Comparative Modeling and Prediction of Carbohydrate Binding Pockets in 3-D Structure of Wild Pulse *Lablab Purpureus* Arcelin

Arumugam.N.
School of Life Sciences
Dept of Zoology
University of Madras
Chennai, India.

Dr. Janarthanan.S.
School of Life Sciences
Dept of Zoology
University of Madras
Chennai, India.

Sakthivelkumar.S. & Veeramani. V.
School of Life Sciences
Dept of Zoology
University of Madras
Chennai, India.

ABSTRACT
Arcelin, a seed protein originally discovered in wild bean accession of *Lablab purpureus* was purified, characterized, and compared to phaseolin, the major seed protein of common bean *Phaseolus vulgaris*. There are several reports available for common bean arcelin from *P. vulgaris* and its defense mechanism against the stored product insect pests, but *L. purpureus* arcelin function is not yet studied well. To understand the molecular function of arcelin in *L. purpureus* the structural knowledge is essential. This work is an attempt to explore the molecular defense mechanism of *L. purpureus* arcelin based on homology modelling and binding pocket analysis to emphasize the structural and functional relationship. The structural template from *P. vulgaris* arcelin [1AVB] is selected for homology modelling of *L. purpureus* arcelin. The 3D structure of *L. purpureus* arcelin was generated using Modeller software. The best model is selected based on Ramachandran Plot, Errat and Energy minimization analysis (Steepest Descent). The overall quality of computed model showed 87.2 % amino acid residues under favored region with 93.5 % overall quality. The putative refined model of *L. purpureus* arcelin was deposited into Protein Model Data Base with ID: PM0076542. Deposited model is used for further active cavity analysis against carbohydrate (sugar) binding sites. These results will help further development of transgenic crops with arcelin for future integrated insect pest management (IPM) programme.

Keywords
Homology Modelling, Binding pockets, *Lablab purpureus* arcelin, Protein Model Data Base, Modeller

1. INTRODUCTION
Arcelin is an insecticidal protein found in some wild species of pulses. It has been shown to confer resistance against infestation of stored product insect pests. Among the plant derived insecticidal proteins, the insecticidal and anti-metabolic properties of arcelin and its variants toward bruchid pests and in particular their inhibitory effect on the larval development of bruchids is considered one of the most important studies in developing an insect resistant bean plant [1]. There are seven different allelic variants (designated Arc 1 – Arc 7) of arcelin proteins which have been described so far with sub-unit molecular weights in the range of 27 – 42 kDa. Every arcelin variant is composed of several polypeptides presumably encoded by a family of different genes [2]. The amino acid sequence comparison showed that arcelins belong to the bean lectin-like family, which includes the two types of phytohemagglutin subunits (PHA-L and PHA-E) and α-amylase inhibitors. Although the members of this protein family display similar structures, they differ in their biochemical properties, glycosylation patterns, secondary structure and sugar binding specificities [3]. It has been reported that even the semi purified form of *L. purpureus* arcelin extract has the potential to control storage pests in cereals [4]. Hence the analysis of 3D structure of arcelin provides essential information for understanding of protein function against insect defense mechanism. Although there are some difficulties exist to establish a protein structure experimentally by NMR or X-ray crystallography studies [5], homology modelling is a very reliable technique that can consistently predict the 3D structure of protein with precision [6]. This technique depends upon the alignment of protein of known structure (target) with of a homologue of known structure (template). Since the study of 3D structure of protein is helpful in recognizing the details of a protein [7], this method is increasingly becoming of wide spread use in the field of bioinformatics. The aim of this work was to construct the 3D model of *Lablab purpureus* arcelin. This study could prove useful in further functional characterization of this important group of protein and its derivatives. The sequence alignment and template structures were then used to produce a structural model of the target. Since protein structures are more conserved than DNA sequences; detectable levels of sequence similarity usually imply significant structural similarity [8]. In the present study, effort was made to generate the 3D structure of the *L. purpureus* arcelin based on the available template structural homologues from Protein Data Bank (PDB) and the model validated with standard parameters. Finally the refined models were deposited in Protein Model Data Base [PMDB]. The best refined model is selected for Active Cavity prediction studies. This is submitted to online Active Cavity prediction site: http://www.modelling.leeds.ac.uk/pocketfinder/help.html to enumerate the number of Active Cavity in the *L. purpureus* arcelin.

2. MATERIALS AND METHODS

2.1 Amino acid sequence comparison

The protein sequence of *Lablab purpureus* arcelin (target) and *Phaseolus vulgaris* arcelin were retrieved from GenBank with accession Number: ABJ16470.1 and GI: 3891966. The proteins with significant amino acid similarity for target were collected by BLAST [9] search of the non-redundant protein sequence database at the NCBI site (http://www.ncbi.nlm.nih.gov/ BLAST). A pair wise sequence alignment of these sequences was produced using http://www.ebi.ac.uk/Tools/emboss/align/ web server.

2.2 Retrieval of the target sequence

The amino acid sequence of isoforms of *L. purpureus* arcelin (Gen Bank DQ985699.1 accession number) and other sequences examined in this study were obtained from the database http://www.ncbi.nlm.nih.gov. The 3D structure of *L. purpureus* arcelin is absent in Protein Data bank, hence the current work was initiated (http://www.rcsb.org/pdb/home/home.do).

2.3 Template selection and Target alignment

The first task in homology modelling technique is recognition of the protein structures linked to the target sequence and subsequently selection of templates [10]. PSI-BLAST was carried out against database specification of PDB proteins available at the National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/blast) web server and an appropriate template was selected on the basis of the quality of the experimental template structure, environmental likeness and phylogenetic similarity. Multiple sequence alignments were performed using Clustal W 1.83 [11] and the alignments were crucially assessed in terms of number, length and using Hierarchical Neural Network (HNN) bin/npsa_automat.pl?page=npsa_nn.html).

2.4 Construction of the model

First, rough 3D models of the *L. purpureus* arcelin protein were constructed by use of the MODELLER 9v7 program [12] based on its alignment with the template protein and satisfaction of spatial restraints [10]. These restraints obtained on the basis of homology, are generally improved by stereochemical restraints on bond lengths, bond angles, dihedral angles, and non-bonded atom-atom contacts that are attained from a molecular mechanics force field. To overcome potential problems, the constructed model was refined by subjection to constraint energy minimization with a harmonic constraint of 100kJ/mol/Å2. The steepest descent (SD) and conjugate gradient (CG) methods were used to remove existing bad sectors between the protein atoms and regularizing the protein structure geometry. All computations were completed *in vacuo* with GROMOS96 43B1 parameters set using the Swiss-PDB Viewer package. Hydrogen bonds were not considered. (http://expasy.org/spdv/program/spdv37sp5.zip) [13].

2.5 Tertiary structure prediction

The tertiary structure of *L. purpureus* arcelin was predicted from the amino acid sequence by the method of GOR IV, based on information-theoretical ideas that are essential for function prediction, protein classification and understanding the structural changes [14].

2.6 Active Cavity prediction

Identification of active cavity or site of protein is a key step for the identification of functional role of protein or enzymes [15]. Active Cavity prediction was achieved by submitting the model in to the online Active Cavity prediction supercomputing facility serverhttp://www.modelling.leeds.ac.uk/pocketfinder/help.html. The functional Active Cavity in association with carbohydrate (sugar) was identified in Castp server [16]. The weighted Delaunay triangulation and the alpha complex were used for shape measurements, which provided identification and measurements of surface accessible pockets as well as interior inaccessible cavities of proteins.

2.7 Deposition of Refined Model in PMDB

The refined model is deposited in PMDB database, which is a public resource, developed to deposit manually build good quality 3D models based on simple stereo chemical check performance PROCHECK [17], ERRAT [18],VERIFY3D [19] and Ramachandran plot analysis [20]. Users can navigate the manually built models for their future analysis. The deposited model is assigned a unique ID. The assigned ID for *L. purpureus* arcelin is PM 0076542. Interested users can directly access the Model for further analysis with template sequence in FASTA format [21].

3. RESULTS AND DISCUSSION

3.1 Amino acid sequence comparison

Two amino acid sequences of *L. purpureus* arcelin (Target) and *Phaseolus vulgaris* arcelin (Template) were compared under Blosum 62 Matrix assessment with 0.0005 Gap extension to achieve accurate comparison between Target and Template (Fig:1) through pairwise sequence alignment. Although pair-wise sequence alignment methods are not useful for identifying very remote relationships, they are capable of producing accurate matching on sequence only for closely related proteins. It can be said that proteins with evolutionary relationship can be identified by pair-wise sequence alignment. Sequence of target protein is aligned to template protein sequence as an input while matching two proteins. Conserved residues in respective alignments are also considered as an input while deciding the match between two proteins. Our output showing the identity of target against template is 60.7% and similarity is 69.9%. This is strongly supporting that the two proteins are closely related family proteins.

```
#=======================================
# Aligned_sequences: 2
# 1: 1AVBB
# 2: ABJ16470.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.0
#
# Length: 229
# Identity:     139/229 (60.7%)
# Similarity:   160/229 (69.9%)
# Gaps:          16/229 ( 7.0%)
# Score: 638.997
#=======================================
1AVBB       1 SNDASFNVETFNKTNLILQGDATVSSEGHLLLTNVKGN-EE--DSMGRAF     47
::::.||.:..||:|||||||:||||.:|||||||..| |.  ||||||.
ABJ16470.1  1 ASETSFEINRFNETNLILQGNATVSFKGHLLLTNVTHNGEPSVDSMGRAL     50

1AVBB      48 YSAPIQINDRTIDNLASFSTNFTFRINAKNIENSAYGLAFALVPVGSRPK     97
||||||:.|.|  ..|||.|||||.|...|...|.||||||||||.|:||
ABJ16470.1 51 YSAPIQMRDST--GNASFDTNFTFNILPANKLMSGYGLAFALVPVDSQPK     98

1AVBB      98 LKGRYLGLFNTTNYDRDAHTVAVVFDTVSNRIEIDVNSIRPIATESCNFG    147
.|||.|||||.:..|.:||||||||||.||||.||||||:||.:...||.
ABJ16470.1 99 KKGRLLGLFNKSENDINAHTVAVVFDTFSNRIGIDVNSIQPIGSVPWNFR    148

1AVBB     148 HNNGEKAEVRITYDSPKNDLRVSLLYPSSEEKCHVSATVPLEKEVEDWVS    197
|..|:.|:|||||||....|.|||.||...::.||||.|.|||.::||||
ABJ16470.1149 HYIGQNADVRITYDSSTKFLAVSLFYPIPRKRYHVSANVELEKVLDDWVS    198

1AVBB     198 VGFSATSGSKKETTETHNVLSWSFSSNFI    226
||||||| : .|  |||:||||
ABJ16470.1199 VGFSATS-A-YE--ETHDVLSW-------    216
#---------------------------------------
#---------------------------------------
```

Fig.1. A pairwise sequence alignment between *L. purperus* arcelin and *Phaseolus vulgaris* arcelin (1AVB). "*" indicates a conserved region (conserved) between two sequences and "-" indicates gaps.

3.2 Tertiary structure prediction

Comparative sequence analysis of *P. vulgaris* arcelin and *L. purpureus* arcelin could lead to identify the evolutionary relationship alone. In order to find out the structural class of these two proteins, proclass web server (http://www.imtech.res.in/ raghava/proclass/) was used. The result of this tool is based on a statistical approach, which revealed that *P. vulgaris* arcelin and *L. purpureus* arcelin belong to same lectin family having beta sheets, helices and loops. Garnier-Osguthorpe-Robson (GOR) approach was adopted for this prediction. The percentage of beta sheets, helices and loops were 59.00, 32.42 and 45.72 in *P. vulgaris* arcelin and 59.00, 32.42 and 45.72 in *L. purpureus* arcelin. From these results, it is confirmed that β sheets were dominated in protein tertiary structures of both Target and Template.

3.3 Prediction of Homology Models

Our PSI-BLAST search found the crystal structure of the *P. vulgaris* protein (PDB ID 1AVB) to be the best template. This template match had the identity of 62% of sequence similarity. The modeled structure of L. *purpureus* arcelin was a homo dimer consisting of 226 amino acids in each chain. Comparable results were obtained for other two proteins as expected based on the conservation of protein contacts [21]. The amino acid Sequences of *L. purpureus* arcelin were submitted to MODELLER 9V7 server and atomic co-ordinates for the proteins were generated. The hypothetical protein models created were stored as PDB output file. The hypothetical proteins were visualized and computed by Swiss-PDB viewer package and Pymol. The best computed 3D structure of the protein was represented by cartoon display in Pymol view (Fig. 2).

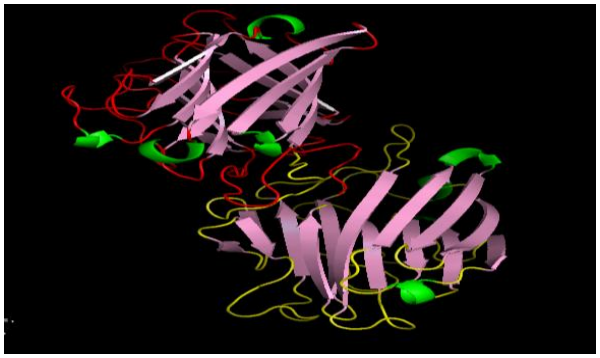


Figure 2. Ribbon structure of *L. purpureus* arcelin Chain – A (yellow), B (Red), Sheets (Pink), Helix (Green)

3.4 Validation of refined model

VERIFY3D [22] was used to validate the refined structure. The 3D structure of the protein was compared to its own amino acid sequence taking into consideration a 3D profile calculated from the atomic coordinates of the structures of correct proteins [23]. The constructed model was evaluated for its backbone conformation using Ramachandran plot. The Auto Deposition Input Tool http://deposit.rcsb.org/validate) (ADIT) was used to inspect favorable and unfavorable properties of the modeled structures. We used SAVES (Structure Analysis and Verification Server) (http://nihserver.mbi.ucla.edu/SAVES/) for the verification of model with PROVE & ERRAT. Finally the refined best models were deposited in Protein Model Data Base (http://mi.caspur.it/PMDB/) with PMDB ID PM0076542 (Fig.3). Active Cavity generated was analyzed online submitting to Active Cavity prediction server–maintained by supercomputing facility of University of Leeds (http://www.modelling.leeds.ac.uk/pocketfinder/help.html) with visualization Tool Jmol.

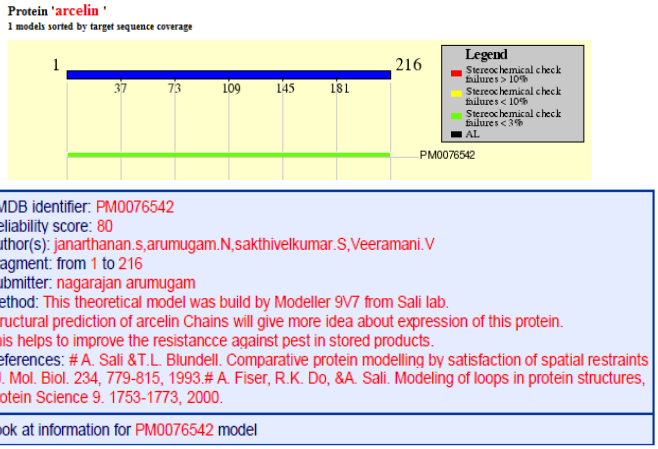


Fig. 3. Deposited Model of *L. purpureus* in PMDB (Blue–Template Bar, Green–Target Bar) with author details

3.5 Validation of *Lablab purpureus* arcelin Model

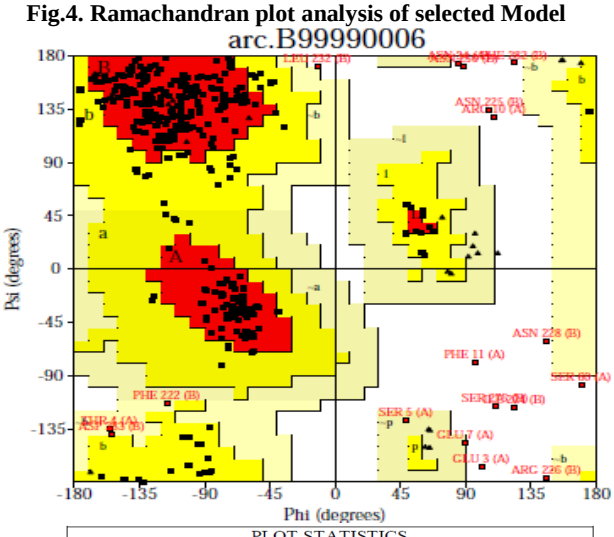
The hypothetical protein models were generated by using MODELLER 9V7. Among the ten models, one model was selected based on the Template energy value (1AVB). Energy value of Template calculated in Swiss-PDB Viewer showed -21701.338 which would act as a reference for the Minimization of energy in the computed models of arcelin at various cycles (Table.1). Based on the reference value Model Number ARC B9990006 at 8th cycle (Minimized Energy is getting coincide the reference energy) was selected for further analysis by VERIFY3D to estimate correctness VERIFY 3D revealed that 93.53% of the residues had an average 3D-1D score >0.2.

Table 1. Energy Minimization Statistics of Computed Models of arcelin by Steepest Descent Algorithms

S.N	Computed Model	Initial energy in Positive	Final energy in negative	% of amino acids in favored region of Ram Plot
1	ARC 990001	27701.338	21800.850	82.82
2	ARC 990002	22147.303	21963.967	81.8
3	ARC 990003	48184.992	21647.996	81.0
4	ARC 990004	11097.016	21602.166	82.1
5	ARC 990005	22826.686	21606.684	82.5
6	ARC 990006	49037.805	21746.818	87.2
7	ARC 990007	26328.736	21691.064	82.3
8	ARC 990008	16148.406	21815.625	82.7
9	ARC 99009	10716.844	21765.367	81.6
10	ARC 99010	26023.900	21715.148	81.4

Ramachandran plot illustrated the backbone confirmation for the modeled protein. On the basis of the analysis 118 structures having resolution of at least 20 angstroms and R factor not greater than 20%, a good quality model is expected to have more than 85% amino acid in the most favored regions of the Ramachandran plot [24] (Fig.4).

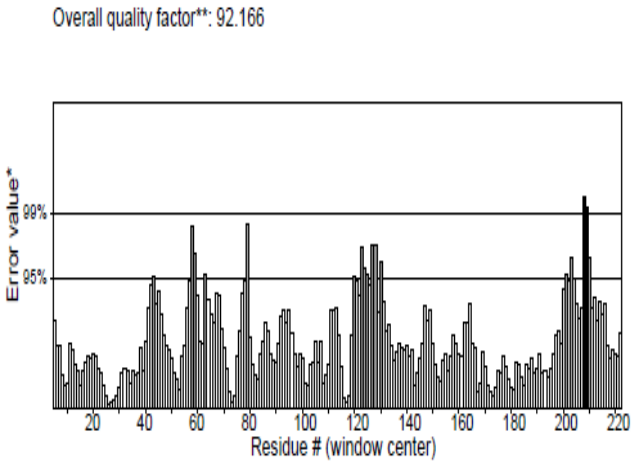
Fig.4. Ramachandran plot analysis of selected Model



PLOT STATISTICS		
Residues in most favoured regions [A,B,L]	310	87.2%
Residues in additional allowed regions [a,b,l,p]	59	10.4%
Residues in generously allowed regions [~a,~b,~l,~p]	4	1.0%
Residues in disallowed regions	6	1.4%
Number of non-glycine and non-proline residues	412	100%
Number of end-residues (excl. Gly and Pro)	29	
Number of glycine residues (in triangles)	26	
Number of proline residues	18	
Total No of Residues	452	

The modeled structures of the *L. purpureus* arcelin strongly satisfied this fact. Out of 432 residues, 87.2% (335) were in most favored regions. The allotment of phi-psi distribution was reliable with right handed alpha helices. ERRAT analysis revealed that overall quality factor of *L. purpureus* arcelin was 92.166. PROVE, VERIFY 3D & ERRAT results for computed model illustrated that the overall quality of the model was good (Fig.5). These results imply that the stereo chemical properties and quality of modeled structure is quiet suitable and consistent and deposited in PMDB data base [ID 0076542] with PMDB reliability score 80.

Fig.5. ERRAT Plot Analysis of Computed *L. purpureus* arcelin



3.6 Active cavity predictions

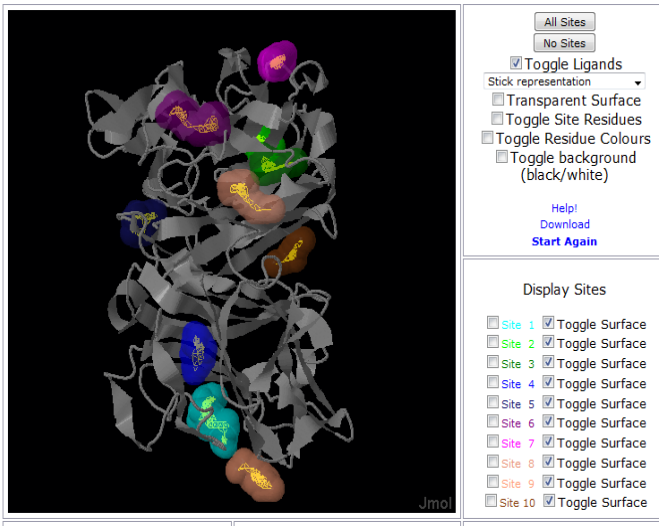
Based on putative protein model of arcelin generated, Active Cavities were predicted online by submitting to Active Cavity prediction server (Q-Site Finder). The server output showed that our predicted model had 10 Active cavities at a distance of 10Å of each amino acid. Amino acids involved in the formation of different active cavities in *L. purperus* arcelin protein are listed (Table 2).

Table: 2 Amino acids involved in an identified active cavity (AC) of *L. purpureus* arcelin

S.No	AC 1		AC 2		AC 3		AC 4		AC 5		AC 6		AC 7		AC 8		AC 9		AC 10	
1	E 3	A	Y 300	B	Y 84	A	F 72	A	I 55	A	R 264	B	N 289	B	I 233	B	I 233	B	M 57	A
2	T 4	A	G 301	B	G 85	A	I 74	A	Q 56	A	V 307	B	I 290	B	Q 235	B	Q 235	B	R 58	A
3	S 5	A	R 318	B	R 102	A	P 76	A	M 57	A	V 311	B	P 292	B	R 264	B	R 264	B	G 62	A
4	F6	A	L 320	B	L 103	A	L 82	A	A 64	A	K 314	B	K 265	B	L 266	B	L 266	B	N 63	A
5	I 271	B	V 338	B	L 104	A	G 85	A	S 65	A	L 322	B	S 298	B	V 307	B	V 307	B	A 64	A
6	Q272	B	N 340	B	F 107	A	L 86	A	F 66	A	F 323	B	T 341	B	P 308	B	P 308	B	Y 161	A
7	M273	B	F 342	B	V 122	A	F 123	A	F 88	A	N 324	B	N 344	B	N 210	B	D 310	B	S 163	A
8	R274	B	S 343	B	N 124	A	D 124	A	L 90	A	K 325	B	F 363	B	S 311	B	S 311	B	E 190	A
9	N279	B	R 345	B	R 129	A	T 125	A	Y 161	A	E 327	B	R 364	B	S 414	B	S 414	B		
10	A280	B	I 346	B	I 130	A	F 147	A	V 199	A	D 329	B	Y 366	B						
11	S281	B	G 347	B	G 131	A	R 148	A	E 219	B	A 332	B	I 367	B						
12	F282	B	I 348	B	N 133	A	Y 150	A	T 220	B	T 334	B	G 368	B						
13	N283	B	D 349	B	I 137	A	I 151	A	S 221	B	N 351	B	Q 369	B						
14	T284	B	I 353	B	P 139	A	G 152	A	F 222	B	S 352	B	Y 390	B						
15	F304	B	P 355	B	S 142	A	Q 153	A												
16	L306	B	S 358	B			Y 174	A												
17	Y377	B					H 211	A												
18	S379	B																		
19	L405	B																		
20	Q406	B																		
21	V413	B																		
22	V415	B																		
23	W432	B																		

Basically arcelin is a lectin-like molecule which will get interaction with (carbohydrate) sugar moiety in the predicted Active Cavity of A & B chains. (Fig.6). These features are expected to play a major role in the toxicity of arcelin protein and its gene regulatory mechanisms against insect pests of stored pulses.

Fig.6. Presence of Active Cavity in Computed L. *purpureus* arcelin in Jmol view



4. CONCLUSION

The 3D structural details of proteins are mainly important for providing insights into their molecular functions. Structural analysis of *Lablab purpureus* arcelin protein was carried out based on the postulation of X - ray crystallographic co-ordinates of X, Y, Z of PDB using Modeller 9v7. The computed model of arcelin has 10 Active Binding Cavity has negatively and positively charged amino acid residues in and around the Cavity surfaces to interact with carbohydrates for defense function against various insect pests in stored grains. Ramachandran plot, Minimized energy value, Errat results revealed the rigidity and quality of the model.

The conserved amino acid residues of target and template also strongly supports that the presence of functional residues in both the structures. This work is a pre-requisite for describing at molecular and structural level studies of arcelin in the seeds of *L. purpureus.* Further conformational and computational analysis of arcelin interaction towards sugar (carbohydrates) in *L. purpureus* arcelin protein will give better knowledge about molecular defense and functional mechanism of arcelin protein in stored grains against insect pests and mammals prior to the development of transgenic crops in the view of Integrated Insect Pest Management (IPM) strategies.

5. ACKNOWLEDGEMENTS

We would like to express deep sense of gratitude to Dr. SURESHKUMAR MUTHUVEL, Assistant Professor, Centre for Bioinformatics, Pondicherry University, Pondicherry, India for his valuable guidance and help. We also extend our thanks to MODELLER developers and various Bioinformatics online server providers and maintenance authorities for successful completion of this preliminary research work.

6. REFERENCES

[1] Osborn, T.C., Alexander, D.C., Sun, S.S.M., Cardona, C. and Bliss, F.A. (1988). Insecticidal activity and lectin homology of arcelin seed protein. Science 240: 207-210.

[2] Acosta-Gallegos, J.A., Quintero, C., Vargas, J. Toro, O., Tohme, J. and Cardona, C. (1998). A new variant of arcelin in wild common bean, *Phaseolus vulgaris* L., from southern Mexico. Gen. Res. Crop. Evol. 45: 235-242.

[3] Mourey, L., Pedelacq, J.D., Fabre, C., Causse, H., Rouge, P. and Samama, J.P. (1997). Small-angle X-ray scattering and crystallographic studies of arcelin-1: An insecticidal lectin-like glycoprotein from *phaseolus vulgaris* L. Proteins: Struct. Funct. Genet. 29: 433-442

[4] Janarthanan, S., Suresh, P., Radk,. G., Morga, D., Oppert, B., (2008). Arcelin from Indian wild pulse, *L. purpureus*, and Insecticidal Activity in Storage Pests. J. Agri. Food Chem. 56: 1676 – 1682.

[5] Othman, R., Wahab, HA., Yosof, R., Rahman, NA (2007). Analysis of secondary structure predictions of dengue virus type 2 NS2B/ NS3 against crystal structure to evaluate the predictive power of the *in silico* methods. In Silico Biol 7:215–224

[6] Martin-Renom, M.A., Stuart, A.C., Fiser, A. Sanchez, R. Melo, F. Sali, A. (2000). Comparative protein structure modelling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291–325.

[7] Paramsivasan, R. Sivaperumal, R. Dhnanjeyan, K.J. Thenmozhi, V. Tyagi, B.K. (2006). Prediction of 3-dimensional structure of salivary odorant-binding protein-2 of the mosquito *Culex quinquefasciatus*, the vector of human lymphatic filariasis. In Silico Biol 7:1–6.

[8] Mourey, L., Pedelacq, J.D., Brick, C., Fabre, C., Rouge, P. and Samama, J.P. (1998). Crystal structures of the arcelin-1 dimer from *Phaseolus vulgaris* at 1.9 A resolution. J. Biol. Chem. 273: 12914-12922.

[9] Altschul SF, Cis W, Millere W, Myers EW and Lipman D (1990). Basic Local Alighnment Search Tool.J.Mol.Biol.215, 403 – 410.

[10] Centeno, N.B,, Planas-Iglesias, J. Oliva, B. (2005). Comparative modelling of protein structure and its impact on microbial cell factories. Microbial Cell Factories 4:20.

[11] Thompson, J.D., Higgins, D.G., Gibson, T. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680.

[12] Sali, A. Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:283–291.

[13] Kaplan, W. Littlejohn, T.G. (2001). Swiss-PDB viewer (Deep View). Brief Bioinform 2:195–197.

[14] Garnier, J. Gibrat, J.F., Robson, B (1996). GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol., 266:540-553.

[15] Zhang. Z, Tang. Y.R., Sheng, Z.Y., and Zhao, D. (2009). An Overview of the *De Novo* Prediction of Enzyme Catalytic Residues. Current Bioinformatics, 2009, 4, 197-206.

[16] Binkowski, T.A., Naghibzadeh, S. Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. Nucl. Acids Res., 31(13): 3352-3355.

[17] Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M., Prochek: A program to check the stereo chemical quality of protein structures (1993). J Appl Cryst, 26: 283-291.

[18] Colovos, C. and Yeates, T.O., Verification of protein structures: patterns of non bonded atomic interactions. Protein Sci. (1993). 2(9):1511- 1519.

[19] Bujnick, J. Rychlewski, L. Fischier, D. (2002). Fold recognition dedects an error in protein Data Bank Vol 18:1391 – 1395.

[20] Ramachandran, G.N., Ramakrishnan, C. Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99.

[21] Castrignanao, T. De Meo, P.D, Cozzetto, D. Talamo I.G., and Tramontano, A .The PMDB Protein Model Data Base. Nucleic acid research, 2006. Vol.34.

[22] Normand, P. Simonet, P. Bardin, R. (1988). Conservation of Cereals sequences in *Phaeseolus* a. Mol Gen Genom 213:238–246.

[23] Eisenberg, D. Luthy, R. Bowie, J.U., (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404.

[24] Rajesh, R. Gunasekaran, K. Muthukumaravel, S Balaraman, K. Jambulingam, P. (2007). *In Silico* analysis of voltage-gated sodium channel in relation to DDT resistance in vector mosquitoes. *In Silico* Biol 7:413–421.