Annals of Phytomedicine 7(1): 69-77, 2018
DOI: 10.21276/ap.2018.7.1.8; Print ISSN : 2278-9839 and Online ISSN : 2393-9885
Ann. Phytomed., 7(1):69-77 (2018)

Original article

In silico structural and functional analysis of copalyl diphosphate synthase enzyme in *Andrographis paniculata* (Burm. f.) Wall. ex Nees: A plant of immense pharmaceutical value

Aayeti Shailaja, Byreddi Bhavani Venkata Bindu, Mote Srinath and Charu Chandra Giri
Centre for Plant Molecular Biology (CPMB), Osmania University, Hyderabad-500007, Telangana State, India

Received April 13, 2018: Revised May 30, 2018: Accepted June 5, 2018: Published online June 30, 2018

Abstract

Andrographis paniculata (Burm. f.) Wall. ex Nees (Acanthaceae) with immense medicinal importance lacks information on its biosynthetic pathway genes and their regulatory role in the production of pharmaceutically important andrographolide. Copalyl diphosphate synthase (CPS) is involved in the production of copalyl diphosphate, a precursor for many bioactive compounds with particular reference to diterpene lactone. In this study, we elucidated the structural and functional aspects of *A. paniculata* CPS (*Ap*CPS). Composition of amino acids and hydrophobic nature of *Ap*CPS were analysed and identified as non trans-membrane protein. A chloroplast transit peptide and mitochondrial targeting peptide in *Ap*CPS were identified. Protein secondary structure prediction has given insight on the distribution of helix (52.52%), loop (45.91%) and strands (1.56%) in *Ap*CPS. The homology modelling of *Ap*CPS was carried out with SWISS MODEL. The validation of 3D model using PROCHECK revealed that 91.74% of the residues have averaged 3D-1D score >= 0.2 which is structurally reliable. In Ramachandran plot, 90.9% amino acid residues were found in most favoured region. Phylogenetic tree was constructed using MEGA 7.0 by taking eudicots, monocots, gymnosperms and fungal species. Among them, *Ap*CPS was clustered within eudicots and closely related to *Sesumum indicum* in Laminales. Protein-protein interaction study using STRING10 revealed that CPS interacts with gibberilic acid and terpene synthase related proteins. In *Arabidopsis thaliana*, CPS coexpression was seen with gibberelic acid related proteins. The present *in silico* analysis will be useful in understanding the structural, functional and evolutionary diversification of *Ap*CPS.

Key words: *Andrographis paniculata* (Burm. f.) Wall. ex Nees, *Ap*CPS protein, motifs and domains, domain linkers, 3D modelling, phylogenetic analysis

1. Introduction

Andrographis paniculata (Burm. f.) Wall. ex Nees (Acanthaceae) is an important medicinal herb and a valuable source for important diterpene lactone, andrographolide and its derivatives. It has immense effect on various diseases and considered to be a valuable source in medicine. It has pharmacological effects such as antimicrobial (Singha *et al*., 2003), anti-inflammatory, anti-cancerous and immuno-stimulatory (Kumar *et al.*, 2004; Subramanian *et al*., 2012; Islam *et al*., 2018), immuno-modulatory and anti-atherosclerotic (Chao and Lin, 2010). This plant has shown effect on suppression of esophageal cancer and metastasis (Li *et al.*, 2018). The demand for such valuable compound diterpene lactones of this plant is very high. However, the detailed mechanisms and the biosynthetic pathway genes are not yet elucidated clearly (Singh *et al.*, 2018).

All the secondary metabolites (specifically diterpenes) of the plants have a common origin from IPP and DMAPP (Figure 1). These can be derived either from MEP or MVA pathway which are interlinked and have connection between them (Chen *et al.*, 2011; Vranová *et al.*, 2013).The enzyme copalyl diphosphate synthase (CPS) catalyzes conversion of geranyl geranyl diphosphate, to copalyl diphosphate (CPP) which serves as intermediate for all diterpenoid reactions (Beale, 1990; Su *et al.*, 2016).This CPS belongs to isomerase super family which involves in the synthesis of terpenoids/isoprenoids. CPP is the direct precursor of gibberellic acid synthesis, other phytoalexins and labdane-related diterpenoids in plants (Prisic *et al.*, 2004; Harris *et al.*, 2005).

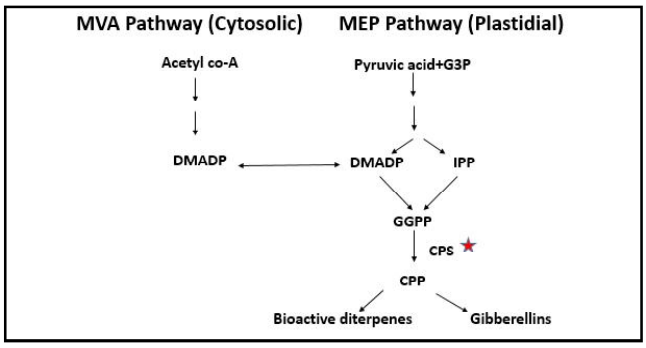


Figure 1: Biosynthetic MEP and MVA pathway showing CPS enzyme for the biosynthesis of diterpenoids.

Author for correspondence: Dr. Charu Chandra Giri
Professor, Centre for Plant Molecular Biology (CPMB), Osmania University, Hyderabad-500007, Telangana State, India
E-mail: giriccin@yahoo.co.in
Tel.: +91-040-27098087

The genomic and metabolite studies when combined with various bioinformatic analysis brings out the unknown diterpenoid scaffolds and the enzyme information (Andersen-Ranberg *et al.,* 2016). In our laboratory, we have a comprehensive research programme on *A.paniculata* distribution, enhanced production of bioactive compound andrographolide and study of key enzymes/proteins (Neeraja *et al*., 2015; Parlapally *et al*., 2015; Zaheer and Giri, 2015; Zaheer *et al.,* 2017a; Zaheer *et al.,* 2017b; ; Bindu *et al.,* 2017; Srinath *et al.,* 2017). In the present communication, *in silico* studies such as structural, functional features and the evolutionary relationship were elucidated using various bioinformatics tools.

2. Materials and Methods

2.1 Analysis of biochemical properties in *Ap*CPS protein

Amino acid sequence of *A. paniculata* ent-copalyl diphosphate synthase (*Ap*CPS) was obtained from NCBI with accession number AEM00024.1. Expasy Compute pI/Mw tool is used for the estimation of pI (isoelectric point) and Mw (molecular weight) for the given protein sequence of *Ap*CPS (Kyte and Doolittle, 1982). For the visualization of hydrophobicity for a peptide sequence the hydropathy plots were developed using Kyte and Doolittle (1982) method for each amino acid. The amino acid composition was shown (in %) using ProtParam tool in Expasy online server (Walker, 2005).

2.2 Prediction of signal peptide sequence

ChloroP was used for finding the choloplast transit peptide (cTP) and TargetP (Emmanuelsson *et al.,* 2000), iPSORT (Bannai *et al*., 2001, 2002) used to find out the subcellular location and signal peptide sequences in *Ap*CPS.

2.3 Elucidation of secondary structure of *Ap*CPS and prediction of trans-membrane helices

The secondary structure of *Ap*CPS representing the families of related proteins was characterised using PredictProtein tool. Solvent accessibility and trans-membrane helix prediction was done by this tool. The analysis of trans-membrane helices in *Ap*CPS was done using HMMTOP (Hidden and Markov Model Topology of Proteins) as per Tusnady and Simon (2001) and Tied Mixture Hidden Markov Model (TMHMM). The secondary structure prediction of *Ap*CPS was carried out using CFFSP prediction server (Chou and Fasman 1974; Dor *et al.,* 2006).

2.4 Domain and motif analysis of *Ap*CPS

The prediction of domain in *Ap*CPS was carried out using Interpro and NCBI-CD search also used for identifying the super family of *Ap*CPS (Marchler-Bauer *et al.,* 2016).

The Motif analysis was done using multiple Em for motif elicitation (MEME: version 4.10.2). Motif Alignment and Search Tool was used for computation of pairwise correlation between each pair of motifs. Motifs with correlations below 0.60 have little effect on the accuracy of the E-values computed by MAST. DLP-SVM Domain prediction tool was used for identifying the Domain linkers (Ebina *et al.,* 2009).

2.5 Tertiary structure prediction / 3D modelling and validation of *Ap*CPS

The tertiary structure (3D) modelling of *Ap*CPS enzyme structure was predicted using SWISS MODEL server (Kiefer *et al.,* 2008) and Phyre2 tool (Kelley *et al.,* 2015). PROCHECK analysis (Laskowski *et al.,* 1996) was carried out for the validation and for analysing the stereochemical reliability of the 3D model using Ramachandran plot. The ligand binding site in the 3D model was predicted using 3D Lig and Site tool (Wass *et al.,* 2010). The 3D model generated by Phyre2 was used for generating ligand binding site.

2.6 Phylogenetic analysis, protein interactions and expression studies

MEGA 7 software was used for phylogenetic tree construction and evolutionary relationship analysis of *Ap*CPS with other CPS enzymes (dicots-33, monocots-4, gymnosprerms-2 and fungi-2) as per Kumar *et al.* (2016). The evolutionary history was inferred using UPGMA method. The optimal tree with the sum of branch length = 6.99757378 is shown next to the branches. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 623 positions in the final dataset, which was employed for construction of phylogenetic tree.

2.6.1 Protein protein interaction and co-expression analysis

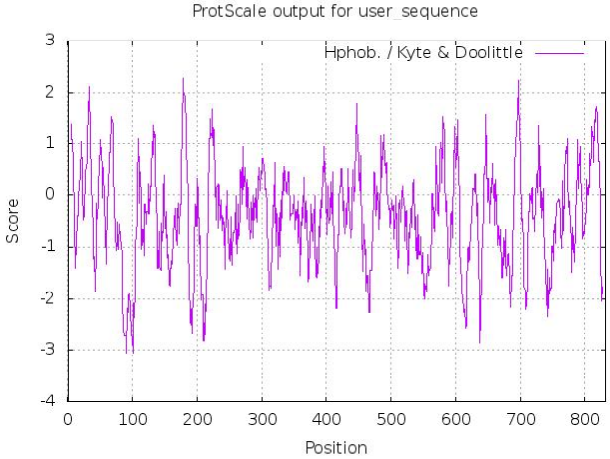
STRING10 tool was used for analysing the retrieval of interacting proteins or genes (Szklarczyk *et al.,* 2015). The model plant *Arabidopsis thaliana* CPS was taken as reference protein sequence for the co expression study of CPS.

3. Results and Discussion

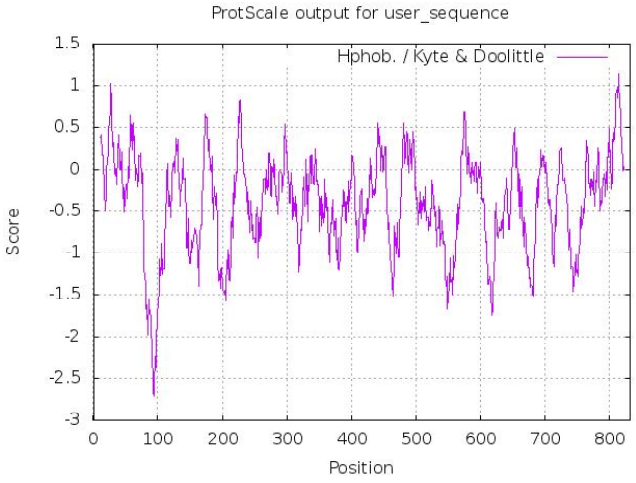
3.1 Biochemical property analysis of *Ap*CPS protein

Amino acid sequence analysis in *A. paniculata* using Compute pI/ Mw tool revealed that *Ap*CPS contained 832 amino acid residues with approximately 2.5 kb size (2496 bp). The theoretical pI value was shown as 7.07 and molecular weight was 95370.83 daltons, *i.e*., 95 Kda. The similar result obtained earlier in *Ap*CPS strengthened our finding in the present study (Garg *et al.,* 2015).

Prot Scale helps in computing and representing the profile produced by any amino acid scale on a selected protein. The hydrophobicity plots for *Ap*CPS with a window size of 9 and 21 shown in Figure 2A and 2B, where the relative weight of the window edges compared to the window center is 100%. The linear weight variation model is used for development of hydropathy plots without normalization. The more positive value of amino acids indicated that isoleucine, valine and leucine are highly hydrophobic at that region (Figure 2C).



(A)



(B)

```
Ala:  1.800  Arg: -4.500  Asn: -3.500  Asp: -3.500  Cys:  2.500  Gln: -3.500
Glu: -3.500  Gly: -0.400  His: -3.200  Ile:  4.500  Leu:  3.800  Lys: -3.900
Met:  1.900  Phe:  2.800  Pro: -1.600  Ser: -0.800  Thr: -0.700  Trp: -0.900
Tyr: -1.300  Val:  4.200  : -3.500  : -3.500  : -0.490
```

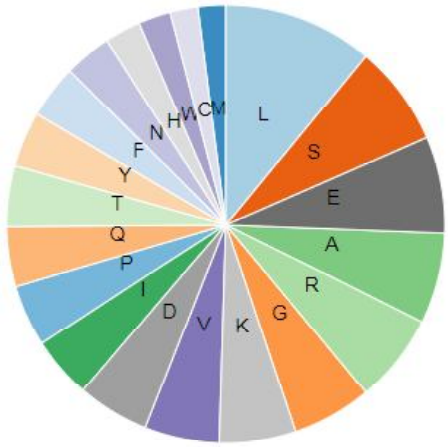
(C)

Figure 2: Hydropathy plot of *Ap*CPS protein showing peaks for each amino acid at window size 9 (A); at window size 21 (B); The individual values for 20 amino acids using the scale Hydrophob./Kyte and Doolittle (C).

The amino acid composition and distribution showed that a highest of 11.1% residues of leucine are present in protein which is highly hydrophobic. After leucine, serine (7.5%) and glutamic acid (7.1%) were having major part of the composition (Figure 3A and 3B). The total number of 103 negatively charged residues of (Asp+Glu) and 102 positively charged residues of (Arg + Lys) were present.

Amino acid composition		
Ala (A)	56	6.7%
Arg (R)	55	6.6%
Asn (N)	29	3.5%
Asp (D)	44	5.3%
Cys (C)	17	2.0%
Gln (Q)	37	4.4%
Glu (E)	59	7.1%
Gly (G)	49	5.9%
His (H)	22	2.6%
Ile (I)	38	4.6%
Leu (L)	92	11.1%
Lys (K)	47	5.6%
Met (M)	17	2.0%
Phe (F)	33	4.0%
Pro (P)	38	4.6%
Ser (S)	62	7.5%
Thr (T)	37	4.4%
Trp (W)	20	2.4%
Tyr (Y)	34	4.1%
Val (V)	46	5.5%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

(A)



(B)

Figure 3: The amino acid composition of *Ap*CPS using ProtParam tool of ExPASy server (3A); diagrammatic representation of amino acid composition (3B) Lys (K); Ser (S); Thr (T); Ile (I); Glu (E); Pro (P); Arg (R); Gln (Q); Phe (F); Tyr (Y) His (H); Asp (N); Met (M); Cys (C); Val (V); Ala (A); Gly (G); Leu (L); Asp (D); Trp (W).

3.2 Prediction of signal peptide sequence

A chloroplast transit peptide (cTP) sequence containing 27 amino acids was identified in *Ap*CPS. iPSORT prediction indicated that the protein sequence has no N-terminal sorting signal but having a mitochondrial targeting peptides in it (Figure 4). In an earlier study the presence of chloroplast transit peptide and mitochondrial targeting peptide in CPS was also observed in *Salvia miltirriza* (Su *et al.*, 2016).

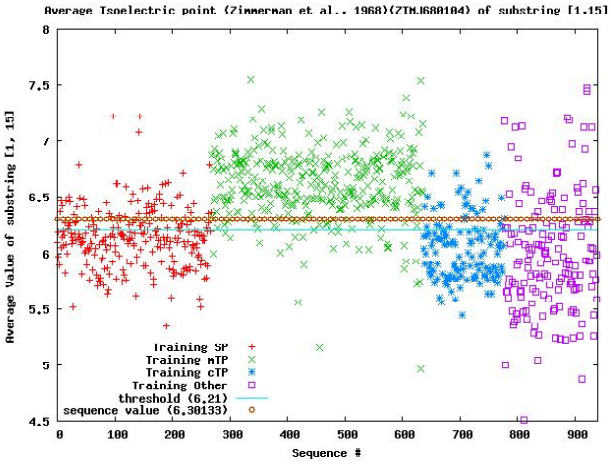


Figure 4: Prediction of mitochondrial targeting peptide using iPSORT.

3.3 *Ap*CPS secondary structure elucidation and prediction of trans-membrane helices

Secondary structure elements and solvent accessibility was predicted using evolutionary information from multiple sequence alignments and a multi-level system (Rost and Sander, 1993). Three states of secondary structure were predicted such as helix (H; includes alpha-, pi- and 3_10-helix), (beta-) strand (E = extended strand in beta-sheet conformation of at least two residues length)

and loop (L). Secondary structure is predicted by a system of neural networks with an expected average accuracy of more than 72% (Rost and Sander, 1994).

PROFsec method in PredictProtein tool was helpful in finding the percentage of amino acid residues covered by helix, loops and strands in the protein. Secondary structure analysis revealed that 611 amino acids are possible for α-helix, 369 residues for extended sheets and 110 residues for turns (Figure 5). About 52.52% of amino acid sequence was covered with helix, 45.91% with loop and 1.56% region was covered with strand (Figure 6). This shows that α-helices and extended sheets were abundant structures in *Ap*CPS. The coils and turns were found to be distributed intermittently in the protein. There were no disulphide bridges and trans-membrane helices found in the *Ap*CPS amino acid sequence.

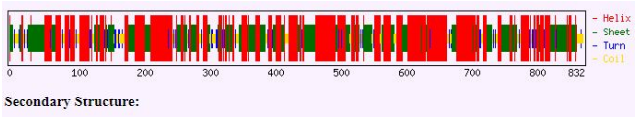


Figure 5: Secondary structure of *Ap*CPS showing helix, sheet, turn and coil.

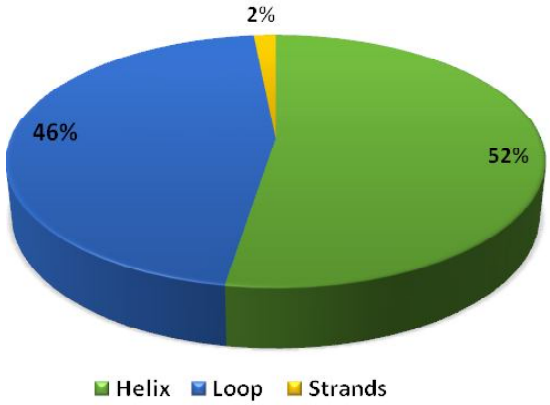


Figure 6: Graphical representation of secondary structure of *Ap*CPS.

Further, analysis of protein revealed, a total of 10 protein binding regions were present in *Ap*CPS (Figure 7). These 10 protein binding regions were at amino acid positions and residues 91(1), 139-141 (3), 189 (1), 213(1), 232 (1), 232(1), 455 (1), 469 (1), 747 (1), 827 (1), 832 (1), respectively (shown in red color: Figure 7). The helices present in the amino acid sequence were shown as brown colored boxes. Amino acid residues were categorized into buried (yellow), intermediate (light colored) and exposed portions (blue) along the sequence. The buried and exposed amino acids covered most of the amino acid composition. There were 4 disorder regions found in the sequence at positions 1(1), 56 (1), 82-99 (18), 744-747 (4) respectively, where the length/size of residues is given in the parenthesis.

There were no trans-membrane helices and surface globular proteins observed in *Ap*CPS based on the results of HMMTOP and TMHMM. Hence, *Ap*CPS is considered as non-trans-membrane protein which is being translocated from nucleus to mitochondria and chloroplast. Although, surface globular proteins were not identified in the present study with *Ap*CPS but these tools are also used in identifying the surface globular proteins in other plant species.

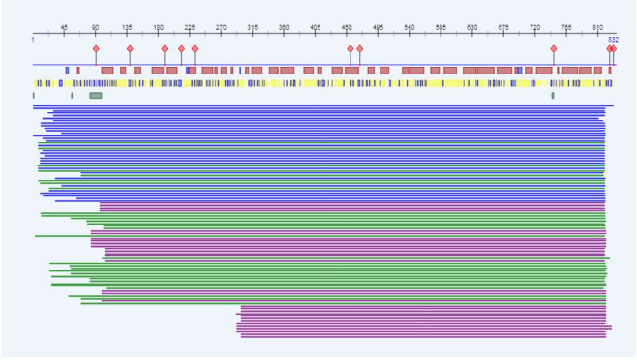


Figure 7: Protein binding, disorder and other regions present in *Ap*CPS.

3.4 Claasification of *Ap*CPS protein and prediction of domains and motifs

*Ap*CPS protein belongs to isoprenoid synthase superfamily (Figure 8). The results with NCBI conserved domain search also revealed that *Ap*CPS belongs to the isoprenoid superfamily and showed a specific hit with terpene synthase (Figure 9).Two domains, one terpene synthase N- terminal domain at 284-489 and one metal binding domain at 533-681 were identified (Figure 8).

Figure 8: Prediction of domains by Interpro

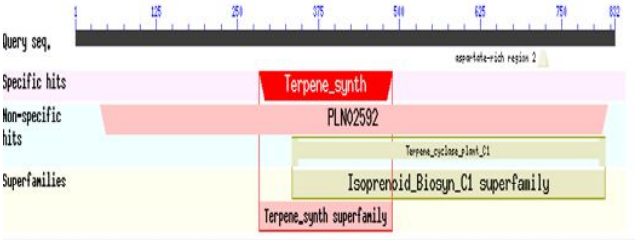


Figure 9: NCBI-CD search of *Ap*CPS showing its superfamily

3.4.1 Predicting domain linkers by loop-length-dependent support vector machine (DLP-SVM) tool

The prediction of structural domains in novel protein sequences is of practical importance. DLP-SVM is a loop-length-dependent support vector machine (SVM) for prediction of domain linkers, which are loops separating two structural domains. A longer domain linker (PQLYIPAASPFPRTSVVAG) was identified at position 15-33 and a short domain linker (GSSPSPPPQ) at position 8-16 (Figure 10). The longer linkers are used when it is necessary to ensure that two adjacent domains do not sterically interfere with one another (Ebina *et al.*, 2009).

The domain linkers were found to be rich in proline residues which contributes for structural confirmation, and flexibility of two domains. They also acted in the prevention of unfavourable communication between the domains. Further, depending on length of the linkers the interaction between the domains varied (Bhaskara *et al.*, 2013). This study is also helpful in protein targeted drug development and other proteomic studies (Shatnawi *et al.*, 2014).

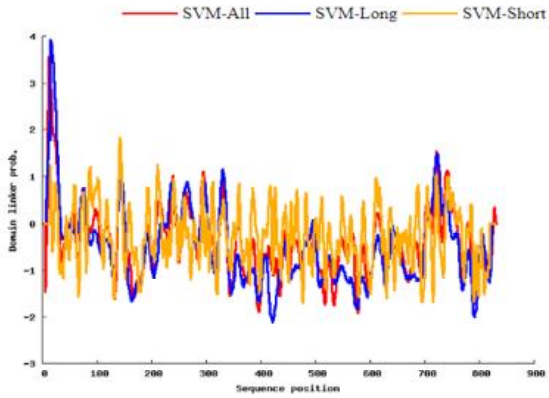


Figure 10: Domain linkers (short and long) identified in *Ap*CPS.

3.4.2 Motif analysis of *Ap*CPS

There were 3 motifs found using MEME with Mast Alignment and Search Tool (Figure 11). The detail analysis of three motifs was characterized and the repetition in the *Ap*CPS sequence detected Figure 12. Each of the sequence has an e value less than 10. The motif YIPAASPF occurred twice at positions 18-25 and 674-689 in the sequence. Motif MHRDWTDKGICW also repeated twice at positions 189-200 and 369-380. The third motif WQKWLRSW occurred at positions 554-561and 674-689 along the sequence.

	Logo	Name	Alt. Name	Width	Motif Similarity Matrix 1.	2.	3.
1.		MHRDWTDKGICW	MEME-1	12	--	0.19	0.19
2.		WQKWLRSW	MEME-2	8	0.19	--	0.19
3.		YIPAASPF	MEME-3	8	0.19	0.19	--

Figure 11: MEME results showing motifs in the *Ap*CPS.

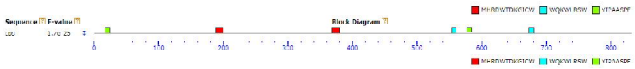


Figure 12: Different motifs and their positions on the sequence of *Ap*CPS (The boxes on the line representing the motifs).

3.5 Elucidation of tertiary structure of *Ap*CPS by 3D homology modelling, validation and ligand binding site prediction

The 3D structure was predicted using Swiss Model which showed 713 amino acid coverage and with 56.81% identity (Figure 13). The elucidated 3D model was structurally validated with SAVES PROCHECK where 91.74% of the residues have averaged 3D-1D score >= 0.2 (Figure 14). This 3D model has shown a overall quality factor (OQF) of 92.1986 (Figure 15).

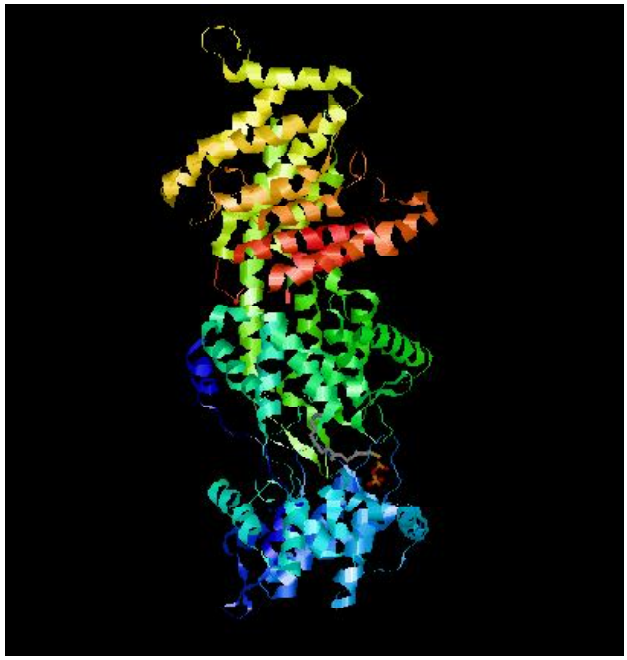


Figure 13: Elucidated 3D structure of *Ap*CPS using Swiss Model.

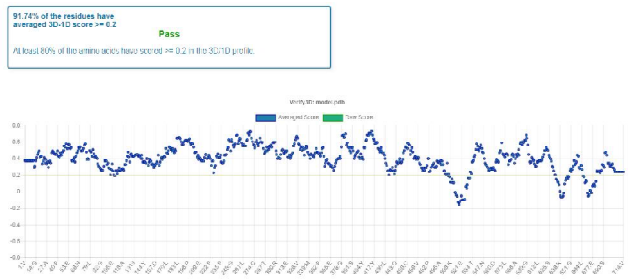


Figure 14: Validation of 3D structure.

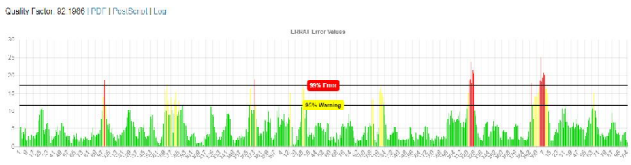


Figure 15: ERRAT result showing the Quality factor for *Ap*CPS.

In Ramachnadran plot, 90.9% amino acid residues were observed in favoured region with 8.4% additional allowed regions (Figure 16). Only 0.2% disallowed regions were found in the plot. This validation studies showing that the predicted 3D model of *Ap*CPS enzyme was structurally reliable. The distribution of main chain bond angles and length were found to be within the limits. Homology modelling and validation of antioxidant proteins were also shown in similar way in Spinach (Sahay and Shakya, 2010). A single ligand binding site was identified in validated 3D structure of *Ap*CPS (Figure 17). It consists of GLY, ASP, VAL residues at (510-528).

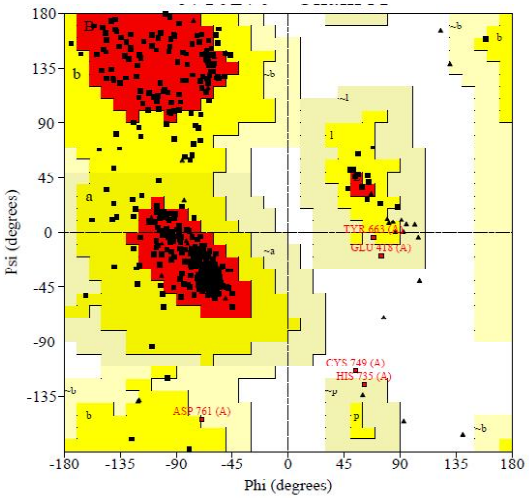


Figure 16: Ramachandran Plot of *Ap*CPS 3D structure. The most favoured regions are represented in red color; additional allowed regions in yellow color.

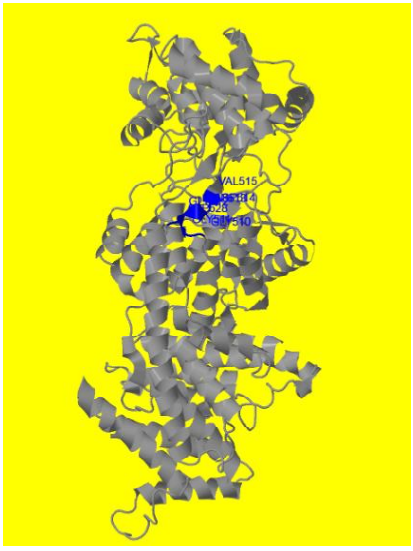


Figure 17: Predicted ligand binding site showed in blue color.

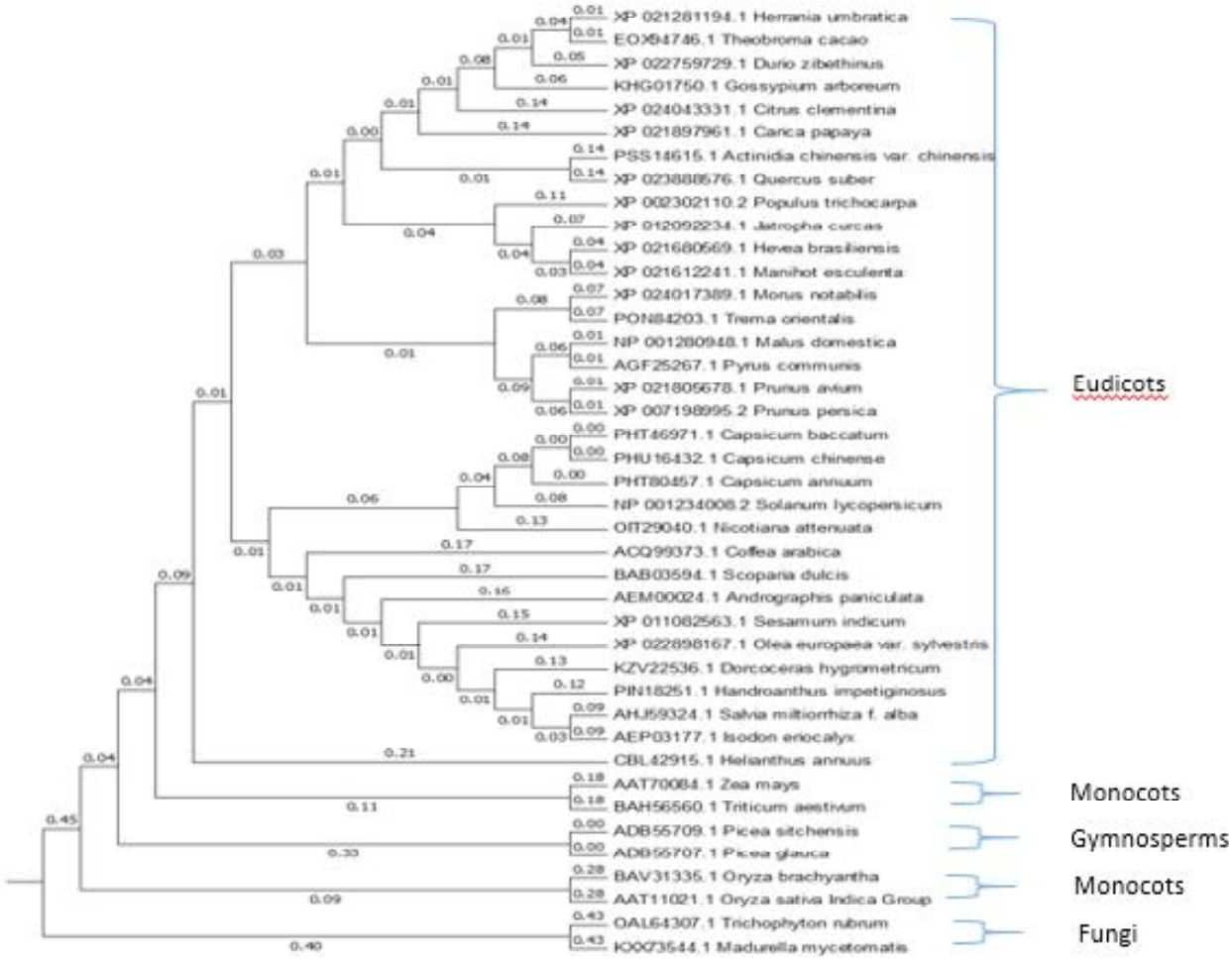


Figure 18: The phylogenetic tree showing the evolutionary relation of *Ap*CPS with other SQS from diverse organisms constructed by neighbour-joining method.

3.6 Phylogenetic analysis of *Ap*CPS

The phylogenetic tree mainly divided into eudicots, monocots, gymnosperms and fungi. Among the 41 sequences taken for the phylogenetic construction, the *Ap*CPS was clustered within the eudicots (Figure 18). *Ap*CPS was shown close relationship with *Sesamum indicum* and *Olea europaea* plants belongs to the same order Lamiales. After Lamiale plants, it was closer to the order Solanales. In an earlier study, comparable result was found with this CPS enzyme by Garg *et al.* (2015).

The evolutionary divergence study of this plant showed close relation to Laminales, followed by Solaneles in *Ap*HMGR and *Ap*DXS enzymes related to the diterpenoid pathways (Bindu *et al.*, 2017; Srinath *et al.*, 2017). After the relation with eudicots, *Ap*CPS was closely linked to monocots particularly *Zea mays* and *Triticum aestivum* and extended to gymnosperms and fungal species.

3.7 Protein interaction study

Arabidopsis thaliana CPS also called as GA1 was used for the protein interaction study (Figure 19). The protein interaction study shown that CPS mainly involved in gibberellic acid mediated signalling pathway. Further, it also involves proteins response to oxygen-containing compound which includes the enzymes such as GA1, RGL1, RGL-2, and GA20OX1, *etc*. The molecular functions such as transcription factor activity and binding included interaction with RGA1, RGL1, RGL2, RGL3 enzymes, iron ion binding included interaction with CYP88A3, KA02. Terpene synthase activity mainly involved the interaction within GA1, GA2, TPS04 enzymes. Most of the interacted proteins shown to be involved in inter cellular and intra cellular membrane bounded organelles. The CPS also shown the relation to proteins involved in diterpene biosynthesis pathway which includes GA1, TPS04, CYP 82G1, GA2, GA3OX1, GA3, GA20OX1, CYP88A3, KAO2 enzymes. These were also involved in plant hormone signal transduction and biosynthesis of secondary metabolites (Figure 20).

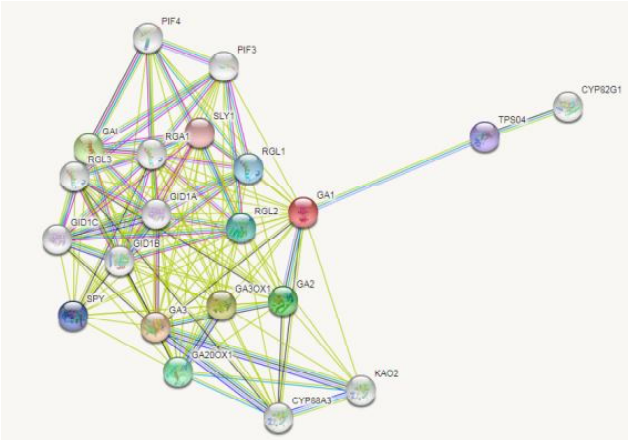


Figure 19: Protein-protein interactions using *Arabidopsis* CPS as reference sequence using STRING 10.

Gene expression and co-expression studies will provide outline about CPS protein where the exchange of molecules and their interaction is high or low (Figure 20). CPS found to be involved in biosynthesis of gibberellin and diterpenoids in various studies (Keeling *et al*., 2010; Yamamura *et al*., 2018). This finding strengthens our result which showed the CPS co-expression mainly with gibberellin and terpenoid synthase biosynthesis genes such as Gibberellin 3-oxidase 1, Gibberellin 20 oxidase 1 and Terpene synthase 04, *etc*.

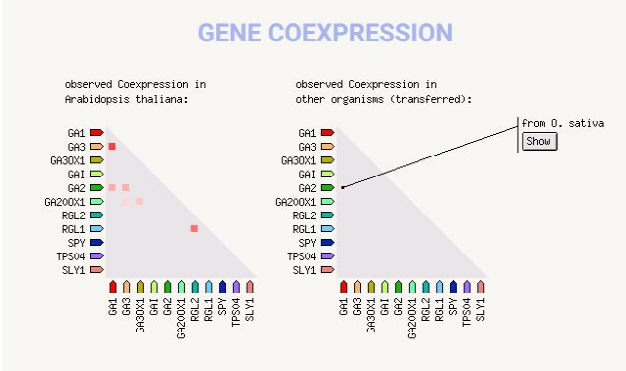


Figure 20: Co-expression analysis of different proteins with reference plant CPS (*Arabidopsis thaliana* CPS). GA1-Gibberillic acid requiring1; GA3 - GA requiring 3; GA3OX1- Gibberellin 3-oxidase 1; GAI- DELLA protein GAI; GA2- GA REQUIRING 2; GA20OX1- Gibberellin 20 oxidase 1; RGL2- RGA-like 2; RGL1- RGA-like 1; SPY- SPINDLY; TPS04- Terpene synthase 04; SLY1- SLEEPY1.

4. Conclusion

The present *in silico* analysis has given structural, functional and physicochemical aspects of protein *Ap*CPS. The secondary structure prediction showed the alpha helices and beta sheets distribution of *Ap*CPS protein folding. The reliable 3D modelling of this enzyme gives information about the protein folding and can be exploited for various docking and drug targeting studies. The protein- protein interaction study will be helpful for understanding the protein at molecular level in various biosynthetic mechanisms. Evolutionary relation will give a scope for molecular biology and gene isolation studies where the primer designing for genes would be easy.

Acknowledgements

We would like to acknowledge the financial support from OU-UGC-CPEPA programme sponsored by University Grants Commission (UGC), New Delhi. Authors also thank OU-UGC-CPEPA programme and UGC-BSR-RFSMS for research fellowships to AS, BBVB, MS.

Conflict of interest

We declare that we have no conflict of interest

References

Andersen Ranberg, J.; Kongstad, K.T.; Nielsen, M.T.; Jensen, N.B.; Pateraki, I.; Bach, S.S. and Møller, B.L. (2016). Expanding the landscape of diterpene structural diversity through stereochemically controlled combinatorial biosynthesis. Angew. Chem. Int. Ed., **55**:2142-2146.

Bannai, H.; Tamada, Y.; Maruyama, O.; Nakai, K. and Miyano, S. (2001). Views: Fundamental building blocks in the process of knowledge discovery. In: Proceedings of the 14th International FLAIRS Conference, pp:233-238.

Bannai, H.; Tamada, Y.; Maruyama, O.; Nakai, K. and Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signals. Bioinformatics, **18**:298-305.

Beale, S.I. (1990). Biosynthesis of the tetrapyrrole pigment precursor, γ-amino levulinic acid, from glutamate. Plant Physiol., **93**:1273-1279.

Bhaskara, R.M.; de Brevern, A.G. and Srinivasan, N. (2013). Understanding the role of domain-domain linkers in the spatial orientation of domains in multi-domain proteins. J. Biomol. Struct. Dyn., **31**:1467-1480.

Bindu, B.B.V.; Srinath, M.; Shailaja, A. and Giri, C.C. (2017). Comparative protein profile studies and *in silico* structural/functional analysis of HMGR (ApHMGR) in *Andrographis paniculata* (Burm. f.) Wall. ex Nees. Ann. Phytomed., **6**: 30-44.

Chao W.W. and Lin B.F. (2010). Isolation and identification of bioactive compounds in *Andrographis paniculata* (Chuanxinlian). Chin Med., **5**:17

Chen, F.; Tholl, D.; Bohlmann, J. and Pichersky, E. (2011). The family of terpene synthases in plants: A mid size family of genes for specialized metabolism that is highly diversified throughout the Kingdom. Plant J., **66**:212-229.

Chou, P.Y. and Fasman, G.D. (1974). Prediction of protein information. Biochem., **13**:222-245.

Dor, O.; Zhou, Y. and Zhou. (2006). Achieving 80% tenfold cross-validated accuracy for secondary structure prediction by large-scale training. Proteins, **66**:838-845.

Ebina, T.; Toh, H.; and Kuroda, Y. (2009). Loop length dependent SVM prediction of domain linkers for high throughput structural proteomics. J. Pept. Sci., **92**:1-8.

Emanuelsson, O.; Nielsen, H.; Brunak, S. and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol., **300**:1005-1016.

Garg, A.; Agrawal, L.; Misra, R.C.; Sharma, S. and Ghosh, S. (2015). *Andrographis paniculata* transcriptome provides molecular insights into tissue-specific accumulation of medicinal diterpenes. BMC Genomics, **16**:659.

Harris, L.J.; Saparno, A.; Johnston, A.; Prisic, S.; Xu, M.; Allard, S. and Peters, R.J. (2005). The maize An2 gene is induced by Fusarium attack and encodes an ent-copalyl diphosphate synthase. Plant Mol Biol., **59**:881-894.

Islam, M.T.; Ali, E.S.; Uddin, S.J.; Islam, M.A.; Shaw, S.; Khan, I. and Gãman, M. A. (2018). Andrographolide, a diterpene lactone from *Andrographis paniculata* and its therapeutic promises in cancer. Cancer Lett., **420**:129-145.

Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N. and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc., **10**:845-858.

Keeling, C.I.; Dullat, H.K.; Yuen, M.; Ralph, S. G.; Jancsik, S. and Bohlmann, J. (2010). Identification and functional characterization of mono-functional ent-copalyl diphosphate and ent-kaurene synthases in white spruce reveal different patterns for diterpene synthase evolution for primary and secondary metabolism in gymnosperms. Plant Physiol., **152**:1197-1208.

Kiefer, F.; Arnold, K.; Künzli, M.; Bordoli, L. and Schwede, T. (2008). The SWISS-MODEL Repository and associated resources. Nucleic Acids Res., **37**:387-392.

Kumar R.A.; Sridevi K.; Kumar N.V.; Nanduri S. and Rajagopal S. (2004). Anticancer and immunostimulatory compounds from *Andrographis paniculata*. J. Ethnopharmacol., **92**: 291-295

Kumar, S.; Stecher, G. and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol., **33**:1870-1874.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol., **1**:105-132.)

Laskowski, R.A.; Rullmannn, J.A.; MacArthur, M.W.; Kaptein, R. and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR, **4**:477-486.

Li, L.; Yue, G.G.L.; Lee, J.K.M.; Wong, E.C.W.; Fung, K.P.; Yu, J. and Chiu, P.W.Y. (2018). Gene expression profiling reveals the plausible mechanisms underlying the antitumor and antimetastasis effects of *Andrographis paniculata* in esophageal cancer. Phytotherapy Res., pp:1-9.

Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S. and Gwadz, M. (2016). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. Nucleic Acids Res., **45**:200-203.

McCaldon, P. and Argos, P. (1988). Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. Proteins: Structure, Function, and Bioinformatics, **4**: 99-122.

Neeraja, C.; Krishna, P.H.; Reddy. C.S.; Giri, C.C.; Rao, K.V. and Reddy, V.D. (2015). Distribution of *Andrographis* species in different districts of Andhra Pradesh. Proc. Natl. Acad. Sci. India, Sect. B. Biol. Sci., **85**:601-606.

Parlapally, S.; Cherukupalli, N.; Bhumireddy, S.R.; Sripadi, P.; Anisetti, R.; Giri, C.C. and Reddy, V.D. (2015). Chemical profiling and antipsoriatic activity of methanolic extract of *Andrographis nallamalayana* J.L. Ellies. Nat. Prod. Res., **30**:1256-1261.

Prisic, S.; Xu, M.; Wilderman, P.R. and Peters, R.J. (2004). Rice contains two disparate ent-copalyl diphosphate synthases with distinct metabolic functions. Plant Physiol., **136**:4228-4236.

Rost, B. and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Boil., **232**:584-599.

Rost, B. and Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. Proteins: Structure, Function, and Bioinformatics, **19**:55-72.

Sahay, A. and Shakya, M. (2010). *In silico* analysis and homology modelling of antioxidant proteins of spinach. J. Proteomics Bioinform., **3**:148-154.

Shatnawi, M.; Zaki, N.and Yoo, P.D. (2014). Protein inter-domain linker prediction using random forest and amino acid physiochemical properties. BMC bioinformatics, **15**:S8.

Singh, S.; Pandey, P.; Ghosh, S. and Banerjee, S. (2018). Anti-cancer labdane diterpenoids from adventitious roots of *Andrographis paniculata*: augmentation of production prospect endowed with pathway gene expression. Protoplasma, pp:1-14.

Singha, P. K.; Roy, S. and Dey, S. (2003). Antimicrobial activity of *Andrographis paniculata*. Fitoterapia, **74**:692-694.

Srinath, M.; Shailaja, A.; Bindu, B.B.V. and Giri, C.C. (2017). Characterization of 1-deoxy-D-xylulose 5-phosphate synthase (DXS) protein in *Andrographis paniculata* (Burm. f.) Wall. ex. Nees: A *in silico* appraisal. Ann. Phytomed., **6**:63-73.

Su, P.; Tong, Y.; Cheng, Q.; Hu, Y.; Zhang, M.; Yang, J and Huang, L. (2016). Functional characterization of ent-copalyl diphosphate synthase, kaurene synthase and kaurene oxidase in the *Salvia miltiorrhiza* gibberellin biosynthetic pathway. Sci. Rep., **6**:230-257.

Subramanian R.; Asmawi M.Z. and Sadikun A. (2012). A bitter plant with a sweet future? A comprehensive review of an oriental medicinal plant: *Andrographis paniculata*. Phytochem. Rev., **11**:39-75

Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta- Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K.P.; Kuhn, M.; Bork, P.; Jensen, L.J. and von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res., **43**:447-452.

Tusnády, G.E. and Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: Applications to topology prediction. J. Mol. Biol., **283**:489-506.

Vranová, E.; Coman, D.; and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu. Rev. Plant Biol., **64**:665-700.

Walker, J.M. (Ed.). (2005). The proteomics protocols handbook. Humana Press.

Wass, M.N.; Kelley, L.A. and Sternberg, M.J. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res., **38**:469-73.

Yamamura, Y.; Taguchi, Y.; Ichitani, K.; Umebara, I.; Ohshita, A.; Kurosaki, F. and Lee, J.B. (2018). Characterization of ent-kaurene synthase and kaurene oxidase involved in gibberellin biosynthesis from Scopariadulcis. J. Nat. Med., pp:1-8.

Zaheer, M. and Giri, C.C. (2015). Multiple shoot induction and jasmonic versus salicylic acid driven elicitation for enhanced andrographolide production in *Andrographis paniculata*. Plant Cell Tissue Organ Cult., **122**:553-563.

Zaheer, M. and Giri, C.C. (2017a). Enhanced diterpene lactone (andrographolide) production from elicited adventitious root cultures of *Andrographis paniculata*. Res. Chem. Intermed., **43**:2433–2444.

Zaheer, M. and Giri, C.C. (2017b). Influence of cotyledon, hypocotyl extracts and authentic andrographolide on selective *Agrobacterium rhizogenes* strains growth: A deterrent to hairy root induction in *Andrographis paniculata* (Burm.f.) Wall. ex Nees. Ann. Phytomed., **6**:51-56.