American Journal of Bioinformatics Research 2013, 3(3): 42-61
DOI: 10.5923/j.bioinformatics.20130303.02

Sequence, Structural and Functional Characterization of Homogentisate-1,2-dioxygenase of *homo sapiens*: An *in silico* Analysis

Sen Gupta Parth Sarthi, Banerjee Shyamashree, Bandyopadhyay Amal Kumar*

Department of Biotechnology, The University of Burdwan, Burdwan, West Bengal, 713104, India

Abstract In silico characterization and molecular modeling of a protein opens wide scope for the prediction of structural and functional information. It is most significant and helpful when very little information about three dimensional (3D) structure of protein(uncharacterized) available in the Protein Data Bank (PDB). So, the present study has been undertaken to carried out *in silico* prediction of structure and function of Homogentisate-1,2-dioxygenase of *homo sapiens*. Primary structure analysis reveals that all the six sequences of Homogentisate-1,2-dioxygenase are hydrophobic in nature (due to the high content of non- polar residues) without transmembrane region. The aliphatic index computation infers that Homogentisate-1,2-dioxygenase of *homo sapiens* can't tolerate wide range of temperature. Secondary structure analysis shows that all of the sequences of Homogentisate-1,2-dioxygenase have predominant sheet and random coil structure. Sub cellular localization prediction suggested that all of the sequences are secretory without transit peptide. 3D structure of the protein is predicted and characterized. Energy minimization, Refinement and validation of the structure is done, which suggest the structure to be a very good quality. The model generated for Homogentisate-1,2-dioxygenase is successfully submitted to the Protein Model Database (PMDB) having PMID PM0079008. Active site of the structure, Protein disorder, disorder propensity and Average Area Buried Upon Folding is also predicted and analyzed. Fold of the protein and motif of the protein is also predicted. The function of the protein is predicted and analyzed. This study highlights the sequence, structural and functional information of the protein.

Keywords *Homo Sapiens,* Molecular Modeling, Homogentisate-1,2-dioxygenase, Disorder, Aliphatic Index

1. Introduction

Homogentisate-1,2-dioxygenase (HGD) is an iron containing enzyme which catalyzes the conversion of homogentisate to 4-maleylacetoacetate. An absence or deficiency of homogentisate-1,2-dioxygenase will result in alkaptonuria (AKU). Homogentisate 1,2-dioxygenase is involved in the catabolism of aromatic rings, more specifically in the breakdown of the amino acids tyrosine and phenylalanine[1].HGD appears in the metabolic pathway of tyrosine and phenylalanine degradation when homogentisate is produced. Homogentisate reacts with HGD to produce maleylacetoacetate, which then is further used in the metabolic pathway. HGD requires the use of Fe^{2+} and O_2 in order to cleave the aromatic ring of homogentisate[2]. AKU is due to the inability of the body to deal with homogentisate, which when oxidized by the body will produce the compound known as the ochronotic pigment. This first of these effects is that the patient's earwax will begin to turn black or red, depends on the patient's diet, since the blood becomes oxidized and thus turns black due to excess of the ochronotic pigment. The other effect of the ochronotic pigment is that it can accumulate in the body's connective tissue leading to degenerative arthritis, as the person grows older[2]. The active site of Homogentisate 1,2-dioxygenase was determined through the crystal structure, which was captured through the work of Titus et al[1]. Borowski et al. propose a mechanism for HGD which is The opening of the aromatic ring in homogentisate is a multi-step process. In the first two steps Fe^{2+} coordinates to the carbonyl and ortho phenol oxygens. The iron atom is also coordinated to His335, His371, and Glu341. O_2 then binds to the iron atom[2], subsequently reacting with the aromatic ring to form a peroxo-bridged intermediate. In the next step, O_2 is cleaved with the formation of an epoxide. This epoxide intermediate allowing radical reactions to eventually open and oxidize the six-membered ring.

Plenty of work has been done on Homogentisate - 1, 2 -dioxygenase of various organisms[3 to 13].The Protein Data Bank (PDB) (www.rscb.org) contain very little information about three dimensional structure of Homogentisate - 1, 2 -

* Corresponding author:
akbanerjee40@gmail.com (Amal Kumar)
Published online at http://journal.sapub.org/bioinformatics

dioxygenase of *Homo sapiens*. Therefore, it is very interesting as well as important to carry out the study of Homogentisate-1,2-dioxygenase of *Homo sapiens*. In the present study Sequence, Structural and Functional Characterization of Homogentisate-1,2-dioxygenase of *Homo sapiens*: An *in silico* analysis. We used various servers and softwares to analyze the sequences and to predict the three dimensional structure and function of the Homogentisate-1,2-dioxygenase. The predicted 3D structure is also successfully submitted in the Protein Model Data Base (PMDB)[14] having PMID PM0079008. The results of this work can further help researcher by providing theoretical basis on enzymological properties, structure and function of Homogentisate-1,2-dioxygenase in future.

2. Materials and Methods

2.1. Sequence Retrieval

A total of six sequences of Homogentisate - 1, 2 - dioxygenase protein were retrieved from the manually curate public protein database Swiss-Prot[15]. Swiss - Prot is scanned for the keyword Homogentisate-1,2-dioxygenase and *Homo sapiens*. The search result yielded 6 sequences of Homogentisate-1,2-dioxygenase (HGD) gene family of *Homo sapiens* (Table 1). All of the six sequences were retrieved in FASTA format and used for further analysis.

Table 1. Sequences of Homogentisate 1,2-dioxygenase of Homo sapiens retrieved from Swiss-Prot database

Accession number	Sequence description	organism
Q93099	Homogentisate 1,2-dioxygenase	Homo sapiens(Human)
C9JTX9	Homogentisate 1,2-dioxygenase	Homo sapiens(Human)
Q8WW71	Homogentisate 1,2-dioxygenase	Homo sapiens(Human)
H7C576	Homogentisate 1,2-dioxygenase	Homo sapiens(Human)
H7C4R8	Homogentisate 1,2-dioxygenase	Homo sapiens(Human)
H7C5G7	Homogentisate 1,2-dioxygenase	Homo sapiens(Human)

2.2. Primary Structure Analysis Using Computational Tools and Servers

The amino acid composition (table 2) of Homogentisate - 1, 2-dioxygenase sequences were computed using the tool CLC *free Workbench* [http://www.clcbio.com/products/clc-sequence-viewer/]. Percentages of hydrophobic and hydrophilic residues were calculated from the primary structure analysis and tabulated in table 3.

Table 2. Amino acid composition (in %) of Homogentisate 1,2 - dioxygenase computed using CLC free Workbench tool

	Accession number					
Amino Acids	Q93099	C9JTX9	Q8WW71	H7C576	H7C4R8	H7C5G7
Ala	6.1	3.5	5.2	9.0	0.0	5.1
Cys	2.7	3.5	2.7	0.0	1.1	3.1
Asp	4.9	6.3	4.9	5.5	3.4	5.1
Glu	5.6	4.9	5.2	2.8	6.8	3.1
Phe	6.1	4.9	5.5	6.9	4.5	7.1
Gly	7.9	5.6	7.3	7.6	10.2	6.1
His	2.9	2.8	3.3	2.8	2.1	3.1
Ile	5.2	7.7	5.8	5.1	9.1	8.2
Lys	5.6	6.3	4.6	4.8	3.4	5.1
Leu	7.2	9.2	8.8	6.2	10.2	7.1
Met	2.2	2.1	1.8	0.7	4.5	3.1
Asn	6.1	6.3	7.0	6.2	2.3	12.2
Pro	7.9	6.3	7.6	9.0	9.1	4.1
Gln	2.9	6.3	3.3	2.1	3.4	3.1
Arg	4.3	2.8	4.6	4.1	5.7	1.0
Ser	7.0	7.0	6.7	2.8	5.7	5.1
Thr	3.8	2.8	3.3	4.1	6.8	4.1
Val	5.8	7.0	6.1	9.7	5.7	8.2
Trp	1.8	2.1	2.1	2.1	1.1	1.0
Tyr	4.0	2.1	4.3	6.9	3.4	5.1

Table 3. Hydrophilic and hydrophobic residues content

Accession Number	Percentage of hydrophobic residues	Percentage of hydrophilic residues	Net hydrophilic residue count
Q93099	52.9	47.1	high
C9JTX9	51.9	47.6	high
Q8WW71	52.9	47.2	high
H7C576	56.3	42.1	high
H7C4R8	55.5	43	high
H7C5G7	53.1	47	high

The physico-chemical parameters, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, half-life, instability index, aliphatic index and grand average hydropathy (GRAVY) were computed using the Expasy's ProtParam (http://us.expasy.org/tools/protparam.html) prediction server and tabulated in table 4.

Table 4. Parameters computed using Expasy's protParam tool

Accession Number	Sequence length	M. wt	pI	-R	+R	EC	II	AI	GRAVY
Q93099	445	49963.7	6.50	47	44	71570	38.50	71.21	-0.357
C9JTX9	142	16120.5	5.74	16	13	21220	37.52	89.86	-0.227
Q8WW71	329	37262.4	6.41	33	30	59860	41.58	79.70	-0.305
H7C576	145	16347.8	8.12	12	13	31400	26.28	85.38	-0.087
H7C4R8	88	9945.8	.0	9	8	9970	41.40	91.82	-0.074
H7C5G7	98	11094.6	9	8	6	13075	29.02	88.47	0.009

The number of cysteine residues and disulfide bond is calculated and analyzed using the tool CYS_REC, which is tabulated in Table 5.

Table 5. No. of CYS and Disulfide (SS) bond pattern of pairs predicted, by CYS_REC (using primary structure analysis)

Accession number	No. of CYS	CYS_REC
Q93099	12	Cys14-Cys126 Cys21-Cys138 Cys35-Cys418 Cys51-Cys146 Cys120-Cys337
C9JTX9	5	Cys79-Cys105
Q8WW71	9	Cys21-Cys138 Cys35-Cys120 Cys51-Cys146
H7C576	00	----------
H7C4R8	1	----------
H7C5G7	3	----------

The secondary structure were predicted using the tools SOPMA[16] (Table 6). The total protein intrinsic disorder and the protein disorder propensity was detected using PrDOS[17] and Globplot (http://glob-plot.embl.de) respectively. Prediction of the average area buried upon folding (AABUF) was calculated using the ExPASy tool ProtScale (http://us.expasy.org/tools/protscale.html). The prediction of sub cellular localization was performed using TargetP software[18]. The transmembrane regions were predicted with TMHMM (Krogh *et al*., 2001) server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/).

Table 6. Secondary structure calculation (in %) of Homogentisate -1 ,2-dioxygenase computed using SOPMA

Accession Number	Secondary Structure			
	Alpha Helix	Beta Sheet	Random Coil	Others
Q93099	17.30	22.02	53.03	7.64
C9JTX9	16.90	23.94	48.59	10.56
Q8WW71	8.81	27.05	54.41	9.73
H7C576	8.97	31.03	52.41	7.59
H7C4R8	6.82	29.55	51.14	12.50
H7C5G7	6.12	44.90	29.59	19.39

2.3. Template Selection and Molecular Homology Modeling

The protein sequence of Homogentisate-1,2-dioxygenase having accession ID Q8WW71 has no three dimensional structure present in Protein Data Bank (PDB).Therefore, the template (PDB ID 1F2V_A) were identified by the BLASTP (http://www.ncbi.nlm.nih.gov:80 /BLAST/) analysis in the Protein Data Bank (PDB). Three dimensional structure of Homogentisate - 1, 2 - dioxygenase having accession ID Q8WW71 was generated using ESyPred3D (http:// fundp. ac. be / urbm/bioinfo/esypred/) automated homology modeling server predicted the homology model based on a package MODELLER and by using Swiss Model[19].

2.4. Energy Minimization, Model Evaluation and Submission

The energy minimization for the 3D structure was carried out using NAMD[20] utilizing CHARMM force field and NOMAD-Ref server[21] which utilizes Gromacs forcefield according to steepest descent, conjugate gradient and L - BFGS methods. Further, refinement of the modeled structure was performed using 3 Drefine server[22]. Structural evaluation, validation and stereochemical analyses were performed using various evaluation tools such as Rampage , Procheck[23], Errat[24], Ramachandran plot 2[25], and Vadar[26]. Protein Quality was checked by Resprox[37]. Furthermore, visualization and analyses of the generated model was performed using UCSF Chimera 1.5.3. The generated 3D model was successfully submitted in the Protein Model Data Base (PMDB) having PMID PM0079008. PMDB Protein Model Database, which collects three dimensional protein models obtained by structure prediction methods. Users can both contribute new models and search for existing ones. The database currently stores all models submitted to the last four editions of the CASP experiment.

2.5. Active Site Prediction

After the final model was built, the active site of the structure was predicted using the tool Meta pocket 2(38).The visualization and analysis of the active site was performed by using UCSF Chimera 1.5.3.

2.6. Identification of Motif, Fold and Functional Domain

Motif of the protein were identified using the tools Prosite [29] and MotifScan[30].The protein fold was predicted using the software PFP-pred[31].

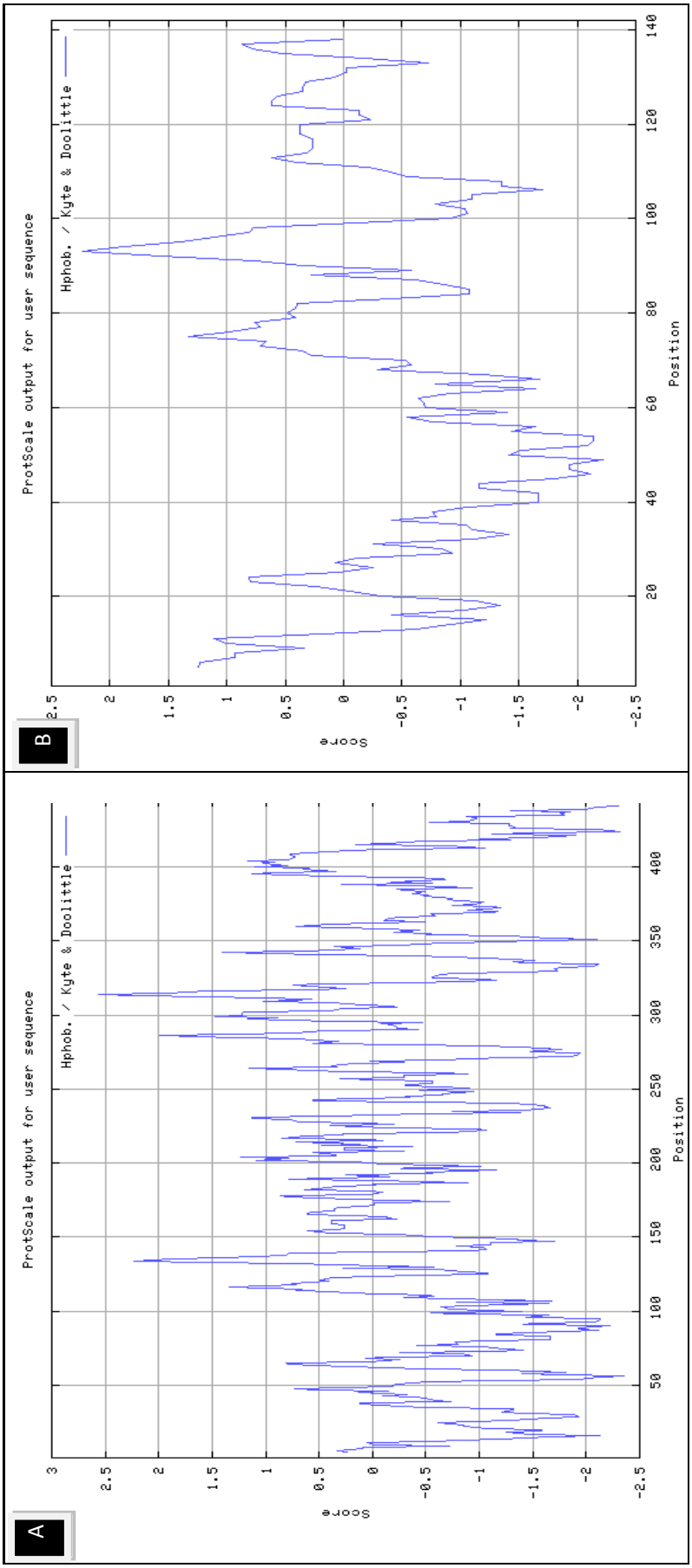
2.7. Function Prediction

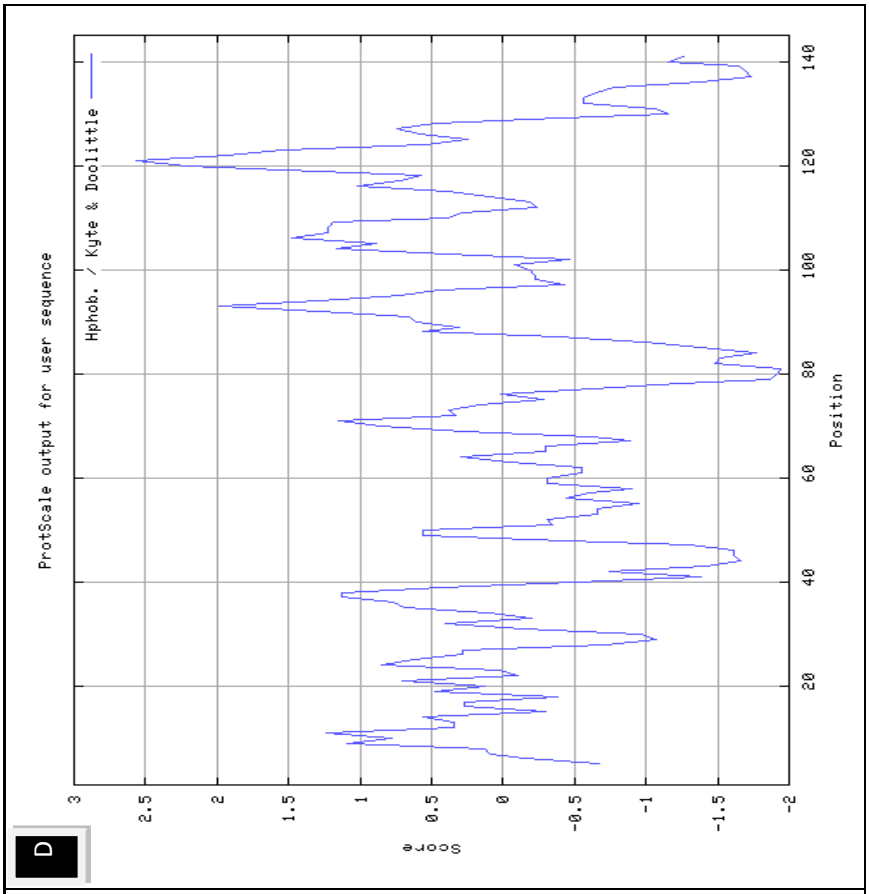
The function of the protein Homogentisate - 1, 2 - dioxygenase of *Homo sapiens* was predicted by using ProFunc 2[32]. The aim of the ProFunc server is to help identify the likely biochemical function of a protein from its three-dimensional structure. It uses a series of methods, including fold matching, residue conservation, surface cleft analysis, and functional 3D templates, to identify both the protein's likely active site and possible homologues in the PDB.

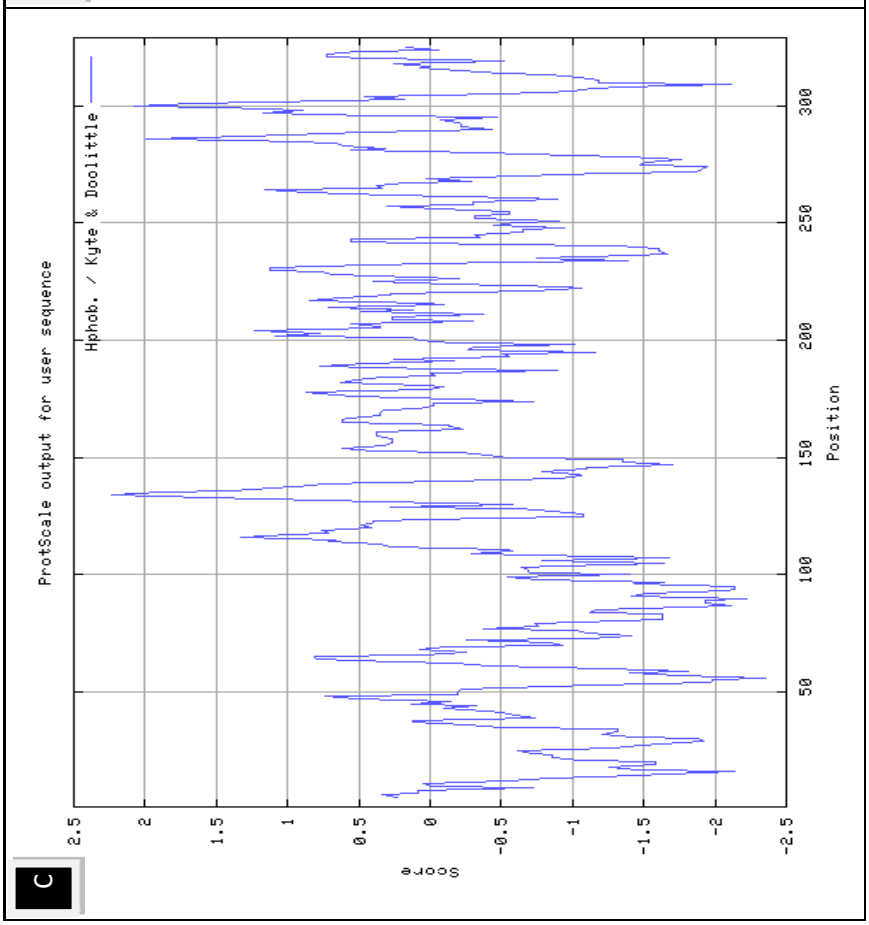
3. Result and Discussion

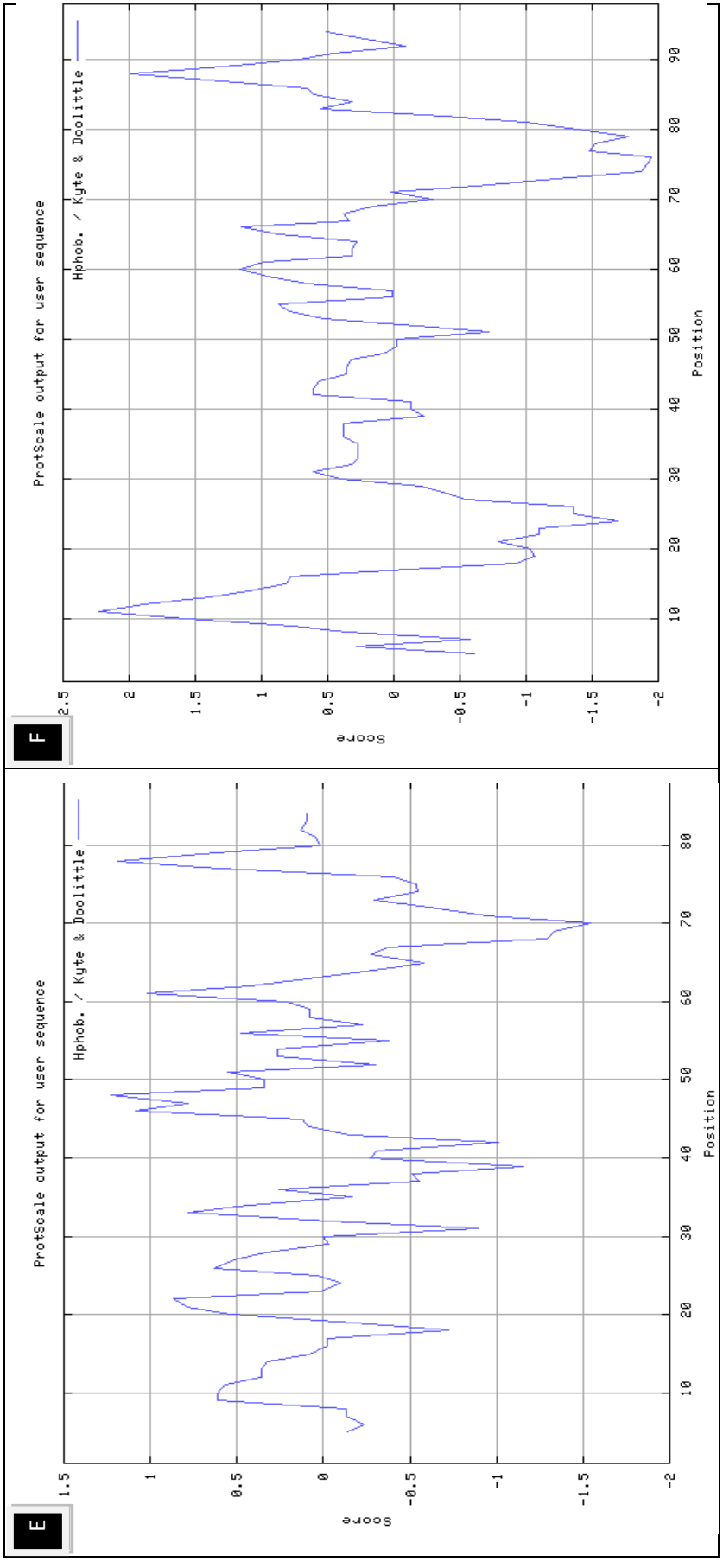
3.1. Primary Structure Analysis

The results of primary structure analysis suggest that all of the sequences of Homogentisate-1,2-dioxygenase are hydrophobic in nature due to the presence of high content of non- polar residues (tables 2 and 3). The kye-dollitle hydrophobicity is analyzed using ProtScale and shown in figure 1.


A
ProtScale output for user sequence
Hphob. / Kyte & Doolittle
Score
Position
B
ProtScale output for user sequence
Hphob. / Kyte & Doolittle
Score
Position


D
ProtScale output for user sequence
Hphob. / Kyte & Doolittle
Score
Position
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
20
40
60
80
100
120
140


C
ProtScale output for user sequence
Hphob. / Kyte & Doolittle
Score
Position
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
50
100
150
200
250
300


F
ProtScale output for user sequence
Hphob. / Kyte & Doolittle
Score
Position
E
ProtScale output for user sequence
Hphob. / Kyte & Doolittle
Score
Position

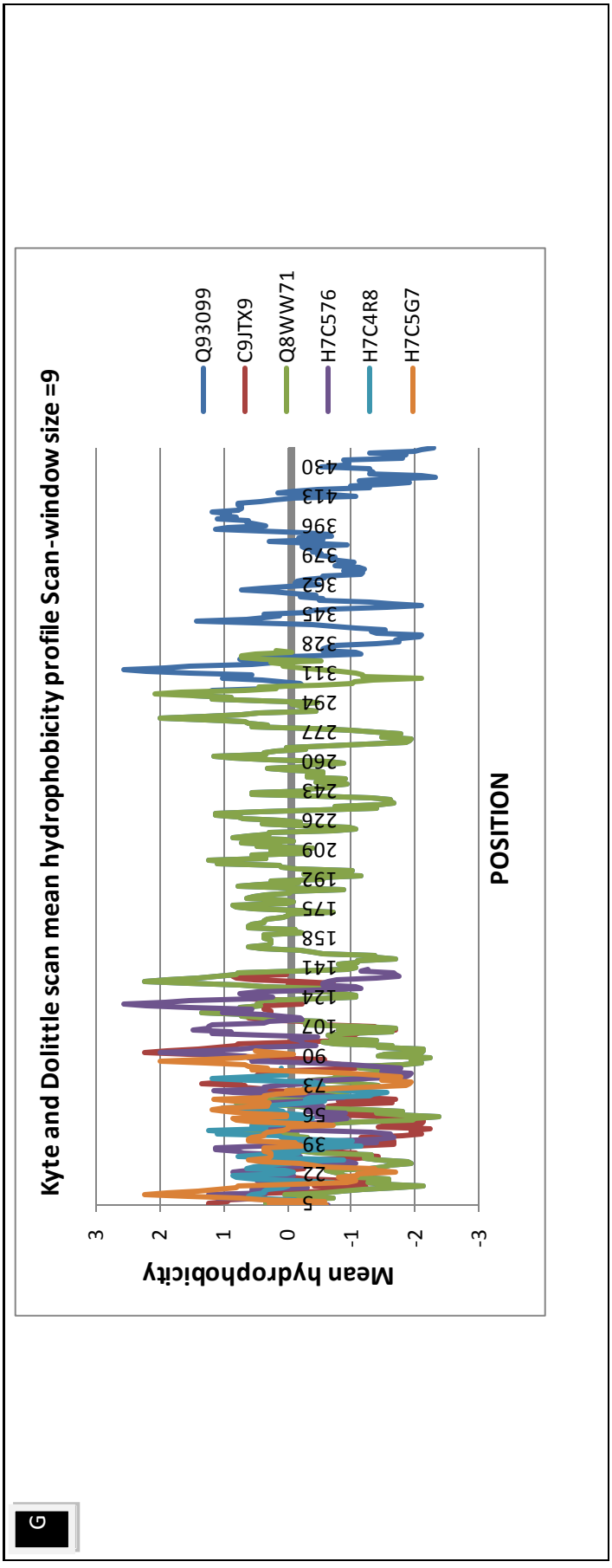


Figure 1. Kyte-doolittle hydrophobicity by ProtScale. (A) For the sequence with accession id Q93099 (B) For the sequence with accession id C9JTX9 (C) For the sequence with accession id Q8WW71 (D) For the sequence with accession id H7C576 (E) For the sequence with accession id H7C4R8 (F) For the sequence with accession id H7C5G7 (G) Kyte-Doolittle scale mean hydrophobicity profile for all the four sequences of Homogentisate-1,2-dioxygenase

The analysis also suggests that the Homogentisate - 1, 2 - dioxygenase contain high residues of acidic and basic amino acid, this might be involves in salt bridge formation. The average molecular weight of Homogentisate -1, 2 - dioxygenase is 20731.16 Da. Isoelectric point (pI) is the pH at which net charge of the protein is zero but the surface of protein is covered with charge. The proteins are generally stable and compact at PI.Isoelectic focusing method is utilized for purification using computed PI by developing buffer systems. most of the sequences of Homogentisate - 1, 2 - dioxygenase have (pI < 7) except H7C576 reveals that all of them are acidic in character. Although Expasy's ProtParam computes the extinction coefficient for a range of (276, 278, 279, 280 and 282 nm) wavelength, but proteins absorb strongly at 280 nm while other substances commonly in protein solutions do not. So, 280 nm of wavelength is favored. With respect to the concentration of Cys, Trp and Tyr, Extinction coefficient of Homogentisate - 1, 2 - dioxygenase at 280 nm is ranging from 9970 to 71570 M^{-1} cm^{-1}. The extinction coefficients and the computed protein concentration help in the quantitative study of protein – protein and protein–ligand interactions in solution. The biocomputed half-life of most of the Homogentisate -1, 2- dioxygenase are 30 h in mammalian reticulocytes, invitro, greater than 20 h in yeast, in vivo and greater than 10h in *E.coli*, in vivo. On the basis of instability index Expasy's ProtParam classifies all the Homogentisate-1,2-dioxygenase as unstable (Instability index > 40). The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chains (A, V, I and L) is regarded as a positive factor for the increase of thermal stability of globular proteins. The lower thermal stability of the Homogentisate-1,2-dioxygenase is indicative of a more flexible structure (table 4). The low aliphatic index of most of the Homogentisate-1,2-dioxygenase infers that it may be thermo labile for a high range of temperature. Grand Average hydropathy (GRAVY) Index of Homogentisate -1, 2 - dioxygenase are ranging from 0.009 to -0.087. The very low GRAVY index of Homogentisate-1,2-dioxygenase infers that these could result in a better interaction with water. The secondary structure predicted with the help of programs SOPM and SOPMA (table 6) infers that the Homogentisate - 1, 2-dioxygenase have rich alanine content and mostly sheets and random coil. Sub cellular localization is predicted by TargetP server which suggested that all of the sequences of Homogentisate-1,2-dioxygenase are secretory without transit peptide (Table 7).

The transmembrane regions were predicted with TMHMM server 2 suggests that Homogentisate - 1, 2 - dioxygenase has no transmembrane region. The total protein intrinsic disorder and the protein disorder propensity calculation shows that protein is mostly maintaining the ordered structure except one place at the middle and at the N terminal and C terminal region (Figure 2).

Prediction of the average area buried upon folding (AABUF) was calculated using the ExPASy tool ProtScale and shown in figure (3)

Table 7. Sub cellular localization by TargetP

Accession number	Sub Cellular Location	MOTIFS binding and phosphorylation site
Q93099	SECRETORY	Protein kinase C Casein kinase II N-myristoylation N-glycosylation cAMP- and cGMP-dependent protein kinase catalytic oxidoreductase dioxygenase
C9JTX9		
Q8WW71		
H7C576		
H7C4R8		
H7C5G7		

3.2. Molecular Homology Modeling, Energy Minimization and Model Evaluation

The protein sequence of Homogentisate-1,2-dioxygenase having accession ID Q8WW71 has no three dimensional structure present in Protein Data Bank (PDB).Therefore, the template (PDB ID 1EYB_A) are identified by the BLASTP analysis in the Protein Data Bank (PDB)(Table 8).

Table 8. PDB templates (first 2 hits with maximum % identity) using BLASTP search against the Protein Data Bank

Accession number	PDB code
Q8WW71	1F2V_A 1EYB_A

The energy minimizations were achieved by NAMD and NOMAD-Ref server. After minimization of the model structure, RMSD of the structure is calculated to be 0.5545, Average surface area calculated is 22745.72and the energy before minimization and after minimization is calculated which is -1676.3Kcal/mol and -14456.9Kcal/mol respectively (Table 9).

Table 9. Energy minimization, average RMSD and average surface area of the modeled structure

Accession number	Energy before minimization Kcal/mol	Energy after minimization Kcal/mol	Average RMSD	Average Surface Area
Q8WW71	-1676.3	-14456.9	0.5545	22745.72

Further, refinement of the modeled structure is performed using 3Drefine server. After refinement out of five refined model, the best refined model structure is chosen. Structural evaluation, validation and stereochemical analyses were performed using various evaluation tools such as Rampage, Procheck, Errat, Ramachandran plot 2, and Vadar. All of the mentioned tools suggest that the modeled structure is to be good model (Figure 4).

According to evaluation analysis, the Ramachandran plot and other parameters (Table 10) were within the standard acceptable limits for the 3D structures modeled using the PDB template 1F2A_A for the (target) protein (Table 11).

Protein Quality is checked by Resprox and the structure fulfill all the criteria of good model (Table 12).

Furthermore, visualization and analyses of the generated model was performed using UCSF Chimera 1.5.3 (Figures 5). The generated 3D model was successfully submitted in the Protein Model Data Base (PMDB) having PMID PM0078945.

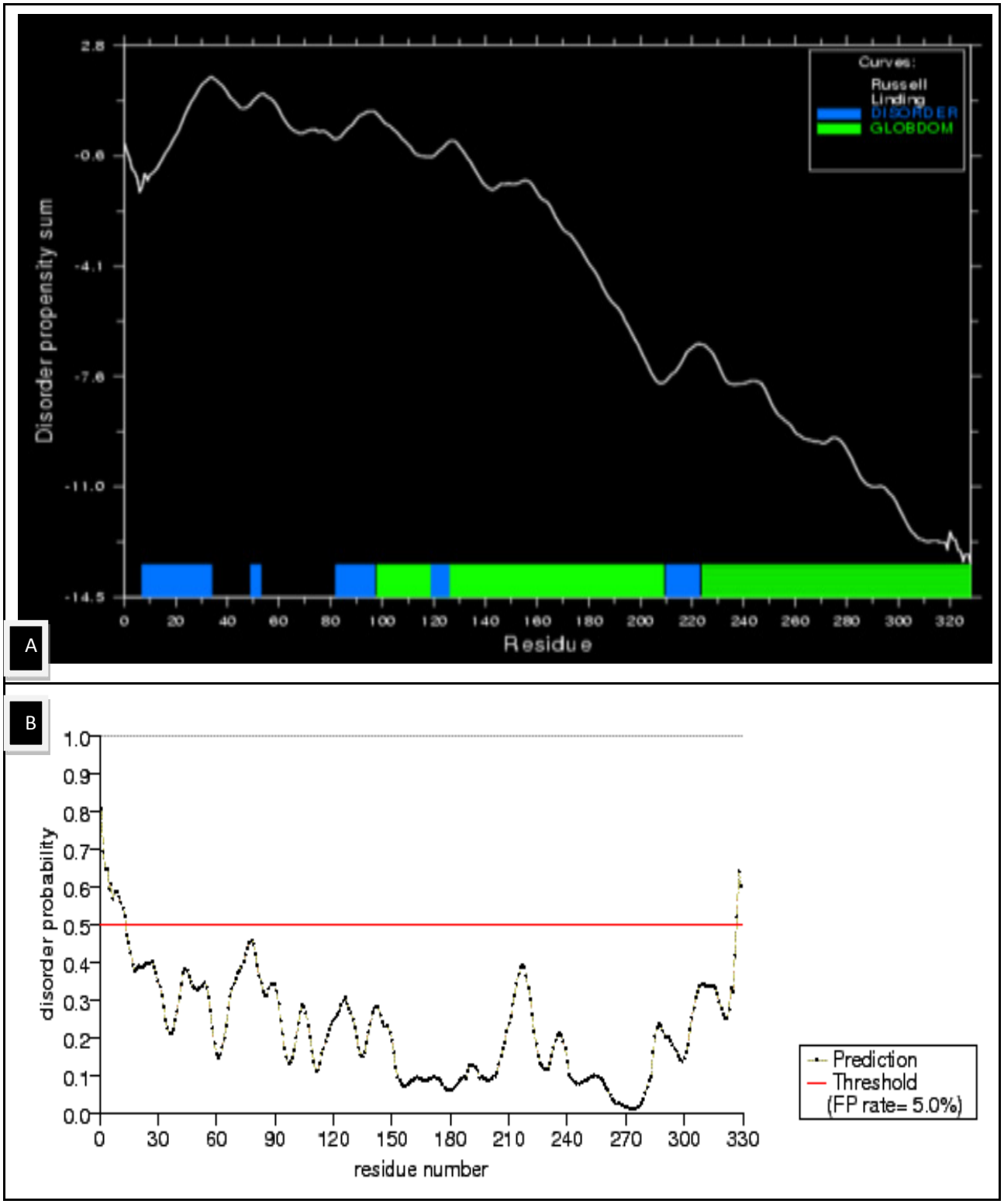


Figure 2. Disordered Regions of the Chalcone Synthase protein (A) Disorder Propensity Sum (B) schematic diagram of disorder Probability over the entire sequence

Table 10. Criteria for a good (model) 3D structure

Percentage of residues In favored region		**ProQ**		**ResProx** Predicted resolution(Å)	Quality of the model
Rampage	**Prochek**	LG score	maxsub		
98	90	>0.1	>1.5	>2.5	Fairly good
		>0.5	>2.5	1.5-2.5	Very good
		>0.8	>4.0	0.0-1.5	Extremely good

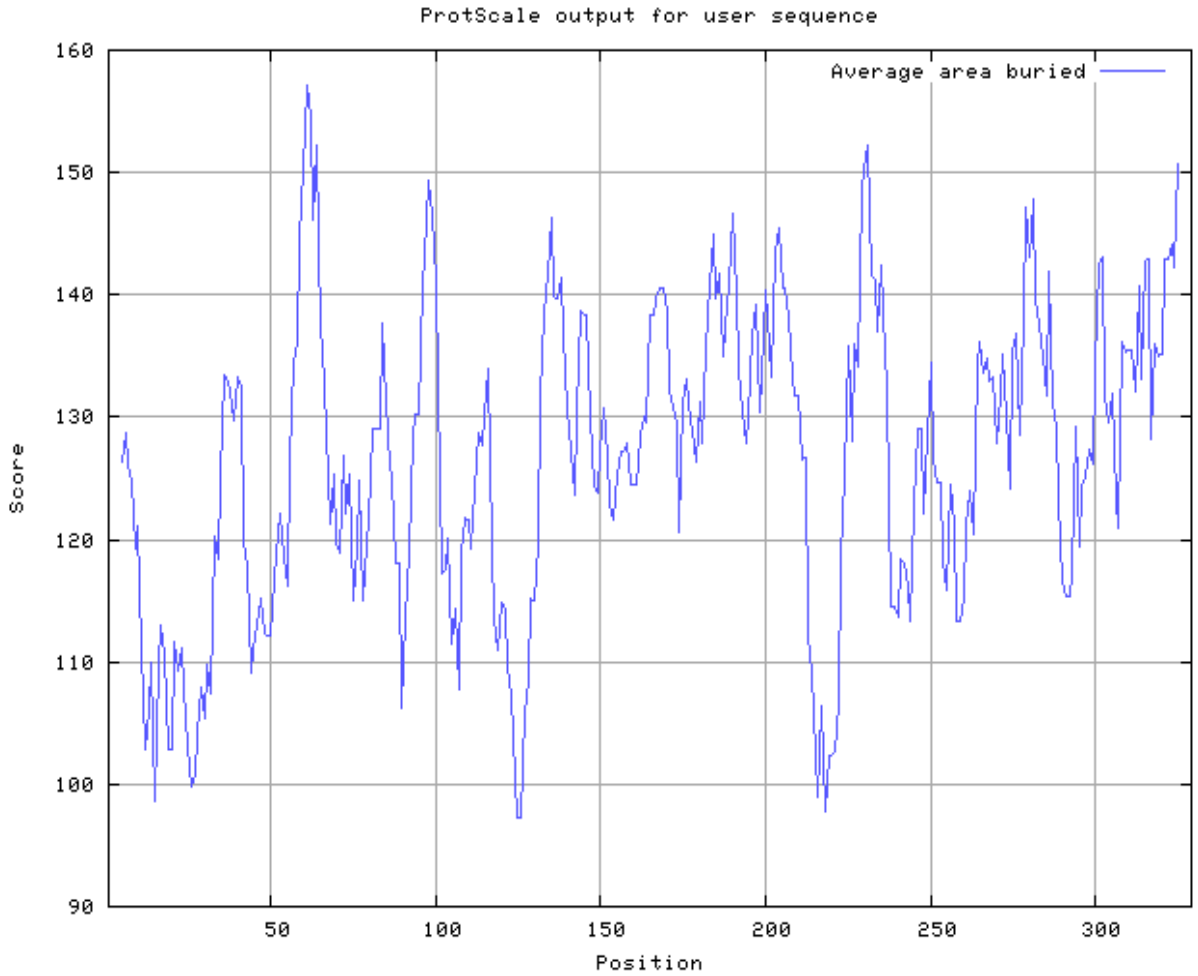
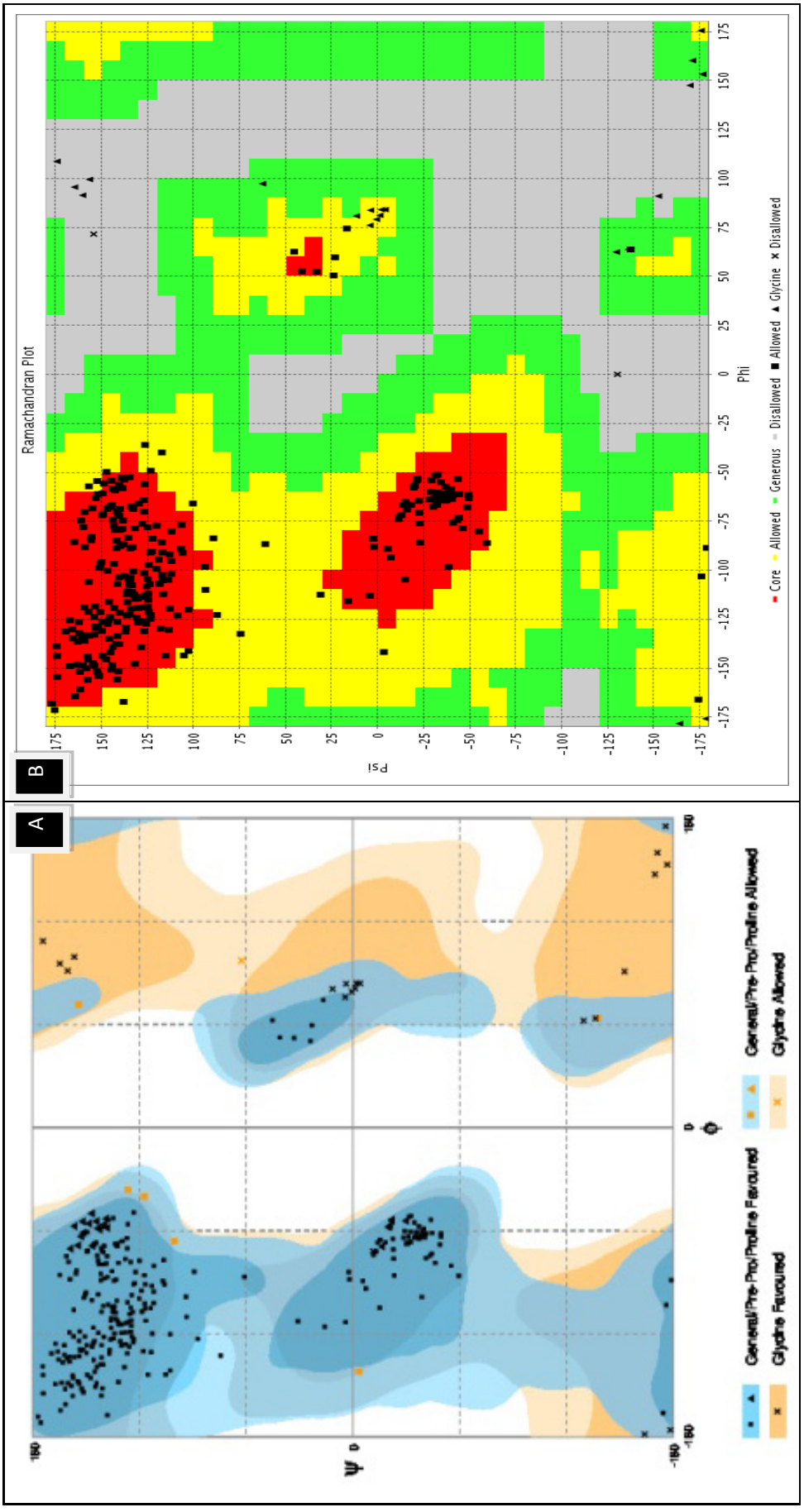
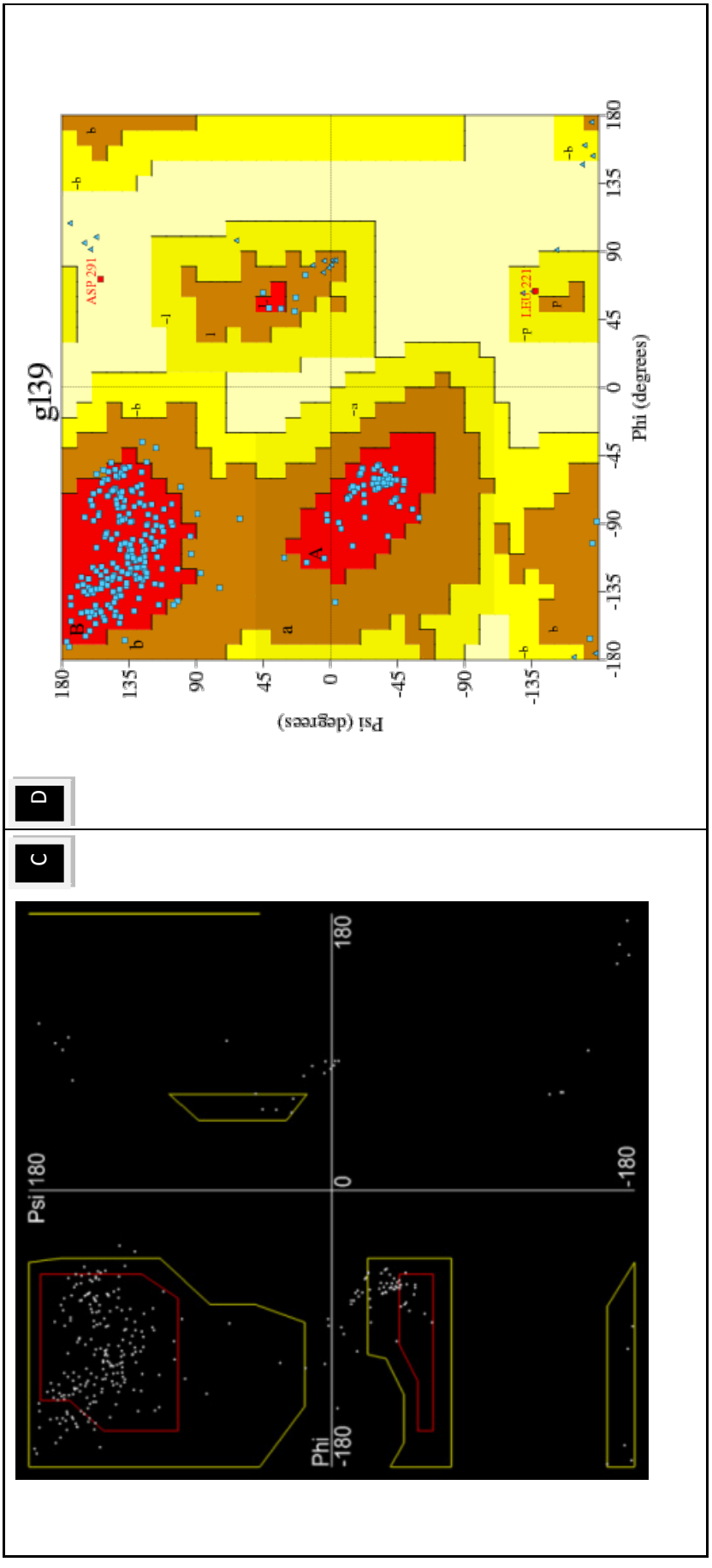


Figure 3. Average Area Buried upon Folding For all the Amino acids residue of the Protein predicted by Protscale from ExPASy


A
B
Ramachandran Plot
Psi
Phi
Core
Allowed
Generous
Disallowed
Allowed
Glycine
Disallowed
General/Pre-Pro/Proline Favoured
Glycine Favoured
General/Pre-Pro/Proline Allowed
Glycine Allowed
Ψ
Φ


D
g139
ASP 291
LEU 221
A
B
a
b
Psi (degrees)
Phi (degrees)
180
135
90
45
0
-45
-90
-135
-180
C
Psi
Phi
180
0
-180

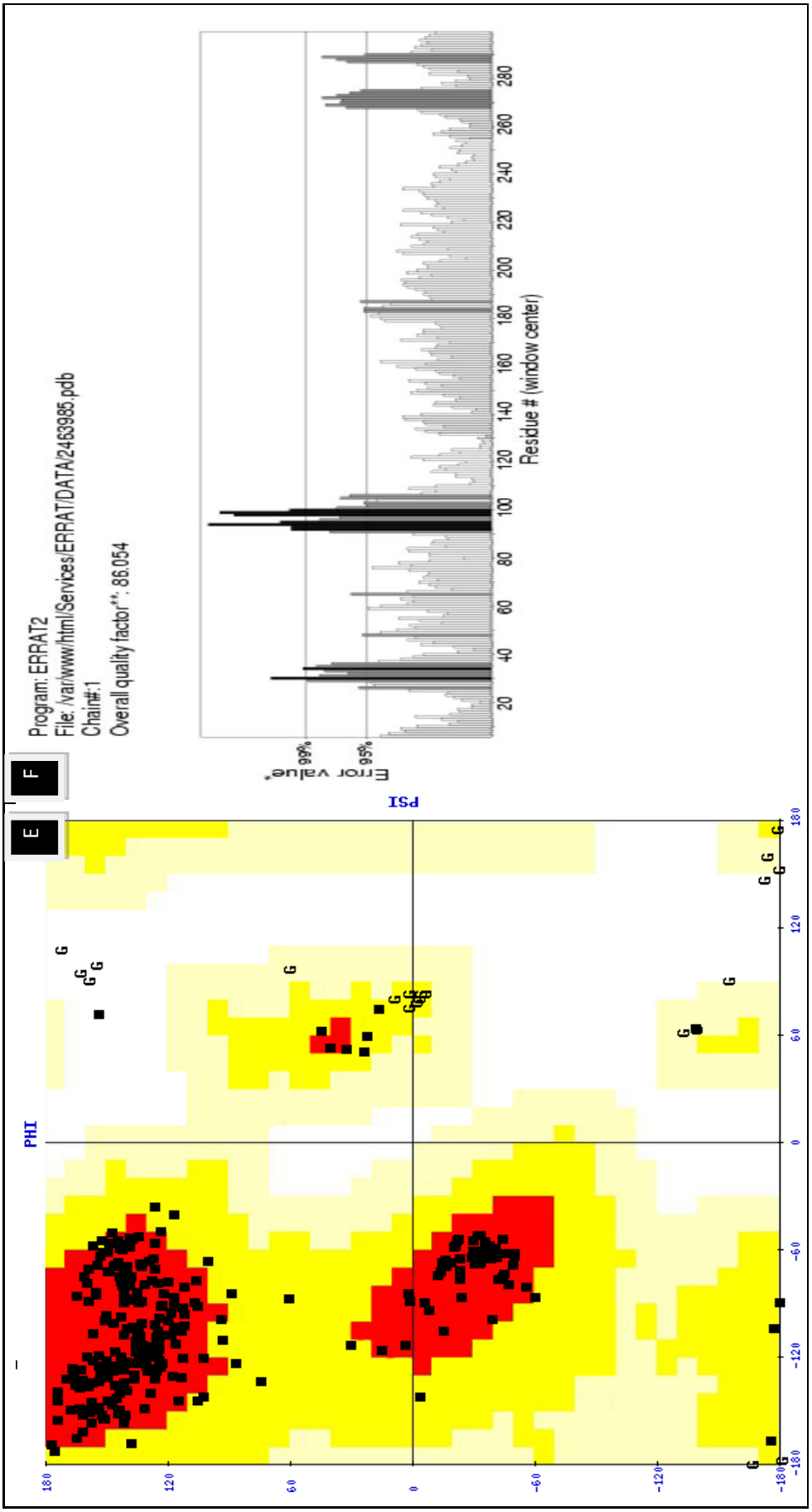


Figure 4. Ramachandran plot using (A)Rampage (B)Vadar (C)Ramachandran Plot 2 (D, E) Procheck (F) Z score using Errat

Table 11. Validation parameters computed for the built 3D structures of target Q8WW71

Target - Q8WW71		
Template(PDB codes) - 1EYB_A		
Rampage Percentage of residues In favored region		97.7
Prochek Percentage of residues In favored region		90.6
ProQ	LG score	4.012
	maxsub	0.235
ResProx Predicted resolution(Å)		1.567

Table 12. Z-score of the Modeled Structure Calculated using Resprox, Z Score Report

Score Name	Quality	Z score	Description	Program
Standard deviation of χ1 pooled	Good	1.37	Standard deviation of the χ1 angles among all 3 (gauche-, gauche+, and trans) configurations.	Vadar
Ramachandran outside of most favored	Good	-0.19	Percentage of residues outside of the most favored regions of the Ramachandran plot.	GeNMR
RosettaHoles 1.0 score	Good	-0.37	A measure of underpacking in the protein core.	Rosetta
Deviation of Θ angles	Good	0.30	Standard deviation of angle between the C-O bond vector of the H-bond acceptor and the O-H(N) bond vector.	PROSESS
Bump score	Good	-0.16	The bump score is calculated from the total number of non-bonded atom contacts below 1.3 Å, divided by the total number of non-bonded contacts in the protein.	GeNMR
Mean H-bond energy	Good	-0.36	The average hydrogen bond energy is calculated using the H-bond energy function used in DSSP program.	Vadar
χ1 score	Good	0.22	Scaled difference between the standard deviation of the observed χ1 angles and the expected one obtained from high quality protein structures.	PROSESS
Radius gyration score	Good	1.03	Scaled difference between the expected radius of gyration and the observed one. The expected radius of gyration is determined using: Rg = 0.395*N**0.6 + 7.257.	GeNMR
Percentage of generously allowed Ω angles	Good	-1.60	This corresponds to the percentage of residues having Ω (omega) angles within 15° to 20° of the ideal trans (180°) and cis (0°).	Vadar
Percentage of packing defects	Good	0.20	This is the percentage of residues with fractional residue volumes greater than 1.20 or less than 0.80. Packing defects indicate the presence of cavities or compressions that are not natural.	Vadar
Percentage of 95% buried residues	Good	0.04	Percentage of residues with fractional accessible areas < 0.05. This score reports the extent of residue burial. Most globular proteins must have a percentage >5% to be stable. Divided by the expected value.	Vadar
Percentage of bad bond angles	Good	-0.69	This parameter is calculated as the number of bond angles (divided by the total number of bond angles in the polypeptde) that exceed, by more than 5 standard deviations, the typical bond angles seen in high resolution, high quality structures.	MolProbity
Ramachandran plot outliers	Good	-0.11	Fraction of residues in the Ramachandran plot that are stereo-chemically not allowed or not observed in high quality structures.	MolProbity

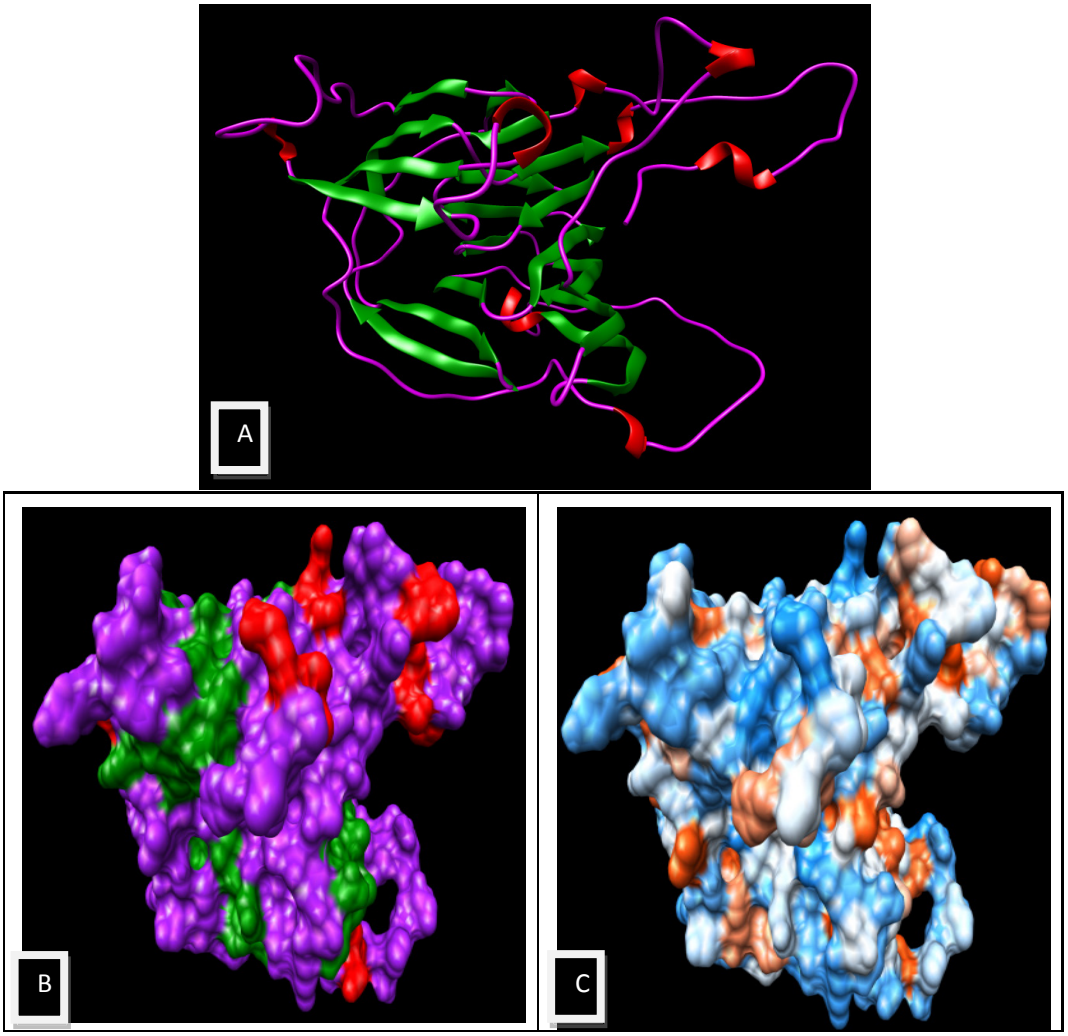


Figure 5. Visualization by using UCSF Chimera 1.5.3 (A) Secondary Structure of the Model (B) Surface View of the model Structure (C) Molecular Surface Colored by amino acid Hydrophobicity (Dodger Blue-most hydrophilic, White- Moderate, Orange Red- most hydrophobic)

3.3. Characterization and Active Site Prediction of the Structure

The modeled structure is characterized by their secondary structure contents and it is found that Homogentisate - 1, 2 - dioxygenase contains high amount of alpha helix and coils (table 13) in respect to beta sheet just like the primary structure analysis (table 6).

The secondary and topological three dimensional structure is calculated using PDBsum and shown in figure 6.

The total protein intrinsic disorder and the protein disorder propensity was detected and calculation shows that protein is mostly maintaining the ordered structure except one place at the middle and at the N terminal and C terminal region also shown in the modeled structure (Figure 7).

For more information just follow this link http:// www. ebi.ac.uk/pdbsum/ by using PDB code: gl39 and Password: 203313 or directly by bookmarking this link: http:/ /www. ebi.ac.uk/thornton-srv/databases/cgi- bin/ pdbsum/ GetPage. pl?pdbcode =gl39 & code =203313 .The visualization and analysis of the active site was performed by using UCSF Chimera 1.5.3 (Figure 8).

The active site of the structure is analyzed using the tool Meta pocket 2 (Table 14).

3.4. Identification of Motif, Fold, Functional Domain and Function Prediction

Motif of the protein is identified using the tools Prosite and MotifScan (Table 16). PFP-pred predicted the protein fold type to be Small inhibitors, toxins, lectins (Table 15).

The function of the protein is identified using ProFunc 2 and it predicted that the Homogentisate-1,2-dioxygenase is involved in many biological and catalytical functions (Table 17).

For more information related to function follow the link to ProFunc home page

http://www.ebi.ac.uk/thornton-srv/databases/profunc using PDB code: gl39 and Password: 203313 or directly by bookmarking this link:

http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/profunc/GetResults.pl?source=profunc&user_id=gl39&code=203313.

Table 13. Modeled Structure is Characterized by their Secondary Structure Contents such as residues, their position and the number of Helices, Sheets and Coils present in the Structure

Accession No	Residues		Secondary Structure		
	Name & Position	Count	Alpha Helix	Beta Sheet	Random Coil
Q8WW71	A,E,LK,Y (2–6)	1	------	------	1
	I,S,G (7 – 9)	3		1	--------
	F,G,N,E (10 –13)	4	------	------	2
	C,S,S,E,D (14–18)	5	------	2	--------
	P,R,C,P,G,S,L,P,E,G,Q,N,N,P,Q, V,C,P,Y,N,L (19–39)	21	------	------	3
	Y,A,E,Q ,L,S,G (40–46)	7	------	3	--------
Q8WW71	S,A,F,T,C,P,R,S,T,N (47–56)	10	------	------	4
	K,R,S,W,L,Y,R,I (57–64)	8		4	---------
	L,P,S,V,S,H,K,P,F (65 – 73)	9	--------	--------	5
	E,S,I(74–76)	3		5	--------
	D, E,G,H,V, T, H,N,W,D,E,V, D,P,N,L,R,T,K,P,F,E,I,P,K,A,S, Q,K,K,V,D (77– 111)	35	------	------	6
	F,V,S(112–114)	3	1	-----	----
	G,L,H,T,L,C,G,D,I (115 – 125)	11	-----	6	----
	K,S (126 – 127)	2	-----	-----	7
	N,N,G,L,A,I,H,I,F,L,C (128–138)	11	--------	7	---------
	N,R,C,F,Y,N,S,D,G,D,F,L,I,V,P, Q (144 – 159)	16	--------	8	---------
	K,G,N(160–162	3	------	------	8
	L,L,I,Y,T(163–167)	5	2	------	-----
	E,F (168 – 169)	2	------	------	9
	G,K,M,L,V(170 – 174)	5	------	9	--------
	Q,P (175 – 176)	2	------	------	10
	N,E,I,C,V,I (177 – 182)	6	--------	10	--------
	Q,R,G,M,R,F (183 – 188)	6	------	------	11
	S,I,D (189 - 191)	3	------	11	-----
	V,F,E,E (192 – 195)	4	------	------	13
	T,R,G,Y,I,L,E,V,Y,G (196 – 205)	10	------	12	-----
	V,H,F,E,L,P,D,L,G,P,I,G,A,N,G, L,A,N,P,R,D (206 – 226)	21	------	------	14
	F,L,I (227 – 229)	3	------	13	-----
	P,I,A,W,Y(230–234	5	------	------	15
	E,D,R,Q ,V(235 – 239)	5		14	--------
	P,G,G,(240– 242)	3	------	------	16
Q8WW71	Y,T,V,I,N,K,Y (243 – 249)	7		15	--------
	Q,G (250 – 251)	2	------	------	17
	K,L,F,A,A,K,D,V,S,P,F,N,V,V,A, W,H,G,N (252 – 271)	20	------	16	--------
	Y,T, P(272 – 274)	3	------	------	18
	Y,K,Y,N,L (275 – 279)	5	------	17	--------
	K,N,F,M,V,I,N,S,V,A,F,D,H,A,D, P (280 – 295)	16	------	------	19
	S,I (296 – 297)	2	3	-------	--------
	F,T,V,L,T,A (298 - 303)	6	-------	-------	20

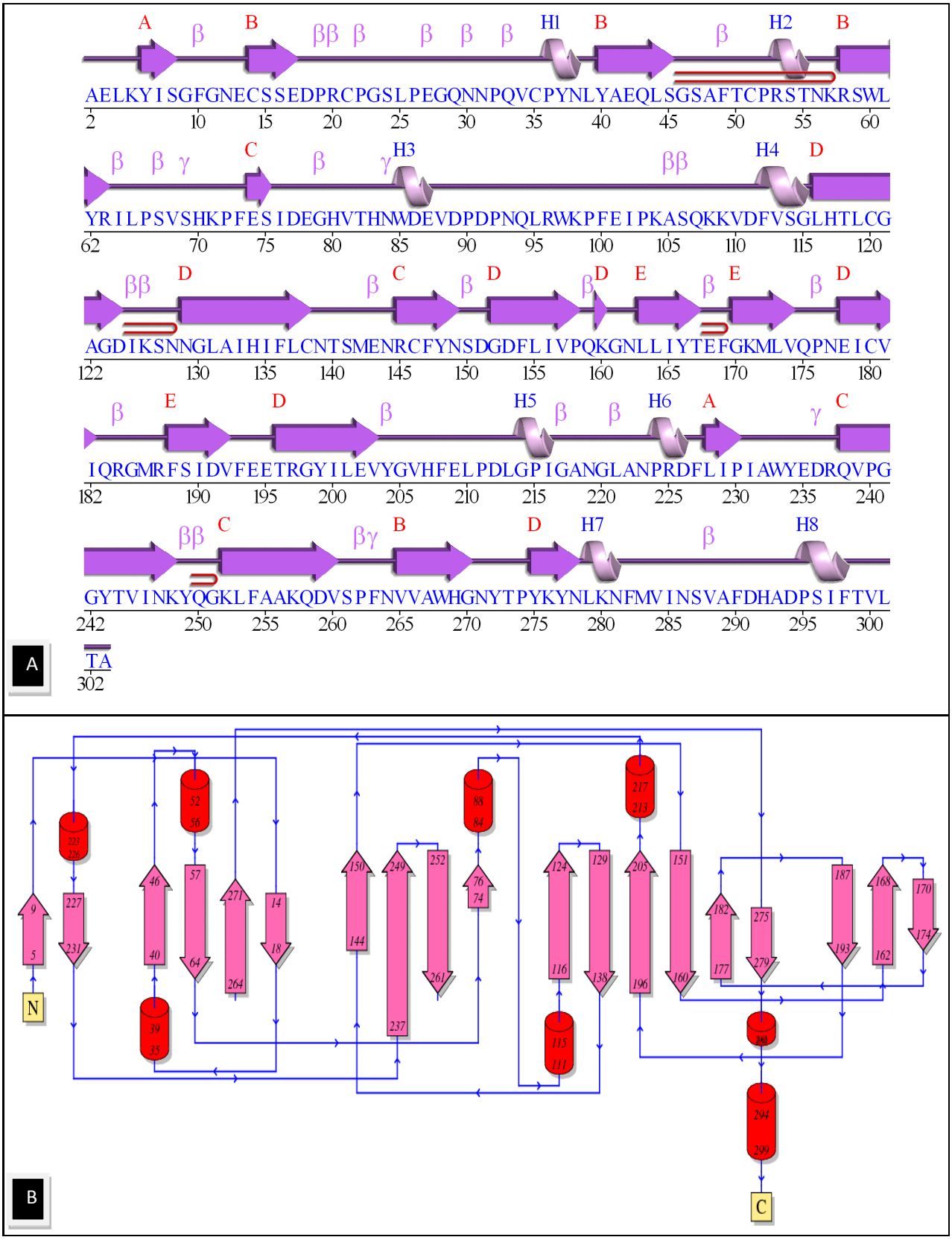


Figure 6. (A) Secondary structure of the Modeled Structure, Helices are labeled as H; Beta turn as β; Gamma turn as χ and Beta hairpin as ⇒ (B) Topology of Secondary Structure of Model, Helices and Sheets are shown in Pink and Red

Figure 7. The total protein intrinsic disorder is shown in the structure (white color)

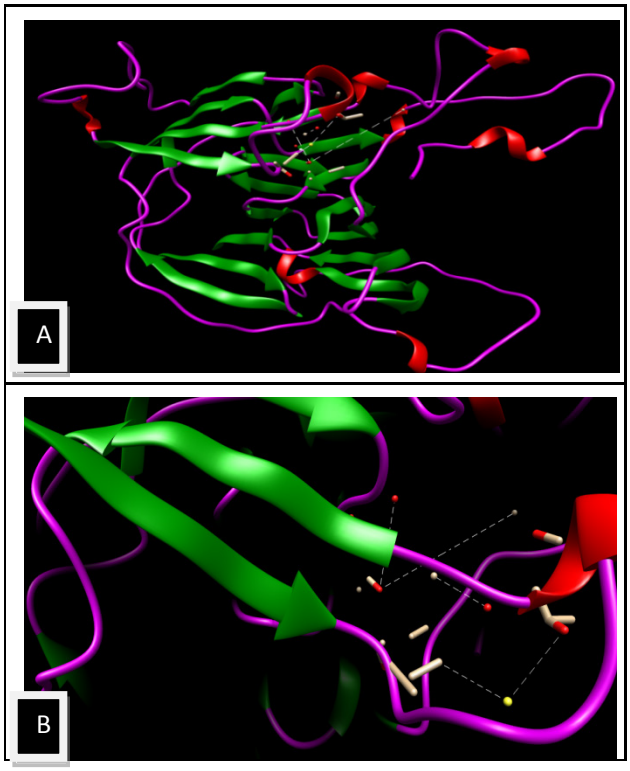


Figure 8. The active site predicted using metapocket 2 A) active site in the modeled structure B) Enlarge view of the active site

Table 14. Analysis of the Active Site of the Structure of Chalcone Synthase Using Meta pocket 2 tool

Residues		Atoms
Name	**Position**	**Position**
SER	8 47 54	57 – 63 394 – 400 454 – 460
CYS	14 51	107 - 113 427 – 433
THR	55 273	461 – 468 2495 – 2502
LYS	57	480 – 489
HIS	207	1885 – 1897
GLU	13 209	97 – 106 1910 - 1919
GLY	9 11 46	64 - 68 81 – 85 389 - 393
ASN	12 271 56	86 - 96 2471 – 2481 469 – 479
ARG	53 58	441 – 453 490 – 502
VAL	181 206	1638 – 1645 1877 – 1884
PHE	10 208	69 – 80 1898 – 1909
LEU	210 228	1920 – 1928 2069 – 2077
GLN	183	1655 – 1666
TYR	272 275	2482 – 2494 2510 - 2522
ASP	153	1379 – 1387
ILE	182 229	1646 - 1654 2078 – 2086
TRP	268	2437 - 2452

Table 15. Protein fold of Q8WW71 is predicted using PFP-pred

Protein	Predicted Fold Type
Q8WW71	Small inhibitors, toxins, lectins

Table 16. Motifs, number of Sites and their Corresponding Residues Identified by using Prosite and MotifScan

Accession No.	Motifs		No. of site	Position
	Pattern	Binding Site & Phosphorilation Site		
Q8WW71	PKC_PHOSPHO_SITE	Protein kinase C phosphorylation site	3	55 – 57 , 69 – 71 , 106 – 108
	CK2_PHOSPHO_SITE	Casein kinase II phosphorylation site	5	15 – 18 , 24 – 27 , 75 – 78 , 140 -143, 150 - 153
	MYRISTYL	N-myristoylation site	3	11 – 16 , 46 – 51 , 185 - 190
	MICROBODIES_CTER	Microbodies C-terminal targeting signal	1	327 - 329

Table 17. Some of the Functions of Chalcone Synthase As predicted by using ProFunc 2 Server

Biological process & Biochemical function	Score
metabolic process	26.44
cellular process	22.88
primary metabolic process	22.83
cellular metabolic process	20.93
catalytic activity	24.74
oxidoreductase activity	14.68
dioxygenase activity	13.62
incorporation of two atoms of oxygen	13.62

4. Conclusions

The sequences of homogentisate-1,2-dioxygenase from *Homo sapiens are* chosen for the present study as because very little or no information of the protein is present currently. Primary analysis suggests that it is hydrophobic in nature and localized as secretory without transit peptide. Physico-chemical characterization provides essential information or data about the protein and its properties. Secondary structure analysis suggests that the protein is dominantly of sheet and random coil. It has no transmembrane region and highly structured as shown by flexibility and disordered studies. Active site analysis suggests that the structure has some highly conserved residues. Identification of motifs showed that it has many binding and phosphorylation site indicating the engagement of the protein in many catalytical processes. Functional analysis suggests that it is involved in protein binding and dimerization activity as well as many other biological functions. The present study provides all the necessary information about sequence, structure and function of the protein to the scientist for the further research in future.

ACKNOWLEDGEMENTS

Authors are grateful to the Department of Biotechnology (DBT), Govt. of India for the computational facility laboratory.

REFERENCES

[1] Titus, Greg P; Mueller, HA; Burgner, J; Rodríguez De Córdoba, S; Peñalva, MA; Timm, DE (2000). "Crystal structure of human homogentisate dioxygenase". Nature Structural Biology 7 (7): 542–546. doi:10.1038/76756. PMID 10876237.

[2] Borowski, Tomasz; Georgiev, V; Siegbahn, PE (2005). "Catalytic Reaction Mechanism of Homogentisate Dioxygenase: A Hybrid DFT Study". Journal of American Chemical Society 127 (49): 17303–17314. doi:10. 1021/ ja054433j. PMID 16332080

[3] Preston AJ, Keenan CM, Sutherland H, Wilson PJ, Wlodarski B, Taylor AM, Williams DP, Ranganath LR, Gallagher JA, Jarvis JC. Ochronotic osteoarthropathy in a mouse model of alkaptonuria, and its inhibition by nitisinone. Ann Rheum Dis. 2013 Mar 19.

[4] Yang YJ, Guo JH, Chen WJ, Zhao R, Tang JS, Meng XH, Zhao L, Tu M, He XY, Wu LQ, Zhu YM. First report of HGD mutations in a Chinese with alkaptonuria. Gene. 2013 Apr 15;518(2):467-9. doi: 10.1016/j.gene.2013.01.020. Epub 2013 Jan 24.

[5] Hu M, Ma HW, Luo Y, Wang L, Song Y, Li F.[Gene diagnosis of alkaptonuria in an infant]. Zhongguo Dang Dai Er Ke Za Zhi. 2012 Oct;14(10):796-7.

[6] Millucci L, Spreafico A, Tinti L, Braconi D, Ghezzi L, Paccagnini E, Bernardini G, Amato L, Laschi M, Selvi E, Galeazzi M, Mannoni A, Benucci M, Lupetti P, Chellini F, Orlandini M, Santucci A. Alkaptonuria is a novel human secondary amyloidogenic disease. Biochim Biophys Acta. 2012 Nov;1822(11):1682-91. doi: 10. 1016 /j. bbadis. 2012.07.011. Epub 2012 Jul 28.

[7] Magesh R, George Priya Doss C. Computational methods to work as first-pass filter in deleterious SNP analysis of alkaptonuria. ScientificWorldJournal. 2012;2012:738423. doi: 10.1100/2012/738423. Epub 2012 Apr 19

[8] Taylor AM, Preston AJ, Paulk NK, Sutherland H, Keenan CM, Wilson PJ, Wlodarski B, Grompe M, Ranganath LR, Gallagher JA, Jarvis JC. Ochronosis in a murine model of alkaptonuria is synonymous to that in the human condition. Osteoarthritis Cartilage. 2012 Aug;20(8):880-6. doi: 10.1016/j.joca.2012.04.013. Epub 2012 Apr 24.

[9] Laschi M, Tinti L, Braconi D, Millucci L, Ghezzi L, Amato L, Selvi E, Spreafico A, Bernardini G, Santucci A. Homogentisate 1,2 dioxygenase is expressed in human osteoarticular cells: implications in alkaptonuria. J Cell Physiol. 2012 Sep;227(9):3254-7. doi: 10.1002/jcp.24018.

[10] Aquaron RR. Alkaptonuria in France: past experience and lessons for the future. J Inherit Metab Dis. 2011 Dec; 34(6):1115-26. doi: 10.1007/s10545-011-9392-7. Epub 2011 Sep 17. Review.

[11] Zatkova A. An update on molecular genetics of Alkaptonuria (AKU). J Inherit Metab Dis. 2011 Dec;34(6):1127-36. doi: 10.1007/s10545-011-9363-z. Epub 2011 Jul 1.

[12] Al-sbou M. Novel mutations in the homogentisate 1,2 dioxygenase gene identified in Jordanian patients with alkaptonuria. Rheumatol Int. 2012 Jun;32(6):1741-6. doi: 10.1007/s00296-011-1868-0. Epub 2011 Mar 25.

[13] Sen Gupta, Parth Sarthi; Bandyopadhyay Amal Kumar. Sequence, Structural and Functional Characterization of Chalcone Synthase of Solanum tuberosum: An in silico analysis; Structural biology(under review)

[14] Castrignanò T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A. (2006). The PMDB Protein Model Database. Nucleic Acids Research, 34: D306-D309.

[15] The UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt) Nucleic Acids Res. 40: D71-D75 (2012).

[16] Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995 Dec;11(6):681-684

[17] Ishida, T and Kinoshita, K, PrDOS: prediction of disordered protein regions from amino acid sequence., Nucleic Acids Res, 35, Web Server issue, 2007

[18] Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.Olof Emanuelsson, Henrik Nielsen, Søren Brunak and Gunnar von Heijne.J. Mol. Biol., 300: 1005-1016, 2000.

[19] Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22,195-201.

[20] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry, 26:1781-1802, 2005.

[21] Lindahl E, Azuara C, Koehl P, Delarue M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis.

[22] Debswapna Bhattacharya and Jianlin Cheng. 3Drefine: Consistent Protein Structure Refinement by Optimizing Hydrogen-Bonding Network and Atomic-Level Energy Minimization. Proteins: Structure, Function and Bioinformatics. 2012 (In Press).

[23] GlobPlot: exploring protein sequences for globularity and disorder.Nucleic Acid Res 2003 - Vol. 31, No.13

[24] Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst., 26, 283-291

[25] Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993 Sep;2(9):1511-9.

[26] Gopalkrishnan K, Sowmiya G, Sheik SS, Sekar K, Ramachandran plot on the web (2.0). PROTEIN PEPT.LETT. (2007), 14,669-671

[27] Leigh Willard, Anuj Ranjan,Haiyan Zhang,Hassan Monzavi, Robert F. Boyko, Brian D. Sykes, and David S. Wishart "VADAR: a web server for quantitative evaluation of protein structure quality" Nucleic Acids Res. 2003 July 1; 31 (13): 3316.3319

[28] Mark Berjanskii, Jianjun Zhou, Yongjie Liang, Guohui Lin and David S. Wishart "Resolution-by-Proxy: A Simple Measure for Assessing and Comparing the Overall Quality of NMR Protein Structures", J Biomol NMR. 2012 Jul;53(3): 167-80

[29] Zengming Zhang, Yu Li, Biaoyang Lin, Michael Schroeder and Bingding Huang (2011), Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics, 27 (15): 2083-2088.

[30] Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. *New and continuing developments at PROSITE* Nucleic Acids Res. 2012; doi: 10.1093/nar/gks1067

[31] *(My Hits, SIB, Switzerland)* includes Prosite, Pfam and HAMAP profiles.

[32] A Hierarchical Approach to Protein Fold Prediction Tabrez Anwar Shamim Mohammad1 and Hampapathalu Adimurthy Nagarajaram2 Laboratory of Computational Biology, CDFD, Bldg.7, Gruhakalpa, Nampally, Hyderabad