Betaine aldehyde dehydrogenase (BADH) from *Hordeum vulgare* - An *In silico* study

Kaushal K. Bhati, Manvendra Singh, & Vijay Kumar Singh

Plant Biotechnology, Rajiv Gandhi South Campus, Banaras Hindu University, India

Abstract-

Glycine Betaine is a quaternary amino compound accumulates in stress conditions mainly abiotic stresses like drought and salinity & synthesized by the two-step oxidation of choline via the intermediate betaine aldehyde, catalyzed by Choline monooxygenase and Betaine aldehyde dehydrogenase (BADH).

The sequence analysis of key enzyme of Glycine betaine biosynthesis i.e. BADH was carried out using various online based proteomic tools available on Ex-pasy and EBL, and then homologous modeling of this enzyme was performed using automated mode SWISS-MODEL & GENO 3D and models were analyzed on QMEAN, this was an attempt to find the possible model of BADH, because no physical model of BADH is still available on PDB.

Soil salinity & drought are two of the major abiotic stresses reducing agricultural productivity, affects large terrestrial areas of the world; the need to produce salt-tolerant crops is evident. The damaging effects of salt accumulation in agricultural soils and drought have influenced ancient and modern civilizations. It is estimated that 20% of the irrigated land in the world is presently affected by salinity. This is exclusive of the regions classified as arid and desert lands (which comprise 25%of the total land of our planet). The loss of farmable land due to salinization is directly in conflict with the needs of the world population, which is projected to increase by 1.5 billion over the next 20 years, and the challenge of maintaining the world food supplies.

The basic resources for biotechnology are genetic determinants of salt tolerance and yield stability. Implementation of biotechnology strategies to achieve this goal requires that substantial research effort be focused to on identify salt tolerance effectors and the regulatory components that control these during the stress episode (Hasegawa *et al.*, 2000b). Further knowledge obtained about these stress tolerance determinants will be additional resource information for the dissection of the plant response to salinity & drought, the cellular response to these stresses is osmotic adjustment. The cytosolic and organellar machinery of glycophytes and halophytes is equivalently Na+ and Cl- sensitive; so osmotic adjustment is achieved in these compartments by accumulation of compatible osmolytes and osmoprotectants (Bohnert, 1995; Bohnert and Jensen, 1996). Increasing the resistance of crops to these osmotic stresses was one of the first objectives of plant metabolic engineering (LeRudulier *et al*., 1984), and remains a major goal today (Sakamoto and Murata, 2001).

GlyBet occurs in some but not all higher plants, as well as in bacteria and other organisms. In all cases, it is synthesized by a two-step oxidation of choline via betaine aldehyde, but different enzymes are involved. In *Escherichia coli*, a membrane-bound, electron transfer-linked choline dehydrogenase (CDH) oxidizes choline to betaine aldehyde. The aldehyde is then oxidized to GlyBet by a soluble, NAD-linked betaine aldehyde dehydrogenase (BADH) (Andresen *et al.*, 1988). In contrast, *Arthrobacter* spp. bacteria have a soluble choline oxidase

(COX) that carries out both oxidation steps and generates H_2O_2 (Ikuta *et al.*, 1977). In plants, the first oxidation is mediated by a ferredoxin-dependent choline monooxygenase (CMO) (Rathinasabapathi *et al.*, 1997) and the second by BADH (Rathinasabapathi *et al.*, 1994). Both plant enzymes are chloroplastic. All these enzymes have been used to engineer tobacco and other plants that lack GlyBet, generally by placing the responsible genes under the control of the CaMV 35S promoter, and in most cases some increase in stress tolerance has been reported. GlyBet is not appreciably degraded in plants (Rhodes and Hanson, 1993; Nuccio *et al*., 1998) so that GlyBet catabolism has not been an important engineering consideration.

The enzyme BADH has been studied in previous works (given in references) as well it has been cloned successfully to prepare salinity tolerance crops moreover the in-silico study of this enzyme in this work reveled its properties as protein. Moreover we tried to develop a 3'D structure for this important protein using tools available on web.

Material & methods

Sequence Analysis

The sequence of BADH protein (505 amino acids) of *Hordeum vulgare* was downloaded for structural modeling from Swiss-prot (Q40024, EC number=1.2.1.8). The physical and chemical properties of the sequence were calculated by the program ProtParam. The computed parameters for a given protein includes the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half life, instability index, aliphatic index and grand average of hydropathicity. Multiple alignments of the related sequences were performed using the online available ClustalW program (http://www.ebi.ac.uk/Tools/clustalw2/index.html) (Thompson *et al*., 1994)

Structure Prediction

X-ray crystallographic or NMR structure of this protein of *H. vulgare* has to be still determined, therefore homology modeling was performed. The process of building a comparative model is conceptually straight forward. The methodology itself can be described in four steps; identifying a suitable template, making an optimal target-template alignment, building the model and validating the model.

SWISS-MODEL(Jain, 2004; Stoermer, 2006) a comparative modeling server having automated mode available for such modeling purpose. Models are built using comparative modeling by satisfaction of spatial restraints as implemented in Modeller (Sali and Blundell, 1993). Homology Modeling is performed in the following steps: Firstly, it searches for suitable template for the submitted protein by using BLAST program. In the second step, it selects the suitable templates with sequence identity of above 25%. The accuracy of a model depends upon the sequence similarity it shares with the template. Models with >50% sequence identity to templates are normally of high quality, with ~1 Å root mean square (RMS) error for main chain atoms (equal to medium-resolution NMR or low resolution x-ray structures). Models that have 30 to 50% sequence identity are normally of medium accuracy with an RMS of ~1.5 Å (Kasteleijn-Nolst *et al*., 2007; Enyedy *et al*., 2001).Secondary program used for Structure development was Geno3D, Geno3D may be accessed at http://pbil.ibcp.fr/htm/. Sequences were entered into the search box and the program automatically identified suitable templates. This

GENO 3D analysis defines inter chain restrains and thus also energetic stability of models. As a result three models were requested and the model of lowest energy is presented here. All models and analysis files were returned by e-mail.

Structure Analysis

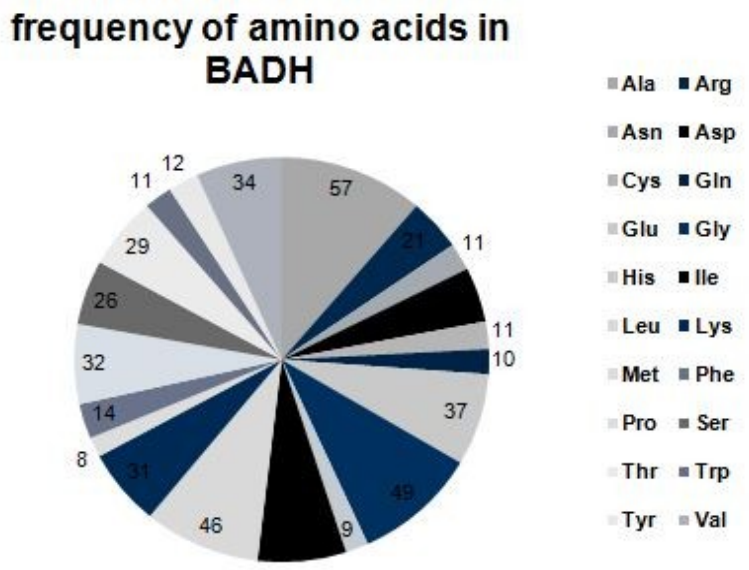
Once the 3D model is created, the important step and last step is evaluating the 3D model of the protein Structure validation on QMEAN server, for model quality. The QMEAN server provides access to two scoring functions for the quality estimation of protein structure models which allow to rank a set of models and to identify potentially unreliable region within these. Either single models (PDB-format) or *tar.gz*-archives with multiple models of the same protein can be uploaded. Additionally, if more than one model is submitted, the full-length sequence of the protein has to be provided. Models were analyzed and viewed either with PyMol(http://pymol.sourceforge.net/index.html)

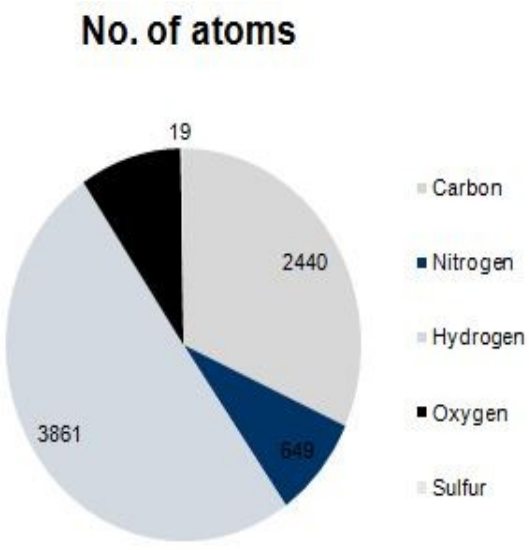
The ProFunc_server was utilized to identify the likely biochemical function of a protein from its three-dimensional structure. It uses both sequence- and structure-based methods (see below) to try to provide clues as the protein's likely or possible function. Often, where one method fails to provide any functional insight another may be more helpful properties of submitted structure.

Results

Sequence analysis-

Primary structure analysis given in pie charts defines the amino acid diversity in BADH protein and the percentage composition of different atoms in BADH molecule.



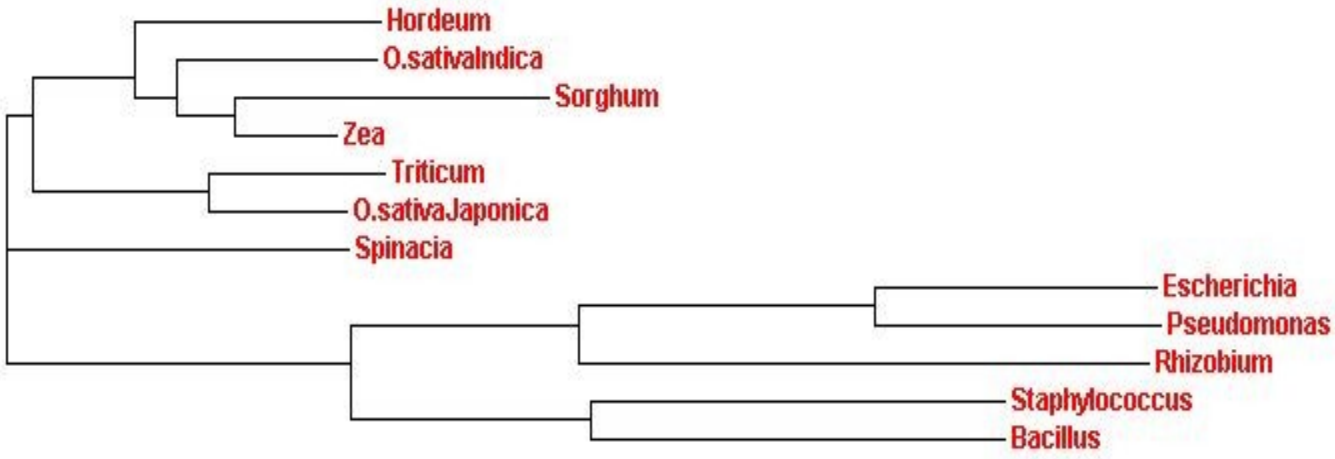


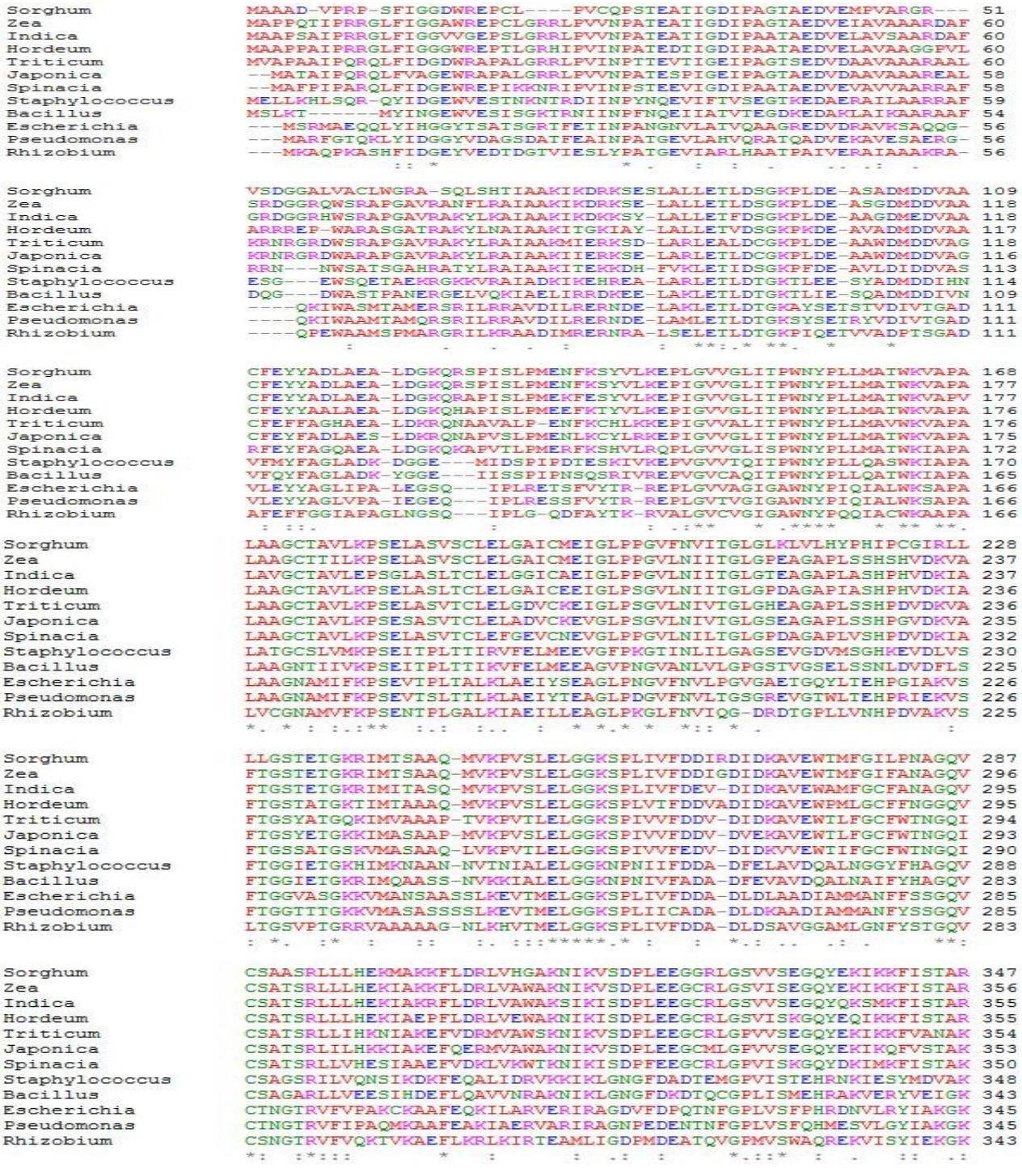
Results of FASTA search against PDB shows maximum similarity with crystal structure of aminoaldehyde dehydrogenase 2 from *P. astivum,* other hits with lower similarity given in table below-

	PDB code	E-value	%-tage id	Overlap	Name
1.	3iwj(A)	2.4e-161	72.200	500	Crystal structure of aminoaldehyde dehydrogenase 2 from pisum sativum (psamadh2)
2.	3iwk(A)	5.8e-155	69.758	496	Crystal structure of aminoaldehyde dehydrogenase 1 from pisum sativum (psamadh1)
3.	3fg0(A)	3.2e-84	43.725	494	1.85 angstrom resolution crystal structure of betaine aldehyde dehydrogenase (betb) from staphylococcus aureus (idp00699) in complex with NAD+
4.	3ed6(A)	3.2e-84	43.725	494	1.7 angstrom resolution crystal structure of betaine aldehyde dehydrogenase (betb) from staphylococcus aureus
5.	1bxs(A)	3.3e-75	41.875	480	Sheep liver class 1 aldehyde dehydrogenase with NAD bound

Sequence alignment and phylogeny of BADH-

The sequence alignment was performed CLASTAL W, and the alignment among amino acid sequences of BADH protein from various plant and microbes shows a large number of amino acid conservation among these, the phylogenetic tree gives position of BADH from *H. vulagare* in cluster with *S. bicolor*, *Zea mays* as shown in picture below-

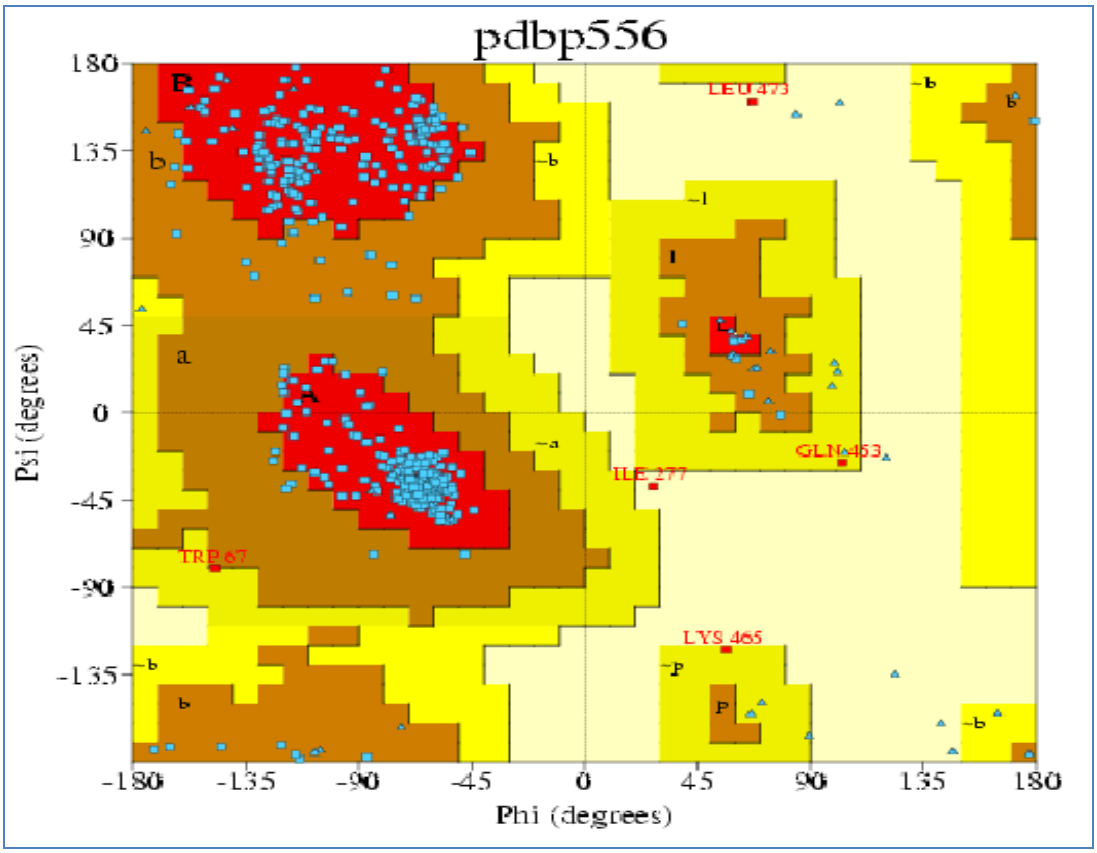




Ramachandran analysis-

The Ramachndran analysis of BADH revealed the presence of 89.7% amino acids with most favored regions and analyzed below statically-

	No. of residues	Percentage
Most favoured regions [A,B,L]	375	89.7%
Additional allowed regions [a,b,l,p]	38	9.1%
Generously allowed regions [~a,~b,~l,~p]	3	0.7%
Disallowed regions [XX]	2	0.5%
Non-glycine and non-proline residues	**418**	100.0%
End-residues (excl. Gly and Pro)	2	
Glycine residues	49	
Proline residues	30	
Total number of residues	**499**	

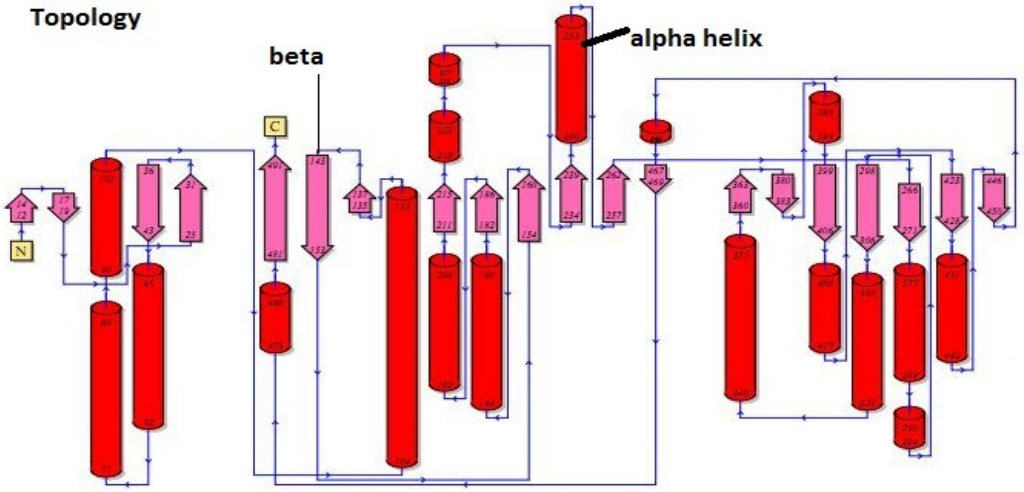


Secondary structure summery

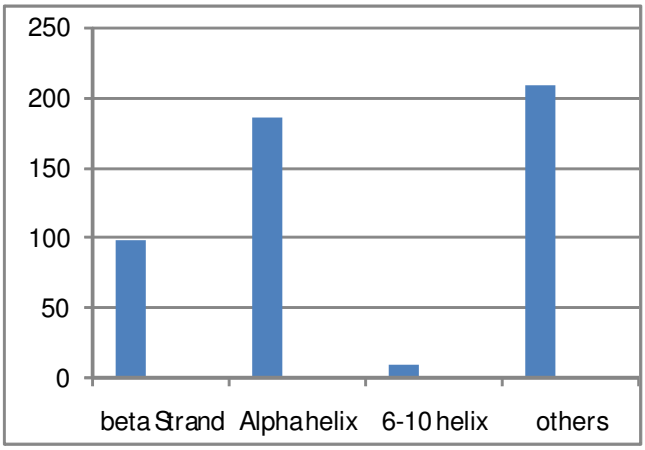
Topological study-

Elaborated in below given chart and diagrammatically shown in topology figure.

Pattern	Frequency	Percentage
beta Strand	98	19.6%
Alpha helix	185	31.1%
6-10 helix	8	1.6%
Others	208	41.7%
Total residues	**499**	



In this topology diagram red colour is showing α Helix while pink colour is indicating the β sheets.



Homologous modeling & Structure assessments

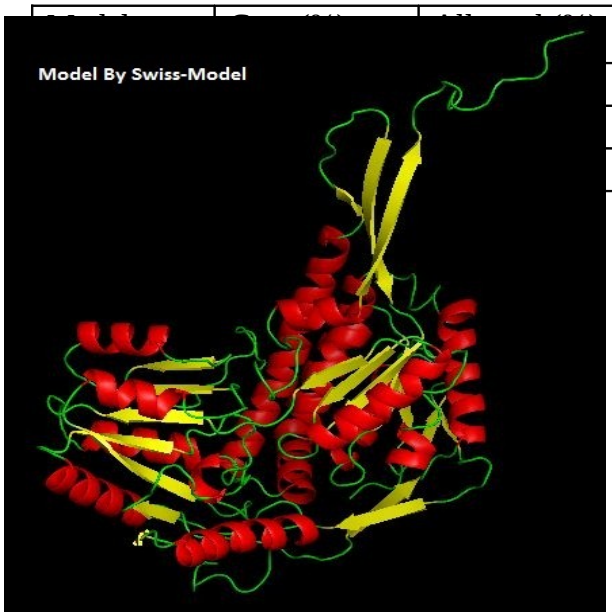
3'D model developed by automated Swiss-modeling using **3IWJB** (crystal structure of Amino aldehyde dehydrogenase 2 from *P. sativum*) as template and automated GENO-3D, using **PDB1CW3** (Human mitochondrial NAD-dependent aldehyde dehydrogenase),
On Geno3D server, the resulted three models were analysed for their energy and Stereochemical quality by PROCHECK, out of these three the Model2 was used for further analysis.

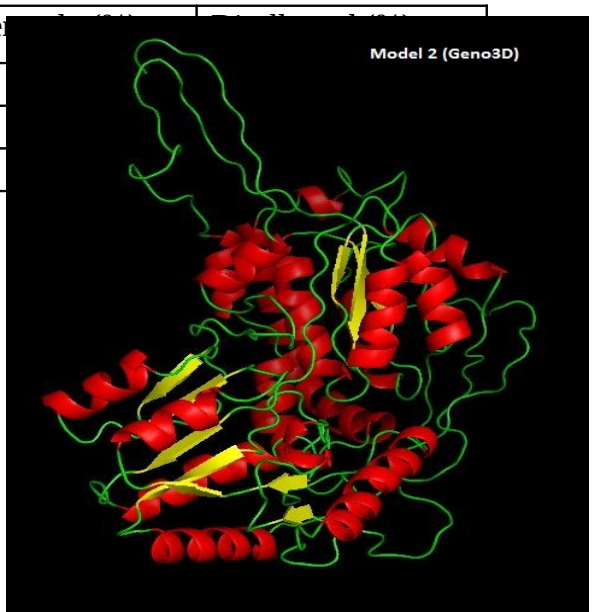
Model energy

Model	Energy (Kcal/mol)
Model 1	-8366.26
Model 2	-20482.40
Model 3	-20276.50

Stereochemical quality of Models

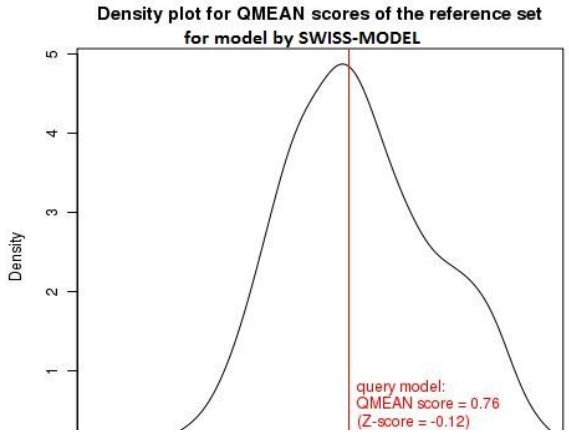
	Gene
	2.7
	2.9
	1.9

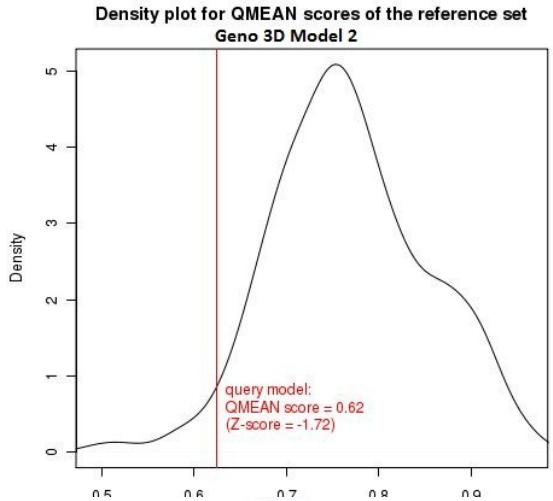




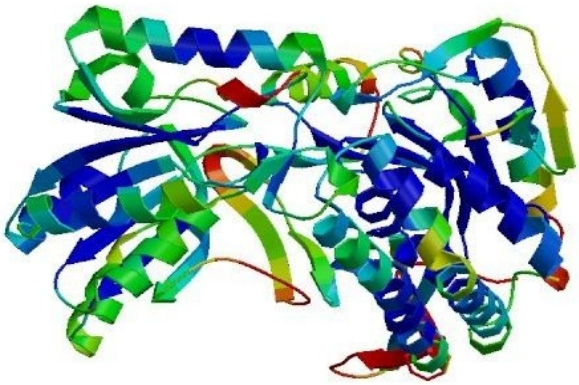
Structure analysis

Structure analysis on QMEAN server of EBI for models obtain from Geno3D and SWISSMODEL revealed structure parameters on different criteria and an comparative assessment is as follows-

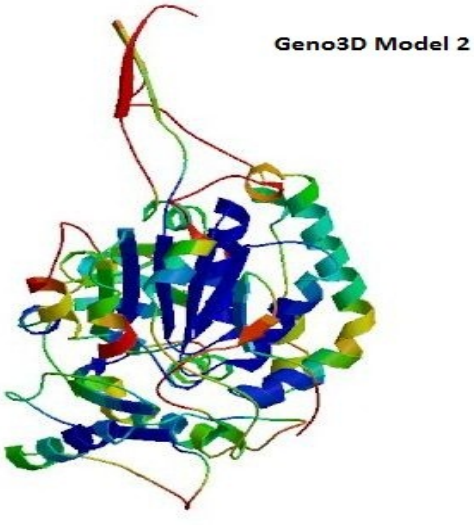




Model by SWISS-MODEL



Geno3D Model 2



Estimated per residue error shown by colour gradient in above structures blue (more reliable) to red(less reliable), combining the QMEAN analysis result is has been concluded that the 3'D structure developed using SWISS-MODEL and **3IWJB** as a template proves to be more reliable and stable as compared to the Model2 given by Geno3D.

ProFunck analysis of functional aspects of BADH Structure

profunk analysis was performed for the BADH model generated using SWISS-MODEL for functional analysis and properties of possible BADH structure.

Active sites

The search for active sites shows the presence of two active sites

Active site	**Name (Amino acids involved)**
PS00687 (260-267)	Aldehyde dehydrogenase glutamic acid active site(LELGGKSP) **Proton Accepter**
PS00070 (289-300)	Aldehyde dehydrogenase cysteine active site(FLNGGQVCSATS)**Nucleophile**

NEST analysis for presence of possible Clefts

Nest analysis on Profunck shows following hits for presence of possible cleft, in below given table hit with red having certain match while those with blue have long shots.

	Score	No. resid.	Ave. access-ibility	Cleft	Ave. conser-vation	Residues
1.	4.000	3	0.00	1	1.000	Val396, Phe397, Gly398
2.	1.000	3	0.00	-	1.000	Glu357, Gly358, Ala359
3.	0.963	3	0.00	-	0.963	Ala179, Gly180, Cys181
4.	0.930	3	0.00	-	0.930	Ile205, Gly206, Leu207
5.	0.914	3	0.00	-	0.914	Ser101, Gly102, Lys103

Ligand binding Templates

Ligand binding templates present on protein BADH are defined in table below given

	Score	Template	PDB	Name
1.	670.625	NADb0124	1bxs	Sheep liver class 1 aldehyde dehydrogenase with NAD bound Het Group NAD
2.	663.750	NADb2053	2onp	Arg475gln mutant of human mitochondrial aldehyde dehydrogenase, complexed with NAD+ Het Group NAD
3.	659.375	NADb1009	1o00	Human mitochondrial aldehyde dehydrogenase complexed with NAD+ and mg2+ showing dual NAD(h) conformations Het Group NAD
4.	648.625	NADb0089	1bpw	Betaine aldehyde dehydrogenase from cod liver Het Group NAD
5.	642.812	NADb1025	1o04	Cys302ser mutant of human mitochondrial aldehyde dehydrogenase complexed with NAD+ and mg2+ Het Group NAD

Discussion-

The role of BADH and its reaction product Glycine betaine is well defined and studies widely. Two basic genetic approaches that are currently being used to improve stress tolerance include: (i) exploitation of natural genetic variations, either through direct selection in stressful environments or through the mapping of quantitative trait loci (QTLs – regions of a genome that are associated with the variation of a quantitative trait of interest) and subsequent marker-assisted selection,(Foolad, M.R. ,2004; Flowers, T.J. 2004) and (ii) generation of transgenic plants to introduce novel genes or to alter expression levels of the existing genes to affect the degree of abiotic stress tolerance(Zhang, J.Z. *et al*. 2004; Chinnusamy, V. *et al*.2005). The genetic engineering of Glycine betaine synthesis has been performed successfully (Sakamoto *et al*,2000). This was an attempt to develop the 3'D structure of BADH and its sequence analysis, structures were developed on SWISS-MODEL and Geno3D and analyzed on QMEAN server which proves the more stability and reliability of structure developed on SWISSMODEL with comparatively high Z & QMEAN score along with that shows more reliable regions on colour gradient suggested by QMEAN. Further the PROTPARAM analysis show presence of two active sites and one possible cleft with four residues Val396, Phe397 & Gly398. The ligand affinity found to most similar to Sheep liver class 1 aldehyde dehydrogenase bound NAD. Sequence analysis revealed that protein BADH have high density of Alpha helix as compared to bets sheets, however unknown hits are most dominants.

References-

Andresen, P.A., Kaasen, I., Styrvold, O.B., Boulnois, G., and Strom, A.R. (1988). Molecular cloning, physical mapping and expression of the *bet* genes governing the osmoregulatory choline-glycine betaine pathway of *Escherichia coli*. *J. Gen. Microb.* **134**, 1737-1746.

Benkert, P., Tosatto, S.C.E. and Schomburg, D. (2008). "QMEAN: A comprehensive scoring function for model quality assessment." Proteins: Structure, Function, and Bioinformatics, 71(1):261-277.

Bohnert, H.J. and Jensen, R.G. (1996) Strategies for engineering water stress tolerance in plants. *Trends in Biotechnology* 14, 89-97.

Bohnert, H.J., Nelson, D.E. and Jensen, R.G. (1995) Adaptations to environmental stresses. *Plant Cell* 7, 1099-1111.

Cavalieiu AJ, Osmoregulatory responses of salt marsh halophytes to the saline environment. PhD thesis(1979). University of South Carolina, Columbia,

Chinnusamy, V. *et al*. (2005) Understanding and improving salt tolerance in plants. Crop Sci. 45, 437–448

Combet C, Jambon M, Deleage G & Geourjon C,(2002) Bioinformatics, 18, 213-214

Flowers, T.J. (2004) Improving crop salt tolerance. J. Exp. Bot. 55,307–319

Foolad, M.R. (2004) Recent advances in genetics of salt tolerance in tomato. Plant Cell Tissue Organ Cult. 76, 101–119

Hasegawa, P.M., Bressan, R.A., Zhu, J.-K. and Bohnert, H.J. (2000b) Plant cellular and molecular responses to high salinity. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. 51, 463-499.

HHpred: Söding, J. (2005) Protein homology detection by HMM-HMM comparison. Bioinformatics 21: 951-960.

Ikuta, S., Imamura, S., Misaki, H., and Horiuti, Y. (1977). Purification and characterization of choline oxidase from *Arthrobacter globiformis. J. Biochem.* **82**, 1741-1749.

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

Laskowski R A, Watson J D, Thornton J M (2005). ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res.*, **33**, W89-W93

LeRudulier, D., Strom, A. R., Dandekar, A. M., Smith, L. T. & Valentine, R. C. (1984) *Science* **224,** 1064–1068.

Nuccio ML, Russell BL, Nolte KD, Rathinasabapathi B, Gage DA, *et al*. (1998). The endogenous choline supply limits glycine betaine synthesis in transgenic tobacco expressing choline monooxygenase. *Plant J.* 16:487–96

Rathinasabapathi, B., Burnet, M., Russell, B.L., Gage, D.A., Liao, P.C., Nye, G.J., Scott, P., Golbeck, J.H., and Hanson, A.D. (1997). Choline monooxygenase, an unusual iron-sulfur enzyme catalyzing the first step of glycine betaine synthesis in plants: prosthetic group characterization and cDNA cloning. *Proc. Natl. Acad. Sci. USA* **94**, 3454-3458.

Rathinasabapathi, B., McCue, K.F., Gage, D.A., and Hanson, A.D. (1994). Metabolic engineering of glycine betaine synthesis: plant betaine aldehyde dehydrogenases lacking typical transit peptides are targeted to tobacco chloroplasts where they confer betaine aldehyde resistance. *Planta* **193**, 155-162.

Rhodes D, Hanson AD (1993) Quaternary ammonium and tertiary sulfonium compounds in higher plants. Annu Rev Plant Physiology Plant Mol Biol **44:** 357–384

Sakamoto, A., and Murata, N. (2001) The use of bacterial choline oxidase, a glycinebetaine-synthesizing enzyme, to create stress-resistant transgenic plants. *Plant Physiol.* **125,** 180-188.

Zhang, J.Z. *et al*. (2004) From laboratory to field. Using information from Arabidopsis to engineer salt, cold and drought tolerance in crops. Plant Physiol. 135, 615–621