Bioinformation
www.bioinformation.net

open access
Hypothesis

Isolation, characterization and structural features of amorpha - 4, 11-diene synthase (ADS_{3963}) from *Artemisia annua* L.

Pravej Alam[1], Usha Kiran[2], M. Mobeen Ahmad[1], Kamaluddin[3], Mather Ali Khan[1], Shalu Jhanwar[4], M. Z. Abdin[1]*

[1]Centre for Transgenic Plant Development, Department of Biotechnology, Faculty of Science, [2]Faculty of Engineering and Interdisciplinary Sciences, Jamia Hamdard, New Delhi, India, [3]Department of Plant Breeding and Genetics, Faculty of Agriculture, Regional Research Station, Wadura, Sher-e-Kashmir University of Agriculture and Technology (J&K), India; [4]School of Information Technology, Jawaharlal Nehru, University, New Delhi, India, email: mzabdin@rediffmail.com; *Corresponding author.

Received January 02, 2010; accepted February 19, 2010; published March 31, 2010

Abstract:
With the escalating prevalence of malaria in recent years, artemisinin demand has placed considerable stress on its production worldwide. At present, the relative low-yield of artemisinin (0.01-1.1 %) in the source plant (*Artemisia annua* L. plant) has imposed a serious limitation in commercializing the drug. Amorpha-4, 11-diene synthase (ADS) has been reported a key enzyme in enhancing the artemisinin level in *Artemisia annua* L. An understanding of the structural and functional correlations of Amorpha-4, 11-diene synthase (ADS) may therefore, help in the molecular up-regulation of the enzyme. In this context, an *in silico* approach was used to study the ADS_{3963} (3963 bp) gene cloned by us, from high artemisinin (0.7-0.9% dry wt basis) yielding strain of *A. annua* L. The full-length putative gene of ADS_{3963} was found to encode a protein consisting of 533 amino acid residues with conserved aspartate rich domain. The isoelectric point (pI) and molecular weight of the protein were 5.25 and 62.2 kDa, respectively. The phylogenetic analysis of ADS genes from various species revealed evolutionary conservation. Homology modeling method was used for prediction of the 3D structure of ADS_{3963} protein and Autodock 4.0 version was used to study the ligand binding. The predicted 3D model and docking studies may further be used in characterizing the protein in wet laboratory.

Keywords: *Artemisia annua*; artemisinin; ADS_{3963} gene; homology modeling; phylogenetic tree; docking

Abbreviations: ADS = Amorpha-4, 11-diene synthase; FPP = farnesyl pyrophosphate; ORF = open reading frame; PCR = polymerase chain reaction

Background:
Multi drug resistance in *P. falciparum* to the commonly used antimalarial agents is becoming more and more widespread and poses a serious threat to conventional and current therapeutic treatments. An alternative class of antimalarial compound, sesquiterpenoid artemisinin from *Artemisia annua* L. is effective against both chloroquine-resistant and sensitive strains of *Plasmodium* species as well as the species causing cerebral malaria. Its relative low yield (0.01-1.1% dry weight) **[1, 2]** however, has caused a serious concern in commercializing the drug **[3]**. The physiological and cell culture studies were carried out to improve the yield of artemisinin, but were mostly unsatisfactory **[4]**. The chemical synthesis of artemisinin is also possible, but it is complicated and economically unviable due to the poor yield **[5]**. Recent reports have highlighted the use of biotechnological approaches such as metabolic engineering and genetic modification of microbe and plants as a feasible alternative for the semi-synthesis of artemisinin and its precursors [**6, 7**]. *S. cerevisiae* had been enginnered to produce artemisinic acid, precursor of artemisinin, at a significantly higher level than in *A. annua* **[8]**. The complete synthesis of artemisinin outside the source plant is however, not achieved yet and has to rely either on biotransformation using plant extract **[7, 9]** or semisynthesis **[10]** to obtain end product. Studies have been conducted in different laboratories to elucidate the biochemical pathway of artemisinin and its regulation with an aim to improve artemisinin content of *A. annua* L. Based on the biosynthetic pathway for artemisinin starting from farnesyl pyrophosphate (FPP) followed by synthesis of germacrane skeleton, dihydrocostunolide, cardinanolide, arteannuin B and artemisinin **[11]**, but these studies did not indicate artemisinic acid as precursor of artemisinin. In another study, two compounds (secocadinane and dihydroxycadinanolide) were isolated and postulated an alternative route for artemisinin biosynthesis **[12]**. According to this postulate, arteannuin B gets converted into a dihydroxycadinanolide, which then undergoes Grobs fragmentation to yield an enolic form of a secocadinane. This further undergoes enzymatic oxygenation yielding artemisitene, which finally reduces to artemisinin. The *in vitro* and *in vivo* transformation of artemisinic acid to arteannuin B and artemisinin with an over all yield of 4.0 % **[13]**, suggesting that artemisinic acid is a common precursor for both arteannuin B and artemisinin. This was confirmed by several studies employing crude and semi purified cell free extracts of leaf homogenates of *A. annua*, where arteannuin B was found an intermediate in the bioconversion of artemisinic acid to artemisinin **[9]**. The transformation of dihydroartemisinic acid into artemisinin by cell free extracts from *A. annua* L. plants **[14]**. These studies thus, indicate that artemisinic acid is converted into artemisinin either via arteannuin B or dihydroartemisinic acid. Recently, amorpha-4, 11-diene synthase (ADS), has been identified as first enzyme of artemisinin biosynthesis. It has also been reported to be the key regulatory enzyme which catalyzes the cyclization of farnesyl pyrophosphate (FPP) into the sesquiterpenoid skeleton, amorpha-4,11-diene **[15, 16]**. Thus, increasing the activity of amorpha-4,11-diene synthase (ADS) enzyme through overexpression of its gene in *A. annua* L. may be a promising approach to enhance artemisinin biosynthesis. Consequently, several groups have reported the cloning, sequencing and expression of Amorpha–4,11-diene synthase genes from different strains of *A. annua* L. **[17-20]**. In this paper, we report cloning and characterization of a putative Amorpha-4, 11- diene synthase gene (ADS_{3963}) from high yielding *A. annua* L. strain (0.7-0.9 % artemisinin dry weight basis).

Methodology:
The seeds of *Artemisia annua* L., a high artemisinin yielding strain (0.7-0.9% artemisinin dry weight basis) was provided by the Ipca Laboratories Limited, Ratlam, MP, India. Plants were grown in the experimental field of Jamia Hamdard, New Delhi, India.

Isolation of genomic DNA
Fresh leaves were collected from *A. annua* L. plants at pre-flowering

stage. The extraction of DNA was done by modified CTAB method **[21]**.

Generation of *A. annua* L. primers

The PCR primers were designed using published Amorpha-4, 11-diene synthase sequence retrieved from NCBI GenBank (Acc. no. AF327527) from *A. annua* L. plant. The primers were derived from total ORF region of Amorpha-4, 11-diene synthase gene and were synthesized by Sigma-Aldhrich Chemicals Pvt. Ltd. with the following sequences: Forward primer, ATGTCACTTACAGAAGAAAAACCTATTC and Reverse primer, TCATATACTCATAGGATAAACG.

Amplification of Amorpha-4, 11- diene synthase (ADS_{3963})

The PCR reaction mixtures containing 100 µl PCR mastermix [(0.2 mmol of each deoxynucleotide, buffer and 1.5 mmol of $MgCl_2$), 1 µl (of 1 µg/µl) of each forward and reverse primer and 0.5 µl (5U) of *Taq* DNA polymerase] was divided into ten 9 µl aliquots. One microlitre of *Artemisia annua* L. genomic DNA was added as either neat, 1 in 5 or 1 in 10 dilution to make the final reaction mixture 10 µl. The cycling parameters were 95 °C for 3 min: 40 cycles of denaturation for 30 s at 94 °C, annealing for 1 min at 55 °C an extension for 1 min at 72 °C, with final extension 72 °C for 5 min.

Cloning of amorpha-4, 11-diene synthase

The resulting PCR product was purified and ligated into the pDrive cloning vector (Qiagen Pvt. Ltd.) following the manufacturer's instruction. The ligation mixture was then used to transform DH5α *E. coli* (100 µl) competent cells. The cells were mixed with 900 µl of Luria broth (LB) medium, incubated at 37 °C for an hour and *E. coli* cell culture (100 µl) was plated onto X-gal/Amp/IPTG/LB plate. Following overnight incubation at 37 °C, the recombinant colonies were picked and used for plasmid extraction. Plasmid DNA was purified using Gene JET™ Plasmid Miniprep Kit (Fermentas Life Science) following the manufacture's instructions. The clones were confirmed through PCR.

Sequencing of Amprpha-4, 11-diene synthase (ADS_{3963})

The sequencing of ADS_{3963} gene cloned in pDrive was carried out through automated sequencer at Bangalore Genei, Bangalore.

Nucleotide and Protein Sequence Accession Number

The genomic sequence of Amorpha-4, 11-diene synthase gene having 3963 bp and its annotation has been submitted to GenBank and was assigned accession numbers FJ432667 for nucleotide and protein ACL15394.

Domain/Motif Search

PROSITE scan at the EXPaSy (http://www.expasy.ch/tools/scanprosite) was used to identify PROSITE motif in amorpha-4 11-diene synthase protein.

Phylogenetic analysis

Nucleotide sequences of ADS_{3963} (FJ432667) from *A. annua* L. and other sesquiterpene synthase genes (EU798693, *Santalum album* sesquiterpene synthase mRNA; EU726270, *Cistus creticus* subsp Creticus germacrene B synthase mRNA; AF441124, *Citrus sinensis* valencene synthase (tps1) mRNA; AF288465, *Citrus junos* terpene synthase mRNA; NM122301, *Arabidopsis thaliana* terpene synthase/cyclase family protein (AT5G23960) mRNA; AF279455, *Lycopersicon hirsutum* sesquiterpene synthase 1 (SSTLH1) mRNA; AY860847, *Artemisia dracunculus* clone 27 sesquiterpene synthase gene; AB247331, *Zingiber zerumbet* zss1 mRNA for alpha-humulene synthase; AY640155, *Cucumis sativus* beta-caryophyllene synthase mRNA; AY397644, *Solidago canadensis* (+)-germacrene D synthase mRNA; AY900123, *Ixeris dentata* leaf sesquiterpene cyclase mRNA; AF174294, *Gossypium arboreum* (+)-delta-cadinene sythase (CAD1-C1) gene and AY860849, *Artemisia dracunculus* clone 27 sesquiterpene synthase gene) were obtained from NCBI GenBank. The evolutionary relationship between nucleotide sequences was developed by MEGA 4.0 program to create Neighbour-Joining tree **[22]**. Bootstrap probability was introduced to assure the statistical significance of groups in the phylogenetic tree **[23]**.

Homology modeling of ADS_{3963}

Template searching

In an attempt to search the suitable template for modeling the ADS_{3963} protein, $(PS)^2$ MODELLER server [http://ps2.life.nctu.edu.tw/], an online tool for searching template based on sequence and structure- wise similarity was used **[24]**. On the basis of homology percentage (39%), single template of the 5-epi- aristolochene synthase [PDB ID: 5eau] was selected [**25]** for 3D modeling of ADS_{3963} protein.

Sequence alignment

Amino acid sequence alignment of target and template protein was done using the Swiss-PdbViewer package (http://www.expasy.ch/spdbv/) with default parameters. The aligned sequences were checked and adjusted manually to minimize the number of gaps and insertions. The homology based modeling was used to predict the 3D structure of ADS_{3963}. The $(PS)^2$ MODELLER was used for alignment as well as modeling. The scripts were also employed to perform an alignment between the target and template sequence. A predicted 3D model was then obtained from the script model default based on the generated alignment.

Refinement and validation of predicted structure

The predicted model constructed was solvated and subjected to constraints of energy minimization with a harmonic constraint of 100 kJ/mol/$Å^2$ applied for all protein atoms, using the steepest descent and conjugate gradient technique to eliminate bad contacts between protein atoms and structural water molecules. Computations were carried out *in vacuo* with the GROMOS96 43B1. The stereo chemical quality of the predicted structure was assessed by VADAR (http://www.redpoll.pharmacy.ualberta.ca/vadar) **[26]** and PROCHECK (http://nihserver.mbi.ucla.edu/). Ramachandran analysis showed that more than 90 % amino acids of ADS_{3963} gene were in most favored region, which supports the reliability of the predicted model.

Docking of Amorpha–4, 11-diene synthase (ADS_{3963}) with Substrate Farnesyl Pyrophosphate (FPP)

After obtaining the final model, the centre co-ordinates of probable binding sites of enzyme were extracted using PASS (Putative Active Sites with Spheres) version pass10_2.0.36 (http://www.ccl.net/cca/software/UNIX/pass). FPP (PDB ID: 1UBY) was used as a substrate for docking studies. The ligand coordinates (FPP) were downloaded from Protein Data bank (http://www.rcsb.org/pdb/home/home.do). The FPP was docked with ADS_{3963} using the Lamarckian Genetic algorithm (LGA) provided by the AutoDock program, version 4 (http://autodock.scripps.edu/). The docking of FPP was performed with respect to all the 17 binding sites of the enzyme. The residues lining the cavities within 10Å region around the ligand molecule were extracted using a GETNEAREST **[27]** and LIGPLOT **[28]** was used to plot protein-ligand interactions.

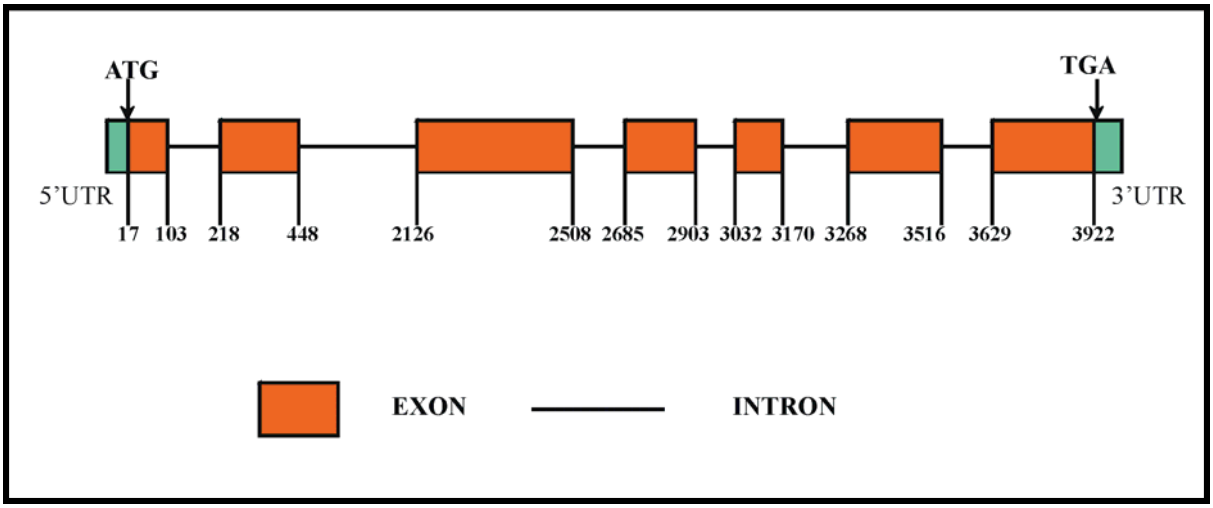

Figure 1: Genomic organization of Amorpha-4, 11-diene synthase gene (ADS3963). Numbers shown below the lines are the start and end positions of exons.

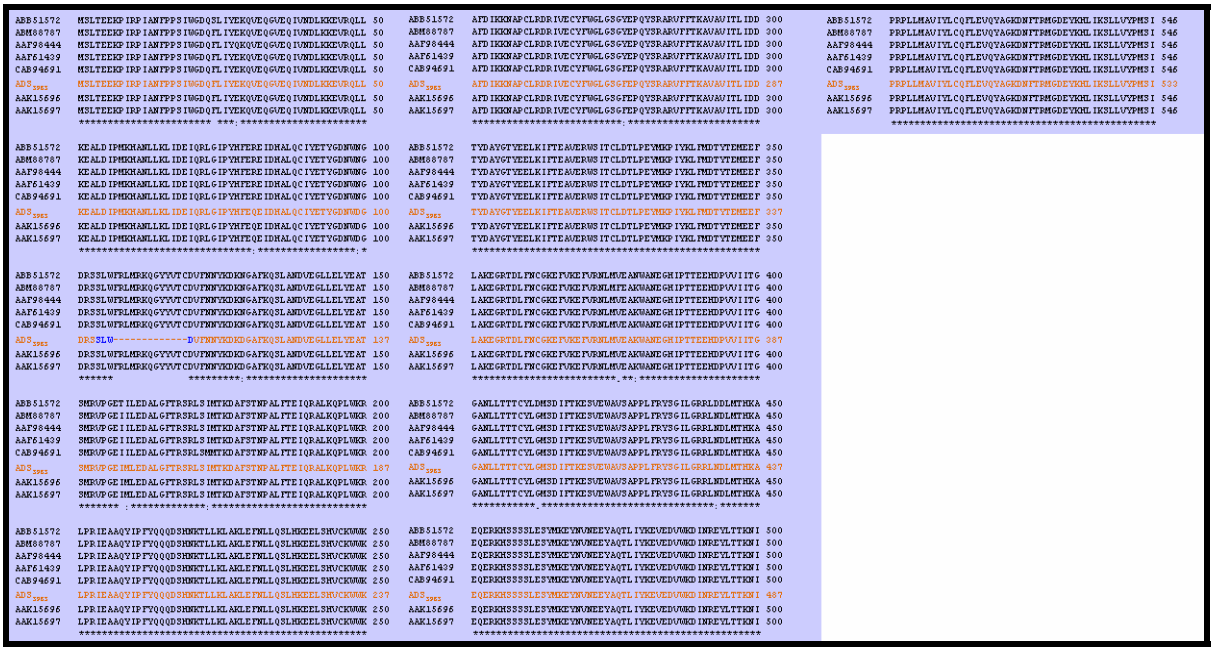

Figure 2: Multiple sequence alignment of different amorpha-4, 11-diene synthase genes.

Results and Discussion:

Amorpha-4, 11-diene synthase belongs to the class of enzymes referred as sesquiterpene synthases (sesquiterpene cyclases). These are very similar in physical and chemical properties. Based on their similar reaction mechanism, conserved structural and sequence characteristics including amino acid sequence homology, conserved sequence motif, introns number and exon size, several groups have suggested that plant sesquiterpene synthases have a common evolutionary origin. Because of lack of the knowledge of structure-function correlation, it was impossible to predict the function of terpene synthases solely on the degree of sequence identity until now **[29]**. The cloned ADS3963 gene sequence (FJ432667) from a high yielding strain of *A. annua* by us, have shown 98 % identity with nucleotide and aminoacid sequences derived from earlier published *A. annua* ADS genes. The genomic organization of ADS3963 gene comprises of 7 exons (17-103; 218-448; 2126-2508; 2685-2903; 3032-3170; 3268-3516 and 3629-3922) and 8 introns (1-16; 104-217; 449-2125; 2509-2684; 2904-3031; 3171-3267; 3515-3628 and 3923-3963), respectively (**Figure 1**). The deduced mass and the pI of the encoded protein are 62.2 kDa and 5.25, respectively. Multiple sequence alignment analysis of ADS3963 gene with seven ADS genes already reported has revealed 98 % homology. ADS3963 (533 aa) protein has shown more than 98 % identity with amorpha-4, 11-diene synthase proteins of other strains of *Artemisia annua* L. with respect to amino acid composition (**Figure 2**). Phylogenetic tree analysis has shown relatively higher homology of ADS3963 gene from *Artemisia annua* L. with sesquiterpene synthases from angiosperms (**Figure 3**). The 6 different types of domain were found in ADS3963 protein (ACL15394) (**Table 1**).

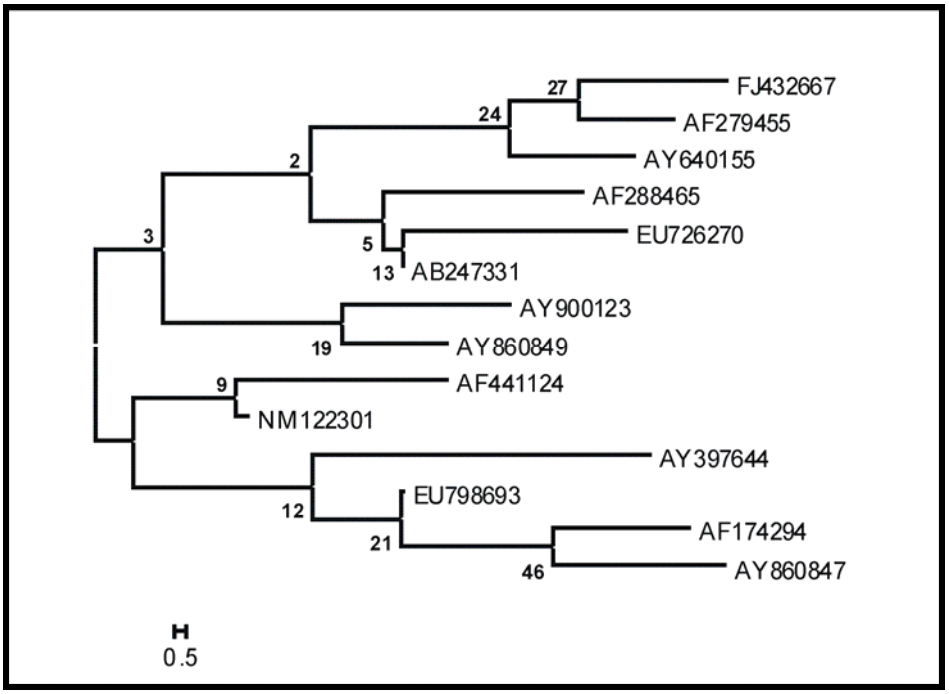

Figure 3: Evolutionary relationships of 13 taxa of nucleotide sequence of ADS3963 and other sesquiterpene synthase genes using the Neighbor-Joining method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site.

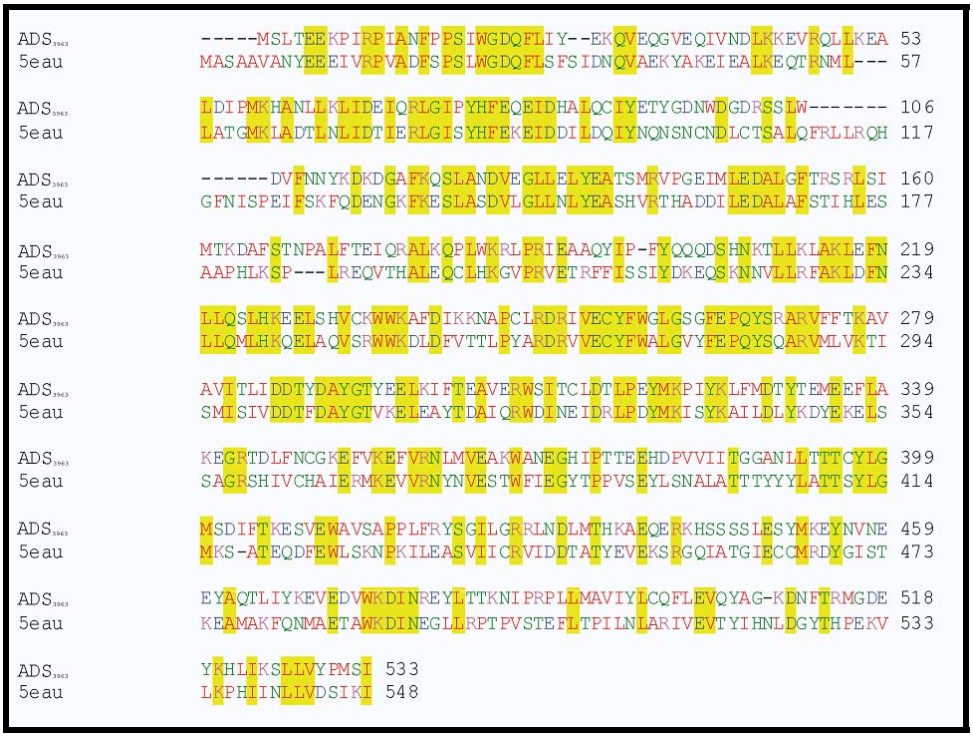

Figure 4: Alignment of deduced amino acid sequences of ASD3963 and epi-aristolochene synthase (5eau) from tobacco. The consensus sequence shows amino acid residues conserved in above sequence.

The highly conserved region, aspartate rich motif (DDxxD position 286-290), is the characteristic of all terpene synthases **[30]**. This motif is involved in the coordination of the substrate bound divalent metal ion (Mg^{2+} and Mn^{2+}) **[25]**. Two basic residues Arg^{264} and Arg^{441} from 5-epi-aristolochene synthase from *Nicotiana tabacum* **[25]** corresponding to Arg^{249} and Arg^{427} in ADS_{3963} protein are brought close to each other by loop movement (**Figure 4**). The deletion of 13 amino acid residues from ADS3963 protein has shown no influence on its catalytic domains, as they may not constitute the essential structure or it may be non catalytic site of protein. The presence of a new motif SlwD, a casein kinase II phosphorylation domain site at position 104-107 in ADS3963 protein, not found in other terpene synthase, is a unique finding of our study. It has been shown that the activity of enzymes of terpenoid biosynthetic pathway in higher eukaryotes is regulated by phosphorylation/dephosphorylation. Enzymes whose activity is regulated by phosphorylation differ both with respect to the spatial relationship between their active and regulatory sites and the mechanisms by which phosphorylation modulates activity. Therefore, the presence of new site present in ADS3963 protein suggested may be part of regulatory region of the enzyme and the DDxxD may be the part of active site for binding of the substrate.

The three-dimensional (3D) structure of protein is of major importance in providing insights into their molecular functions. The result showed that the putative ADS3963 protein contains 72 % of α–helices, 23 % β-turns and 26 % of random coils (**Figure 5**). Penetrating through most part of the secondary structure, α - helices and random coils are the most abundant structural elements of ADS3963, while β turns are intermittently distributed in protein. The total energy values of the predicted 3D model of ADS protein were calculated as 93 % before energy minimization and 90 % after energy minimization from Ramachandran plot. The refined model of ADS3963, analyzed by VADAR for the evaluation of the Ramachandran plot quality, was found to be satisfactory based on expected values representing those numbers which would be expected for highly refined X-Ray and NMR protein structure (**Table 2**). The description of reaction regulation in enzymes responsible for activating and catalyzing small molecules requires identification of ligand movement into the binding site and out of the enzyme through specific channels and docking sites.

Recent studies have revealed that the core sequences of many proteins were nearly optimized for stability by natural evolution. Surface residues, by contrast, were not so optimized, presumably because protein function is mediated through surface interactions with other molecules. Here, we sought to determine the extent to which the sequences of protein ligand-binding and enzyme active sites could be predicted by optimization of scoring functions based on protein ligand-binding affinity rather than structural stability **[32]**. In an attempt to find the possible binding sites of FPP on Amorpha-4, 11-diene synthase (ADS3963), PASS was performed. The output of PASS contains center coordinates for 17 binding sites. Docking of FPP was performed with respect to all the 17 binding sites of the enzyme (**Table 3**). The aim of molecular docking is to achieve an optimized conformation for both the protein, ligand and relative orientation between protein and ligand such that the free energy of the overall system is minimized. Lower energy corresponds to better binding therefore; initial six of these binding sites (**Table 3**) were studied for the interaction with FPP (**Figure 6**). Based on the docking studies the following stretches are Ala 321-Ala324 Lys-398; Ala 234-Val237-Phe283-Thr286-Tyr-287; Ser94-Met95-Trp141-Trp430-Asn434; Ser94-Arg96- Glu104-Leu107-Lys142-Lys431; Lys137-Arg143-Ile147-Ala150-Gln151-Leu478 and Ser218-Gly219- Tyr224-Arg228-Cys352-Met356-Aln450 may constitute to substrate binding. These results may have implications for understanding the role of amorpha–4, 11-diene synthase in cyclization of FPP.

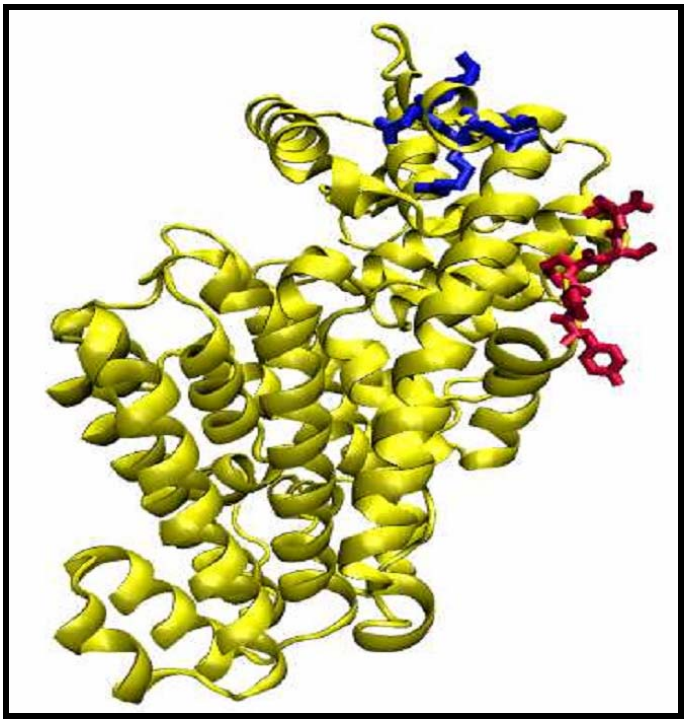

Figure 5: Prediction of 3D model of Amorpha-4, 11-diene synthase showing N (blue) and C (Red) terminal isolated from *Artemisia annua* L. plant.

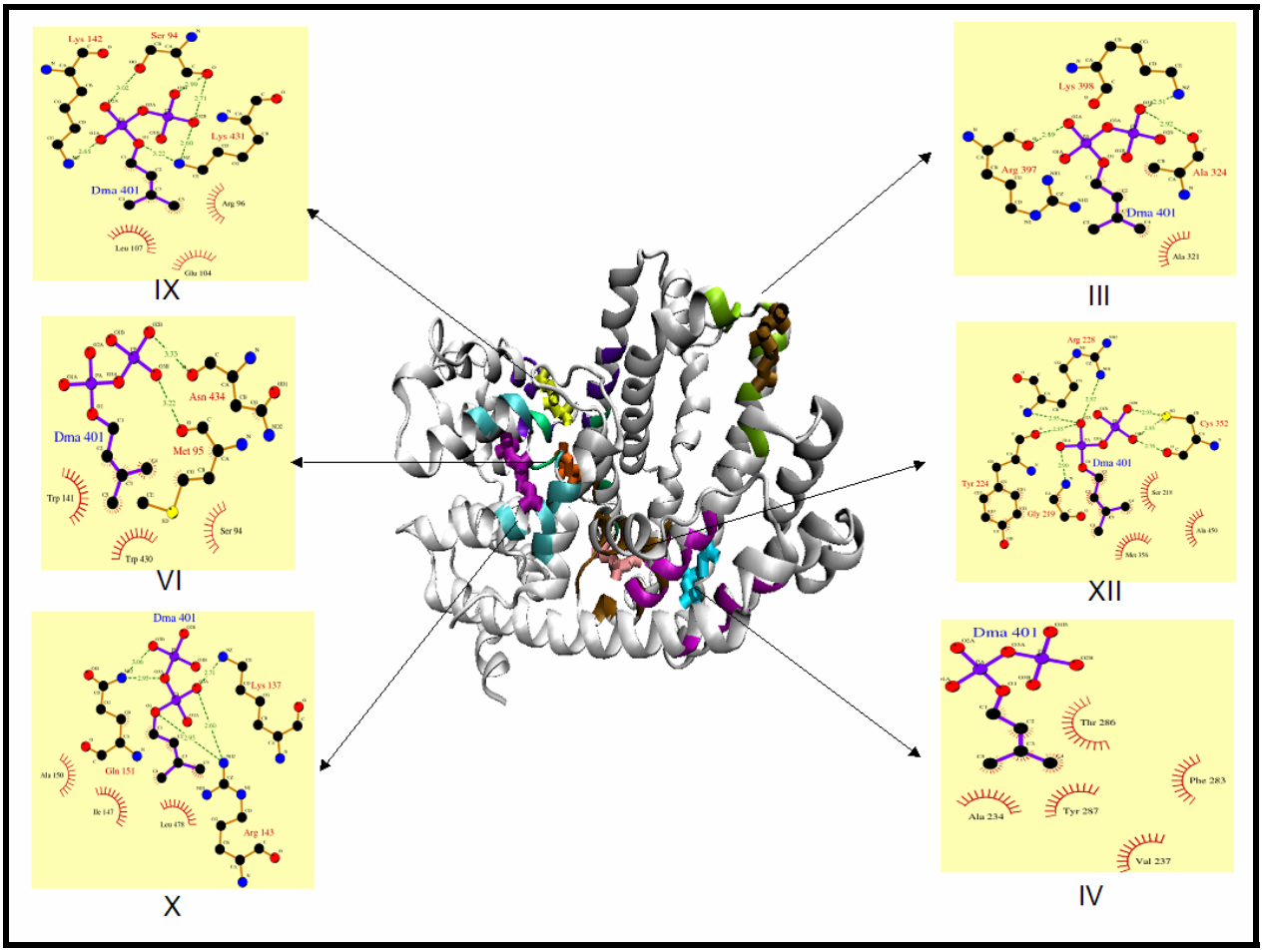

Figure 6: Figure showing the receptor molecule with the ligand docked in six top ranking binding sites. All the six figures in different colour correspond to six binding sites containing the ligand molecule FPP docked into it.

Conclusions:

Amorpha-4,11-diene synthase in *A. annua* L. plants is reported to be a key regulatory enzyme, catalyzing the rate-limiting step in the biosynthesis of artemisinin. In this study, we conclude that the putative amorpha–4,11-diene synthase gene (ADS3963) cloned by us from high artemisinin (0.7-0.9 % dry wt basis) yielding strain of *A. annua* L. plant is evolutionary conserved, as suggested by phylogenetic analysis. It encodes a protein of 533 amino acid residues with conserved domain DDxxD. The absence of thirteen amino acids in this protein has resulted into the formation of a new motif, SlwD, which might have role in regulating the active state of the enzyme. This finding of our study also indicates that ADS3963 protein may be an isoform. These results, however, has to be corroborated with wet lab data. Further, if the structure and function of ADS3963 protein is understood, it can serve as model to study other related enzymes and could be utilized to develop new functional approaches for over-expression of this key enzyme leading to enhanced synthesis and accumulation of artemisinin in *A. annua* L. plants.

Acknowledgement:

P.A. thanks to Jamia Hamdard for financial support in the form of a studentship. We are highly thankful to Andrew M. Lynn Director, School of Information Technology, Jawaharlal Nehru, University, New Delhi, India for providing me docking facilities in their lab. We thank Ipca Laboratories Limited, Ratlam, India for providing the seeds of high yielding strain of *A. annua* L.

References:

[1] Abdin *et al*., *Planta Med* 2003; **69:** 289. PMID: 12709893
[2] Akhila *et al*., *Phytochemistry* 1990; **29:** 2129
[3] Bharel *et al*., *J Nat Prod* 1998; **61:** 633 PMID: 9599263
[4] Bhattacharya *et al*., *Mendeleev Communications* 2007;**1:** 27
[5] Bouwmeester *et al*., *Phytochemistry* 1999; **52:** 843. PMID: 10626375
[6] Bowman *et al*., *Proc Natl Acad Sci* 1990; **8:** 9052.
[7] Brown GD. *J Nat Prod* 1993; **55:** 1756
[8] Chakrabarti *et al*., *Proc. Natl. Acad. Sci.* USA. 2005:**102:** 10153 PMID: 15998733
[9] Chang *et al*, *Arch Biochem Biophys* 2000:**383,** 178. PMID: 11185551
[10] Chen *et al*., *Nucleic Acids Res* 2006; **34:**W152.
[11] Delabays *et al*., *Acta Hort* 1993; **330:** 203.
[12] Dhingra *et al*., *Life Sci* 2000; **66:** 279. PMID: 10665980
[13] Khan *et al*., *Afr J Biotechnol* 2007; **6:** 175
[14] Kim NC, Kim SU. *J Kor Agric Chem Soc Rev* 1992; **35:** 106.
[15] Kuntz *et al*., *J Mol Biol* 1982; **161:** 269
[16] Laughlin JC. *Trans Royal Soc Trop Med Hyg* 1994; **88**: 21. PMID: 8053017
[17] Liu *et al*., *J Integr Plant Biol* 2006; **48:** 1486.
[18] Martin *et al*., *Nat. Biotechnol.* 2003:**21,** 796. PMID: 12778056
[19] McGarvey DJ, Croteau R. *Plant Cell* 1995; **7:** 1015.
[20] Mercke *et al*., *Arch Biochem Biophys* 2000; **381:** 173. PMID: 11032404
[21] Posner *et al*., *Med Chem* 1995; **38:** 607.

[22] Ravindranathan *et al*., *Tetrahedron Lett* 1990; **31:** 755
[23] Ro *et al*., *Nature* 2007; **440:** 940. PMID: 16612385
[24] Sangwan *et al., Phytochemistry* 1993; **34:** 1301
[25] Starks *et al*., *Science*1997; **277:** 1815.
[26] Tamura *et al*., *Mol Biol Evol* 2007; **24:** 1596.
[27] Tamura *et al*., *Proc Natl Acad Sci* (USA) 2004; **101:** 11030.
[28] Trapp SC, Croteau RB.*et al*., *Genetics* 2001; **158:** 811.
[29] Van *et al*., *Trends Pharmacol Sci* 1999; **20:** 199
[30] Wallaart *et al*., *Planta* 2001; **212:** 460. PMID: 1128961
[31] Willard *et al*., *Nucleic Acids Res* 2003; **31:** 3316.
[32] http://www.expasy.ch/tools/scanprosite
[33] http://ps2.life.nctu.edu.tw/
[34] http://www.expasy.ch/spdbv/
[35] http://www.redpoll.pharmacy.ualberta.ca/vadar
[36] http://www.ebi.ac.uk/Tools/clustalw2/index.html

Edited by P. Kangueane
Citation: Alam *et al.,* Bioinformation 4(9): 421-429 (2010)

Supplementary material:

Table 1: Domain position and description

Site	Position	Domain	Function
Casein kinase II phosphorylation site	2 –5 93 – 96 **104 – 107*** 262 – 265 283 – 286 294 – 297 311 – 314 315 – 318 330 – 333 332 – 335 375 – 378 446 – 449	SltE TygD **SlwD*** SgfE TliD TyeE TclD TlpE TytE TemE TteE SslE	Casein kinase II (CK-2) is a multifunctional protein serine/threonine kinase that has been implicated in a variety of cellular processes and functions, including mitosis and cellular transformation. CKII is highly conserved among eukaryotes, including yeast, *Dictyostelium discoideum*, *C. elegans*, *Drosophila*, plants, bovines and humans.
Protein kinase C phosphorylation site	138 – 140 434 – 436 483 – 485	SmR ThK TtK	Protein kinase C is actually a family of protein kinases. In vivo, protein kinase C exhibits a preference for the phosphorylation of serine or threonine residues found close to a C-terminal basic residue. PKC is involved in receptor desensitization, in modulating membrane structure events, in regulating transcription, in mediating immune responses,in regulating cell growth, and in learning and memory.These functions are achieved by PKC mediated phosphorylation of other proteins.
N-glycosylation site	207 – 210 511 – 514	NKTL NFTR	Potential N-glycosylation sites are specific to the consensus sequence Asn-Xaa-Ser/Thr. The presence of the consensus tri-peptide is not sufficient to conclude that an Asn residue is glycosylated, due to the folding of the protein which plays an important role in the regulation of N-glycosylation. The presence of proline between Asn and Ser/Thr will inhibit N-glycosylation.
N-myristoylation site	259 – 264	GLgsGF	Eukaryotic proteins are acylated by the covalent addition of myristate (a C14-saturated fatty acid) to their N-terminal residue via an amide linkage. The sequence specificity of the enzyme responsible for this modification, myristoyl CoA:protein N-myristoyl transferase (NMT), has been derived from the sequence of known N-myristoylated proteins.
Amidation site	425 – 428	lGGR	The precursor of hormones and other active peptides which are C-terminally amidated is always directly followed by a glycine residue which provides the amide group, and most often by at least two consecutive basic residues (Arg or Lys) which generally function as an active peptide precursor cleavage site. Although all amino acids can be amidated, neutral hydrophobic residues such as Val or Phe are good substrates, while charged residues such as Asp or Arg are much less reactive.
cAMP- and cGMP- dependent protein kinase	441 – 444	Rkhs	Both type of kinases share an inclination to phosphorylate serine or threonine residues found close to at least two consecutive N-terminal basic residues.
Aspartate rich domain	286-290	DDxxD	Cordination of the substrate bound metal ion (Mg^{2+} and Mn^{2+})

Table 2: Statistical values of ADS$_{3963}$ protein of *Artemisia annua* L.

Statistical	Observed	Expected
Expected values for highly refined X-ray and NMR protein structures		
Helix	356 (72 %)	-
Beta	3 (0 %)	-
Coil	130 (26 %)	-
Turn	116 (23 %)	-
Hydrogen bonds (H-bonds)		
Mean bond distance	2.2 ± 0.3 Å	2.2 ± 0.4 Å
Mean bond energy	-1.5 ± 1.0	-2.0 ± 0.8
Residues with H-bonds	430 (87 %)	366 (75 %)
Dihedral angles		
Mean Helix Phi	-64.2 ± 5.1	-65.3 ± 11.9

Mean Helix Psi	-37.5 ±19.1	-39.4 ±25.5
Residues with Gauche + Chi	186 (44%)	231 (55%)
Residues with Gauche- Chi	62 (14%)	84 (20%)
Mean Chi Gauche+	-66.0 ± 6.9	-66.7 ± 15.0
Mean Chi Gauche-	64.1 ± 7.7	64.1 ± 15.7
Mean Chi Trans	172.9 ± 6.7	168.6 ± 16.8
Standard deviation of chi pooled	6.94	15.70
Mean Omega (|ω |>90)	-177.6 ± 3.3	180.0 ± 5.8
Residues with | ω |<90	1 (0%)	-
3D profile quality index		
Resolution		-
R-value		-
Residues in phi-psi core	456 (93%)	440 (90%)
Residues in phi-psi allowed	30 (6%)	34 (7%)
Residues in phi-psi generous	1 (0%)	5 (1%)
Residues in phi-psi outside	0 (0%)	0 (0%)
Residues in omega core	471 (96%)	469 (96%)
Residues in omega allowed	17 (3%)	15 (3%)
Residues in omega generous	0 (0%)	0 (0%)
Residues in omega outside	0 (0%)	5 (1%)
Packing defects	85	34
Free energy of folding	-468.12	-468.09
Residues 95% buried	113	203
Buried charges	8	0

Table 3: Six top ranking binding sites of with ADS_{3963} with lowest energy of interaction between protein and ligand molecules

Binding Sites on ADS_{3963}	Amino acid lining the cavity	Energy Value
3	YDAANGRLH	-2.55
4	YDAANGALH	-3.22
6	SMWLWINWLLI	-2.40
9	TSMRLGDALTVGWLN	-4.15
10	RLGRIEANHLSL	-4.77
12	GLSGFGWSARFTCYMPMAY	-3.04