 ,2011,1(4) :22-27.