Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 5, Issue1, 2014, pp8-22
http://www.bipublication.com

AN INSIGHT INTO THE STRUCTURE AND FUNCTION OF CHALCONE SYNTHASE FROM SEQUENCE OF *SOLANUM TUBEROSUM*

Parth Sarthi Sen Gupta[§], Shyamashree Banerjee[§], Sudipta Mondal[§], Buddhadev Mondal[#] and Amal K. Bandyopadhyay[§]*

[§]The Department of Biotechnology,
The University of Burdwan, Golapbag, Burdwan, 713104, West Bengal, India.
[#] Burdwan Raj College, Burdwan University, Burdwan, 713104, West Bengal, India
***Correspondence: Amal K. Bandyopadhyay. Email: akbanerjee@biotech.burunib.ac.in

[Received-08/01/2014, Accepted-11/02/2014]

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 there are no three dimensional(3D) structure of protein 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 chalcone synthase of *solanum tuberosum.* Primary structure analysis reveals that all the four sequences of chalcone synthase are hydrophilic in nature (due to the high content of polar residues) without transmembrane region. The aliphatic index computation infers that chalcone synthase of *solanum tuberosum.* Can't tolerate wide range of temperature. Secondary structure analysis shows that all of the sequences of chalcone synthase have predominant α-helical structure. Sub cellular localization prediction suggested that all of the sequences are cytosolic 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 chalcone synthase is successfully submitted to the Protein Model Database (PMDB) having PMID PM0078945.Active site of the structure is predicted and the analysis shows that it contains a conserved catalytic triad of Cys164, His303 and Asn336. Protein disorder, disorder propensity and Average Area Buried Upon Folding is also predicted and analyzed. Fold of the protein is predicted to be (TIM)-barrel 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: *solanum tuberosum, molecular modeling, chalcone synthase, disorder, aliphatic index*

1. INTRODUCTION:

Chalcone synthase is an enzyme generally present in the higher plants belongs to a family of polyketide synthase enzymes (PKS) known as type III PKS. It is the first committed enzyme in

flavonoid biosynthesis [1]. The enzyme catalyzes the conversion of 4-coumaroyl-CoA and malonyl-CoA to naringenin chalcone, serving as general precursor of various flavonoids. Flavonoids are important plant secondary metabolites that serve various functions in higher plants. These include pigmentation, UV protection, fertility, antifungal defense and the recruitment of nitrogen-fixing bacteria [2]. Chalcone synthase also facilitate the production of isoflavonoid-type phytoalexins and other metabolites to protect the plant from stress. Expression of chalcone synthase is also involved in the salicylic acid defense pathway. It is also involved in a phenylpropanoid pathway which serve as precursors to a range of plant metabolites important to human health such as antioxidants, anti-inflammatory agents, antiallergens, and even antioncogenic products [3]. Chalcone synthase is localized in the cytosol, associated with the endoplasmic reticulum membrane [4].Crystal structure of *medicago sativa* had been analyzed and showed that the active site of the enzyme contains four highly conserved residue $Phe_{202,}$ $His_{304,}$ $Asn_{340,}Cys_{164}$[5,6] . The active site contains a conserved catalytic triad of Cys164, His303 and Asn336. These residues aid in multiple decarboxlyation and condensation reactions, with Cys164 acting as the active site nucleophile. Phe215 and Phe265 are two other important amino acids that act as "gatekeepers" to block the lower protein of the opening between the CoA-binding tunnel and the active site cavity. This limits the access of water to the active site while accommodating substrates and intermediates of varying shapes and sizes. Phe215 also orients the substrates at the active site during elongation of the polyketide intermediate.

Plenty of work has been done on chalcone synthase of various organisms [7to16] but no or little work has been done on *solanum tuberosum*. The Protein Data Bank (PDB) (www.rscb.org) does not contain a single three dimensional structure of chalcone synthase of *solanum tuberosum.* Therefore, it is very interesting as well as important to carry out the study of chalcone synthase of *solanum tuberosum.* In the present study Sequence, Structural and Functional Characterization of Chalcone Synthase of *Solanum tuberosum*: 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 chalcone synthase. The predicted 3D structure is also successfully submitted in the Protein Model Data Base (PMDB)[17] having PMID PM0078945.The results of this work can further help researcher by providing theoretical basis on enzymological properties, structure and function of chalcone synthase in future.

2. MATERIALS AND METHODS:

2.1 Sequence Retrieval

A total of four sequences of chalcone synthase protein were retrieved from the manually curate public protein database Swiss-Prot [18].Swiss-Prot is scanned for the keyword chalcone synthase and *solanum tuberosum.* The search result yielded 4 sequences of chalcone synthase (CHS) gene family of *solanum tuberosum* (Table 1). All of the four sequences were retrieved in FASTA format and used for further analysis.

Table 1. Sequences of chalcone synthase of *solanum tuberosum* retrieved from Swiss-Prot database.

Accession number	Sequence description	Organism
Q43188	**chalcone synthase 2**	*solanum tuberosum* (Potato)
Q41436	**chalcone synthase 1A**	*solanum tuberosum* (Potato)
Q43163	c**halcone synthase 1B**	z*solanum tuberosum* (Potato)
G3E943	**chalcone synthase**	*solanum tuberosum* (Potato)

2.2 Primary Structure Analysis using Computational Tools and Servers

The amino acid composition (table 2) of chalcone synthase 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 chalcone synthase computed using CLC free Workbench tool.

	Accession number			
Amino Acids	Q43188	Q41436	Q43163	G3E943
Ala	7.5	7.5	7.7	7.7
Cys	1.8	2.1	2.1	2.1
Asp	5.1	4.4	4.4	4.4
Glu	6.9	7.5	7.5	7.2
Phe	3.3	3.3	3.3	3.3
Gly	8.2	8.7	8.7	8.7
His	2.1	2.1	2.1	2.1
Ile	5.7	5.1	5.1	5.1
Lys	6.7	6.7	6.9	6.7
Leu	10.3	11.3	11.3	11.3
Met	3.6	3.1	3.1	2.8
Asn	2.6	2.8	2.6	2.8
Pro	5.1	4.6	4.6	4.6
Gln	3.6	3.3	3.3	3.6
Arg	3.6	4.1	4.1	4.1
Ser	6.9	6.4	6.4	6.4
Thr	6.2	5.9	5.9	5.9
Val	7.7	7.7	7.5	7.7
Trp	1.0	1.0	1.0	1.0
Tyr	2.1	2.3	2.3	2.3

Table 3. Hydrophilic and hydrophobic residues content.

Accession Number	Q43188	Q41436	Q43163	G3E943
Percentage of hydrophobic residues	46	45.7	45.7	45.6
Percentage of hydrophilic residues	54	54.2	54.2	54.2
Net hydrophobic residues content	low	low	low	low

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	**Q43188**	**Q41436**	**Q43163**	**G3E943**
Sequence length	389	389	389	389
M. wt	42476.0	42562.2	42548.2	42501.1
pI	5.71	6.12	6.28	6.28
-R	47	46	46	45
+R	40	42	43	42
EC	34295	35910	35910	35910
II	39.60	37.30	47.90	37.17
AI	91.98	93.98	93.50	94.24
GRAVY	-0.067	-0.060	-0.067	-0.060

The secondary structure were predicted using the tools SOPMA [19] (Table 5).The total protein intrinsic disorder and the protein disorder propensity was detected using PrDOS [20] 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 [21].The transmembrane regions were predicted with TMHMM (Krogh *et al.*, 2001) server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/).

Table 5. Secondary structure calculation (in %) of chalcone synthase computed using SOPMA

Accession number	**Secondary Structure**			
	Alpha Helix	Beta Sheet	Random Coil	Others
Q43188	44.47	16.71	32.13	6.68
Q41436	46.02	14.91	32.13	6.94
Q43163	46.27	15.17	30.59	7.97
G3E943	47.30	14.14	31.88	6.68

2.3 Template Selection and Molecular Homology Modeling

The protein sequence of chalcone synthase having accession ID Q43188 has no three dimensional structure present in Protein Data Bank (PDB).Therefore, the template (PDB ID 1CGZ_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 chalcone synthase having accession ID Q43188 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 [22].

2.4 Energy Minimization, Model Evaluation and Submission

The energy minimization for the 3D structure was carried out using NAMD [23] utilizing CHARMM force field and NOMAD-Ref server [24] which utilizes Gromacs forcefield according to steepest descent, conjugate gradient and L-BFGS methods . Further, refinement of the modeled structure was performed using *3Drefine* server [25]. Structural evaluation, validation and stereochemical analyses were performed using various evaluation tools such as Rampage , Procheck [26], Errat [27], Ramachandran plot 2 [28], and Vadar [29]. Protein Quality was checked by Resprox [30].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 PM0078945. 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 Characterization of the Modeled Structure-

The modeled structure is characterized by their secondary structure contents. The model structure is further analyzed using PDBsum (http://www.ebi.ac.uk/pdbsum/).The PDBsum is a pictorial database that provides an at-a-glance overview of the contents of each 3D structure deposited in the Protein Data Bank (PDB).It shows the molecule(s) that make up the structure (*i.e.* protein chains, DNA, ligands and metal ions) and schematic diagrams of their interactions. Entries are accessed either by their 4-character PDB code, or by one of the two search boxes provided on the PDBsum home page . We got the 4-character PDB code of the chalcone synthase protein and the code is gk99.Since it is password protected so to see the result this Password: 002820 is required.

2.6 Active Site Prediction

After the final model was built, the active site of the structure was predicted using the tool Meta pocket 2[31].The visualization and analysis of the active site was performed by using UCSF Chimera 1.5.3.

2.7 Identification of Motif, Fold and Functional Domain

Motif of the protein were identified using the tools Prosite [32] and MotifScan [33].The protein fold was predicted using the software PFP-pred [34].

2.8 Function Prediction

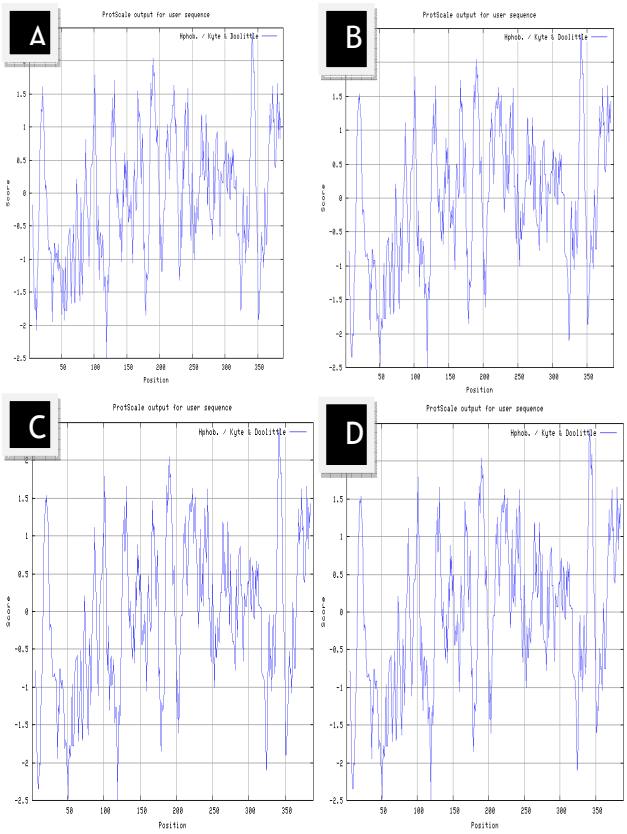
The function of the protein chalcone synthase of solanum tuberosum was predicted by using ProFunc 2 [35]. The aim of the ProFunc server is to help identify the likely biochemical function of a protein from its three-dimensional structure. It uses a series of methods, including fold matching, residue 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 chalcone synthase are hydrophilic in nature due to the presence of high content of polar residues (tables 2 and 3). The kye-

dollitle hydrophobicity is analyzed using ProtScale and shown in figure 1.



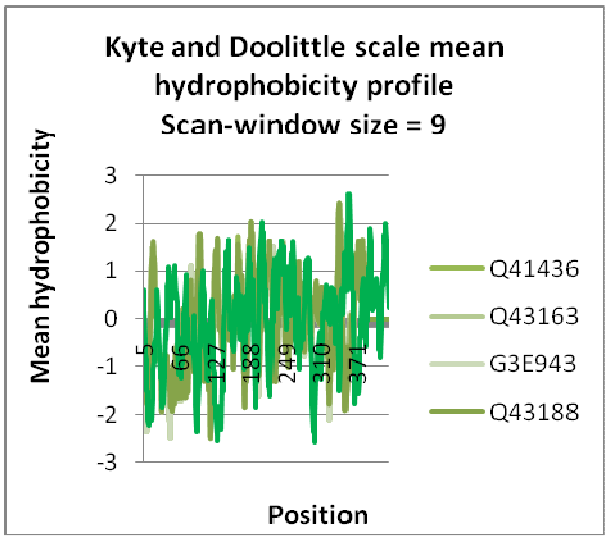


Fig: 1. Kyte-doolittle hydrophobicity by ProtScale. (A) For the sequence with accession id Q43188 (B) For the sequence with accession id Q41436 (C) For the sequence with accession id Q43163 (D) For the sequence with accession id G3E943 (E) Kyte-Doolittle scale mean hydrophobicity profile for all the four sequences of chalcone synthase**.**

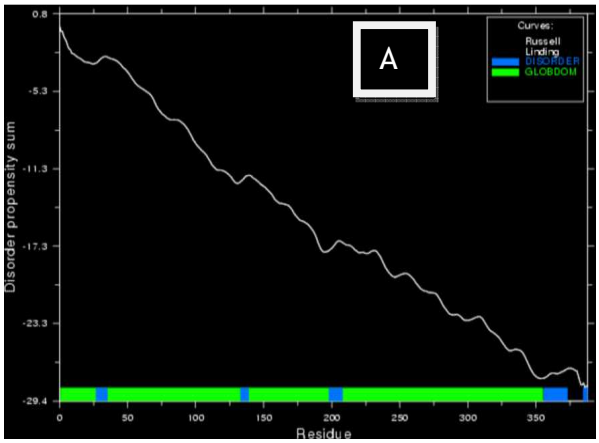
The analysis also suggests that the chalcone synthases contain high residues of acidic and basic amino acid, this might be involves in salt bridge formation. The average molecular weight of chalcone synthase is 42521.85 Da. The proteins are generally stable and compact at PI. Isoelectic focusing method is utilized for purification using computed PI by developing buffer systems. All of the four sequences of chalcone synthase Q43188, Q41436, Q43163 and G3E943 have (pI < 7) 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 chalcone synthase at 280 nm is ranging from 34295 to 35910 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 chalcone synthases 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 chalcone synthases 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 chalcone synthase is indicative of a more flexible structure (table 4). The low aliphatic index of most of the chalcone synthases infers that it may be thermo labile for a high range of temperature. Grand Average hydropathy (GRAVY) Index of chalcone synthases are ranging from –0.060 to -0.067. The very low GRAVY index of chalcone synthase

infers that these could result in a better interaction with water. The secondary structure predicted with the help of programs SOPM and SOPMA (table 5) infers that the chalcone synthase have rich alanine content and mostly α-helices. Sub cellular localization is predicted by TargetP server which suggested that all of the sequences of chalcone synthase are cytosolic without transit peptide (Table 6).

Table 6. Sub cellular localization by TargetP

Accession No.	Sub Cellular Location	MOTIFS binding and phosphorylation site
Q43188	**Cytosol**	**Malonyl-CoA**
		Protein kinase C
		Casein kinase II
		N-myristoylation
		Chalcone and stilbene synthases active site
Q41436		**N-glycosylation**
		cAMP- and cGMP-dependent protein kinase
		Prenyl group binding site (CAAX box)
Q43163		**Prokaryotic membrane lipoprotein lipid attachment**
		Leucine zipper pattern
G3E943		**Microbodies C-terminal targeting signal**

The transmembrane regions were predicted with TMHMM server 2 suggests that chalcone synthase 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).



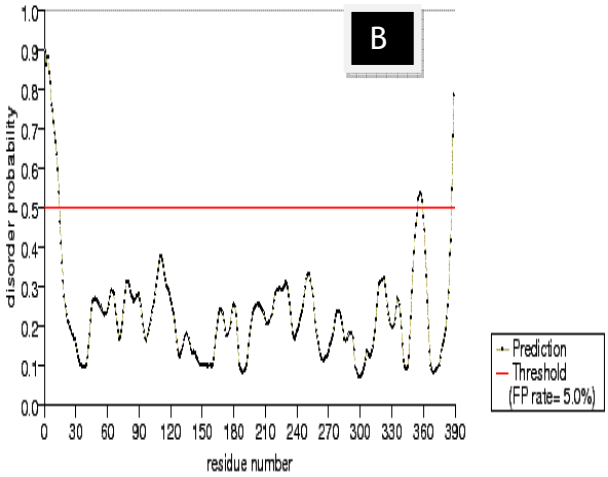


Fig: 2. Disordered Regions of the Chalcone Synthase protein (A) Disorder Propensity Sum (B) schematic diagram of disorder Probability over the entire sequence.

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

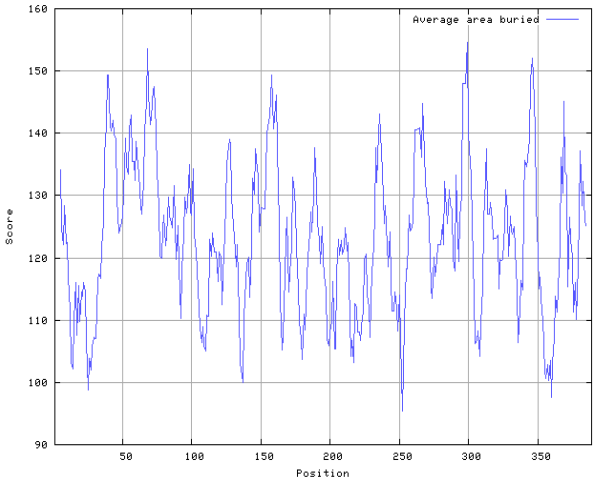


Fig :3. Average Area Buried upon Folding For all the Amino acids residue of the Protein predicted by Protscale from ExPASy.

3.2 Molecular Homology Modeling, Energy minimization and Model Evaluation

The protein sequence of chalcone synthase having accession ID Q43188 has no three dimensional structure present in Protein Data Bank (PDB).Therefore, the template (PDB ID 1CGZ_A) are identified by the BLASTP analysis in the Protein Data Bank (PDB)(Table 7).

Accession number	PDB code
Q43188	1CGZ_A

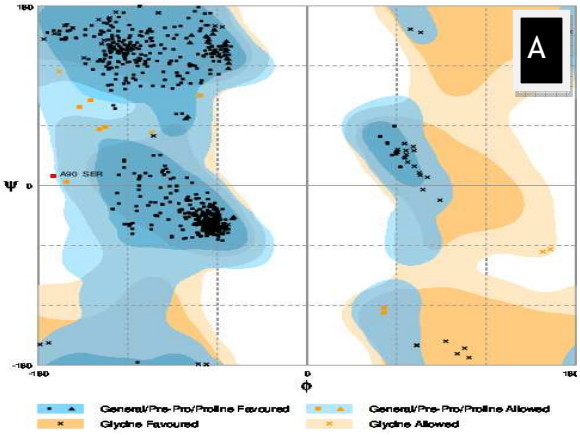
Table 7. PDB template using BLASTP search against the Protein Data Bank.

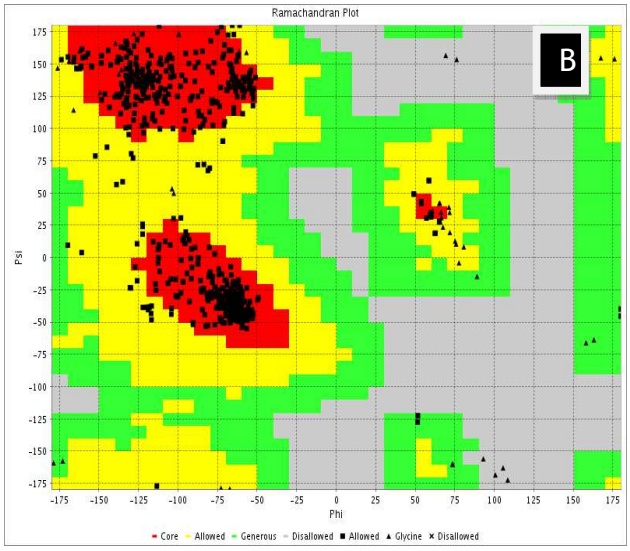
The energy minimization was achieved by NAMD and NOMAD-Ref server. After minimization of the model structure, RMSD of the structure is calculated to be 0 .6545, Average surface area calculated is 23749.92 and the energy before minimization and after minimization is calculated which is -1876.3 Kcal/mol and -15656.9Kcal/mol respectively (Table 8).

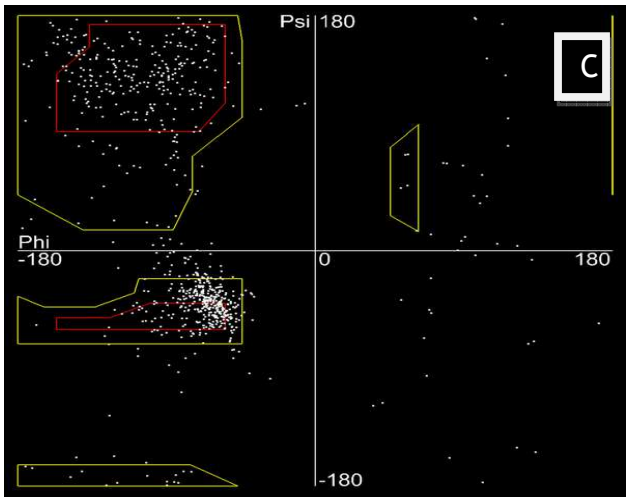
Table 8. Energy before and after minimization, Average RMSD and Average Surface Area of the Modeled Structure

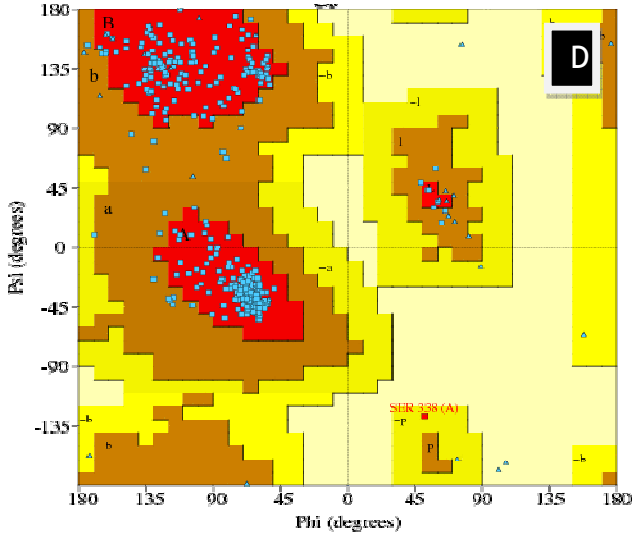
Accession No.	Energy before minimization Kcal/mol	Energy after minimization Kcal/mol	Average RMSD	Average Surface Area
Q43188	**-1876.3**	**-15656.9**	**0.6545**	**23749.92**

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 suggests that the modeled structure is to be good model (Figure 4).









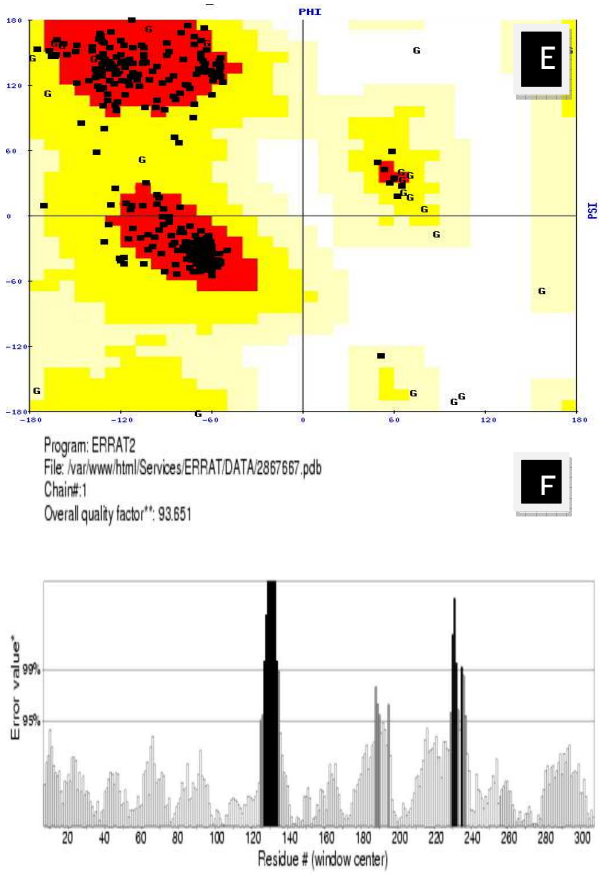


Fig: 4. Ramachandran plot using (A) Rampage (B) Vadar (C) Ramachandran Plot 2 (D, E) Procheck (F) Z score using Errat

According to evaluation analysis, the Ramachandran plot and other parameters (Table 9) were within the standard acceptable limits for the 3D structures modeled using the PDB template 1CGZ_A for the Q43188 protein (Table 10).

SOFTWARE		Quality of Model		
		Fairly Good	**Very Good**	**Extremely Good**
ProQ	LG score	>0.1	>0.5	>0.8
	Maxsub	>2.5	1.5 – 2.5	0 – 1.5
ResProx	Predicted resolution(Å)	>1.5	>2.5	>4.0
Rampage	Percentage of residues in favored region	98		
PROCHEK	Percentage of residues in favored region	90		

Table 9 Criteria for a good (model) 3D structure.

Target		Q43188
Template(PDB Codes)		**1CGZ_A**
SOFTWARE		SCORE
ProQ	LG score	**7.439**
	Maxsub	**0.617**
ResProx	Predicted resolution(Å)	**1.567**
Rampage	Percentage of residues in favored region	**98.7**
PROCHEK	Percentage of residues in favored region	**91.6**

Table 10. Validation parameters computed for the built 3D structures of target Q43188.

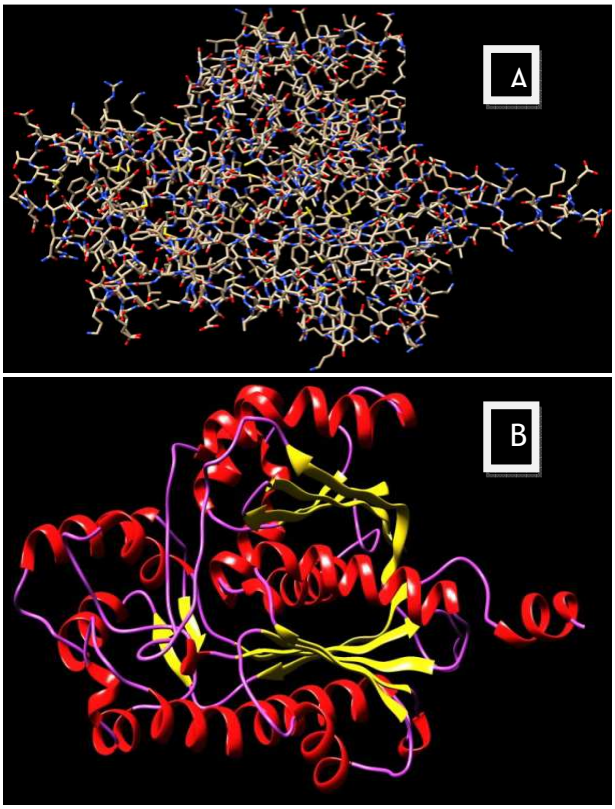
Protein Quality is checked by Resprox and the structure fulfill all the criteria of good model (Table 11)

Table 11. 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	1.02	Standard deviation of the χ1 angles among all 3 (gauche-, gauche+, and trans) configurations.	Vadar
Clash score	Good	-0.07	Number of non-hydrogen bond atomic overlaps > 0.4 Å per thousand atoms.	MolProbity
Mean H-bond energy	Good	-0.27	A measure of underpacking in the protein core.	Rosetta
χ1 score	Good	1.00	Standard deviation of angle between the C-O bond vector of the H-bond acceptor and the O-H(N) bond vector.	PROSESS
Radius gyration 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
Percentage of generously allowed Ω angles	Good	-0.36	The average hydrogen bond energy is calculated using the H-bond energy function used in DSSP program.	Vadar
Percentage of packing defects	Good	1.23	Scaled difference between the standard deviation of the observed χ1 angles and the expected one obtained from high quality protein structures.	PROSESS
Percentage of 95% buried	Good	0.75	Scaled difference between the expected radius of gyration and the observed one. The	GeNMR

residues			expected radius of gyration is determined using: Rg = 0.395*N**0.6 + 7.257.	
Percentage of bad bond length	Good	0.52	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 bad bond angles	Good	0.75	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
Ramachandran plot outliers	Good	2.47	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

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.



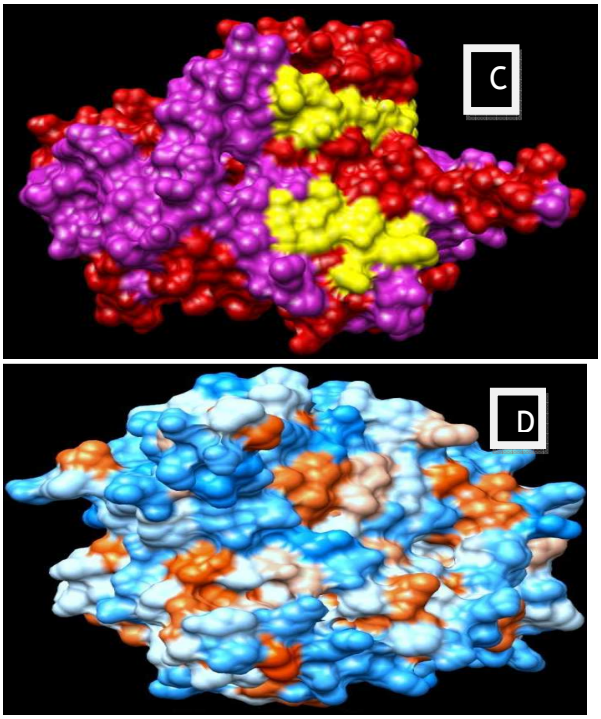


Fig: 5. Visualization by using UCSF Chimera 1.5.3 (A) Stick Model of the Structure (B) Secondary Structure of the Model (C) Surface View of the model Structure (D) Molecular Surface Colored by amino acid Hydrophobicity (Dodger Blue-most hydrophilic, White-Moderate, Orange Red- most hydrophobic)

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 Chalcone synthase contains high amount of alpha helix and coils (table 12) in respect to beta sheet just like the primary structure analysis (table 5).

Table 12. 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
Q43188	**3** (T)	**1**	--------	------	**1**
	4 – 11 (V,E,E,V,R,K,A,G)	**8**	**1**	------	----
	12 – 16 (R,A,K,G,P)	**5**	--------	-------	**2**
	17 – 25(A,T,I,M,A,I,G,T,A)	**9**	--------	**1**	----
	26 – 28 (T,P,S)	**3**	--------	-------	**3**
	29 – 33 (N,C,V,N,Q,)	**5**	--------	**2**	----
	34 – 35 (S,T)	**2**	--------	-------	4

	36 – 43 (Y,P,N.Y,Y,F,R,I)	**8**	**2**	-------	--
	44 – 49 (T,N,S,E,H,M,)	**6**	---------	------	**5**
	50 – 62 (T,E,L,K,E,K,F,L,R,M, C,N,L,)	**13**	**3**	-------	--
	63 – 65 (S,M,I)	**3**	--------	-------	**6**
	66 – 70 (N,K,R,Y,M)	**5**	--------	**3**	--
	71 – 73 (H,L,T)	**3**	--------	-------	**7**
	74 – 78 (E,E.I,L,K)	**5**	**4**	-------	---
	79 – 81 (E,N,P)	**3**	--------	-------	**8**
	82 – 84 (N,I,C)	**3**	**5**	-------	---
	85 – 90 (E,Y,M,A,P,S)	**6**	--------	-------	**9**
	91 – 117 (L,N,A,R,Q,N,I,V,V,V, E,V,P,K,L,G,K,E,A,A, Q,K,A,I,K,E,T)	**27**	**6**	-------	--
	118 – 124 (G,Q,P,K,S,K,I)	**7**	--------	-------	**10**
	125 – 131 (T,H,V,V.F,C,T)	**7**	--------	**4**	--
	132 – 139 (T,S,G,V.D,M,P,G)	**8**	--------	-------	**11**
	140 – 148 (A,D,Y,Q,L,T,K,L,L)	**9**	**7**	-------	--
	149 – 154 (G,L,R,P,S,V)	**6**	--------	-------	**12**
Q43188	**155 – 161** (K,R,L,M,M,Y,Q)	**7**	--------	**5**	-------
	162 – 164 (Q,G,C)	**3**	--------	-------	**13**
	165 – 179 (F,A,G,G,T,V,I,R,L,A, K,D,L,A,E)	**15**	**8**	-------	--
Q43188	**180 – 184** (N,N,K,G,A)	**5**	--------	-------	**14**
	185 – 192 (R,V,L,V,V,C,S,E,)	**8**	--------	**6**	--
	193 – 194 (I,T)	**2**	--------	-------	**15**
	195 – 197 (A,V,T)	**3**	**9**	-------	--
	198 – 205 (F,S,G,P,S,D,T,H,)	**8**	--------	-------	**16**
	206 – 214 (L,D,S,M,V,G,Q,A,L,)	**9**	**10**	-------	--
	215 – 217 (F,G,D)	**3**	--------	-------	**17**
	218 – 225 (G,A,A,A,M,I,I,G)	**8**	--------	**7**	--
	226 – 236 (S,D,P,L,P,E,V,E,R,P, L)	**11**	--------	-------	**18**
	237 – 247 (F,E,L,V,S,A,A,Q,T,L, L)	**11**	--------	**8**	--
	248 – 252 (P,D,S,E,G,)	**5**	--------	-------	**19**
	253 – 259 (A,I,D,G,H,L,R)	**7**	---------	**9**	--
	260 – 261 (E,V)	**2**	---------	-------	**20**
	262 – 268 (G,L,T,F,H,L,L,)	**7**	---------	**10**	------
	269 – 270 (K,D)	**2**	---------	-------	**21**
	271 – 286 (V,P,G,L,I,S,K,N,I,E,K ,S,L,I,E,A)	**16**	**11**	-------	--
	287 – 298 (F,Q,P,L,G,I,S,N,W,N, S,I,)	**12**	--------	-------	**22**

	299 – 303 (F,W,I,A,H)	**4**	--------	**11**	--
	304 – 307 (P,G,G,P)	**4**	--------	-------	**23**
	308 – 317 (A,I,L,N,Q,V,E,L,K,L)	**10**	**12**	-------	--
	318 – 324 (G,L,K,P,E,K,L,)	**7**	--------	-------	**24**
	325 – 333 (Q,A,T,R,Q,V,L,S,D)	**9**	**13**	-------	---
	334 – 339 (Y,G,N,M,S,S)	**6**	--------	-------	**25**
	340 – 352 (A,C,V,L,F,I,L,D,E,M, R,K,A)	**13**	**14**	-------	--
	353 – 358 (S,S,K,E,G,L,)	**6**	--------	-------	**26**
	359 – 361 (S,T,T)	**3**	--------	**12**	--
	362 – 365 (G,E,G,L)	**4**	--------	-------	**27**
	366 – 374 (D,W,G,V,L,F,G,F.G)	**9**	--------	**13**	--
	375 – 376 (P,G)	**2**	--------	-------	**28**
	377 – 388 (L,T,V,E,T,V,V,L,H,S, V,S)	**12**	--------	**14**	--

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

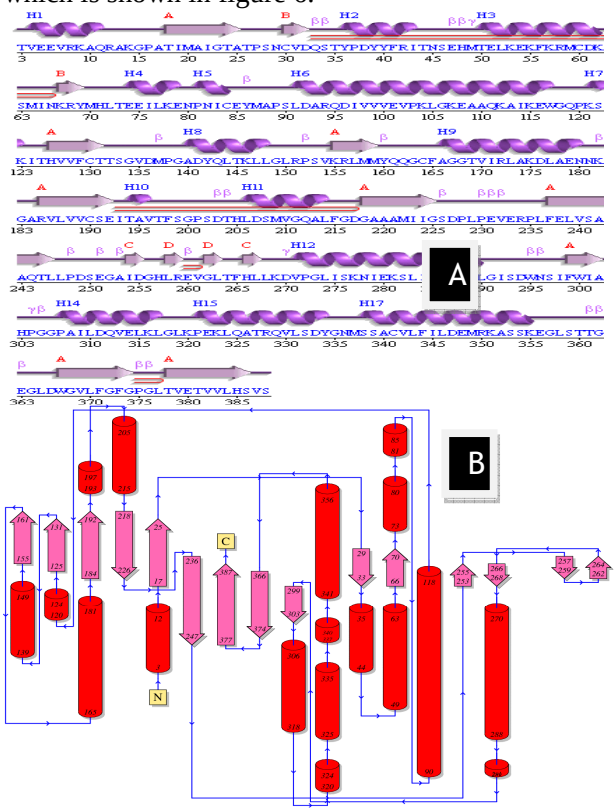


Fig: 6 (A) Secondary structure of the Modeled Structure, Helices are labeled as H; Beta turn as β; Gamma turn as χ and Beta hairpin as ⇒ (B) Topology of Secondary Structure of Model, Helices and Sheets are shown in Pink And Red.

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

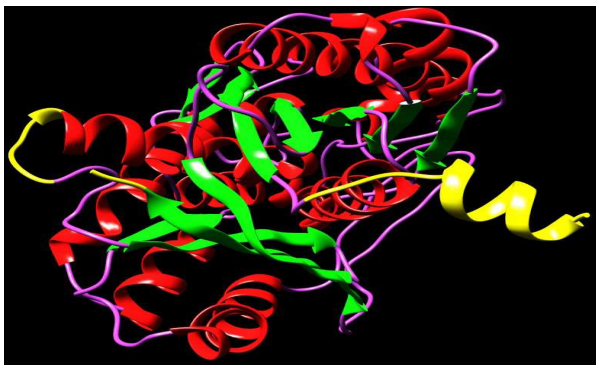


Fig: 7. Disordered Portions of the Protein is shown in Three Dimensional Structure, Disordered Regions are shown in Yellow Color.

For more information just follow this link http://www.ebi.ac.uk/pdbsum/ by using PDB code: gk99 and Password: 002820 or directly by bookmarking this link: http://www.ebi.ac.uk/thornton-srv/databases/cgi-in/pdbsum/GetPage.pl?pdbcode=gk99&code=002820.The visualization and analysis of the active site was performed by using UCSF Chimera 1.5.3 (Figure 8).

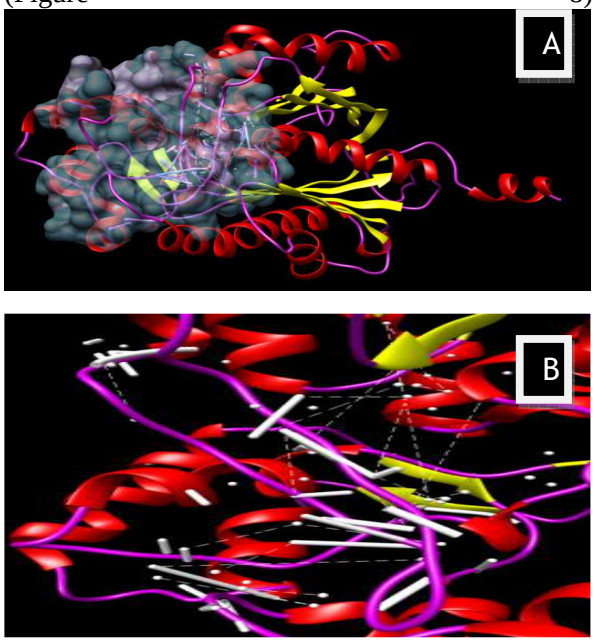


Fig: 8. Visualization by using UCSF Chimera 1.5.3 (A) Surface View of the Active Site (B) Enlarge View of the Active Site

The active site of the structure is analyzed using the tool Meta pocket 2 and showed that the active site of the enzyme contains a conserved catalytic triad of Cys164, His303 and Asn336(colored in Red), Phe215 and Phe265(colored in Orange) are two other important amino acids that act as "gatekeepers" to block the lower protein of the opening between the CoA-binding tunnel and the active site cavity. (Table 13)

Table 13. Analysis of the Active Site of the Structure of Chalcone Synthase Using Meta pockets 2 tools.

Residue		**Atom**
Name	Position	Position
ARG	58	442 – 448
	68	526, 528 – 529
	94	740
MET	59	450, 452 – 454, 456
	64	486, 490 - 493
	209	1602 - 1603
LYS	62	472 - 478
	55	410, 412
	269	2045 – 2046
ALA	88	195, 213, 694, 2346 – 2347, 2350
	195	1502
	213	1634
	308	2346 – 2347, 2350,
VAL	31	202, 206 – 207
	196	1510, 1512 - 1513
	210	1611, 1615 - 1616
	261	529, 740, 750, 752 – 753
	271	2060, 2064 – 2065
ASP	32	208, 211
	61	465, 466
	217	1665
	255	1932 - 1934
	270	2052 – 2053, 2056 – 2058
PRO	272	2072 - 2073
	307	2343 - 2345
GLN	161	1255, 1257 – 1263,
	162	1265 - 1267
	212	1622, 1624, 1626 - 1629

	312	2383
SER	63	481
	133	1041 – 1042
	199	1536 – 1537
	208	1599 – 1601
	338	2581, 258 – 2586
LEU	72	560, 562 – 567
	77	607
	214	1640 – 1642
	258	1952
	263	1991, 1993 – 1998
	267	2029 – 2034
	377	2871
PHE	165	2871
	198	1522, 1525, 1528, 1530
	215	1646 - 1653
	265	2006, 2009 – 2010, 2012, 2014, 2016
	373	2841 – 2843, 2848
ASN	29	192 - 193
	336	2568, 2572
CYS	30	198
	60	459 - 460
	164	1277, 1280 - 1282
GLY	163	1273 - 1274
	211	1618 - 1620
	216	1657
	200	1541
	256	1938 - 1941
	262	1988 – 1989
	305	2332 - 2334
	306	2335 - 2336
	374	2335 - 2336
ILE	254	1925, 1928 – 1929
	309	2356 – 2358
THR	73	569 – 570
	131	1028
	132	1033 – 1035, 1037
	194	1496, 1999 - 1501
	197	1516 - 1520
	264	2000, 2002 - 2005
HIS	205	1574
	257	1943
	266	2023, 2025 - 2026
	303	2318, 2320 -2321
ILE	43	315,317
	193	1487, 1489 - 1490
GLU	74	575,579
	85	667
	192	1482, 1484 - 1885
TYR	36	244, 246, 249
	39	266, 269 - 276
	40	278
	86	673, 676 – 677, 679 -684

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

Motif of the protein is identified using the tools Prosite and MotifScan (Table 14). PFP-pred predicted the protein fold to be (TIM)-barrel

Table 14. Motifs, number of Sites and their Corresponding Residues Identified by using Prosite and MotifScan

Accession No.	Motifs		No. of site	Residues
	Pattern	Binding Site & Phosphorilation Site		
Q43188	ASN_GLYCOSYLATION	N-glycosylation site	1	NMSS (336 - 339)
	cAMP_PHOSPHO_SITE	cAMP & cGMP dependent protein kinase	1	RKAS (350 - 353)
	PKC_PHOSPHO_SITE	Protein kinase C phosphorylation site	2	SVK (153 - 155) SSK (353 - 355)
	CK2_PHOSPHO_SITE	Casein kinase II phosphorylation site	8	TVEE(3 - 6)TYPD(35 - 38)TNSE(44 - 47) SGVD(133 - 136)THLD(204 - 207) SLIE(282 - 285)SSKE(353 –356) TTGE(360 - 363)
	MYRISTYL	N-myristoylation site	8	GQPKSK(118 - 123)GLRPSV(149 - 154) GCFAGG(163 - 168)GAIDGH(252 - 257) GNMSSA(335 - 340)GLSTTG(357 - 362) GLDWGV(364 - 369)GVLFGF(368 - 373)
	PRENYLATION	Prenyl group binding site (CAAX box)	1	CVLF (341 - 344)
	MICROBODIES_CTER	Microbodies C-terminal targeting signal	6	SKI(122 - 124)THV(125 - 127)TKL(145 - 147)ARV(184 - 186)THL(204 - 206)GHL(256 - 258)
	LEUCINE_ZIPPER	Leucine zipper pattern	1	LDQVELKLGLKPEK LQATRQVL (310 - 331)
	CHALCONE_SYNTH	Chalcone and stilbene synthases active site	1	RLMMYQQGCFAGG TVIR (156 – 172)

The function of the protein is identified using ProFunc 2 and it predicted that the chalcone synthase is involved in many biological and catalytical functions (Table 15).

Table 15. Some of the Functions of Chalcone Synthase As predicted by using ProFunc 2 Server.

Catalytic process	Score
catalytic activity	141.85
transferase activity	137.20
transferase activity\, transferring acyl groups	137.20
transferase activity\, transferring acyl groups other than amino\-acyl groups	130.61
acyltransferase activity	27.99
3\-oxoacyl\-[acyl\-carrier\-protein]synthase activity	27.35
fatty\-acid synthase activity	27.35
beta\-ketoacyl\-acyl\-carrier\-protein synthase III activity	21.29
naringenin\-chalcone synthase activity	17.85
binding	13.89
protein binding	13.89
identical protein binding	13.12
protein homodimerization activity	9.72
protein dimerization activity	9.72

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: gk99 and Password: 002820 or directly by bookmarking this link: http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/profunc/GetResults.pl?source=profunc&user_id=gk99&code=002820.

4. CONCLUSION:

The sequences of chalcone synthase from *solanum tuberosum* is choosen for the present study as because very little or no information is present currently. Primary analysis suggests that it is hydrophilic in nature and localized in cytosol 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 alpha helices. 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 such as Cys164, His303 and Asn336 and act as catalytic triad. 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 Flavonoid and chalcone synthesis 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.

ACKNOWLEDGEMENT:

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

REFERENCES:

1. Tohge T, Yonekura-Sakakibara K, Niida R, Wantanabe-Takahasi A, Saito K (2007). "Phytochemical genomics in Arabidopsis thaliana: A case study for functional identification of flavonoid biosynthesis genes". Pure and Applied Chemistry **79** (4): 811–23. doi:10.1351/pac200779040811
2. Cain CC, Saslowsky DE, Walker RA, Shirley BW (October 1997). "Expression of chalcone synthase and chalcone isomerase proteins in Arabidopsis seedlings". Plant Mol. Biol. **35** (3): 377–81. doi:10.1023/A:1005846620791. PMID 9349261
3. Choi O, Wu CZ, Kang SY, Ahn JS, Uhm TB, Hong YS (2011). "Biosynthesis of plant-specific phenylpropanoids by construction of an artificial biosynthetic pathway in Escherichia coli". Journal of industrial microbiology & biotechnology **38** (10): 1657–65. doi:10.1007/s10295-011-0954-3. PMID 21424580
4. Hzardina G, Jensen RA (1992). "Spatial organization of enzymes in plant metabolic pathways". Annu Rev Plant Physiol Plant Mol Biol **43**: 241 – 67.
5. Shao H, Dixon RA, Wang X. Crystal structure of vestitone reductase from alfalfa (Medicago sativa L..J Mol Biol. 2007 May 25;369(1):265-76. Epub 2007 Mar 21.
6. Wang X, He X, Lin J, Shao H, Chang Z, Dixon RA.Crystal structure of isoflavone reductase from alfalfa (Medicago sativa L.). J Mol Biol. 2006 May 19;358(5):1341-52. Epub 2006 Mar 29.

7. Rühmann S, Pfeiffer J, Brunner P, Szankowski I, Fischer TC, Forkmann G, Treutter D. Induction of stilbene phytoalexins in grapevine (Vitis vinifera) and transgenic stilbene synthase-apple plants (Malus domestica) by a culture filtrate of Aureobasidium pullulans. Plant Physiol Biochem. 2013 Mar 23. doi:pii: S0981-9428(13)00099-5. 10.1016/j.plaphy.2013.03.011.
8. Huang L, Wang H, Ye H, Du Z, Zhang Y, Beerhues L, Liu B. Differential expression of benzophenone synthase and chalcone synthase in Hypericum sampsonii. Nat Prod Commun. 2012 Dec;7(12):1615-8.
9. Dare AP, Tomes S, Jones M, McGhie TK, Stevenson DE, Johnson RA, Greenwood DR, Hellens RP. Phenotypic changes associated with RNA interference silencing of chalcone synthase in apple (Malus × domestica). Plant J. 2013 May;74(3):398-410. doi: 10.1111/tpj.12140. Epub 2013 Mar 14.
10. Füssy Z, Patzak J, Stehlík J, Matoušek J. Imbalance in expression of hop (Humulus lupulus) chalcone synthase H1 and its regulators during hop stunt viroid pathogenesis. J Plant Physiol. 2013 May 1;170(7):688-95. doi:10.1016/j.jplph.2012.12.006. Epub 2013 Feb 8.
11. Sen Gupta, Parth Sarthi., Mondal, Buddhadev and Bandyopadhyay Amal Kumar. *In silico* characterization of human tyrosinase Using computational tools and servers. Int J Pharm Bio Sci 2013 July; 4(3): (B) 181 - 193
12. Ghasemzadeh A, Jaafar HZ, Karimi E. Involvement of Salicylic Acid on Antioxidant and Anticancer Properties, Anthocyanin Production and Chalcone Synthase Activity in Ginger (Zingiber officinale Roscoe) Varieties. Int J Mol Sci. 2012 Nov 13;13(11):14828-44. doi: 10.3390/ijms131114828.
13. Katsuyama Y, Ohnishi Y. Type III polyketide synthases in microorganisms. Methods Enzymol. 2012;515:359-77. doi: 10.1016/B978-0-12-394290-6.00017-3.
14. Wakimoto T, Morita H, Abe I. Engineering of plant type III polyketide synthases. Methods Enzymol. 2012;515:337-58. doi: 10.1016/B978-0-12-394290-6.00016-1.
15. Weng JK, Noel JP. Structure-function analyses of plant type III polyketide synthases. Methods Enzymol. 2012;515:317-35. doi: 10.1016/B978-0-12-394290-6.00014-8.
16. Sen Gupta Parth Sarthi, Banerjee Shyamashree, Bandyopadhyay Amal Kumar, Sequence, Structural and Functional Characterization of Homogentisate-1,2-dioxygenase of *homo sapiens*: An *in silico* Analysis, *American Journal of Bioinformatics Research*, Vol. 3 No. 3, 2013, pp. 42-61. doi: 10.5923/j.bioinformatics.20130303.02.
17. Castrignanò T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A. (2006). The PMDB Protein Model Database. Nucleic Acids Research, 34: D306-D309.
18. 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
19. Ishida, T and Kinoshita, K, PrDOS: prediction of disordered protein regions from amino acid sequence., *Nucleic Acids Res*, **35**, Web Server issue, 2007
20. 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.
21. 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.
22. 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.
23. Lindahl E, Azuara C, Koehl P, Delarue M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis.
24. 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).

25. GlobPlot: exploring protein sequences for globularity and disorder.Nucleic Acid Res 2003 - Vol. 31, No.13
26. 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.
27. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993 Sep;2(9):1511-9.
28. Gopalkrishnan K, Sowmiya G, Sheik SS, Sekar K, Ramachandran plot on the web (2.0). PROTEIN PEPT.LETT.(2007), 14,669-671
29. 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
30. 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
31. 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.
32. 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
33. *(MyHits, SIB, Switzerland)* includes Prosite, Pfam and HAMAP profiles.
34. 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
35. Roman A. Laskowski, James D. Watson, and Janet M. Thornton.ProFunc: a server for predicting protein function from 3D structure Nucleic Acids Res. 2005 July 1; 33