International Journal of ChemTech Research
CODEN (USA): IJCRGG ISSN : 0974-4290
Vol.6, No.9, pp 4285-4306, September 2014

RTBCE 2014[12th August 2014]
Recent Trends in Biotechnology and Chemical Engineering

Insilico Structural and Functional Analysis of Nitrogen Fixing Proteins from *Trichodesmium thiebautii.*

Kalpana J[1], Vivek P[1]*, Santhana Bharathi N[1], Sathishpandiyan S[1], Prathap S[1], Karunakaran S[2], Saravanan A[2]

[1]Department of Biotechnology, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Avadi, Tamilnadu Chennai -600 062, India.
[2]Department of Biotechnology, Vivekanandha College of Engineering for Women, Elayampalayam, Namakkal 637205, Tamilnadu,India.

***Corres.author: vivek@velhightech.com**

Abstract: Some proteins of Trichodesmium thiebautii namely Glutamine synthetase, Heterocyst differentiation protein (HetR), Nitrogenase Molybdenum iron protein (NifD) and Ribulose 1, 5 bisphosphate carboxylase (RuBisCO) play an important role in the oceanic nitrogen fixation and have unique properties that are rarely found in any other organisms. Insilico study of the selected proteins of Trichodesmium thiebautii can be used as a model for the in vitro studies and research. Amino acid sequence of these proteins were retrieved and modeled using the bioinformatics tools. Primary, Secondary, tertiary structures and functions of nitrogen fixing proteins of*Trichodesmium thiebautii* were predicted.

PROTPARAM was used for the prediction of primary structure of the proteins. GOR V was used for the secondary structure of the proteins, SWISS MODEL (Homology Modelling) and I-TASSER (Protein Threading) were used for predicting the tertiary structures and functions of the proteins. QMOL and RASMOL were used for the molecular viewing and editing. BLASTp search was also done for the above mentioned proteins.All the above mentioned proteins are involved in the nitrogen fixation characters of the organism. From the study done it is seen that the HetR and the NifD protein of *Trichodesmium thiebautii* widely vary in their structure and functions of any other known protein from the Protein Data bank.
Keywords: *Trichodesmium thiebautii,* Glutamine synthetase, Heterocyst differentiation protein, Ribulose 1, 5 bisphosphate carboxylase, Nitrogenase Molybdenum iron protein.

Introduction

Prediction of structures and functions of the proteins makes in vitro studies easier and can be performed in a short run. Properties of the proteins are predicted and that gives us an idea of the metabolic activity of Trichodesmium thiebautii in the marine environment[1]. The involvement of the proteins in nitrogen fixation can be predicted from the models generated which can help us in the study of nitrogen cycle in the marine environment. The heterocyst differentiation protein of Trichodesmium thiebautii, which is a non-heterocyst organism, helps in the activity of diazocytes (group of cells involved in the nitrogen fixation) of Trichodesmium thiebautii[1, 2].

Studies on diazocytes are still in their pro- stage, so prediction of HetR structures and functions can be helpful in the study of diazocytes.The Nitrogenase molybdenum iron protein and heterocyst differentiation protein of Trichodesmium thiebautii is found to be structurally vastly dissimilar to any other protein in the protein data bank[3]. Modeling of these two proteins can be helpful in predicting their original structures, their relations and differences from other proteins.The data predicted from this study can be useful in forming a phylogenetic tree and predicting the evolutionary changes of this Trichodesmium thiebautii. The ultimate scope of this project is to predict the structures and functions of unknown proteins of Trichodesmium thiebautii for the use of in vitro studies[2, 3].

Cyanobacteria

Cyanobacteria, also known as blue-green algae, blue-green bacteria or cyanophyta are a phylum of bacteria that obtain their energy through photosynthesis. The name "cyanobacteria" comes from the color of the bacteria[4]. They are a significant component of the marine nitrogen cycle and an important primary producer in many areas of the ocean, but also found in habitats other than the marine environment.In particular cyanobacteria are known to occur in both freshwater and hyper-saline lakes[3, 4].

Cyanobacteria include unicellular and colonial species. Colonies may form filaments, sheets or even hollow balls. Some filamentous colonies show the ability to differentiate into several different cell types: vegetative cells, the normal, photosynthetic cells that are formed under favorable growing conditions; akinetes, the climate resistant spores that may form when environmental conditions become harsh; and thick-walled heterocysts, which contain the enzyme nitrogenase, vital for nitrogen fixation[5]. Heterocyst-forming species are specialized for nitrogen fixation and are able to fix nitrogen gas, which cannot be used by plants into ammonia (NH_3), nitrites (NO^{2-}) or nitrates (NO^{3-}), which can be absorbed by plants and converted to protein and nucleic acids[4, 5].

Trichodesmium Thiebautii

Trichodesmium thiebautii is a toxin producing non-heterocystous, Diazotropic cyanobacteria ubiquitous in tropicaland sub -tropical and temperate seas. Trichodesmium is known for its ability to fix nitrogen and for its massive blooms; as a result, it is considered the major component of oceanic primary production and global nitrogen cycling[5, 6]. The Primary, secondary and tertiary structure of the following proteins of this cyanobacteria has been predicted. The proteins that were selected are

- ✓ Heterocyst Differentiation Protein
- ✓ Glutamine synthetase
- ✓ Nitrogenase Molybdenum iron Protein
- ✓ Ribulose 1, 5 bisphosphate Carboxylase

Heterocyst Differentiation Protein

As it is difficult to reconcile concomitant oxygenic photosynthesis and oxygen labile nitrogen fixation cyanobacteria have evolved different adaptation strategies to overcome this anomaly. A confinement of nitrogenase into special micro-aerobic cell is seen for heterocystous species. Cyanobacteria lacking this cell type confide in other behavioral strategies. The genus Trichodesmium is the only exception to the above strategies. This marine cyanobacterium fixes nitrogen aerobically only during the light phase. Its nitrogen fixation activity is restricted to the phosphatase and nitrogenase is confined to subsets of cells termed Diazocytes, arranged consecutively and constituting 7-20% of all cells. The protein modeled here can be helpful in the understanding of the functions, properties and for in vitro experiments on HetR of Trichodesmium thiebautii[7-10].

Glutamine Synthetase

Glutamine synthetase (GS) (EC 6.3.1.2) is an enzyme that plays an essential role in the metabolism of nitrogen by catalyzing the condensation of glutamate and ammonia to form glutamine.

The enzyme glutamine synthetase is a key enzyme controlling the use of nitrogen inside cells. Glutamine, as well as being used to build proteins, delivers nitrogen atoms to enzymes that build nitrogen-rich molecules, such as DNA bases and amino acids. So, glutamine synthetase, the enzyme that builds glutamine, must be carefully controlled. When nitrogen is needed, it must be turned on so that the cell does not starve. But when the cell has enough nitrogen, it needs to be turned off to avoid a glut. Glutamine synthetase acts like a

tiny molecular computer which monitoring the amounts of nitrogen–rich molecules. It watches levels of amino acids like glycine, alanine, histidine and tryptophan and levels of nucleotides like AMP and CTP. If too much of one of these molecules is made, glutamine synthetase senses this and slows production slightly. But as levels of all of these nucleotides and amino acids rise, together they slow glutamine synthetase more and more. Eventually, the enzyme grinds to a halt when the supply meets the demand[7-10].

Ribulose 1, 5 Bisphosphate Carboxylase

Ribulose-1, 5-bisphosphate carboxylase/oxygenase, most commonly known by the shorter name RuBisCO, is an enzyme (4.1.1.39) that is used in the Calvin cycle to catalyze the first major step of carbon fixation, a process by which the atoms of atmospheric carbon dioxide are made available to organisms in the form of energy rich molecules such as sucrose. RuBisCO catalyzes either the carboxylation or the oxygenation of Ribulose 1, 5 bisphosphate (also known as RuBP) with carbon dioxide oroxygen.

RuBisCO is very important in terms of Biological impact because it catalyzes the mostcommonlyused chemical reaction by which inorganic carbon enters the biosphere.RuBisCO is also the most abundant protein in leaves, and it may be the most abundantprotein on Earth[7-10].

Nitrogenase Iron Molybdenum Protein

The enzyme responsible for nitrogen fixation, the nitrogenase, shows a high degree ofconservation of structure, function, and amino acid sequence across wide phylogeneticranges. All known Mo-nitrogenase consist of two components, component I (also called dinitrogenase or Fe-Mo protein), an alpha2beta2 tetramer encoded by the nifDand nifK genes, and component II (dinitrogenase reductase, or Fe protein) a homodimerencoded by the nifH gene. Two operons, nifDK and nifEN, encode a tetrameric(alpha2beta2 and N2E2) enzymatic complex. Nitrogenase contains two unusual raremetal clusters, one of them is the iron molybdenum cofactor (FeMo-co), which isconsidered to be the site of dinitrogen reduction and whose biosynthesis requires theproducts of nifNE and of some other nif genes. It has been proposed that NifNE mightserve as a scaffold upon which FeMo-co is built and then inserted into component I[7-10].

Tools and Methodology

NCBI

The amino acid sequences of the proteins were retrieved from the National Center for Biotechnology Information (NCBI) database (Table- 1).

Table 1: Selected proteins from NCBI

Protein	File Number
Heterocyst differentiation Protein	gi|3986486|gb|AAC95054.1|
Glutamine synthetase	gi | 930363 | gb | AAA77684.1
Nitrogenase molybdenum iron protein	gi|20141524|sp|P26254.2|
Ribulose 1,5 bisphosphate carboxylase	gi | 4928660 | gb | AAD33675.F

BLASTp

The amino acid sequences of the selected proteins were submitted to the BLASTp of NCBI database and the results were retrieved. The results include the color keys, Alignment scores and the percentage of structural similarity.

Structure Prediction

Using the above referred amino acid sequences, the primary, secondary and tertiary structure prediction was done using the following tools.

Primary Structure:

Protparam

The amino acid sequences were submitted to PROTPARAM server and the following properties of the protein were predicted: Number of amino acids, Molecular weight, Theoretical pi, Amino acid composition, Total number of negatively charged residues (Asp + GLU), Total number of positively charged residues (Arg + Lys), Atomic composition, Formula, Extinction coefficients, Estimated half-life, Instability index, Aliphatic index and Grand average of Hydropathicity.

Secondary Structure

GOR V

The amino acid sequences were submitted to GOR server. Alpha helix, Pi helix, Beta bridge, extended strand, Beta turn, Bend region, Random coil, Ambiguous states, other states and Secondary structure graph was predicted.

Tertiary Structure

Swissmodel

The amino acid sequences were submitted to the SWISSMODEL server and the Tertiary structure was predicted through Homology modeling. The output gives a protein model and graph from the Protein Data Bank.

I-Tasser

The amino acid sequences were submitted to the serverI-TASSER and Up to five full-length atomic models (ranked based on cluster density), estimated accuracy of the models (including a confidence score of all models, and predicted TM -score and RMSD for the first model), GIF images of the models, secondary structures, Top 10 proteins in PDB which are structurally closest to the models, EC numbers and the confidence score, GO terms and the confidence score, ligand- binding sites and the confidence score, An image of the ligand-binding sites were predicted

Results and Discussion

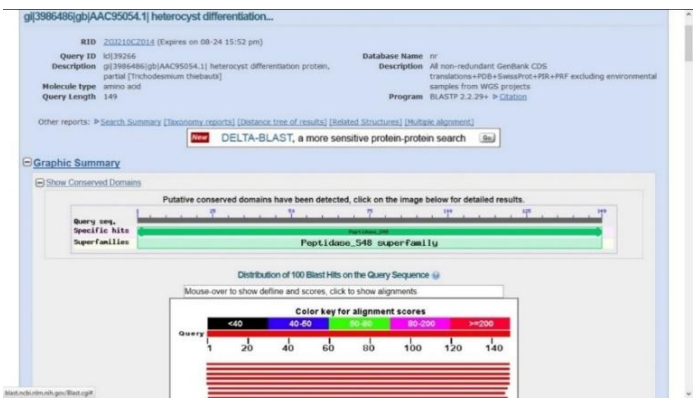
BLASTp

The following screenshots shows the Query id, description, molecular type, Query length, name of the database that was used, the program that was run on the query sequence. The graphical summary shows the color keys. According to the result the submitted query proteins has alignment score greater than 200 (red).

HetR

The HetR of Trichodesmium thiebautii is 29% similar to the Dihydropicolinate Reductase In complex with Nadh and 2, 6 Pyridine Dicarboxylate of Escherichia coli. This shows the wide variation of the structure from other proteins from the protein data bank (Figure- 1).

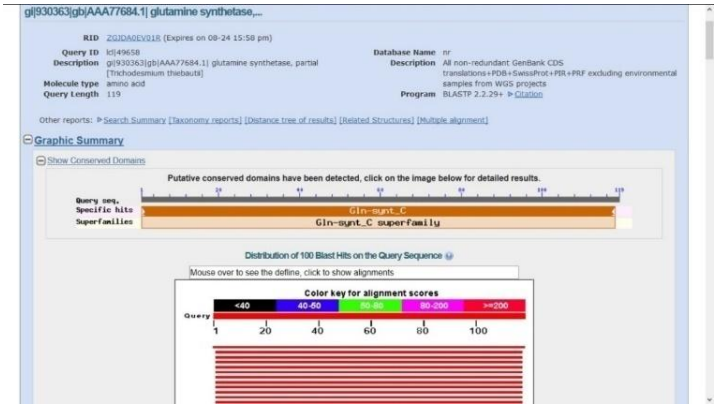
Figure1: Screenshot of blast result for Hetr



Glutamine Synthetase

It is found out that the protein is structurally 76% similar to the Glutamine synthetase of Salmonella typhimurium (Figure 2).

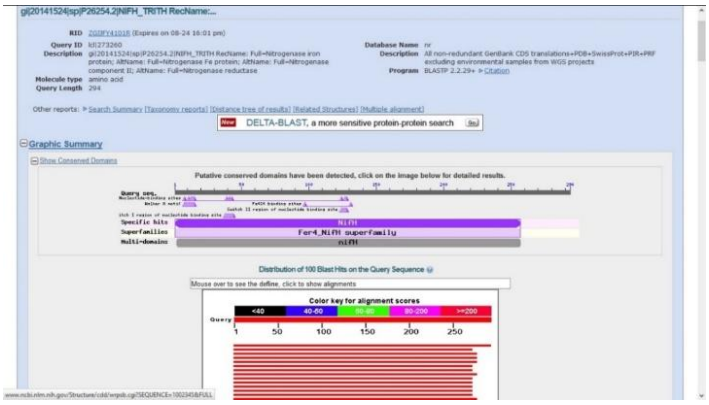
Figure 2: Screenshot of blast result for GS



NifD

The protein is closest to the NifD protein of Klebsiella pneumonia, which has 59% identity with it (Figure 3).

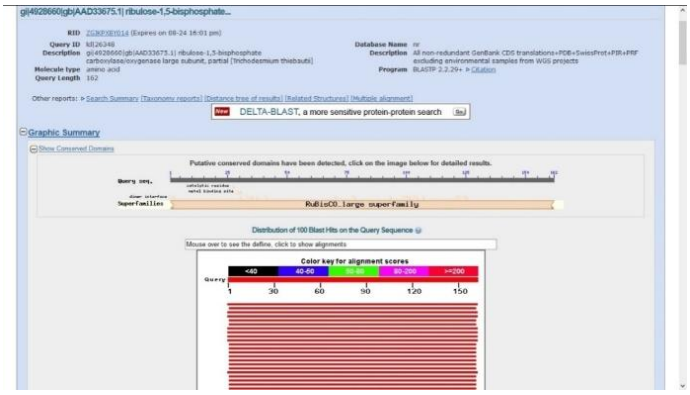
Figure 3: Screenshot of blast result for Hetr



RuBisCO

It is similar to the RuBisCO protein of synechoccous which 88% identical (Figure4).

Figure 4: Screenshot of blast result for RuBisCO



Primary Structure

Protparam

Results from Protparam server was tabulated below (Table 2).

Table 2: Result comparison of Protparam.

Properties		**Protein**							
		HetR		Gln syn		NifD		RuBisCO	
No of amino acid		149		119		294		162	
Molecular weight (da)		17187.9		13074.7		32276.1		18101.8	
Theoretical PI		7.98		8.67		5.84		9.32	
Amino acid composition	Residue	No Of Residue	%	No Of Residue	%	No Of Residue	%	No of residue	%
	A	5	3.4%	14	11.8%	29	9.9%	14	8.6%
	R	11	7.4%	3	2.5%	16	5.4%	12	7.4%
	N	4	2.7%	8	6.7%	15	5 .1%	6	3.7%
	D	3	2.0%	4	3.4%	13	4 .4%	7	4.3%
	C	2	1.3%	2	1.7%	6	2.0%	5	3.1%
	Q	7	4.7%	5	4.2%	16	5.4%	4	2.5%
	E	16	0.7%	6	5.0%	25	8.5%	9	5.6%
	G	10	6.7%	10	8.4%	27	9.2%	17	10.5 %
	H	5	3.4%	8	6.7%	5	1.7%	9	5.6%
	I	9	6.0%	9	7.6%	22	7.5%	9	5.6%
	L	17	11.4%	5	4.2%	26	8.8%	14	8.6%
	K	9	6.0%	9	7.6%	18	6.1%	11	6.8%
	M	4	2.7%	2	1.7%	10	3.4%	4	2.5%
	F	4	2.7%	4	3.4%	7	2.4%	7	4.3%
	P	13	8.7%	3	2.5%	9	3.1%	5	3.1%
	S	7	4.7%	7	5.9%	7	2.4%	4	2.5%
	T	7	4.7%	5	4.2%	14	4.8%	10	6.2%
	W	2	1.3%	0	0.0%	0	0.0%	2	1.2%
	Y	5	3.4%	7	5.9%	8	2.7%	4	2.5%
	V	9	6.0%	8	6.7%	21	7.1%	9	5.6%
	O	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	U	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Total number of negatively charged residues (Asp+Glu)		19		10		38		16	
Total number of positively charged residues (Arg+Lys)		20		12		34		23	
Atomic composition	Carbon	772		580		1410		801	
	Hydrogen	1228		896		2297		1267	
	Nitrogen	214		166		401		239	
	Oxygen	218		172		431		223	
	Sulfur	6		4		16		9	
Formula		$C_{772}H_{1228}N_{214}O_{218}S_6$		$C_{580}H_{896}N_{166}O_{172}S_4$		$C_{1410}H_{2297}N_{401}O_{431}S_{16}$		$C_{801}H_{1267}N_{239}O_{223}S_9$	
Extinction coefficient		18575		10555		12295		17210	

Estimated half life	2.8 hours (mammalian reticulocytes, in vitro). 10 min (yeast, in vivo). 2 min (Escherichia coli, in vivo).	>20 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo).	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).	1.3 hours (mammalian reticulocytes, in vitro). 3 min (yeast, in vivo). 3 min (Escherichia coli, in vivo).
Instability index	70.50	42.75	37.92	53.14
Aliphatic index	88.93	77.14	94.25	80.12
Grand average of hydropathicity	-0.493	-0.365	-0.200	-0.323

Extinction Coefficients

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for following a protein which a spectrophotometer when purifying it. It has been shown that it is possible to estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition. From the molar extinction coefficient of tyrosine, tryptophan and cystine (cysteine does not absorb appreciably at wavelengths >260 nm, while cystine does) at a given wavelength, the extinction coefficient of the native protein in water can be computed using the following equation:

E (Prot) = No of Tyr*Ext (Tyr) + NO of Trp*Ext (Trp) + No of Cys*Ext (Cys)

Where (for proteins in water measured at 280 nm)

- Ext (Tyr) = 1490
- Ext (Trp) = 5500
- Ext(Cys) = 125

The absorbance (optical density) can be calculated using the following formula:

Absorb (Prot) = E (Prot) / Molecular weight

Two values are produced by Protparam based on the above equations, both for proteins measured in water at 280 nm. The first one shows the computed value basedon the assumption that all cysteine residues appear as half cystines, and the secondone assuming that no cysteine appears as half cystine. Experience shows that thecomputation is quite reliable for proteins containing Trp residues, however there may bemore than 10% error for proteins without Trp residues. The extinction coefficients of the proteins are given in the above table.

In Vivo Half-Life

The half-life is a prediction of the time it takes for half of the amount of protein in a cellto disappear after its synthesis in the cell. Protparam relies on the "N -end rule", whichrelates the half-life of a protein to the identity of its N -terminal residue; the prediction isgiven for 3 model organisms (human, yeast and E.coli). The N -end rule originated fromthe observations that the identity of the N -terminal residue of a protein plays animportant role in determining its stability in vivo. The rule was established fromexperiments that explored the metabolic fate of artificial beta-galactosidase proteinswith different N -terminal amino acids engineered by site -directed mutagenesis. Thebeta -gal proteins thus designed have strikingly different half-lives in vivo, from morethan 100 hours to less than 2 minutes, depending on the nature of the amino acid at theamino terminus and on the experimental model (yeast in vivo; mammalian reticulocytesin vitro, Escherichia coli in vivo). In addition, it has been shown that in eukaryotes, theassociation of a destabilizing N -terminal residue and of an internal lysine targets theprotein to ubiquitin-mediated proteolytic degradation. Note that the program gives anestimation of the protein half-life and is not applicable for N -terminally modified proteins.

Instability Index (II)

The instability index provides an estimate of the stability of your protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed that there are certain dipeptides, the occurrence of Mitch is significantly different in the unstable proteins compared with those in the stable ones. The authors of this method have assigned a weight value of instability to each of the 400 different dipeptides (DIVVV). Using these weight values it is possible to compute an instability index (11) which is defined as:

i=L-1
II = (10/L) * Sum DIVVV(x[i]x[i+1])

- Where: L is the length of sequence
- DIVVV(x[i]x[i+1]) is the instability weight value for the dipeptide starting in position i.

Protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. From the results obtained it is predicted that HetR is highly unstable, whereas the other 3 proteins can be stable due to their values closer to 40.

Aliphatic Index

The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermo stability of globular proteins. The aliphatic index of a protein is calculated according to the following formula:

Aliphatic index = X (Ala) + a * X (Val) + b * (X (Ile) + X (Leu))

- Where X (Ala), X (Val), X (11e), and X (Leu) are mole percent (100 X mole fraction)of alanine, valine, isoleucine, and leucine.
- The coefficients a and b are the relative volume of valine side chain (a = 2.9)
- And of Leu/lle side chains (b = 3.9) to the side chain of alanine.

Gravy (Grand Average of Hydropathicity)

The GRAVY value for a peptide or protein is calculated as the sum of hydropathy valuesof all the amino acids, divided by the number of residues in the sequence. Since theabove proteins are lesser than 0, they are predicted to be Hydrophobic in character.

Secondary Structure
Gor V

GOR V is the fourth version of GOR secondary structure prediction methods based on the information theory. There is no defined decision constant. GOR V uses all possible pair frequencies within the window of 17 amino acid residues (Table 3).

Table 3: results of GOR V server

Prpperties	**Proteins**			
ALPHA HELIX (%)	32.21	30.25	41.33	37.65
3_{10} HELIX (%)	0	0	0	0
Pi HELIX (%)	0	0	0	0
BETA BRIDGE (%)	0	0	0	0
EXTENDED STRAND (%)	20.81	26.05	26.67	20.37
BETA TURN (%)	0	0	0	0
BEND REGION (%)	0	0	0	0
RANDOM COIL (%)	46.98	43.70	32.00	41.98
AMBIGUOUS STATE (%)	0	0	0	0
OTHER STATES (%)	0	0	0	0

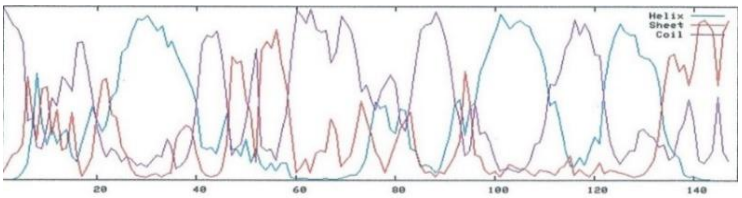
From the result predicted it is seen that only the percentage of alpha helix, extended strand and random coil are obtained. This is due to the fragmented amino acid sequences used and also due to the absence of other structural properties.

Graphs

In these graphs (Figure 5, 6, 7 and 8) the alpha helix is shaded in blue, extended strand or sheet is shaded in red and the random coil is shaded in purple.

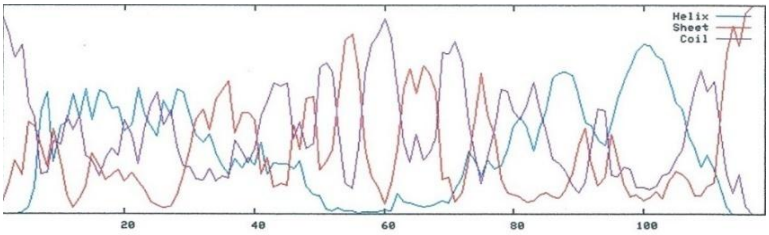
Heterocyst Differentiation Protein (Figure 5)

Figure 5: Heterocyst Differentiation Protein



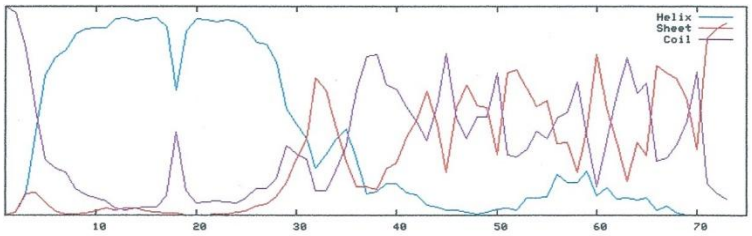
Glutamine Synthetase (Figure 6)

Figure 6: Glutamine Synthetase



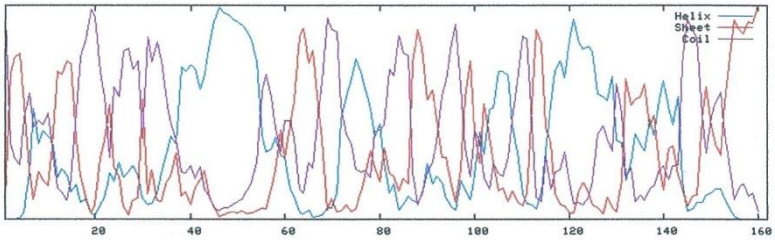
Nitrogenase Molybdenum Iron Protein (Figure 7)

Figure 7: Nitrogenase Molybdenum Iron Protein



Ribulose 1, 5 Biphosphate Carboxylase (Figure 8)

Figure 8: Ribulose 1, 5 Biphosphate Carboxylase



Tertiary Structure

I-Taser

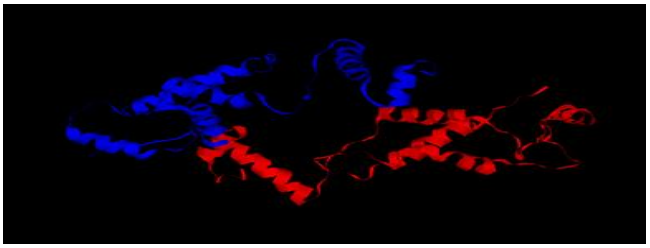
Hetr

Since this protein is not closely similar to any other protein to predict the structure by homology modeling, it is predicted through I-TASSER by protein threading.

Top 5 Structure Model

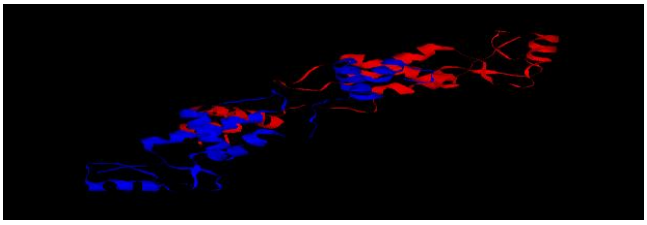
(1)C score = -4.12 (**Figure9)**

Figure 9: C score value of matching model 1



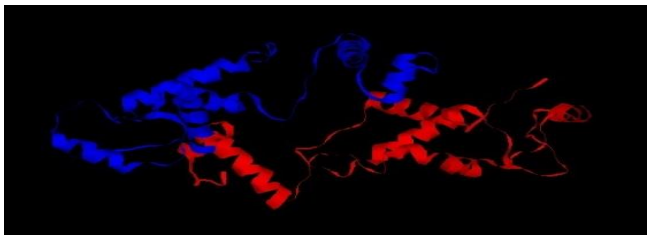
(2) C score = -4.45(**Figure10)**

Figure 10: C score value of matching model 2



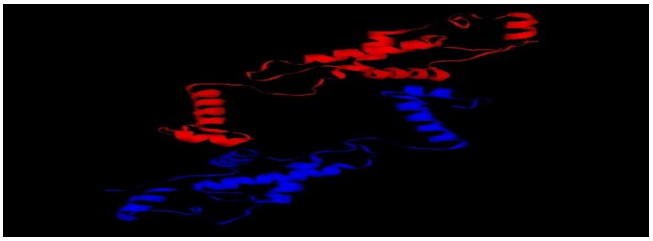
(3) C score = -4.75(**Figure11)**

Figure 11: C score value of matching model 3



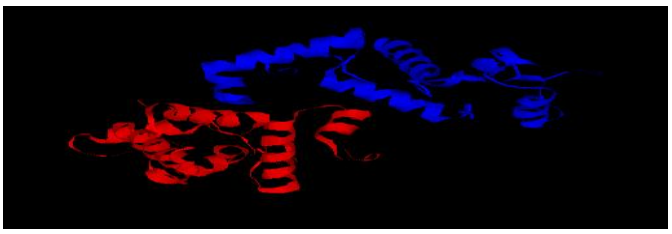
(4) C score = -5.21(**Figure12)**

Figure 12: C score value of matching model 4



(5) C score = -4.51(**Figure13)**

Figure 13: C score value of matching model 5



C -score is a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations (Figures 9, 10, 11, 12 and 13). C -score is typically in the range of [-5, 2], where

a C -score of higher value signifies a model with a high confidence and vice -versa. From the C Scores obtained it is predicted that the protein modeled is within the limits of the confidence score (Table 4).

Table 4: Top Ten structurally similar proteins

Rank	TM- score	RMSDa	IDENa	Cov.	PDB hit
1	0.5675	3.57	0.07	0.81	1jr3A
2	0.5525	3.57	0.07	0.79	1jr3A
3	0.5132	3.31	0.08	0.71	1a5Ta
4	0.5127	4.99	0.08	0.86	1t33A
5	0.5117	3.50	0.11	0.72	1a5tA
6	0.5088	4.70	0.06	0.85	2fbqA
7	0.5037	4.43	0.09	0.79	2d6yA
8	0.5005	4.61	0.06	0.83	1ej6B
9	0.4984	3.78	0.10	0.74	1jr3D
10	0.4980	4.53	0.06	0.81	1ej6B

Ranking of proteins is based on TM -score of the structural alignment between Model 1 and the PDB structures in our template library. TM -score is a recently proposed scale for measuring the structural similarity between two structures the purpose of proposing TM -score is to solve the problem of RMSD which is sensitive to the local error. Because RMSD is an average distance of all residue pairs in two structures, a local error will arise a big RMSD value although the global topology is correct. In TM -score, however, the small distance is weighted stronger than the big distance which makes the score insensitive to the local modeling error. A TM -score >0.5 indicates a model of correct topology and a TM-score<0.17 means a random similarity. These cutoffs do not depend on the protein length. RMSDa is the RMSD between residues that are structurally aligned by TM -align. IDENa is the percentage sequence identity in the structurally aligned region. Cov, represents the coverage of the alignment by TM -align and is equal to the number of structurally aligned residues divided by length of the model.

3.4.1.1.2. Function Prdiction

Ranking is based on EC -score. RMSDa is the RMSD between models and the PDB structure in the structurally aligned regions by TM -alignment. IDENa is percentage sequence identity in the structurally aligned region. Cov, represents the coverage of the alignment and is equal to the number of structurally aligned residues divided by length of model (Table 5).

Table 5: Top 5 EC numbers

Rank	**TM-score**	**RMSDa**	**ITENa**	**Cov.**	**EC no.**	**EC-score**	**PDB hit**
1	0.1564	3.22	0.19	0.20	3.6.3.14	0.7665	1pyvA
2	0.0984	1.50	0.14	0.11	2.3.1.61	0.7091	2cyuA
3	0.865	1.66	0.14	0.09	2.3.1.48	0.6735	2prgC
4	0.3739	4.71	0.16	0.61	3.4.13.9	0.6709	1wy2A
5	0.2711	4.74	0.16	0.44	3.1.27.5	0.6687	1rraA

EC -Score is a confidence score for the Enzyme Classification (EC) Number prediction, which is defined based on the C -score of the structure prediction and the TM -score, RMSDa, IDENa of the structural alignment by TM -align between the predicted models and the PDB structures. A prediction with an EC -score >1 signify a prediction with high confidence (up to 3 digit numbers of EC) and vice versa. The first ranking EC score is ATP Synthase. Hence the protein is predicted to functionally closer to ATP synthase.

The GO term is based on three structured controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species -independent manner. GO -Score is defined as a relative frequency of the GO terms appearing in the top 10 functional homologs. A prediction with a GO -score >0.6 signify a prediction with high confidence and vice versa.According to the GO terms, this protein is predicted to be involved in DNA binding which is a molecular function (Table 6).

Table 6: predicted GO terms

Rank /protein	HetR 1-GO score
1	0003677-0.6
2	0006355-0.3
3	0008152-0.3
4	0006350-0.3
5	0003700-0.3
6	0045449-0.3
7	0055114-0.2
8	0006260-0.2
9	0046872-0.2
10	0009360-0.2

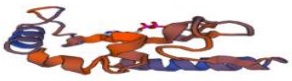
3.4.2. Swiss – Model

Tertiary structure of the protein were modeled by the SWISS- MODEL server and it is viewed with Qmol pdb viewer (Figures 14, 20, 23 and 26). Here the coils are colored in red, Strands in blue, a helix in orange, turns in brown and bulges in pink.

3.4.2.1. Glutamine Synthetase

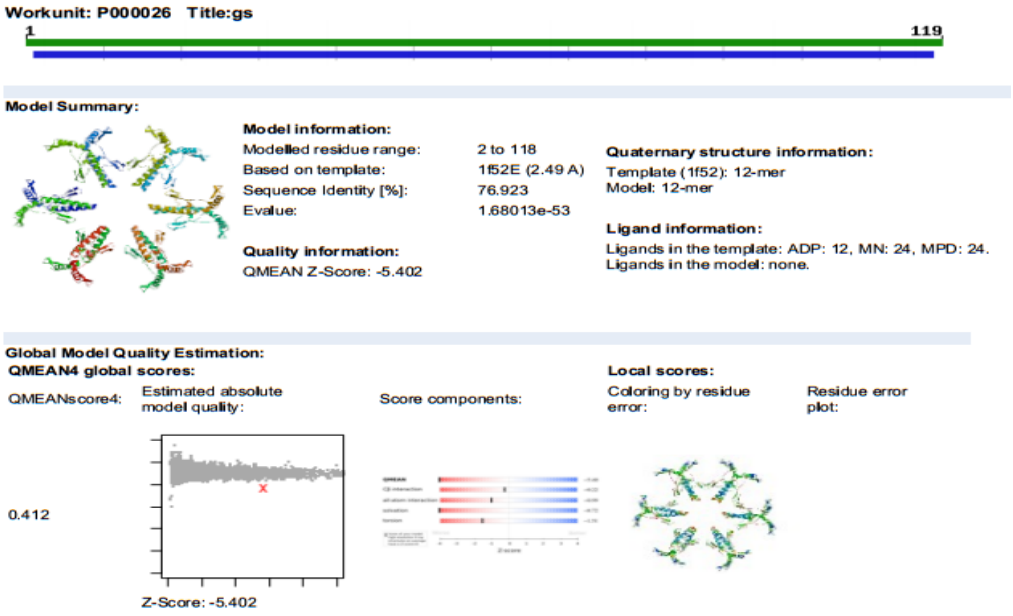
SWISS MODEL of Glutamine synthetase (Figure14)

Figure 14: Swiss modeling of Glutamine synthetase



Swiss Model Workspace Screenshot (Figure 15)

Figure 15: Swiss model workspace screenshot of Glutamine synthetase



This is the screenshot of the result for this particular protein from the SWISS – MODEL server. Here the query sequence submitted was named as "gs". The corresponding work unit allocated by the server was P000026. The results show the modeled residue range, basing on which template the protein was modeled, sequence identity with thetemplate and the E value of the model (Figure 15).

This below graph here represents the ANOLEA, GROMOS and Qmean predictions of the protein (Figures 16, 17, 18 and 19).

Figure 16: Results from swiss model for Glutamine synthetase

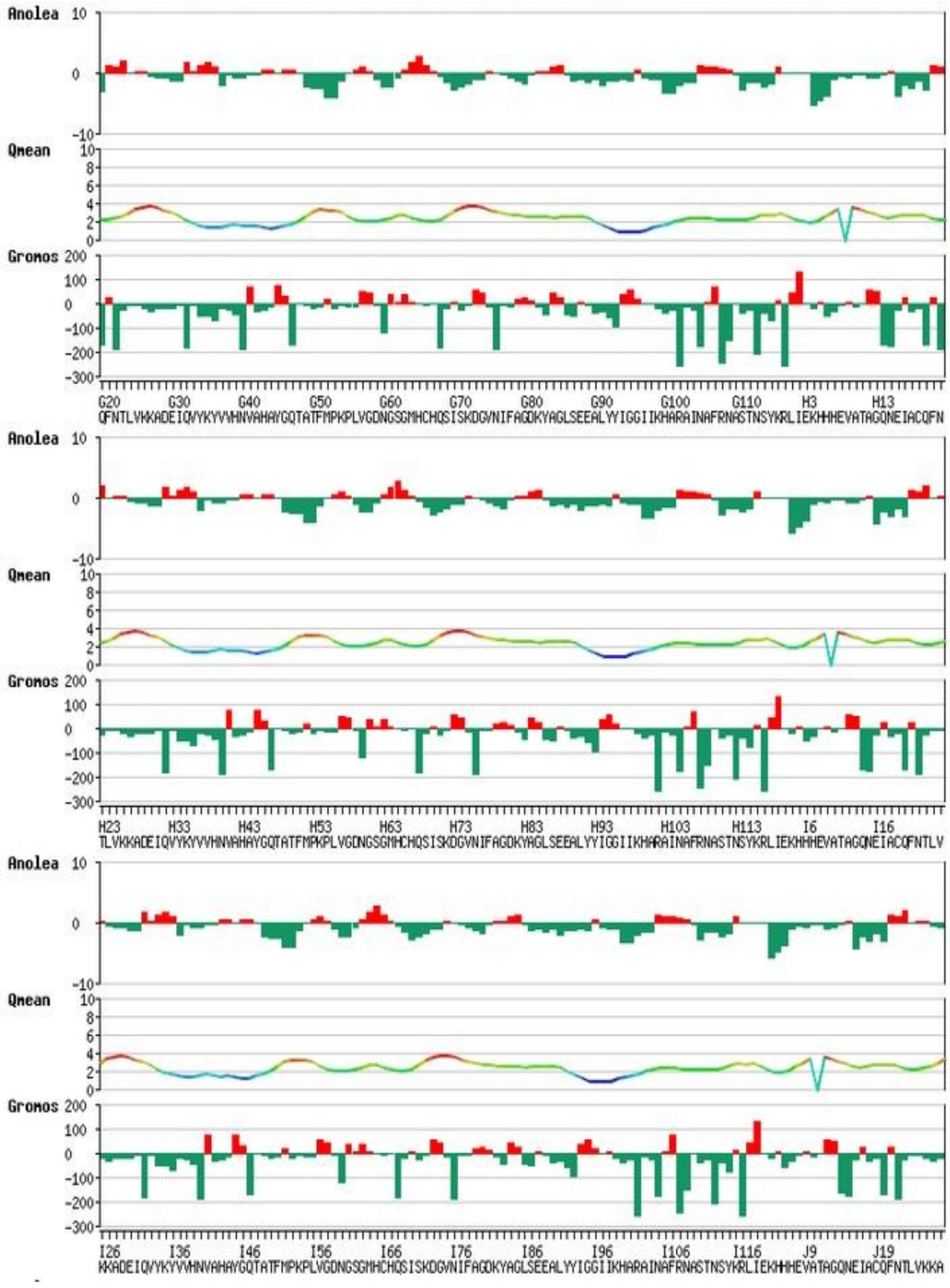


Figure 17: Results from swiss model for Glutamine synthetase

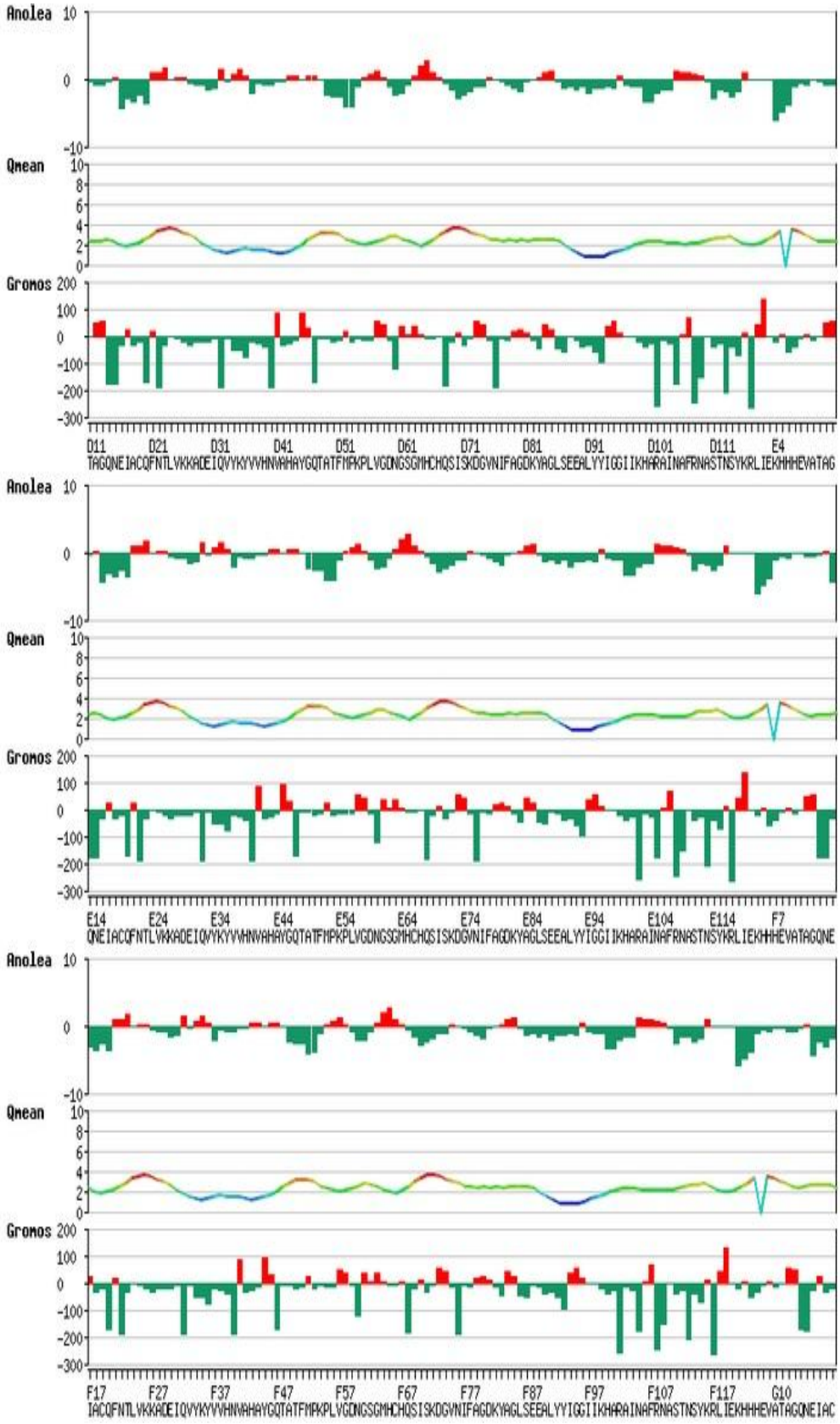


Figure 18: Results from swiss model for Glutamine synthetase

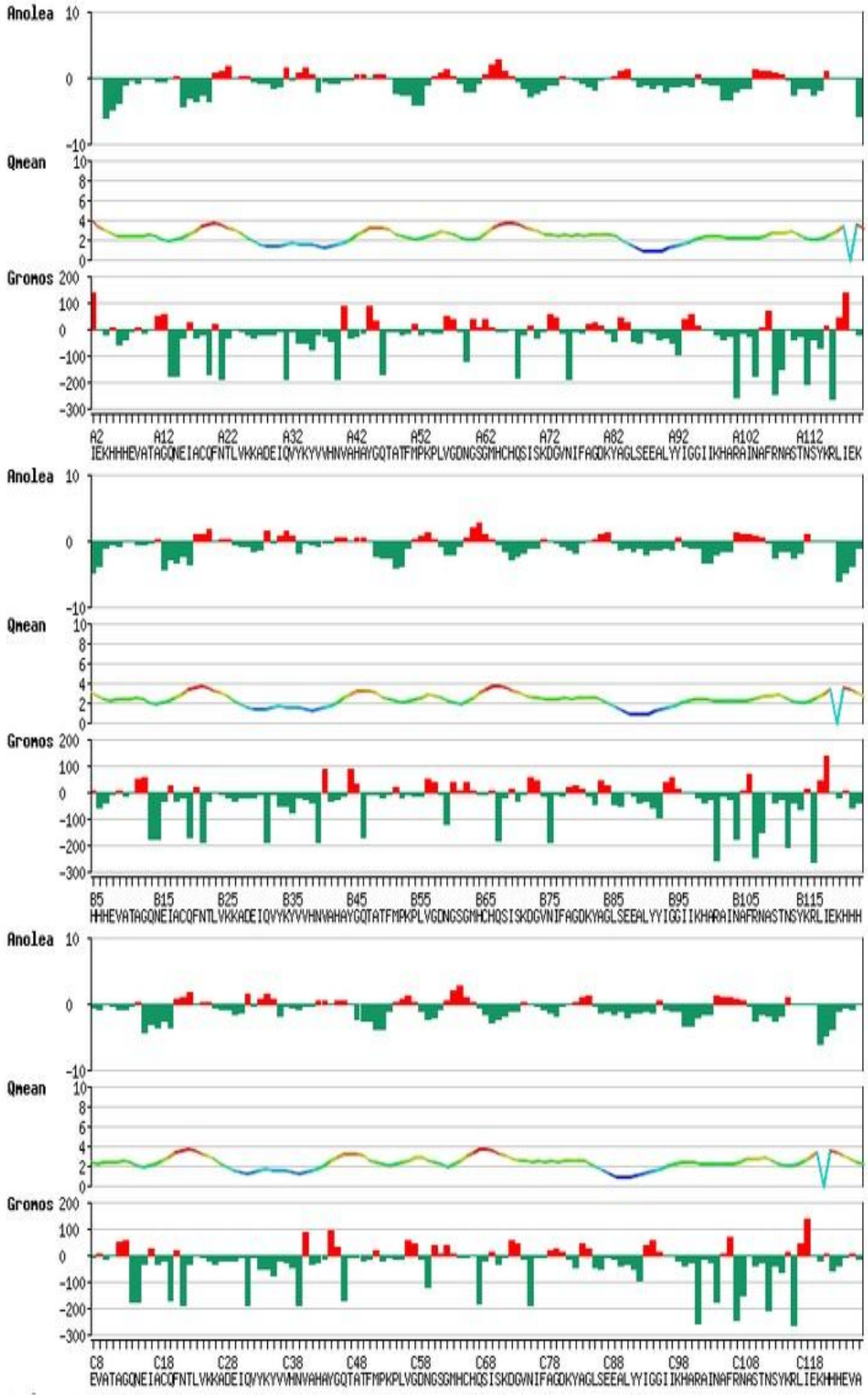
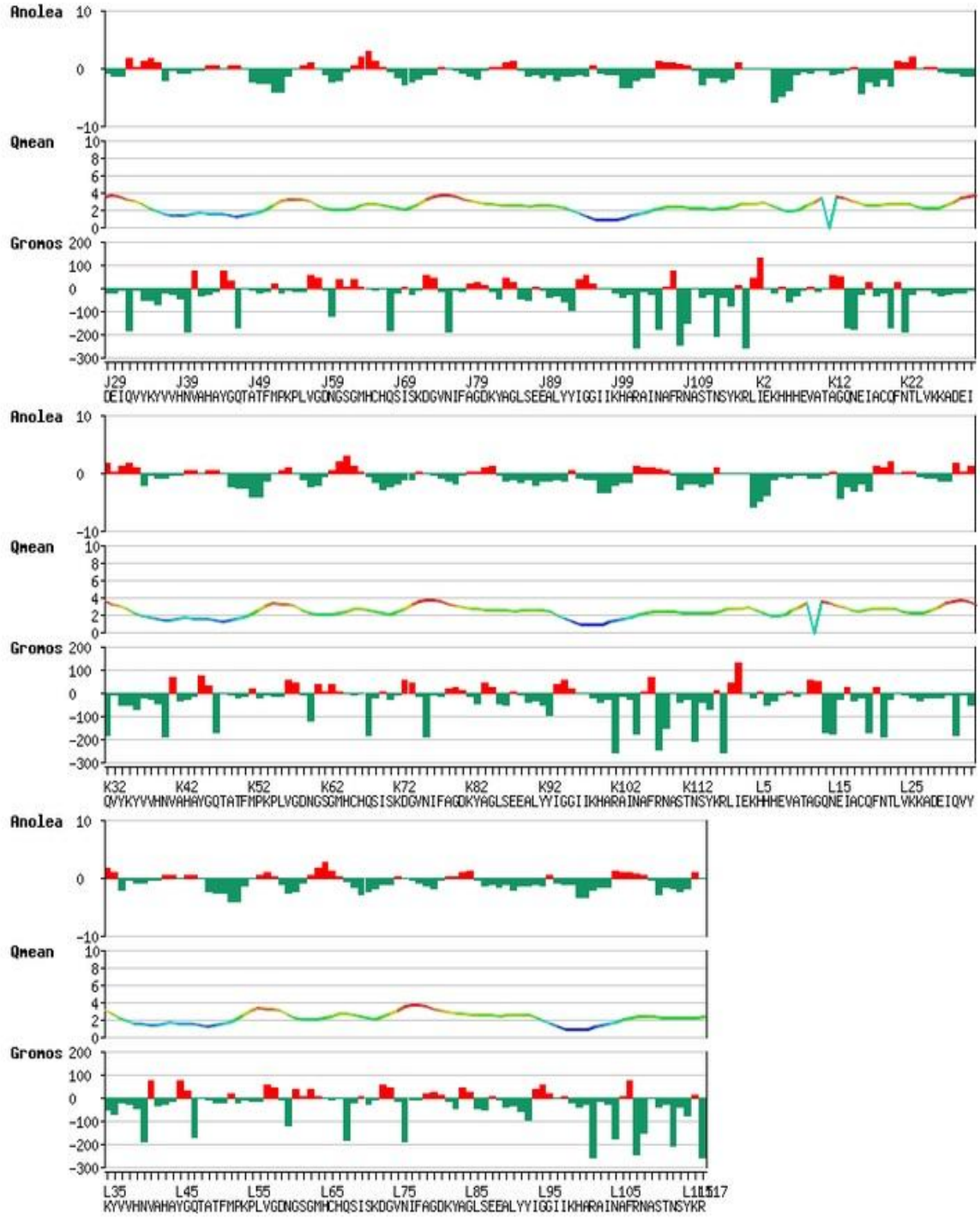


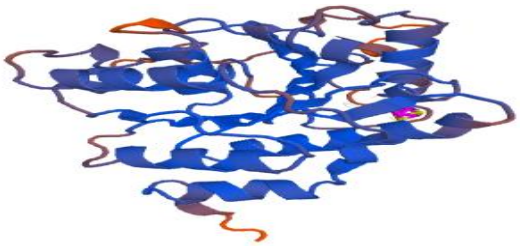
Figure 19: Results from swiss model for Glutamine synthetase



3.4.2.2. Nitrogenase Molybdenum Iron Protein

Swiss Model of Nitrogenase Molybdenum Iron Protein (Figure 20)

Figure 20: Swiss Modeling of Nifd



Swiss Model Workspace Screenshot (Figure 21)

Figure 21: Swiss Modeling workspace for Nifd

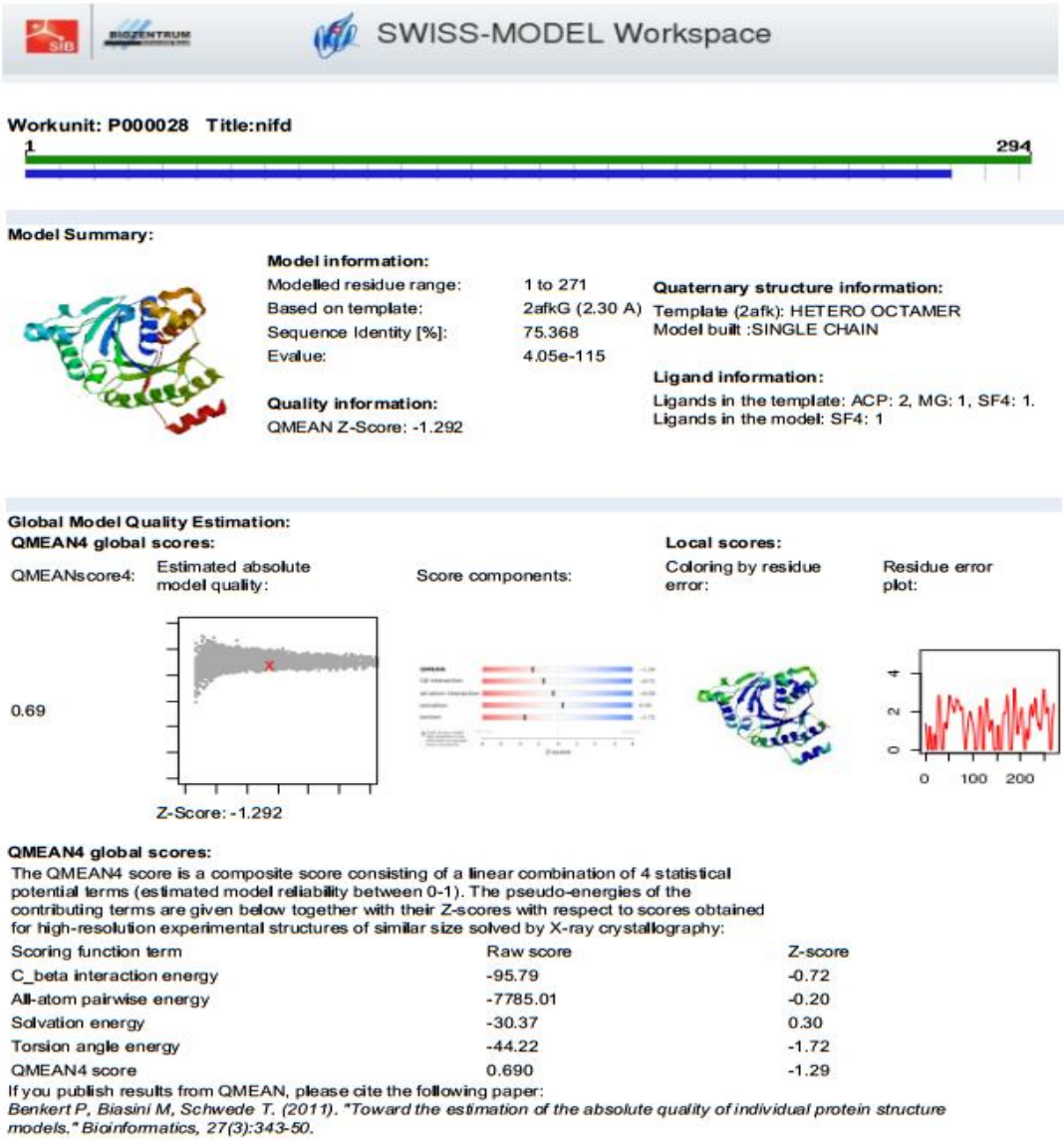
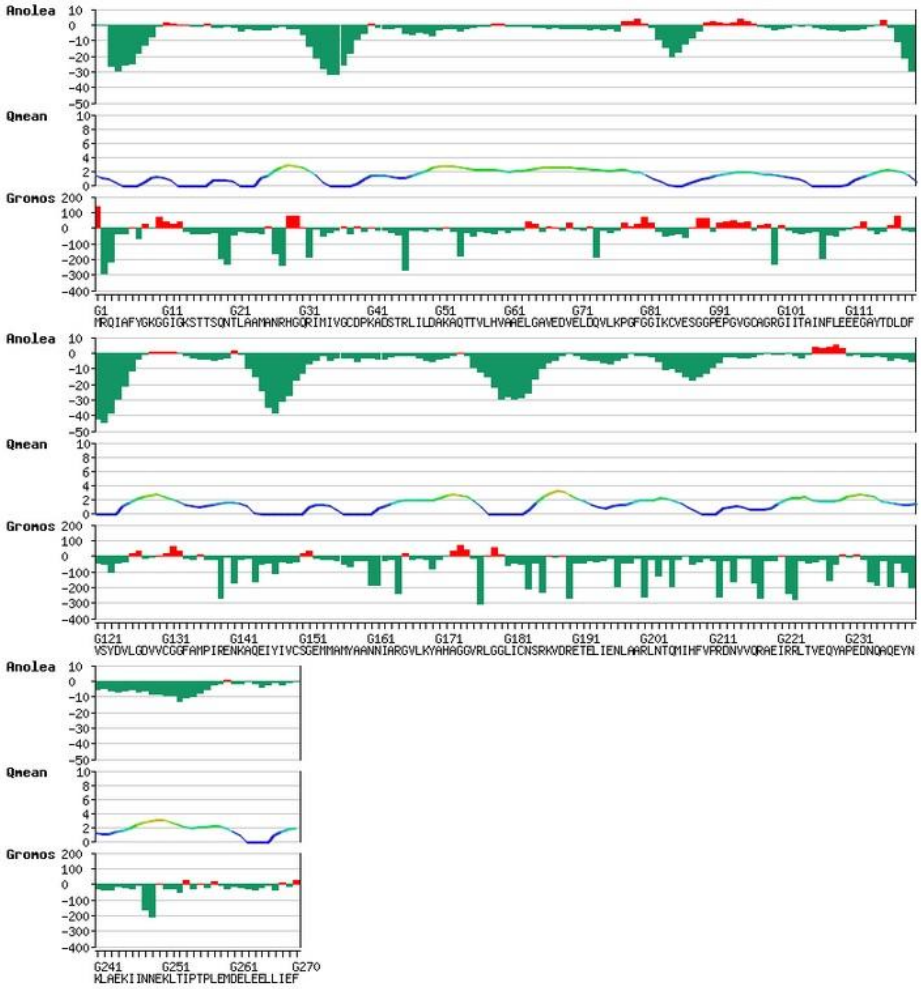


Figure 22: Swiss Modeling result for Nifd

This is the screenshot of the result for this particular protein from the SWISS – MODEL server. Here the query sequence submitted was named as "nifd". The corresponding work unit allocated by the server was P000028. The results show the modeled residue range, basing on which template the protein was modeled, sequence identity with the template and the E value of the model (Figure 21).

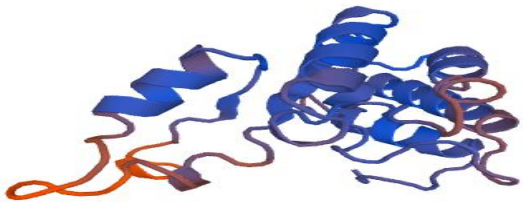
Graph here represents the Anolea, Gromos and Qmean predictions of the protein (Figure 22).



3.4.2.3. Ribulose 1, 5 Biphosphate Carboxylase

Swiss Model of Rubisco (Figure 23)

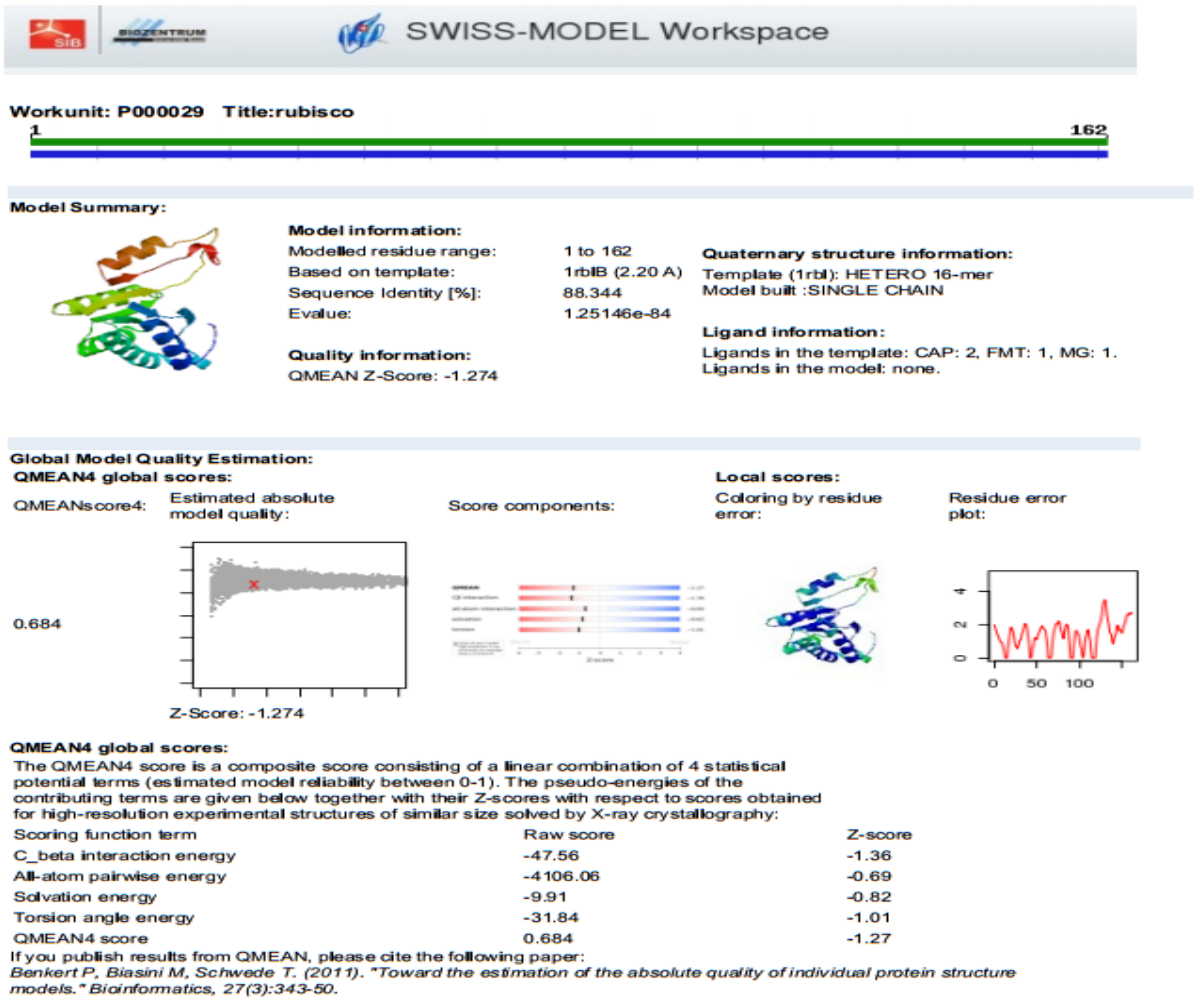
Figure 23: Swiss Modeling of RuBisCO



Swiss Model Workspace Screenshot (Figure 24)

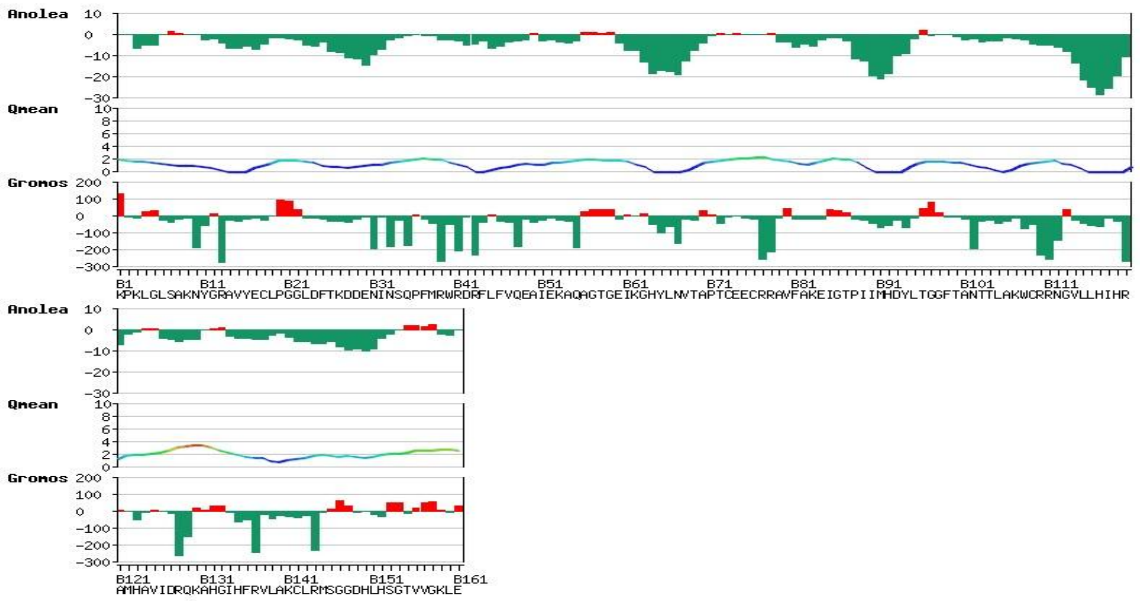
This is the screenshot of the result tor this particular protein from the SWISS – MODEL server. Here the query sequence submitted was named as “rubisco”. The corresponding work unit allocated by the server was P000029. The results show the modeled residuerange, basing on which template the protein was modeled, sequence identity with thetemplate and the E value of the model (Figure 24).

Figure 24: Swiss model workspace for RuBisCO



Graph here represents the Anolea, Gromos and Qmean predictions of the protein (Figure 25).

Figure 25: Swiss model result for RuBisCO



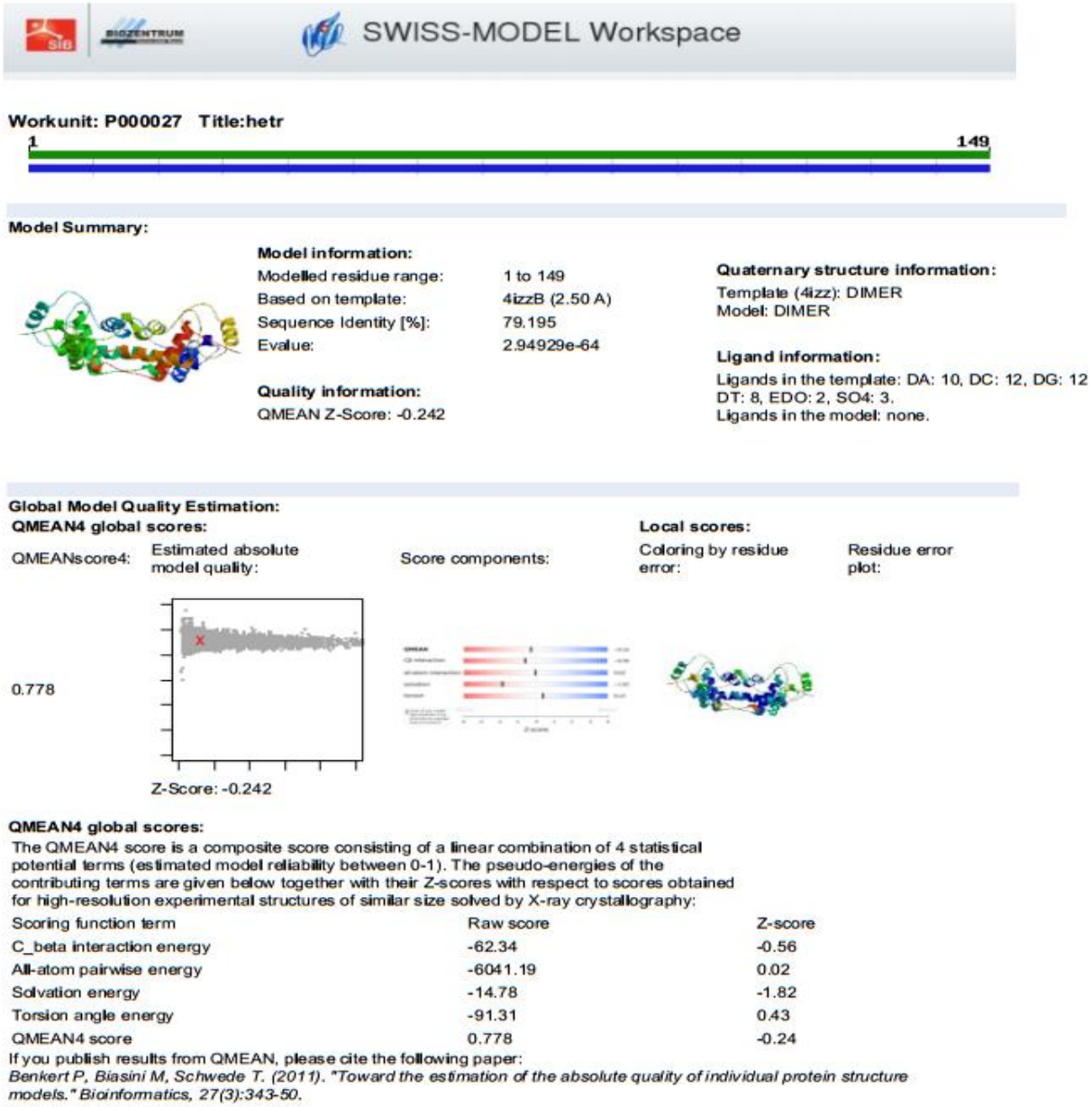
3.4.2.4. Heterocyst Differentiation Protein

Swiss Model Of Hetr (Figure 26)

Figure 26: Swiss modeling of HetR

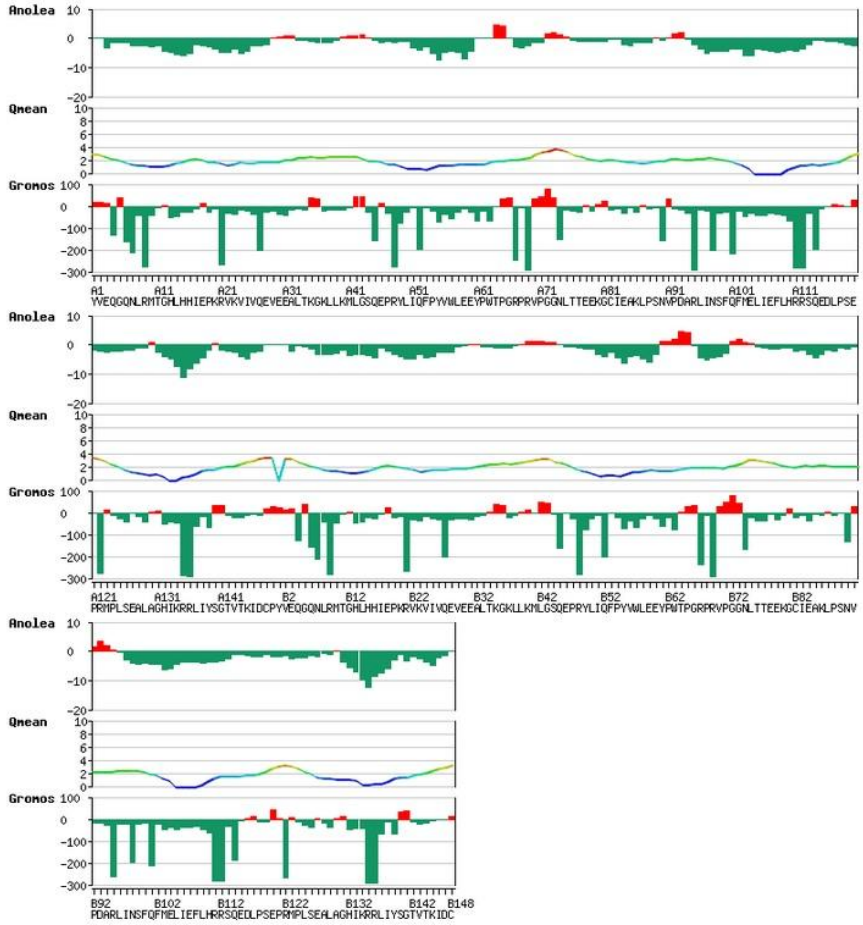
Swiss Model Workspace Screenshot (Figure 27)

Figure 27: Swiss model work space for HetR



This is the screenshot of the result for this particular protein from the SWISS – MODEL server. Here the query sequence submitted was named as Hetr. The corresponding work unit allocated by the server was P000010. The results show the modeled residue range, basing on which template the protein was modeled, sequence identity with the template and the E value of the model (Figure 27).

Figure 28: Swiss model result for HetR



Graph here represents the Anolea, Gromos and Qmean predictions of the protein (Figure 28).

4. Conclusion

Four different amino acid sequence was retrieved from NCBI database namely heterocyst differentiation protein, Glutamine synthetase, Nitrogenase molybdenum iron protein and Ribulose 1,5 bisphosphate carboxylase of the cyanobacteria- Trichodesmium thiebautii.The retrieved proteins were submitted to BLASTp of NCBI and the results were taken.

The primary, secondary and tertiary structure of the proteins waspredicted using the following Insilco tools of Bioinformatics.

The Primary structure of the proteins was predicted by using PROTPARAM – anExPASy server. The molecular weight, amino acid composition, atomiccomposition, extinction coefficient, estimated halt -life, instability index, aliphatic indexand grand average of hydropathicity (GRAVY) of the selected proteins were predictedand the results were tabulated.

The secondary structure of the selected proteins was predicted by using GOR IV – anExPASy server. The Alpha helix, Pi helix, Beta Bridge, Extended strand, Beta turn, Bendregion, Random coil, Ambiguous states, other states and Secondary structure graph ofthe protein were predicted and the results were tabulated.

The tertiary structure of the protein, Heterocyst differentiation protein was predicted byI-TASSER server. The protein was modeled by protein threading due to its widevariation from any other known structure from the Protein Data Bank. Up to five full-length atomic models (ranked based on cluster density), estimated accuracy of themodels (including a confidence score of all models, and predicted TM –scoreand RMSD for the first model), Top 10 proteins inPDB which are structurally closest to the models, EC numbers and the confidence score, GO terms and the confidence score were predicted for this particular protein and the results were tabulated.

The tertiary structure of other proteins - Glutamine synthetase, Nitrogenase molybdenum iron protein and Ribulose 1, 5 bisphosphate carboxylase was predicted by using SWISS - MODEL server from ExPASy_3D model of the proteins, graph with Anolea, Gromos and Qmean was predicted.

From the study done here, it is concluded that the Proteins Heterocyst differentiation protein and Nitrogenase molybdenum iron protein, structurally vary widely from any other known protein from the Protein Data Bank. It is also found out that all the predicted models have a major role in the Nitrogen fixation Mechanism of the Cyanobacteria - Trichodesmium thiebautii.

In vitro studies and further research on Trichodesmium thiebautii may reveal the uniqueness and significant properties of the proteins and toxins present in it.

References

1. Bergman, B., and E. J. Carpenter. 1991. Nitrogenase confined to randomly distributed trichomes in the marine cyanobacterium Trichodesmium thiebautii. J. Phycol. 27: 158-165.
2. Capone, D. G., J. M. O'neil, J. Zehr, And E. J. Carpenter. 1990. Basis for diet variation in nitrogenase activity in the marine planktonic cyanobacterium Trichodesmium thiebautii. Appl. Environ. Microbial. 56: 3532-3536.
3. Matthijs, J. C. P., and H. J. Lubberd1ng. 1988. Dark respiration in cyanobacteria, p. 131-145. Zn Biochemistry of the algae and cyanobacteria. Proc. Phytochem. Sot. Eur. Clarendon
4. Carpenter, E.J., D.G. Capone, and J.G. Reuter [Eds.]. 1992. Marine pelagic cyanobacteria: Trichodesmium and other diazotrophs. Kluwer. AND OTHERS. 1990. Re-evaluation of nitrogenase oxygen -protective mechanisms in the planktonic marine cyanobacterium Trichodesmium. Mar. Ecol. Prog. Ser. 65: 151-158.
5. Hoch, G., 0. H. Owens and B. Kok. 1963. Photosynthesis and respiration. Arch. Biochem. Biophys. 101: 171-180.
6. Miller, A. G., G. S. Esp1e, and D. T. Canv1n. 1988. Active transport of inorganic carbon increases the rate of 0, photoreduction by the cyanobacterium Synechococcus UTEX 625. Plant Physiol. 88: 6-9.
7. Ohk1, K., P. G. Falkowski, J. G. Rueter, and Y. Fujita. 199 1. Experimental study of the marine cyanophyte Trichodesmium sp., a nitrogen -fixing phytoplankton in tropical and subtropical sea area, p. 205-2 16. In Marine biology, its accomplishment and future prospect.
8. Hokucensha and Y. Fujita. 1988. Aerobic nitrogenase activity measured as acetylene reduction in the marine non-heterocystous cyanobacterium Trichodesmium spp. grown under artificial conditions. Mar. Biol. 98: 111-114.
9. Paerl, H. W., and B. Bebout. 1992. Oxygen dynamics in Trichodesmium spp. aggregates, p. 43-59. Zn E. J. Carpenter et al. [eds.], Marine pelagic cyanobacteria: Trichodesmium and other diazotrophs. Kluwer.
10. Bunt, J. S., K. E. Cooksey, M. A. Heed, C. C. Leiz, and B. F. Taylor. 1970. Assay of algal nitrogen fixation in the marine subtropics by acetylene reduction. Nature (Lond.)227: 1163-1164.

*****