In silico analysis and homology modeling of strictosidine synthase involved in alkaloid biosynthesis in *catharanthus roseus*

Archna Sahay¹,²*, Anil Piprodhe¹ and Mashitha Pise¹

**Abstract**

**Background:** In this study, strictosidine synthase (STR) from *Catharanthus roseus* that plays an important role in alkaloid biosynthesis was selected. The purpose of this work was to perform in silico analysis and to predict the three-dimensional structure of this protein that is not available.

**Results:** Physicochemical characterization was performed by Expasy’s Protparam server. The computed theoretical isoelectric point (pI) found to be less than 7 indicates the acidic nature of this protein. The aliphatic index 73.04 indicates the thermal stability of the protein. Grand average hydropathy (GRAVY) was predicted to be −2.85; this lower value of GRAVY shows the possibility of better interaction of this protein with water. Functional analysis of these proteins was performed by SOSUI server which predicted the transmembrane helix. Secondary structure analysis was carried out by SOPMA that revealed that Alpha helix dominated among secondary structure elements followed by random coil, extended strand, and beta turns. The modeling of the three-dimensional structure of the STR was performed by Swiss model. The model was validated using protein structure checking tools PROCHECK and PROVE.

**Conclusions:** This study reveals in silico analysis by Expasy Protparam server, SOPMA, and SOSUI server. Homology modeling of STR was performed by Swiss model.

**Keywords:** *Catharanthus roseus*, STR, SOSUI, PROCHECK, PROVE

**Background**

*Catharanthus roseus* (*C. roseus*) is an important medicinal plant used to treat various diseases. The plant is distributed throughout the world. It is known to produce modern chemotherapeutic agent for pain-relieving properties [1]. One of the important types of alkaloid is the vinblastine produced from *C. roseus* which was reported due to its antitumor activity and has wide pharmaceutical use [2]. It has been reported that accumulation of free radicals can cause pathological conditions such as ischemia, asthma, arthritis, inflammation, neurodegeneration, Parkinson’s diseases, mongolism, aging process, and perhaps dementia [3]. The flower petals, seeds, and other parts of *Catharanthus roseus* exhibit antioxidant properties. Thus, phenolic compounds have redox properties that act as reducing agents, hydrogen donors, singlet oxygen quenchers, or metal chelator. *C. roseus* was of enormous pharmaceutical interest because it contains more than 120 terpenoid indole alkaloids (TIAs). Terpenoid indole alkaloids (TIAs) are among the most important secondary metabolites in plants that play important roles in the growth and reproductive development of plants [4]. Over 100 different TIAs were discovered in *Catharanthus roseus* (periwinkle) [5]. Strictosidine was found to be the key biosynthetic precursor of TIA. It was presented in a wide variety of higher plants [6]. This important molecule is generated by strictosidine synthase (STR) from tryptamine and secologanin and catalyzes the synthesis of 3-a(S)-strictosidine. STR 1 is the first STR gene isolated.
from *Rauvolfia serpentina*. STR catalyzes the Pictet-Spengler reaction between tryptamine and secologanin and is a key enzyme for the biosynthesis of alkaloids [7, 8]. The Pictet-Spengler reaction was a two-part reaction. First, an electron-rich aromatic amine and an aldehyde condense to form an iminium species. Second, an electrophilic aromatic substitution reaction occurs in which the ayl amine attacks the electrophilic iminium to yield a positively charged intermediate which is then deprotonated to yield the β-carboline product(s). In nonenzymatically catalyzed reactions, two enantiomers are typically formed, but strictosidine synthase catalyzes the asymmetric synthesis of the strictosidine 3 diastereomer [9].

This was the central reaction in the biosynthesis of the entire family of monoterpenoid indole alkaloids in plants that includes vincamine, ajmaline, raubasine, quinine, vinblastine, reserpine, vincamine, C-toxiferen, I, and camptothesine. STR proteins are involved in different physiological and biochemical pathways. The monoterpenoid-derived indole alkaloids were one of the structurally largest and pharmacologically most diverse alkaloid families in higher plants. These alkaloids have medical applications and therapeutics include the treatment of cancer (vinblastine or the camptothecin-derivative topotecan), malaria (quinine), hypertension (raubasine and reserpine), schizophrenia (reserpine in high dosage), disturbed cerebral blood flow (vincamine), and antiarrhythmic heart disorders (ajmaline), from the Indian medicinal plant *Rauvolfia serpentina* [10].

The major drawbacks of experimental method used to characterize protein was involvement of high cost and time consumption. Thus, these methods were not amenable to high throughput techniques. In silico approaches provide a viable solution to these problems. Computational tools provide researchers to understand physicochemical and structural properties of proteins. The amino acid sequence provides most of the information required for determining and characterizing the molecule’s function and physical and chemical properties. In this study, the in silico analysis and homology modeling studies on STR involved in alkaloid biosynthesis were reported. The three-dimensional structure for this protein was yet not available. Hence to describe its structural features and to understand molecular function, the model structures for this protein was constructed. Physicochemical characterization was performed by computing theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index, and grand average hydropathy (GRAVY). Functional analysis and secondary structure prediction of these proteins was performed by Swiss model. The model was validated using protein structure checking tools PROCHECK and PROVE.

**Methods**

**Ethics approval**

This article does not contain any studies with human participants or animals performed by any of the authors.

**Sequence retrieval**

Sequences of proteins involved in alkaloid biosynthesis of *C. roseus* were retrieved from the SWISSPROT, a public domain protein database [11]. Table 1 shows the protein sequences considered in this study. The protein sequences were retrieved in FASTA format and used for further analysis.

**Physicochemical characterization**

For physicochemical characterization, theoretical pI, molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index, and grand average hydropathy (GRAVY) [15] were computed using the Expasy’s ProtParam server [16]. The results were shown in Table 2.

**Functional characterization**

The SOSUI server performed the identification of transmembrane regions. Table 3 represents the transmembrane region identified for these proteins. Prosite is a database of protein families and domains [17]. Table 4 represents the output of Prosite that was recorded in terms of the length of amino residues of protein with specific profiles and patterns.

**Secondary structure prediction**

SOPMA [18] was employed for calculating the secondary structural features of the selected protein sequences considered for this study. The results were presented in Table 5.

**Model building and evaluation**

The modeling of the three-dimensional structure of the proteins was performed by Swiss model [19]. The overall stereochemical property of the protein was assessed by Ramachandran plot analysis [20]. The validation for structure models obtained from the two software tools was performed by using PROCHECK [21]. The models were further checked with PROVE. The results of PROCHECK and PROVE analysis was shown in Tables 6 and 7.
Results
In this study, STR involved in the first step of alkaloid biosynthesis present in C. roseus was selected. Sequence was retrieved from UNIPROT in FASTA format and used for further analysis. Table 1 shows the protein sequences considered.

Expasy’s Protparam tool was used to compute parameters to analyze physicochemical properties of this protein. The results were shown in Table 2. The computed isoelectric point value of P18417 was less than 7 (pI < 7). The instability index value for P18417 was found to be 47.42. Aliphatic index for the selected protein sequence was found to be 73.04. The grand average hydropathy (GRAVY) was found to be –0.285.

Table 2 Parameters computed using Expasy’s Protparam

| Protein                | Accession number | M. wt     | pI   | II    | AI    | GRAVY |
|------------------------|------------------|-----------|------|-------|-------|-------|
| Strigolactone synthase  | P18417           | 39093.72  | 5.14 | 47.42 | 73.04 | –0.285|

Swiss Model. The final modeled structures were visualized by Swiss PDB Viewer that was shown in Fig. 1. For validation of generated structure, Ramachandran plot and PROVE plot were constructed. The Phi and Psi distribution of the Ramachandran map generated by non-glycine, non-proline residues were summarized in Table 6 and Fig. 2.

Discussions
After retrieval of STR sequence, physicochemical properties were studied by Expasy’s ProtParam server (Table 2). The calculated isoelectric point (pI) will be useful because at pI, solubility is least and mobility in an electro-focusing system is zero. At pI, proteins were stable and compact. The computed isoelectric point value of P18417 was less than 7 (pI < 7) which indicates that this protein was considered acidic. The computed instability index will be useful for developing a buffer system for purification by isoelectric focusing method. The instability index value for P18417 (47.42) was found to be near 40 indicating that it was stable. The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chains (A, V, I, and L) is regarded as a positive factor for the increase of thermal stability of globular proteins. Aliphatic index for the selected protein sequence was found to be 73.04. This higher value indicates the thermal stability of the protein. The grand average hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of

Table 3 Transmembrane regions identified by SOSUI server

| Protein | Accession No. | Types         | Transmembrane helix |
|---------|---------------|---------------|---------------------|
| STR     | P18417        | Membrane protein | One transmembrane helix |

Table 4 Functional characterization of proteins of C. roseus at Prosite

| Prototypr | Accession number | Motif found | Position in the protein |
|-----------|------------------|-------------|-------------------------|
| STR       | P18417           | NO hits     |                         |

Table 5 Secondary structure prediction by SOPMA SERVER

| Proteins secondary structure | P18417 |
|-----------------------------|--------|
| Alpha helix                 | 15.91% |
| 310 helix                   | 0.00%  |
| Pi helix                    | 0.00%  |
| Beta bridge                 | 0.00%  |
| Extended strand             | 34.09% |
| Beta turn                   | 12.78% |
| Bend region                 | 0.00%  |
| Random coil                 | 37.22% |
| Ambiguous states            | 0.00%  |
| Other states                | 0.00%  |
The hydropathy values of all the amino acids, divided by the number of residues in the sequence. GRAVY indices of the selected protein was found to be −2.85 which indicates. This lower value of GRAVY shows the possibility of better interaction with water.

The SOSUI server classifies the STR protein P18417 in the form of membrane protein. The transmembrane regions are rich in hydrophobic amino acids (Table 3).

Sequence of a particular cluster of residue types, which is variously known as a pattern, motif, signature, or fingerprint. These motifs, typically around 10 to 20 amino acids in length, arise because specific residues and regions thought or proved to be important to the biological function of a group of proteins are conserved in both structure and sequence during evolution [23]. Prosite analysis suggested the functionality of these proteins with profiles and patterns identified for characteristic functionality were represented in Table 4. In protein P18417, no motif was found in any range.

The secondary structure as predicted using SOPMA was represented in Table 5. The results show that random coil dominated among secondary structure elements followed by extended strand, alpha helix, and beta turns. Homology modeling is the method of choice for obtaining 3D coordinates of proteins. SWISS-MODEL workspace was an integrated web-based modeling expert system. For the STR, experimental protein structures were searched to identify suitable templates. On the basis of a sequence alignment between the STR and the template structure, a three-dimensional model for the STR was generated. STR catalyzes a Pictet-Spengler-type reaction and represents a novel six-bladed b-propeller fold in plant proteins [10]. STR from R. serpentina is a monomeric precursor protein with 344 amino acids that exhibits 100, 79, and 58% identity to STR1 from Rauvolfia munnii, C. roseus, and Ophior rhizapumila, respectively. STR has been detected and biochemically characterized from cell suspension culture of apocynaceae plant C. roseus [24] and the Indian medicinal plant Rauvolfia serpentina [25] as well as from Rubiaceae ophiollahiza [26]. Three-dimensional structures are predicted for proteins where such data is unavailable. There is lack of experimental structures for this protein considered. Swiss model was used to generate the homology model for STR that has not been reported in literature yet. The protein-ligand interaction was found to be very important. 3D structure enables us to understand the binding specificities of ligands with protein.

Validation of STR model
Homology modeling was the most accurate computational method to generate reliable structural models. It was used in many biological applications. Model quality assessment tools were used to estimate the reliability of the models. The stereochemical quality of
the predicted model and accuracy of the protein model was evaluated after the refinement process using Ramachandran map calculations computed with the PROCHECK program. The assessment of the predicted model generated by Swiss model was shown in Fig. 2. The main chain parameters plotted are Ramachandran plot quality, peptide bond planarity, bad non-bonded interactions, main chain hydrogen bond energy, C-alpha chirality, and overall G factor. In the Ramachandran plot analysis, the residues were classified according to its regions in the quadrangle. The red regions in the graph indicate the most allowed regions whereas the yellow regions represent allowed regions. Glycine is represented by triangles and other residues are represented by squares. The result revealed that the modeled structure for P18417 has 88.8% residue in allowed region. Such figures assigned by Ramachandran plot represent a good quality of the predicted models.

PROVE plot was used to calculate the atoms in the modeled structure of strictosidine synthase, which shows the particular result in Table 7. The RMS Z score was found to be more than 1 that also supports good quality of model structure.

**Conclusion**

In this study, STR that plays an important role in alkaloid biosynthesis of *C. roseus* was selected. Physicochemical characterization was performed by computing theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index, and grand average hydropathy (GRAVY). Functional analysis of this protein was performed by SOSUI server which predicted the transmembrane helix. Secondary structure analysis revealed that Alpha helix dominated among secondary structure elements followed by random coil, extended strand, and beta turns. The modeling of the three-dimensional structure of the Strictosidine synthase (P18417) was performed by Swiss model. The model was validated using protein structure checking tools PROCHECK and PROVE.

**Abbreviations**

*C. roseus*: Catharanthus roseus; STR: Strictosidine synthase; GRAVY: Grand average hydropathy; AI: Aliphatic index; II: Instability index; pI: Isoelectric point; SOPMA: Self-Optimized Prediction Method with Alignment

**Acknowledgements**

The authors are grateful to the P. G. Department of Biochemistry, Hislop College, Nagpur, and Dr D Y Patil ACS College, Pune, for providing the necessary facilities and encouragement.
Authors’ contributions
Study conception and design: Archna Sahay and Maspitha Pise. Acquisition of data: Anil Piprodhe. Analysis and interpretation of data: Archna Sahay. Drafting of manuscript: Archna Sahay. The authors read and approved the final manuscript.

Funding
Not applicable

Availability of data and materials
Not applicable

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Received: 14 April 2020 Accepted: 8 July 2020

Published online: 28 August 2020

References
1. Kumari K, Gupta S (2013) Phytopotential of Catharanthus Roseus L.(G.) Don. Var. “Rosea” and “Alba” Against Various Pathogenic Microbes In vitro. International Journal of Research in Pure and Applied Microbiology 3(3):77–82
2. Rischer H, Oresic M, Seppanen-Laakso T, Katajamaa M, Lammertyn M, Ardiles-Diaz W (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in catharanthus roseus cells. Proceedings of the national Academy of sciences of the United States USA 103:5614–5619. https://doi.org/10.1073/pnas.0601027103
3. Sarabjot Kaur, Poonam Mondal. Study of Total Phenolic and Flavonoids Content, Antioxidant Activity and Antimicrobial Properties of Medicinal Plants. Journal of microbiology and experimentation. 1(1): 1-6 (2014) DOI: https://doi.org/10.15406/jmen.2014.01.00005
4. Yamazaki Y et al (2003) Metabolite profiling of alkaloids and stricostidine synthase activity in camptothecin producing plants. Phytochemistry 62:461–470. https://doi.org/10.1016/S0031-9422(02)00543-5
5. Kutchan T. M. Alkaloid Biosynthesis[MdS99]The Basis for Metabolic Engineering of Medicinal Plants. The plant cell, 1059–1070 (1995). doi: https://doi.org/10.10101/tpc.7.7.1059
6. Smith, G. N. Stricostidine: a key intermediate in the biogenesis of indole alkaloids. Chem. Commun. 912–914, doi:10.1039/C19680000912 (1968).
7. Kutchan TM (1989) Expression of enzymatically active cloned stricostidine synthase from the higher plant Rauvolfia serpentina in Escherichia coli. FEBS Lett 257:127–130. https://doi.org/10.1016/S0014-5793(89)81802-2
8. Kutchan TM, Hampp N, Lottspeich F, Beyreuther K, Zenk MH (1988) The cDNA clone for stricostidine synthase from Rauvolfia serpentina. DNA sequence determination and expression in Escherichia coli. FEBS Lett 237: 40–44. https://doi.org/10.1016/0014-5793(88)81067-4
9. Justin J. Mares, Lesley-Anne Giddings, Anne Friedrich, Elke A. Loren, Santosh Panjikar, Bernhardt L. Trout, Joachim Stöckigt, Baron Peters, and Sarah E. O’Connor. Stricostidine Synthase: Mechanism of a Pictet–Spengler Catalyzing Enzyme. 130, 710–723 (2008). DOI: 10.1021/ja077190z
10. Ma X, Panjikar S, Koeple J, Loren E, Stockigt J (2006) The Structure of Rauvolfia serpentina Stricostidine Synthase Is a Novel Six-Bladed b-Propeller Fold in Plant Proteins. Plant Cell 18:907–920 https://doi.org/10.1105/tpc.105.038018
11. Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28:45–48. https://doi.org/10.1093/nar/28.1.45
12. Gill SC, Von Hippy PH (1989) Extinction coefficient. Anal Biochem 182: 319-328. doi:10.1016/0003-2697(89)9002-7
13. Guruprasad K, Reddy BVP, Pandit MW (1990) Ca protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 4:155–164. https://doi.org/10.1093/protein/4.2.155
14. Ikai, A 1980. Thermostability and aliphatic index of globular proteins. J Biochem. 1980 Dec;88(6):1895-1898. doi.org/10.1093/oxfordjournals.
15. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132. https://doi.org/10.1016/0022-2836(82)90515-0
16. Gasteiger, Elisabeth & Hoogland, Christine & Gattiker, Alexandre & Duzaud, Severine & Wilkins, Marc & Appel, Ron & Bairoch, Armos. (2007). Protein Identification and Analysis Tool on the ExPASy Server. DOI: https://doi.org/10.1385/1-59259-890-0:571
17. Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ A et al (2002) The PROSITE database, its status in 2002. Nucleic Acids Res 30:235–238. https://doi.org/10.1093/nar/30.1.235
18. Goujon C, Deléage G (1995) SOPMA: signifi cant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11:681–684. https://doi.org/10.1093/141958-017-07064-4
19. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-Ma web-based environment for protein structure homology modelling. Bioinformatics 22: 195-201.ODEL workspace: dohttps://doi.org/10.1093/bioinformatics/bt6770
20. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99. https://doi.org/10.1016/0022-2836(63)80023-6
21. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486. https://doi.org/10.1007/bf00228148
22. Gues N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis, 19(15):2714–2723. DOI: https://doi.org/10.1021/oe0701150
23. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch P (2002) PROSITE: a documented database using patterns and profiles as model descriptors. Brief Bioinform 3(3):265–274. https://doi.org/10.1093/bib/3.3.265
24. Stockigt, J., and Zenk, M.H. (1977). Stricostidine (isovincoside): The key intermediate in the biosynthesis of monoterpoid indole alkaloids. J. Chem. Soc. Chem. Commun. 466–468. DOI: https://doi.org/10.1039/C9770000646
25. Hanpp N, Zenk MH (1988) Homogenous stricostidine synthase from cell suspension cultures of Rauvolfia serpentina. Phytochemistry 27:3811–3815. https://doi.org/10.1016/0031-9428(88)80323-1
26. Yamazaki Y, Sudo H, Yamazaki M, Ainm N, Saito K (2003) Camptothecin biosynthetic genes in hairy roots of Ophorhniza pumila: Cloning, characterization and differential expression in tissues and by stress compounds. Plant Cell Physiol 44:395–403. https://doi.org/10.1093/pcp/pcg051

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.