Octa Journal of Environmental Research
International Peer-Reviewed Journal
Oct. Jour. Env. Res. Vol. 4(1): 063-072
Available online http://www.sciencebeingjournal.com

Jan – Mar., 2016
ISSN 2321 3655

Research Article

Octa Journal of Environmental Research
(Oct. Jour. Env. Res.) ISSN: 2321-3655
Journal Homepage: http://www.sciencebeingjournal.com

INSILICO APPROACH FOR SINAPYL ALCOHOL DEHYDROGENASE (SAD) PROTEIN ANALYSIS IN *Cedrus deodara*

Goyal, N. and Ginwal, H.S.
Genetics and Tree Propagation Division, Forest Research Institute, Dehradun Uttarakhand (India)
Correspondence Author's Email:**ginwalhs@icfre.org**
Received: 24th Mar 2016 **Revised**: 27th Mar. 2016 **Accepted**: 30th Mar. 2016

Abstract: Deodar (*Cedrus deodara*) is an important timber tree belongs to family *Pinaceae*, due to the high lignin content. It grows in the high altitude with the elevation range of 1200m to 3050m. The sinapyl alcohol dehydrogenase participates in the formation of syringyl lignin through phenylpropanoid pathway of monolignols. The protein product of *SAD* gene is responsible for high lignin content which provides benefit in timber industry. Here the phylogeny analysis, three dimensional protein model followed by the validation to show more emphasis on evolutionary relationship, structure and function of sinapyl alcohol dehydrogenase (*SAD*). The model structure showed the suitability with a member of NADP (H)-dependent dehydrogenase family which catalyzes the formation of monolignols from *Populus tremuloides*. The model revealed the stable confirmation during the study of atomic flexibility. In the future perspective molecular docking can be study for the response in defence mechanism against the pathogens and it will more visualize the self-defence pathway of the deodar tree.
Keywords: Evolutionary Relationship; Lignin Biosynthesis; Phenylpropanoid Pathway; Sinapyl Alcohol Dehydrogenase;Tertiary Structure.
Postal Address: Dr. H.S. Ginwal, Genetics and Tree Propagation Division, Forest Research Institute, Dehradun Uttarakhand (India).

INTRODUCTION

The Pinaceae family was considered as the largest group among the conifers which provided the evolutionary history among them (Lin *et al.,* 2010). The phylogenetic analysis reveals about the convergence or divergence among the genomes within the family (Wang *et al.,* 2000). Deodar (*Cedrus deodara*) is one of the model species of the family and considered as major timber yielding tree and grown in temperate region. *Cedrus deodara* has evaluated for its preliminary phytochemical, pharmacognostical and physio-chemical properties from heartwood (Jain *et al*., 2014). The lignin formation as most abundant component of wood in conifer species deodar which accounting the 25% of plant biomass (Lewis and Yamamoto, 1990) and provide the mechanical support, water and solute transport, biotic and abiotic stress resistance, increases stiffness and reduces the viscoelasitic damping (Whetten *et al.,* 1998; Boerjan *et al.,* 2003; Peter and Neale 2004, Vanholme *et al.,* 2008). The lignin are aromatic heteropolymers which deposited in the secondary cell wall thickness and composed by three hydroxylcinnamyl alcohol monomers named *p*-coumaryl (H), guaiacyl (G) and syringyl (S) as the building blocks of lignin (Barcelo *et al.,* 2007). The monolignols synthesized through the phenylpropanoid pathway by phenylalanine which has been derived *via* shikimate pathway in plastid (Rippert *et al.,* 2009). The syringyl lignin is the derivative of monolignol precursor known as sinapyl alcohol where as S subunit of lignin has been methylated on both 3' and 5'- hydroxyl moieties and indicates the polymeric cross-linking through the more reliable ether bonds at 4-hydroxyl position (Dixon *et al.,* 1996 and Li *et al.,* 2000). The mechanical versatility of wood

makes clear understanding of biomechanics and proof the wood as good material for engineering and construction purpose (Koheler and Telewski, 2006). The distribution of microsatellite in the coding and non-coding regions of chloroplast genomes from various members of pinaceae family which includes *Cathaya argyrophylla*, *Cedrus deodara*, *Larix deciduas*, *Picea morrisonicola*, *P. sitchensis*, *Pseudotsuga sinesis* var. *wilsoniana* (Filiz and Koc, 2014). The evolution rate of alcohol dehydrogenase was fast in angiosperm rather than gymnosperms (Yokoyama and Harry, 1993). The alcohol dehydrogenases performs the enzyme activity which represent in different type of proteins included the long chain with zinc active site as catalytic domain while the small chain have not The long subunit of ADH found as the conserved region of genome among the different species of plants, mammals and yeast (Jornvall *et al.,*1987). On the other hand, SAD was considered as the homolog of CAD and concluded as the key enzyme in sinapyl monolignol biosynthesis in aspen (*Populus tremuloides*) and *Arabidopsis* (Li *et al.,* 2001; Kim *et al.,* 2007). The evolution and functional history of SAD gene family was been understood due to the reliable gene phylogeny. The phylogeny history was retrieved and proved through efficient approach of comparative genomics. Homology modelling was done to predict the tertiary structure of target protein. The structure prediction has been done through the sequence or sequence profile comparison (Skolnick *et al*., 2000). The generated models have high resolution and satisfy the spatial restraints as in template structure. The scarcity of experimental tertiary structure of deodar SAD protein is constraint for the molecular study of phenylpropanoid pathway. The homology modelling is the accurate approach to identify the structural information while no crystal structure is available for SAD in deodar. The phylogeny analysis of SAD in deodar with the other pinaceae members reveals the origin of SAD evolution. The detailed analysis of comparative modelling and molecular phylogeny of SAD encodes for the catalytic unit of alcohol dehydrogenase to explore the structure, function and evolution. The refined and stable tertiary structure of SAD will be useful to insight the protein-protein and protein-ligand interaction to understand the lignin content in wood and its durability at the molecular level.

EXPERIMENTAL

The reviewed amino acid sequence is encoded by the *SAD* (Sinapyl Alcohol Dehydrogenase) gene in *Cedrus deodara* (GenBank Accession: HM185284.1). The sequence was retrieved from the NCBI database. The protein family prediction and architecture of domain was analyzed by InterPro Scan tool (http://www.ebi.ac.uk/tools/pfa/iprscan) (Zdobnov and Apweiler, 2001) and same was verified through various databases *viz*. CDD (http://www.ncbi.nlm.nih.gov/Strucutre/cdd/cdd) (Marchler-Bauer *et al.,* 2011) Pfam (http://pfam.janelia.org) (Fin *et al.,* 2010). Signal Peptide prediction was explored through the web server SignalP (http://www.cbs.dtu.dk/services/SignalP/)(Peter sen *et al.,* 2011). To analyze the homology of SAD protein BLASTP (Atschul *et al.,* 1990) was carried out against the non redundant database of NCBI. On certain criteria (cut –off >= 85% identity) sequences were selected including the query sequence and were aligned by CLUSTALW (Larkin *et al.,* 2007).The phylogenetic analysis was performed through the evolutionary tree with Neighbour-Joining (NJ) method (Saitou and Nei 1987) was included in BIONJ. Protein distances was calculated through poisson correction, using the gap opening penalty of 10, gap extension penalty of 0.1 and distance of gap separation 4, employing Blocks Substitution weight matrix with no hydrophilic penalty (Thompson *et al.,* 1994).The bootstrap value shows that the robustness of constructed evolutionary tree by using 1000 iterations.

Primary Structure Prediction

To understand the chemistry on SAD, study of primary structure of protein was carried out by using the ProtParam tool (http://expasy.org/cgi-bin/protparam) (Gateiger *et al.,* 2005) available at Expasy proteomic server. This server contains various parameters such as molecular

weight, theoretical iso-electric point, instability index, aliphatic index, and grand average hydropathy (GRAVY) to analyze the primary structure.

Secondary and Tertiary Structure analysis

Secondary structure of SAD protein was analyzed through PSIPRED server (Buchan *et al.,* 2010). PSIPRED server is widely use due to easy and simple secondary structure prediction tool. It incorporates the neural network based analysis tool for further analysis of PSI-BLAST output. The selected template and target SAD protein of deodar constituted of 313 amino acids. Delta-BLAST was performed for SAD protein analysis against PDB (http://www.rcsb.org/) to identify best template and it was further used for comparative protein structure modelling and function prediction. GeneSilico Metaserver (http://genesilico.pl/meta2) (Kurowski and Bujnicki, 2003) is used to check the sensitivity and accuracy of template through consensus approach to predict the template and to build the model of protein. The applied methodology shows that 1YQD–chain A was appropriate template to build the model with the high sequences similarity, query coverage and less E-value. The target–template alignment was done with the CLUSTALX and results were displayed through ESPript format (http://espript.ibcp.fr/ESPript/ESPrtipt/) (Gouet *et al.,* 2003). Modeller 9.12 was used to predict the models of SAD protein on the basis of target-template alignment as comparative modelling engine (Sali *et al.,* 1995). The designed models were ranked on the criteria of DOPE energy score. The model refinement was conduct by using the 3D-refinment server.

Quality Assessment for Predicted model

The structural consistency and reliability of SAD model was evaluated through number of tools to assess the quality. PROCHECK (Laskowski *et al.,* 1993) was used to analyze the residue quantity in different zones at Ramachandran plot and to assess the stereo chemical quality of model. ERRAT (Colovos and Yeates, 1993) tool was used to check the statistics of nonbonded interactions between different atom types through the overall quality factor of protein. VERIFY-3D program was used to determine the compatibility of the atomic model (3D) with its own amino acid sequence (1D) (Luthy *et al.,* 1992). All the specified analysis was carried out using Structural Analysis and Verification Sever (SAVeS) (http://nihserver.mbi.ucla.edu/SAVES/). WHAT-IF web server has been used to calculate the bond length and bond angle of SAD model (Hekkelman *et al.,* 2010) (http://swift.cmbi.ru.nl.whatif/). Similarly, to analysis the atomic contact during any steric problem within the molecules and calculate the dihedral angle diagnostics, the MolProbity server was used in quality validation of the 3D model. The native protein folding energy was evaluated by ProSA web server (Weiderstein and Sippl, 2007) through comparison with the energy of known protein models which having the potential mean force.

RESULTS AND DISCUSSION

The published primary amino acid sequence of the target protein SAD (313aa) from *Cedrus deodara* which play an important role in lignin biosynthesis pathway was downloaded from public database Genbank. SAD protein possess the N-terminal ADH–N (Alcohol Dehydrogenase) domain which belongs to the GroES superfamily (Taneja and Mande, 1999) and zinc binding dehydrogenase (ADH zinc–N) revealed by Pfam database. The Alcohol dehydrogenase (ADH) catalyzes the reversible oxidation of alcohols along their acetaldehyde or ketone with the association of NAD reduction. It was divided in three categorizes namely zinc-containing 'long-chain' alcohol dehydrogenases, insect-type, or 'short-chain' alcohol dehydrogenases and iron-containing alcohol dehydrogenases. Zinc-containing ADH's are dimeric or tetrameric enzymes that bind two atoms of zinc per subunit. One of the zinc atoms is essential for catalytic activity while the others are not. Both zinc atoms are coordinated by either cysteine or histidine residues; the catalytic zinc is coordinated by two cysteines and one histidine (Jornvall *et al.,* 1987). Zinc-dependent alcohol dehydrogenases are closely related to other zinc dehydrogenases namely another gene of

Xylitol in *Ptchiu stipitis* (Perrson *et al.,* 1993). The results from above specified tool are perfectly similar with the result through InterproScan (Figure 1). CDD also reveals the same. SignalP results show that there is no signal peptide cleavage site in SAD protein.

Evaluation of Primary Structure

The molecular weight of SAD protein was calculated through ProtParam server and found 34161.4 KD/Dalton. The aliphatic index (AI) can be explained positive factor for the increase of thermal stability of globular protein. (Ikai, 1980) AI is the relative volume of aliphatic side chains of the protein namely alanine, valine, isoleucine and leucine. AI value of *SAD* was found to be high value i.e. 84.53, it shows that *SAD* protein can be stable at wide range of temperature. The isoelectric point (pI) explains pH where the protein surface covers the charge while net charge of protein is zero. It indicated that stability and compatibility of protein. The pI of protein *SAD* was 6.42 which shows the acidic nature of protein (pH<7.0). The instability index provides an estimate of the stability of protein in the test tube. If instability index is <40, then protein is more stable. If its value increases as well as protein unstability increases (Guruprasad *et al.,* 1990). The instability index of SAD was 38.77 which indicated more stability of protein. The protein affinity with water was calculated by GRAVY indices. *SAD* contains very low GRAVY indices *i.e*. -0.037 which indicates its high affinity for water.

Phylogeny Analysis of SAD

BLAST was used for the comparative analysis search of *SAD* gene against the non-redundant database which revealed that *SAD* was closely related to the GroES superfamily members with the ADH-Zn domain from various gymnosperm tree species and showed the evolutionary relationship with pinaceae. SAD query sequence produced the significant alignment with sequences on the specific criteria of query coverage (90%), identity (95%) with E-value< $3e^{-20}$. Multiple sequence alignment through CLUSTALX was passed by total 8 sequences including SAD of deodar. The alignment denotes the highly conserve SAD throughout the process of evolution. Phylogenetic analysis using the Neighbour-joining method in MEGA 5.0 demarcated that *SAD* gene family lies among the 7 coniferous species and a *Populus tremuloides* as angiospermic tree (Figure 2). The clusters show the species-specific divergence with the strong bootstrap values on the nodes. The tree represents the distinct tree-specific clustering of sequences and shows the diversity among the long chain of alcohol dehydrogenase containing members. The sinapyl alcohol dehydrogenase (SAD) of deodar contains high identity percentage with SAD in *Pinus and Picea* sp while it showed 80% similarity with *Populus* sp. The phylogram reflects that these are still evolving and show evolved differentially.

Homology Modelling of SAD protein

3D structure prediction performed through the various techniques such as comparative modelling and ab-initio modelling. Homology modelling considered as the feasible method to predict the suitable model for different application on wide spectrum (Bodade *et al.,* 2010). The method worked on the similarity search between the sequence of target protein and known 3D structure in PDB database. This technique works on the hypothesis that two proteins share the same tertiary structure if their sequences are related to each other. This approach gave the more reliable alignment between the target sequence and template structures with the high level of sequence similarity. Comparative analysis has been done through the BLAST search which revealed the three putative templates with chain A in all of them (PDB id: 1YQD, 2CF5 and 3TWO). The selected template show the higher similarity values with the good query coverage showed in Table 1. These templates are the crystallographic structures of NADP (H)-dependent dehydrogenase family that the formation of monolignols in *Populus tremuloides* (Bomati and Noel, 2005), *Arabidopsis thaliana* (Youn *et al.,* 2006) and *Helicobacter pylori* (Seo *et al.,* 2012). The best template for comparative modelling is 1YQD with the resolution 1.65 Å (Table 2). The SAD of deodar and template sequence has been pairwise aligned and generated the alignment using MUSCLE and displayed in ESPript

shown in (Figure 3). Target-template alignment reflects the perfect homology between them. On the basis of alignment Modeller 9.12 generated the 10 models of *SAD*. The model with the lowest dope energy value has been considered to be thermodynamically stable and analyzed further for the refinement and validation through ProSA and SAVeS (Figure 4).

Table 1. Templates Selected for Comparative Model Building of SAD from BLAST search against PDB

Templates (PDB with their chain)	Total score	Query Coverage (%)	Identity (%)	Resolution (A°)
IYQD	412	99	63	1.65A^0
2CF5	301	100	54	1.76A^0
1UUF	275	99	49	2.0A^0

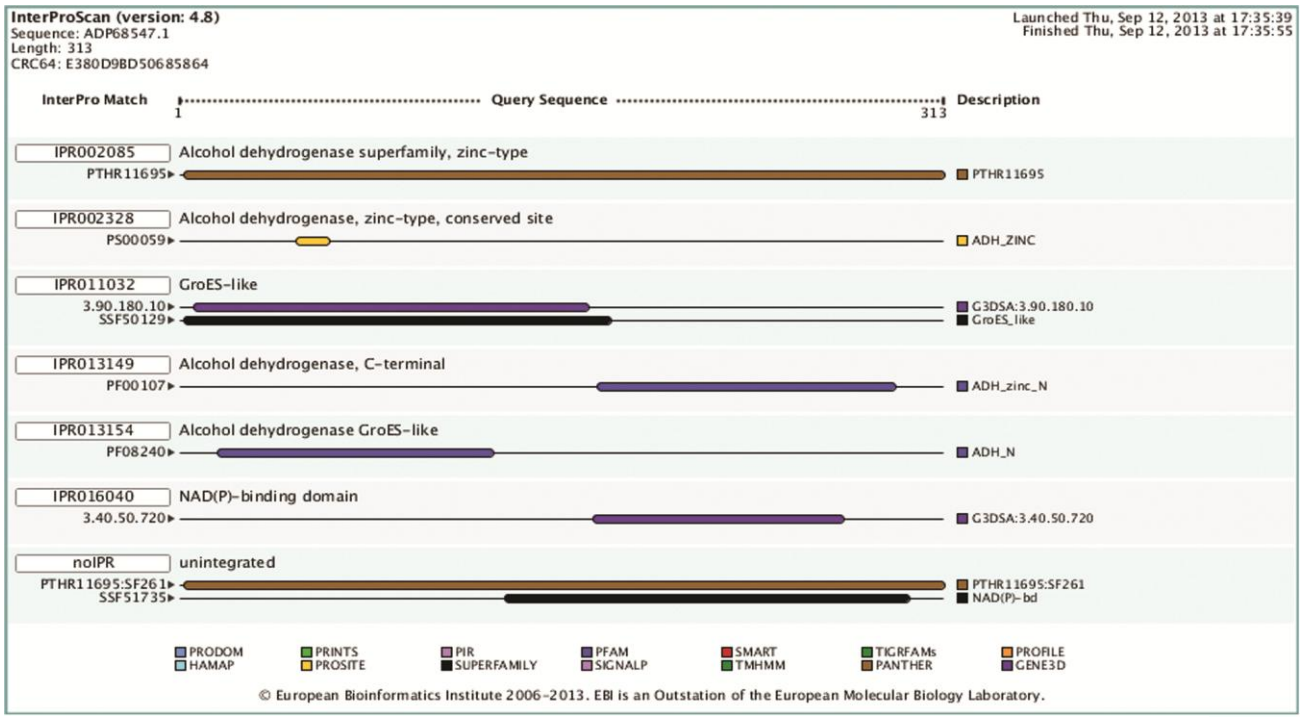

Figure 1. Domain Analysis for SAD protein of Deodar spp. through the InterProScan web server

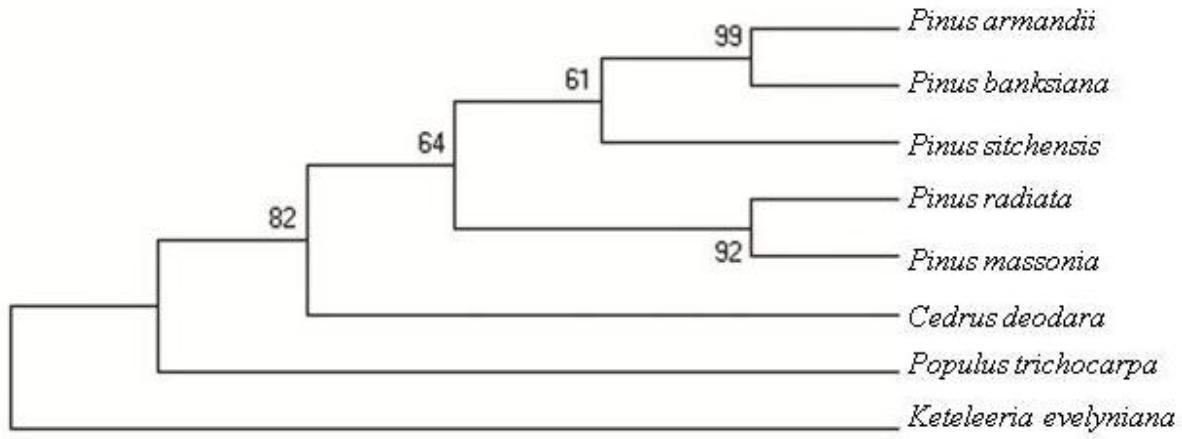

Figure 2. Neighbour-joining tree from SAD and alcohol dehydrogenase containing gymnosperm trees in MEGA 5.0. The numbers of nodes showed the percentage of bootstrap value obtained from 1000 sampling bar

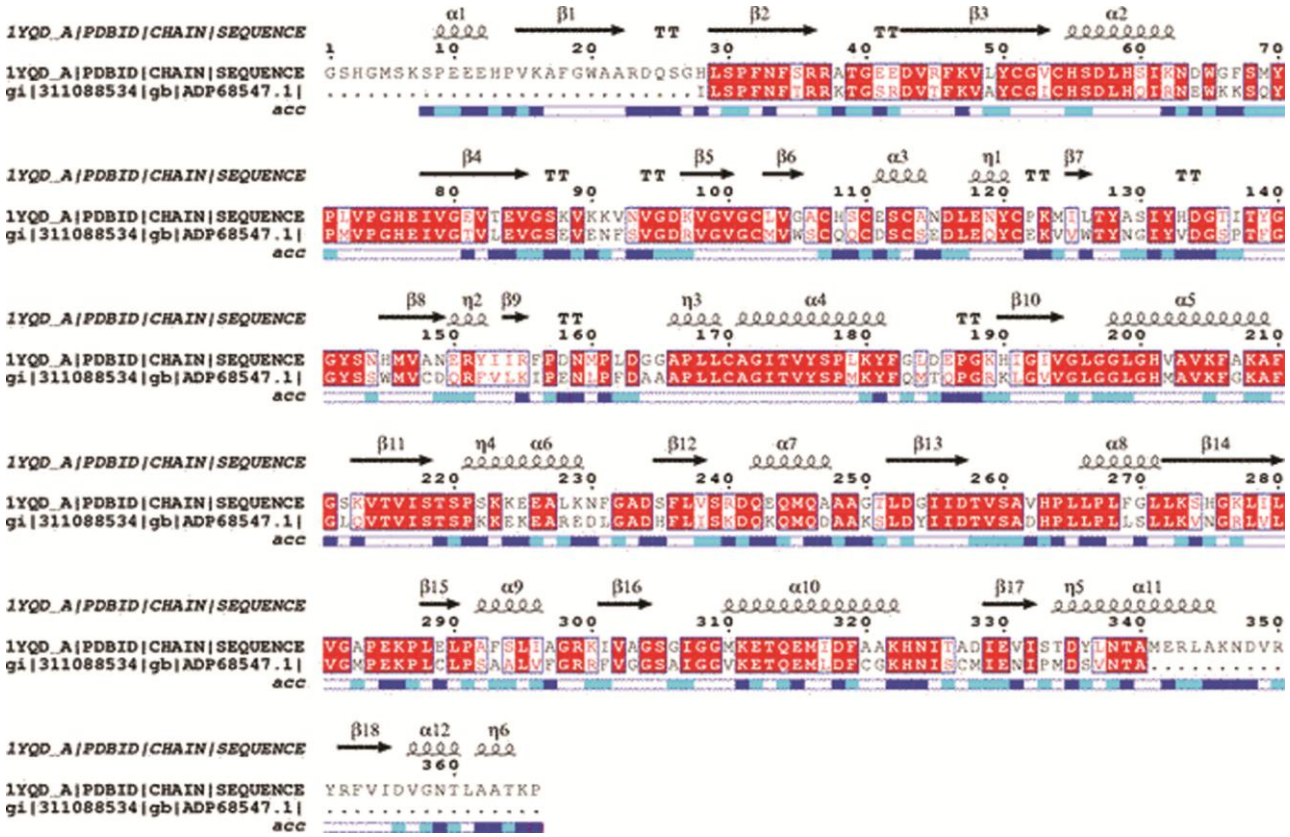

Figure 3. Sequence Alignment between the target SAD and chain-A of 1YQD

The secondary structural elements were identified from the 1YQD structure using ESPript. The α-helices, η-helices, β-sheets and β-turns are denoted α, η, β and TT respectively. Similar amino acids are indicated by *red letters* and completely conserved residues are showed as *white letters* in *white* and *red background* respectively.

Table 2. Secondary Structure Comparison of SAD and Template by PSIPRED

Template Target	Number of amino acid residues (%)			
	Turn	Helix	Strand	Total
SAD (target)	51.4%	28.1%	20.4%	313
1YQD	44.3%	29%	26.7%	359

Table 3. Comparison of Ramachandran plot Statistics of SAD model and its template 1-YQD by PROCHECK

Ramachandran plot statistics	Homology model structure of *SAD*		X-ray Crystallographic structure of template-Chain A-1YQD	
	Residues	Percent	Residues	Percent
Residue in most favoured region [A,B,L]	238	89.8	542	89.7
Residues in additionally allowed region[a,b,l,p]	26	9.8	60	9.9
Residues in generously allowed region[-a,-b,-l,-p]	1	0.4	1	0.2
Residues in disallowed region	0	0.0	1	0.2
Number of non-glycine and non-proline residue	265	100	604	100
Number of end residues (excl Gly & Pro)	2		2	
Number of Glycine residue (as Triangle)	30		76	
Number of Proline residue	16		36	
Total number of Residues	313		718	

Model assessment and Validation

The reliability of backbone of torsion angles phi and psi of designed model, which reflects the residual positions in available zones of Ramachandran plot as shown in Figure 5. Ramachandran plot reveals that 95.8% residue occurs in most favoured regions, 3.5% residues fell in additional allowed region, 0.7% residues fell in allowed region and no residue in disallowed region (Table 3). The model quality was further analyzed through the Errat server which indicates the acceptable protein environment with the good errat score 81.84% (Colovos *et al.,* 1993). The SAD model was analyzed through the Verify -3D , the results showed the 84.5 % amino acids had an

average 3D-1D score >0.2 and indicates the reliability of the model. The average magnitude of the volume irregularities in terms of Z-score root mean square deviation of model was predicted through the PROVE server The Z-Score root mean square (RMS) values of the model and the template were 1.498 and 28.308 respectively, Z-score RMS value of ~1.0 indicates the good resolution of structure. What-If server was used to predict the different criteria such as the coarse packing quality, anomalous bond length & angles, planarity, packing quality and the collision with the symmetry axis, distribution of omega angles, proline puckering of the model protein reflects the acceptance of good quality. ProSA programme was used to predict the energy profile of the model and Z-score value which reflects the model quality through the measurement of the total energy of the structure. It also calculates the interaction energy per residue on the basis of distance based pair potential. The negative ProSA energy shows more reliability of model and reflects the good quality of it. So, Z-score of the model is -7.41 and -9.42 for template has been done through the ProSA analysis shown in (Figure 6).

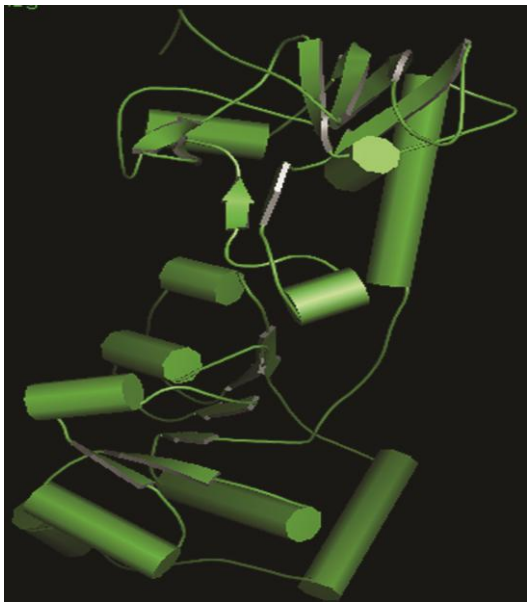

Figure 4. Protein 3D model designed by Modeller 9v 12 and visualize as cartoon through Pymol

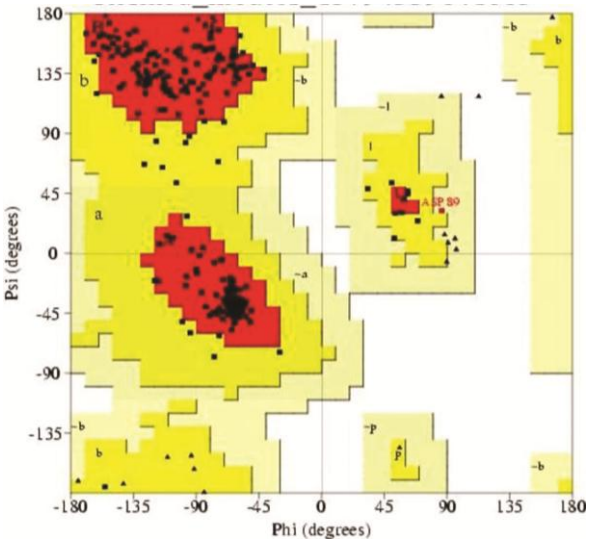

Figure 5. Ramachandran Plot of SAD model
(Plot was generated by PROCHECK program)

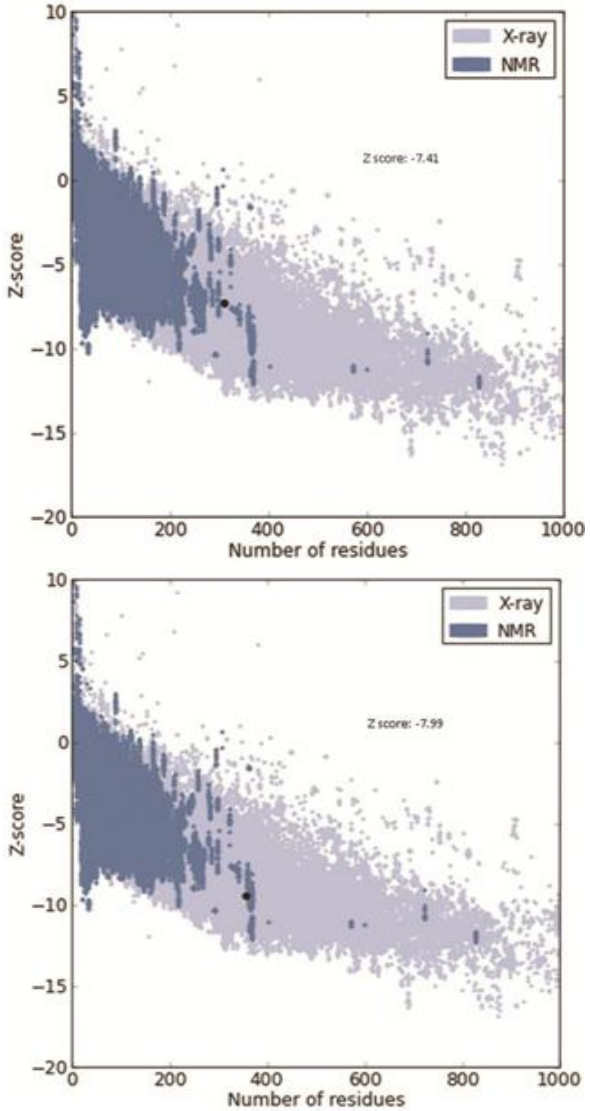

Figure 6. Protein Structure Analysis (ProSA) of model SAD and template 1YQD showing z-score -7.41 and -7.99 respectively

CONCLUSION

SAD, a gene belongs to the CAD/SAD gene family of various tree species which encodes for the catalysis of alcohol dehydrogenase and contains Zn domain in its active site. Although SAD represents the long chain of alcohol dehydrogenases in plants, it shares the significant similarity with the members of conifers. Despite it shows the structural similarity with the *Populus tremuloides*. The molecular evolutionary analysis revealed that SAD evolved differentially among the angiosperms and gymnosperms. The result showed that SAD in gymnosperms has functional divergence with poplar. So, it may be evolved due to the gene duplication. The SAD model has been generated by using the 1YQD to analyze the structural divergence among deodar and poplar but the result reflected that SAD in deodar also have the long chain of alcohol dehydrogenase as in poplar. The structural analysis showed that SAD model of deodar was well fitted with the poplar and indicated that it govern the novel phenylpropanoid pathway of lignin biosynthesis *via* shikimate pathway. Further study will involve the protein-protein and protein-ligand interaction to explore the stability and diversity in phenylpropanoid pathway in molecular aspects.

Acknowledgements: The authors thankfully acknowledge the support of DST-Inspire (Govt. of India), New Delhi and Forest Research Institute (Govt. of India), Dehradun to provide the facility to work.

REFERENCES

Altschul S.F, Gish W., Miller W., Myers E.W., and Lipman D.J. (1990). Basic Local Alignment search tool. *J Mol Biol* 215:403-410.

Barcelo A., Ros Gomez Ros L.V. and Carrasco A.E (2007). Looking for Syringyl peroxidises. Trends Plant Sci, 12:486-491.

Bodade R.G., Beedkar S.D., Manwar A.V., and Khorbragade C.N (2010). Homology modeling and docking study of xanthine oxidase of arthrobacter sp. XL26. *Int J Biol Macromol*, 47: 298-303.

Boerjan W., Ralph J. and Baucher M (2003). Lignin biosynthesis. *Annu Rev Ecol Syst,* 54:519-546.

Bomati E.K. and Noel J.P (2005). Structural and Kinetic basis for substrate selectivity in Populus tremuloides sinapyl alcohol dehydrogenase. *Plant Cell,* 17:1598-1611.

Buchan D.W., Ward S.M., Lobley A.E., Nugent T.C., Bryson K., and Jones D.T (2010). Protein Annotation and modelling servers at University College London. *Nucleic Acids Res*, 38:W563-W568.

Colovos C. and Yeates T.O (1993). Verification of Protein structures: Patterns of non-bonded atomic interactions. *Protein Sci,* 2:1511-1519.

Dixon R.A., Lamb C.J., Masoud S., Sewalt V.J.H. and Paiva N.L (1996). Metabolic engineering: Prospects for crop improvement through the genetic manipulation of phenylpropanoid biosynthesis and defense responses—a review. *Gene*,179:61–71.

Filiz E. And Koc I (2014). Assessment of chloroplast microsatellite from pine family (Pinaceae) by using bioinformatics tools. Indian J. of Biotech, 13:34-40.

Finn R.D., Mistry J., Tate J (2010). The Pfam protein families database. *Nucleic Acids Res,* 38:D211-222.

Gasteiger E., Hoogland C., Gattiker A. (2005). Protein Identification and analysis tools on the ExPASy server. In: John M. Walker Ed., The Proteomics Protocols Handbook, Humana Press, pp.572-607.

Gouet P., Robert X., and Courcelle E (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from Atomic Structures of proteins. *Nucleic Acids Res*, 31:3320-3323.

Guruprasad K., Reddy B.V. and Pandit M.W (1990). Correlation between stability of protein and its dipeptide composition: A novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Prot Eng*, 4:155-164.

Hekkelman M.L., Beek T.T., Pettifier S.R., Thorne D., Attwood T.K., and Virend G (2010). WIWS: A protein structure bioinformatics Web service collection. Nucleic Acids Res, 38, W719-W723.

Ikai A.J (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem,* 88:1895-1898.

Jain A.,Jain A., Vaidhya A., Kumar D., Jain V (2014). Preliminary phytochemical pharmacognostical and physio-chemical evaluation of Cedrus deodara heartwood. *J. of Pharm. and Phytochem*, 3(1):91-95.

Jornvall H., Persson M., and Jeffrey J (1987). Characteristics of alcohol/polyol ehydrogenases: the zinc-containing long-chain alcohol dehydrogenases. *Eur. J. Biochem,* 167:195-201.

Kim S.J., Kim K.W., Cho M.H., Franceschi V.R., Davin L.B. and Lewis N.G (2007). Expression of cinnamyl alcohol dehydrogenases and their putative homologues during Arabidopsis thaliana growth and development: lessons for database annotations? *Phytochemistry*, 68:1957-1974.

Kurowski M.A. and Bujnicki J.M (2003). GeneSilico protein structure prediction meta server. *Nucleic Acids Res,* 31:3305-3307.

Larkin M.A., Blackshields G., Brown N.P (2007). CLUSTAL-W and CLUSTALX version 2.0. *Bioinformatics*, 23:2947-2948.

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

Lewis NG, Yamamoto E. (1990). Lignin: occurrence, Biogenesis and biodegradation. *Annual Review of Plant Physiology and Plant Molecular Biology*, 41:455–496.

Li L., Popko J.L., Umezawa T. and Chiang V.L (2000). 5-hydroxyconiferyl aldehyde modulates enzymatic methylation for syringyl monolignol formation, a new view of monolignol biosynthesis in angiosperms. *J Biol Chem,* 275:6537–6545.

Li L.G., Cheng X.F., Leshkevich J., Unezawa T., Harding S.A. and Chiang V.L (2001). The Last step of Syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. *Plant Cell,* 13:1567-1585.

Lin C.P., Huang J.P., Chung S.W., Chih Y.H., and Chaw S.M. (2010). Comparative Chloroplast Genomics Reveals the Evolution of Pinaceae Genera and Subfamilies. *Genome Biol Evol*, 2:504–517.

Lothar K. and Telewski F.W. (2006). Biomechanics and transgenic wood. *Am. J. Bot,* 93:1433–1438.

Luthy R., Bowie J.U. and Eisenberg D (1992). Assessment of protein models with three dimensional profiles. *Nature*, 356:83-85.

Marchler B.A., Lu S., Anderson J.B (2011). CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res,* 39:D225-229.

Persson B., Johan H., Walfridsson M., Hagerdal B.H., Keranen S., Penttila M. and Jornavall H (1993). Dual relationships of xylitol and alcohol dehydrogenases in families of two protein types. FEBS, 324:9-14.

Peter G. and Neale D (2004). Molecular basis for the evolution of xylem lignification. *Curr Opin Plant Biol,* 7:737-742.

Petersen T.N., Brunak S., Heijne G., and Nielsen H (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods*, 8:785-786.

Rippert P., Puyaubert J., Grisollet D., Derrier L., Matringe M (2009). Tyrosine and phenylalanine are synthesized within the plastids in Arabidopsis. *Plant Physiol,* 149:1251–1260.

Saitou N. and Nei M (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol,* 4:406-425.

Sali A., Potterton L., Yuan F., Van-Vlijmen H., and Karplus M (1995). Evaluation of comparative protein modeling by MODELLER. *Proteins*, 23:318-326.

Seo K.H., Zhuang N.N., Chen C., Song J.Y., Kang H.L., Rhee K.H. and Lee K.H (2012). Unusal NADPH conformation in crystal structure of cinnamyl alcohol dehydrogenase from Helicobacter pylori in complex with NADP(H) and substrate docking analysis. *Febs Lett,* 586:337-343.

Skolnick J., Fetrow J.S. and Kolinski A (2000). Structural genomics and its importance for gene function analysis. Nature Biotech, 18”283–287.

Taneja B. and Mande S.C (1999). Conserved Structural features and sequence patterns in GroES fold family. *Protein Eng,* 12:815-818.

Thompson J.D., Higgins D.G. and Gibson T.J (1994). CLUSTAL W: improving the sensitivity of progressive multiple

sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nuc. Ac. Res,* 22:4673-4680.

Vanholme R., Morreel K., Ralph J. and Boerjan W (2008). Lignin engineering. *Curr Opin Plant Biol*, 11, 278–285.

Wang X.Q., Tank D.C. and Sang T (2000). Phylogeny and Divergence Times in Pinaceae: Evidence from Three Genomes. *Mol. Biol. Evol*, 17:773–781.

Weiderstein M. and Sippl M.J (2007). ProSA-web: Interactive web service for the recognition of errors in three dimensional structures of proteins. *Nucleic Acids Res*, 35:W407-W410.

Whetten R.W., Mackay J.J. and Sederoff R.R (1998). Recent advances in understanding the lignin biosynthesis. *Annu Rev Plant Phys*, 49:585-609.

Yokoyama S and Harry D.E (1993) .Molecular phylogeny and evolutionary rates of alcohol dehydrogenases in vertebrates and plants. *Mol Biol Evol*,10:1215-1226.

Youn B., Camacho R., Moinuddin S.G.A., Lee C., Davin L.B., Lewis N.G., Kang C.H (2006). Crystal structure and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org Biomol Chem,* 4:1687-1697.

Zdobnov E.M. and Apweiler R (2001). InterProScan-An integration platform for the signature recognition methods in InterPro. *Bioinformatics*, 17:847-848.

Source of Support: DST-Inspire (Govt. of India).
Conflict of interest: None. Declared.