©1996-2012 All Rights Reserved. Online Journal of Bioinformatics . You may not store these pages in any form except for your own personal use. All other usage or distribute on is illegal under international copyright treaties. Permission to use any of these pages in any other way besides the before mentioned must be gained in writing from the publisher. This article is exclusively copyrighted in its entirety to OJB publications. This article may be copied once but may not be, reproduced or re-transmitted without the express permission of the editors. This journal satisfies the refereeing requirements (DEST) for the Higher Education Research Data Collection (Australia). Linking:To link to this page or any pages linking to this page you must link directly to this page only here rather than put up your own page.

OJB™

Online Journal of Bioinformatics©

Volume 13(1):120-129, 2012.

In silico analysis and homology model of legume antioxidant proteins.

Sarika[1], M. A. Iquebal[2] and Anil Rai[1]

[1]Centre for Agricultural Bioinformatics, [2]Division of Biometrics and Statistical Modelling, Indian Agricultural Statistics Research Institute, PUSA, New Delhi, INDIA

ABSTRACT

Sarika M, Iquebal A, Rai A., In silico analysis and homology model of legume antioxidant proteins, Online J Bioinform, 13(1):120-129, 2012. The availability of plant genome data has stimulated *in silico* homology modelling of plant proteins which are structurally unknown. The conserved domain database search analysis to determine domain and motif showed that the antioxidants were found to belong to Cu-Zu Superoxide-Dismutase super family. In this study physico-chemical and functional characterization of antioxidant proteins in chickpea, rajmash, pea and *vigna* were determined. Self-optimized prediction with alignment was used to calculate the secondary structural features of protein sequences. The 3-D structure was determined by GENO3D and SWISS-MODEL. SWISS-MODEL proved to be more precise using SAVS. Stereochemical properties and quality of the modelled structures were then assessed by Ramchandran plot analysis validated with PROCHECK. The modelled structures could be used as a foundation for functional analysis of experimentally derived crystal structures.

Keywords: Antioxidants, Chickpea, Homology modelling, Ramachandran plot, Swiss-Model

INTRODUCTION

Widely available digital information about plant genomes and its products has triggered the use of *in-silico* identification of important proteins in crop. Antioxidant proteins are known to interrupt uncontrolled oxidation in certain organelles (Shigeoka *et al*., 2002). The antioxidant defence is primarily constituted by the actions of glutathione peroxidase (GPX), superoxide

dismutase, catalase and ascorbate peroxidase (Barbehenn, 2002). Experimental methods used to characterize a protein involve high cost and time frame. The *in-silico* approaches provide a viable solution to this problem. Further, computational tools play an important role to understand physicochemical and structural properties of proteins.

A large number of computational tools are available for identification and structure prediction of proteins. Since, amino acids are the molecular building blocks of proteins, it provides information required for determining and characterizing molecular function, physical and chemical properties. Determination of protein structure through X-ray crystallography or nuclear magnetic resonance spectroscopy is time consuming and very costly. Protein Data Bank (PDB) is a repository for three-dimensional structural data of large biological molecules submitted by biologists and biochemists from around the world. Still, majority of protein sequences have no structural information. Computational methods for protein structure prediction have gained popularity in recent years. Normally, similarity between two proteins at the sequence level imply structural similarity of these proteins. Although similarity search may not always provide positive result, many structure-function relationships can be deduced from a reasonable model, which may further be used for successful drug design (Yadav *et al*., 2011).
In view of above facts, the *in-silico* identification and analysis of antioxidant proteins in legumes has been carried out. The antioxidant proteins from important legume crops like chickpea, rajmash, pea and *vigna* were screened for which the three dimensional structures were not available at PDB. Conserved Domain Database (CDD) search was performed to find out the domain organization of the proteins in order to describe its structural features and to understand molecular function, the model structures for these proteins were constructed.

MATERIALS AND METHODS

Screening and retrieval of antioxidant proteins in legumes

Antioxidant proteins of legumes were retrieved from SWISS-PROT (Bairoch and Apweiler, 2000) with standard settings (Table 1). Amino acid sequences of Q9ZNQ4, Q9FE12, C4P7F3, D5FUD7, F2WVP3, C4P7F2, P11964 and Q02610 were taken up for this study. The sequences were retrieved in FASTA format and processed further for its physico-chemical characterisation, functional characterization, structure prediction, model building and evaluation. The approach adopted for homology modelling is shown in Figure 1. Antioxidant protein sequences of legumes were aligned using the SIM-alignment tool (http://expasy.org/tools/sim-prot.html) to study the pairwise sequence similarity of the proteins. The CDD analysis was done in order to determine the domain organization of the proteins. SOPMA (Geourjon and Deleage, 1995), the secondary structure prediction tool was used for predicting secondary structural elements. This sequence analysis information was used to study the similarity among the antioxidant proteins taken for the study.

3D Structure generation and Evaluation of model

The 3-D structures of antioxidant proteins under study are not available at Protein Data Bank. 3D models of the proteins were constructed with Geno3D (Combet *et al*., 2002) and SWISS-MODEL (Arnold *et al*. 2006). Geno3D is an automatic web server for molecular modelling of

protein which performs homology modelling in steps, starting with identification of homologous proteins with known 3D structures using PSI-BLAST followed by alignment of both target and template proteins and finally construction of 3D structure of the protein using distance geometry approach. SWISS-MODEL tool may be accessed via ExPASy web server, or from the DeepView (Swiss Pdb-Viewer) programme. In the last step of homology modelling, the modelled structures were subjected to a series of tests for testing its internal consistency and reliability. Backbone conformation was evaluated by the inspection of the Psi/Phi Ramachandran plot (Ramachandran *et al*., 1963) obtained from PROCHECK (Laskowski *et al*., 1996) and WHAT IF (Vriend, 1990) under Structural Analysis and Verification Server (SAVS).

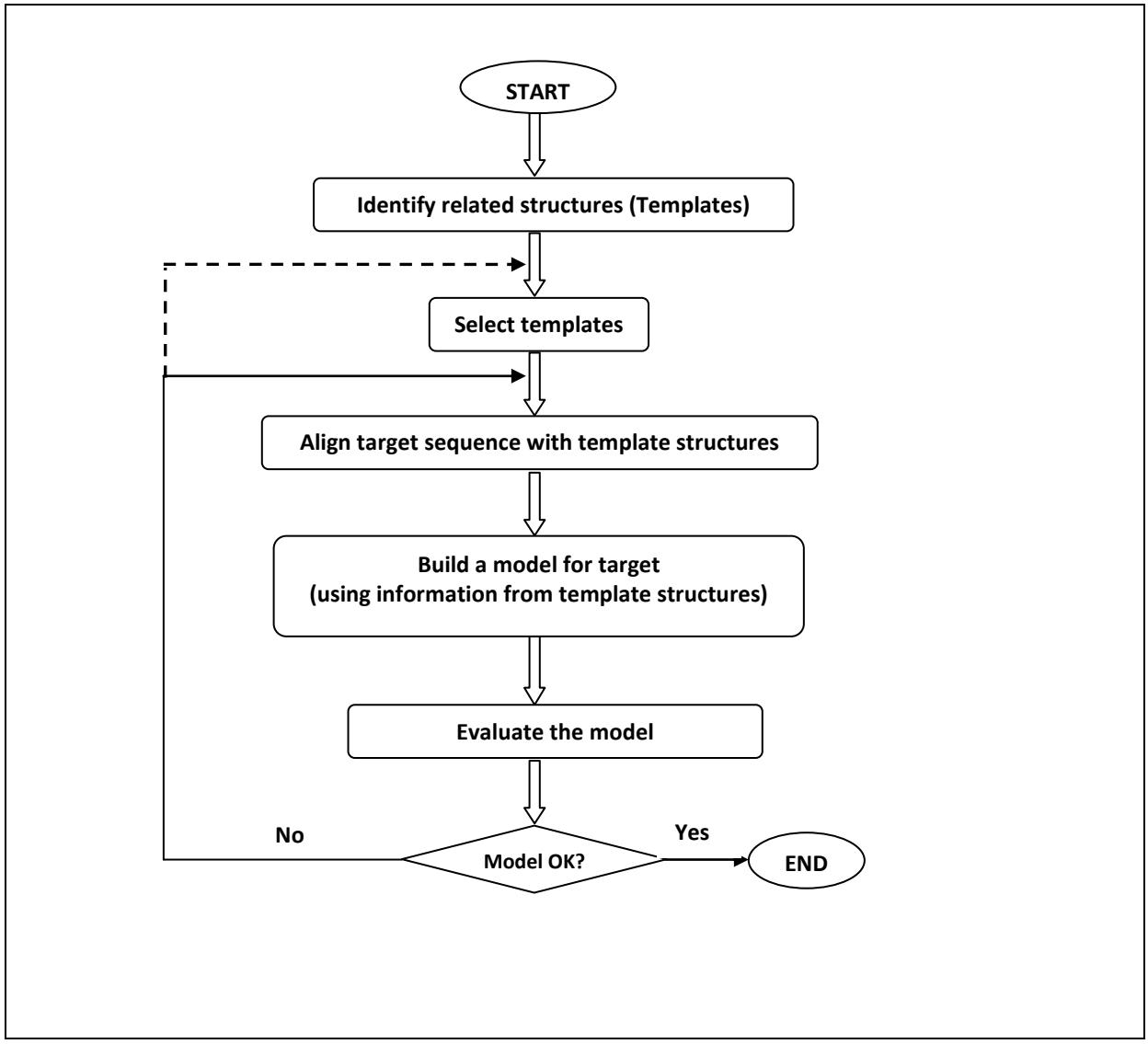


Figure 1: The general schema adopted for homology model

RESULTS AND DISCUSSIONS

In order to study various physico-chemical properties of antioxidant proteins in legumes in terms of molecular weight, isoelectric point (pI), number of positively and negatively charged residues (+R and –R), extinction coefficient (EC), instability index (II), aliphatic index (AI) and grand average of hydropathy (GRAVY) EXPASY'S ProtParam server (Gasteiger *et al*., 2005) (*http://us.expasy.org/tools/protparam.html*) was used and results are shown in Table 1. The computed isoelctric point (pI) for all was less than 7, indicating acidic nature of protein. This may be useful for developing buffer system for purification by isoelectric focusing. Antioxidants C4P7F2, P11964 and Q02610 were found to be unstable in view of high instability index (II). Aliphatic Index (AI) of all antioxidant protein sequences except C4P7F2, P11964 and Q02610 indicates that these antioxidant proteins may be stable for a wide temperature range. Low GRAVY indices of antioxidants under study indicate the possibility of better interaction with water.

Table 1: Physico-chemical properties of antioxidants present in legumes

Source	Accession	Length	Mol Wt	-R	+R	pI	EC	II	AI	GRAVY
Cicer arietinum	Q9ZNQ4	152	15221.7	15	8	5.44	125	13.53	80.79	-0.206
Phaseolus vulgaris	Q9FE12	260	28622.5	29	24	5.18	23045	41.54	87.38	-0.094
Vigna radiate	C4P7F3	135	13745.8	15	7	5.42	0	6.70	66.44	-0.514
Vigna radiate	D5FUD7	261	28593.5	29	25	5.50	21555	40.96	88.54	-0.105
Vigna radiate	F2WVP3	152	15255.7	15	8	5.59	1615	12.49	74.534	-0.278
Vigna luteola	C4P7F2	133	13496.6	15	7	5.42	0	69.62	6.65	-0.479
Pisum sativum	P11964	202	20626.2	17	12	5.94	1615	94.06	20.19	0.031
Pisum sativum	Q02610	152	15322.8	15	8	5.59	125	78.88	11.36	-0.247

Table 2: Disulphide (SS) bond pattern of pairs predicted, by CYS_REC

Antioxidant	Cys-rec	Motif found	Profile
Q9ZNQ4	56-145	43 - 53 Sod_Cu_Zn_1 137-148 Sod_Cu_Zn_2	-
Q9FE12	7,113,235	-	67-226 Thioredoxin_2
C4P7F3	49	36-46 Sod_Cu_Zn_1	-
D5FUD7	114-236	-	68-227 Thioredoxin_2
F2WVP3	56-145	43-53 Sod_Cu_Zn_1 137-148 Sod_Cu_Zn_2	-
C4P7F2	46	33-43 Sod_Cu_Zn_1	-
P11964	105-194	-	92-102 Sod_Cu_Zn_1 186-197 Sod_Cu_Zn_2
Q02610	56-145	43-53 Sod_Cu_Zn_1 137-148 Sod_Cu_Zn_2	-

Table 3 : Secondary structural features of the antioxidant protein sequences under study using SOPMA (Geourjon and Deleage, 1995)

	Q9ZNQ4	Q9FE12	C4P7F3	C4P7F4	D5FUD7	F2WVP3	C4P7F2	P11964	Q02610
Alpha helix	5.26	24.62	5.93	7.69	25.29	4.61	4.51	9.41	5.92
310 helix	0	0	0	0	0	0	0	0	0
Pi helix	0	0	0	0	0	0	0	0	0
Beta bridge	0	0	0	0	0	0	0	0	0
Extended strand	36.84	16.54	27.41	30.77	19.54	34.21	30.08	35.15	33.55
Beta turn	8.55	6.54	10.37	6.223	6.13	8.55	9.02	9.41	8.55
Bend region	0	0	0	0	0	0	0	0	0
Random coil	49.34	52.31	56.30	52.31	49.04	52.63	56.39	46.04	51.97
Ambigous states	0	0	0	0	0	0	0	0	0
Other states	0	0	0	0	0	0	0	0	0

Disulphide linkages, motifs and profiles were predicted and functional analysis of proteins was performed by SOSUI server. CYS_REC (http://sunI.softberry.com/berry.phtml?topic) predicted SS-bonding states of cysteines and located disulphide bridges (Table 2). From the domain and motif identification, the antioxidants were found to belong to Cu-Zu Superoxide-Dismutase super family. Graphical view of conserved domain of *Cicer arietinum* (Q9ZNQ4) may also be observed in Figure 2.

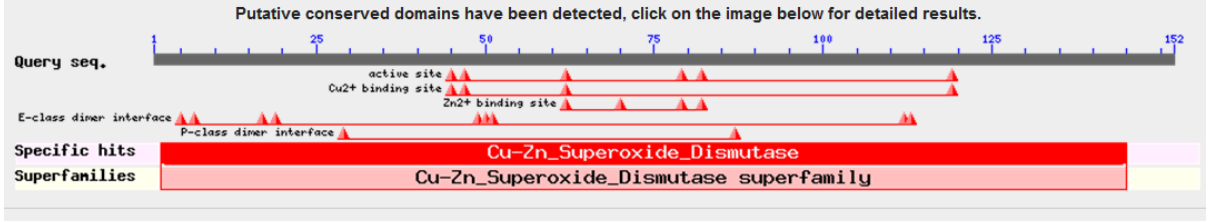


Figure 2: Graphical view of conserved domain of *Cicer arietinum* (Q9ZNQ4) under study

The antioxidant proteins were further undergone BLAST with the non-redundant database to find the template protein sequence. Since, some of them were found to be unstable ones, the final four proteins were taken up for final structure prediction and model development. The selected proteins were undergone BLASTP (*http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins*) for the identification of template proteins of their respective target antioxidants. BLASTP searches protein database using a protein query. The identified target proteins Q9ZNQ4, Q9FE12, C4P7F3 and P11964 reported 84.2% identity with PDB 2Q2LA, 62.4% identity with PDB 1QMVA, 65.0% identity with PDB 1SRDA and 91.0% identity with PDB 3KM2A respectively. The following Figure shows the pairwise alignment of the target and template proteins.

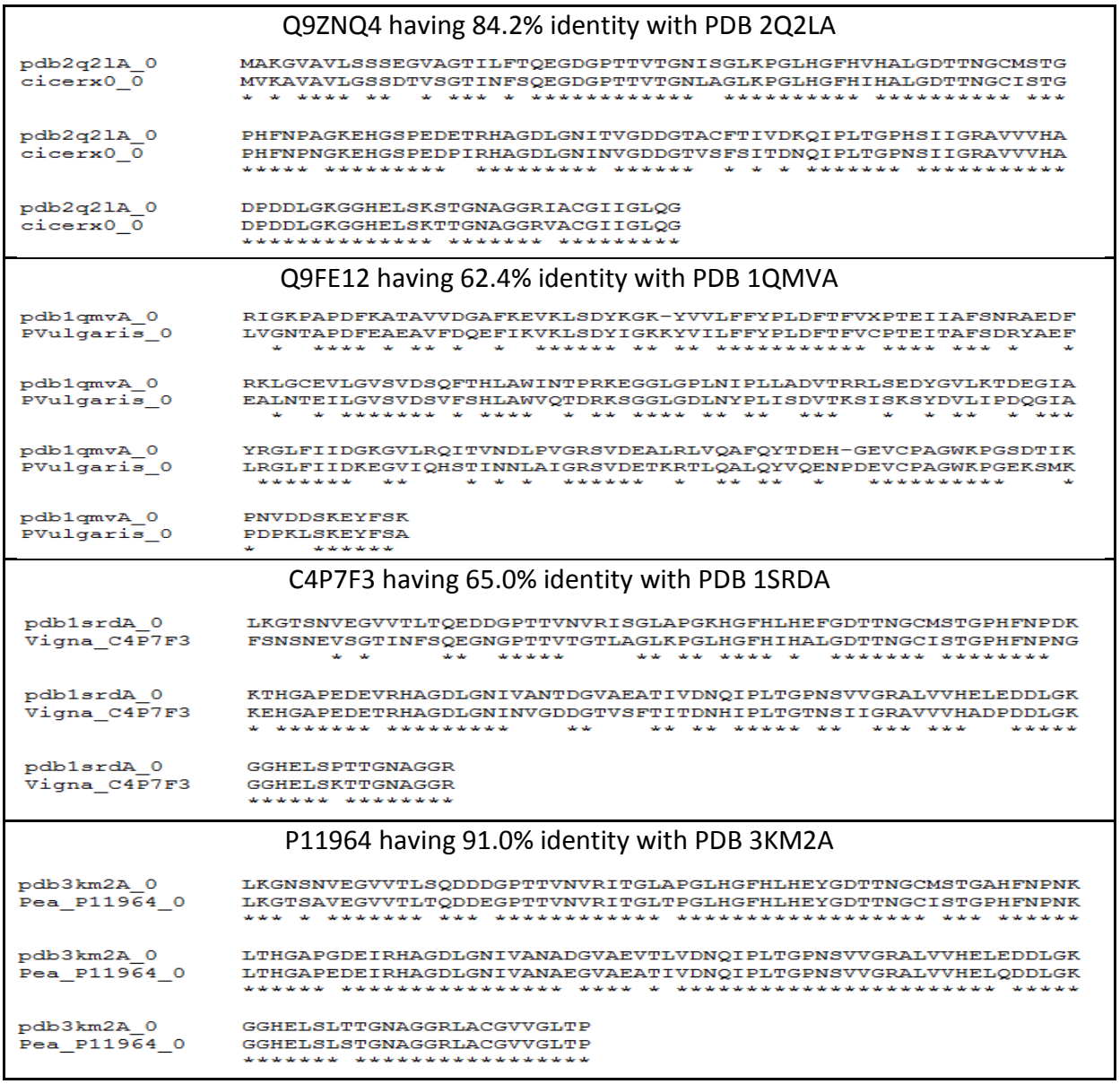


Figure 3: Sequence alignment of target proteins with PDB templates

Further, table 3 shows predicted secondary structural elements for proteins undertaken for study using SOPMA method. Random coils were found to dominate secondary structure elements followed by extended strand and alpha helix for all the antioxidants in this study sequences considered.

Modelling of the sequence

The comparative results from Geno3D and SWISS-MODEL servers are presented in Table 4. For Q9ZNQ4 protein, Geno3D and SWISS-MODEL showed 78.9 and 96.0% of the total residues fall in the most favoured region respectively. Similarly, 0.9 and 0.0% of the residue fall in the disallowed region respectively. The result from SWISS-MODEL was found to be the best for the

residues as the percentage contribution of the residues in generously allowed and disallowed regions were 0.0% for Q9ZNQ4 protein. Similar trend for most favoured region was also observed in all other antioxidants under study.

Table 4: Comparative analysis of the models from GENO3D and SWISS-MODEL

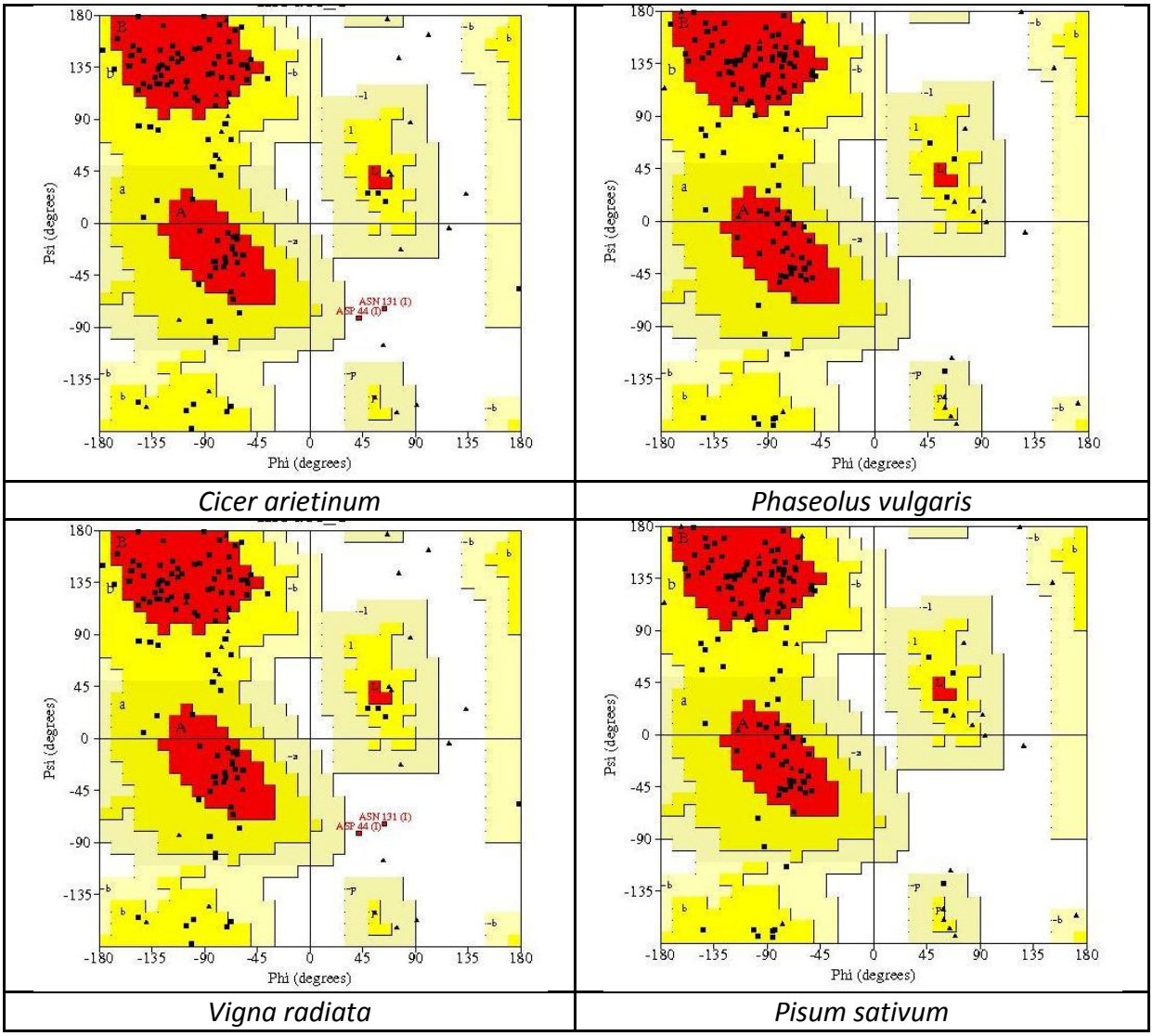
	Antioxidant	**Q9ZNQ4**	**Q9FE12**	**C4P7F3**	**P11964**
Geno3D	Most favored	78.9%	81.0%	70.6%	75.2%
	Additionally allowed	19.3%	17.3%	25.5%	23.9%
	Generously allowed	0.9%	1.2%	2.0%	0.9%
	Disallowed	0.9%	0.6%	2.0%	0.0%
SWISS-MODEL	Most favored	96.0%	94.7%	84.6%	95.7%
	Additionally allowed	4.0%	4.9%	14.0%	4.3%
	Generously allowed	0.0%	0.0%	0.0%	0.0%
	Disallowed	0.0%	0.5%	1.4%	0.0%

Validation of the model

Evaluation of model quality is the next crucial step in homology modelling. For this, final models were inspected using validation tools in order to confirm the consistency of model's stereochemistry with typical values found in crystal structures. SAVS was employed for evaluation of model quality. The overall stereochemical property of the proteins as well as the quality of the modelled structures was assessed by Ramchandran plot analysis in PROCHECK validation package. The Ramachandran plot (Figure 4) shows the phi-psi torsion angles for all residues in the structure (except those at the chain termination).

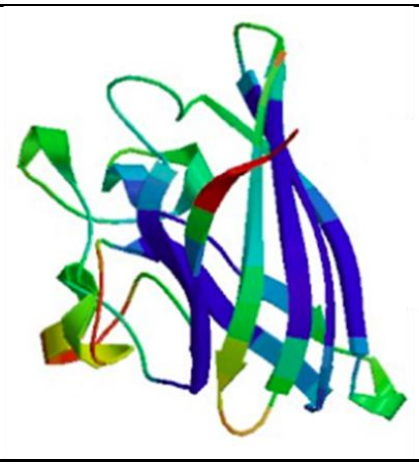
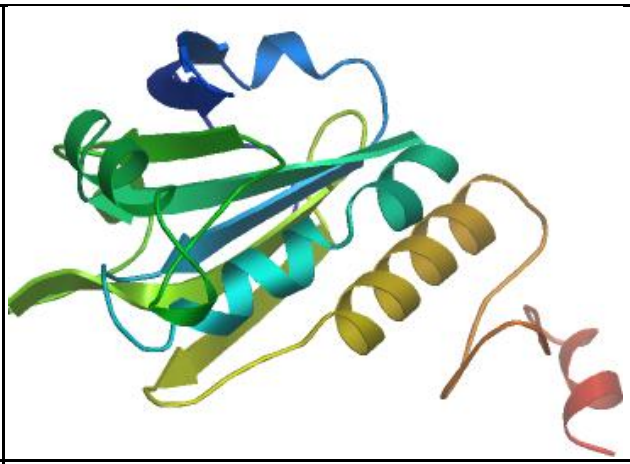
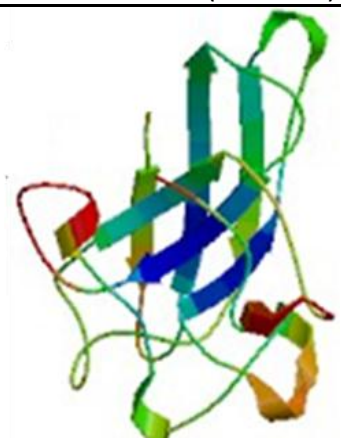
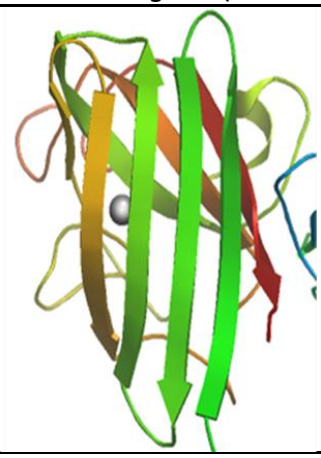
The darkest areas correspond to the "core" regions representing the most favorable combinations of phi-psi values. Ideally, over 90% of the residues are desired in these "core" regions. The percentage of residues in the "core" regions is one of the best guides to stereo-chemical quality. The final validation for structure models obtained from the software tools was performed by using PROCHECK and WHAT IF. The average Z score for Q9ZNQ4, Q9FE12, C4P7F3 and P11964 antioxidants are 1.27, -0.16, -0.25 and 0.30 respectively as computed in PROVE of SAVS.

A comparison of the results obtained showed that the model generated by SWISS-MODEL was more acceptable in comparison to Geno3D. The final modelled structures (Figure 5) for all the four antioxidant protein under study were visualized by Rasmol (www.RasMol.org) software.



Cicer arietinum	*Phaseolus vulgaris*
Vigna radiata	*Pisum sativum*

Figure 4: Ramachandran Plot determining protein conformations

	
Cicer arietinum (Q9ZNQ4)	*Phaseolus vulgaris* (Q9FE12)
	
Vigna radiata (C4P7F3)	*Pisum sativum* (P11964)
Figure 5: Modelled Structures of antioxidant proteins	

References

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace:a web-based environment for protein structure homology modelling. Bioinformatics 22: 195-201.

Ayyagari R, Narasinga BS, Roy DN (1989) Lectins, trypsin inhibitors, BOAA and tannins in legumes and cereals and the effect of processing. Food Chemistry 34: 229–238

Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28: 45-48.

Barbehenn RV (2002) Gut-based antioxidant enzymes in a polyphagous and graminivorous grosshopper. J Chem Ecol 28: 1329-1347.

Combet C, Jambon M, Deleage G, Geourjon C (2002) Geno3D: Automatic comparative molecular modelling of protein. Bioinformatics 18: 213-214.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D. and Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker ed, The Proteomics Protocols Handbook, Humana Press, 571-607

Geourjon, C. and Deleage, G. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comp Appls Biosci. 11(6):681-684.
Guex N, Manuel CP (1997) Data Modeling, Analysis and Classification SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis 18: 2714-2723.
Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8: 477-486.
Ramachandran GN, Ramakrishnan C, Sasisekhran V (1963) Stereochemistry of polypeptide chain confi guarations. J Mol Biol 7: 95-99.
Sali A, Blundelll TL (1993) Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 234: 779-815.
Shigeoka S, Ishikawa T, Tamoi M, Miyagawa Y, Takeda T, et al. (2002) Regulation and function of ascorbate peroxidase isoenzymes. J Exp Bot 53: 1305-1319.
Vriend G (1990) WHAT IF: A molecular modeling and drug design program. J Mol Graph 8: 52-56.
Yadav K Navneet, Sarika, Iquebal MA and Akram, M. (2011) *In-silico* analysis and homology modelling of coat-protein of Mungbean Yellow Mosaic India Virus. J Food Legumes 24: 138-141.