

# Structure Prediction of Dihydroflavonol 4-Reductase and Anthocyanidin Synthase from Spinach

Archna Sahay<sup>1,\*</sup>, Madhvi Shakya<sup>2</sup>

<sup>1</sup>Department of Bioinformatics, MANIT, Bhopal- 462051, Madhya Pradesh, India; <sup>2</sup>Department of Mathematics, MANIT, Bhopal- 462051, Madhya Pradesh, India; Archna Sahay - Email: sahayarchna@gmail.com, archnasahay@yahoo.com; Phone: +91-0755-2670327-1622; Fax: +91-0755-2670562;  
\*Corresponding author.

Received July 20, 2010; Accepted August 20, 2010; Published November 27, 2010

**Abstract:**

Spinach is an important dietary vegetable associated with beneficial health effects. Flavonoids have various biological activities such as antioxidant, antibacterial, and anticancer effect. Flavonoid including anthocyanin provides brilliant and colored pigments in different plant tissues. Anthocyanidin synthase and dihydroflavonol 4-reductase are responsible for anthocyanin biosynthesis. They contributed in plant protection against UV-B radiation, microbial and herbivore pathogens. A 3D structures of anthocyanidin synthase and dihydroflavonol 4-reductase from spinach are constructed in this study through homology modeling. The homology modeling is done by using the MODELLER 9v7 software. The energy of models was minimized by applying molecular mechanics method. The root mean square deviation (RMSD) for C atoms between the template and the homology-modeled structures was estimated by CE program. The final models were assessed by PROCHECK and WHATCHECK which showed that the final refined models are reliable.

**Keywords:** spinach; anthocyanidin synthase; dihydroflavonol 4-reductase; flavonoid, structure prediction.

**Background:**

Spinach (*Spinacia oleracea*) is a flowering plant in the family of Amaranthaceae. Though Spinach is most often commonly used as a food, it has medicinal value as well. Dietary supplementation with blueberries, spinach, or spirulina reduces ischemic brain damage [1]. Natural antioxidant mixture (NAO), a water-soluble extract obtained from spinach leaves has been shown to have anti-inflammatory [2], antiproliferative and antioxidative properties [3] in biological systems. A detailed analysis of the protein sequences which are responsible for these properties, their probable structures and mode of action has yet to be accomplished.

Flavonoids have various biological activities such as antioxidant, antibacterial, and anticancer effect [4]. Flavonoids comprise a class of abundant secondary metabolites and the most common group of plant polyphenols. It provides much of the flavor and color to fruits and vegetables and contributes in many ways to the growth and subsistence of plants. Flavonoids act as antioxidants against harmful free radicals in animals. More than 5000 different flavonoids have been described. The six major subclasses of flavonoids include the flavones (e.g., apigenin, luteolin), flavonols (e.g., quercetin, myricetin), flavanones (e.g., naringenin, hesperidin), catechins or flavanols (e.g., epicatechin, gallocatechin), isoflavones (e.g., genistein, daidzein) and anthocyanidins (e.g., cyanidin, pelargonidin), [5]. Anthocyanin pigments and the related oligomeric proanthocyanidins are under investigation for their medicinal potential [6].

The red colors of flowers are mainly produced by anthocyanin and betacyanin. They serve essential functions in plant reproduction as flower and fruit colors by recruiting pollinators and seed dispersers. Dihydroflavonol 4-reductase (DFR) is the first enzyme committed to anthocyanin biosynthesis in the flavonoid pathway. DFR catalyses the

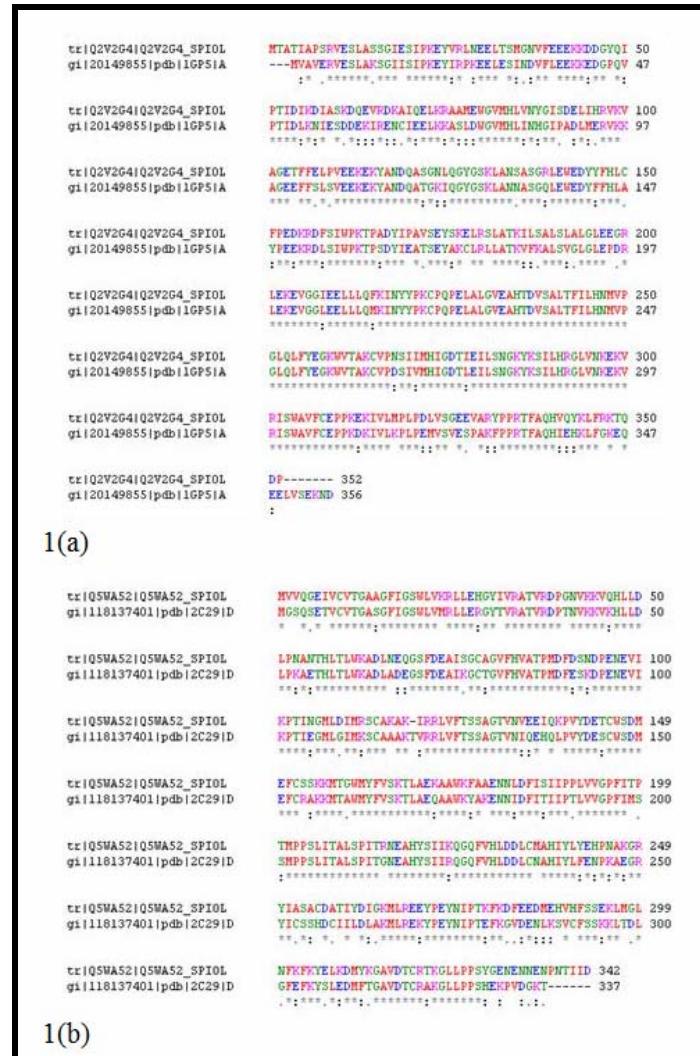
NADPH-dependent reduction of dihydroflavonols into leucoanthocyanidins which is considered as the immediate precursor of anthocyanidin. Anthocyanidin synthase (ANS) is one of the four dioxygenases (DOX) of the flavonoid biosynthetic pathway. ANS is a 2-oxoglutarate iron-dependent oxygenase, which catalyses the next step to anthocyanidins. It is responsible for the formation of the colored anthocyanidins from the colorless leucoanthocyanidins [7]. The leucoanthocyanidins are precursors for catechins and proanthocyanidins, which are involved in plant resistance and influence food and feed quality of plant products [8]. They are also direct precursors of one of the most conspicuous flavonoid classes, the anthocyanins, which are found in fruits, flower petals, and leaves exhibiting a wide range of functions such as attraction of pollinators and seeds dispersers, UV light damage protection, and plant defense against pathogen attack. In humans, the antioxidant properties of anthocyanins have been involved in protection against oxidative stress and certain tumors, and age related and cardiovascular diseases [9].

The structure of the spinach ANS and DFR proteins have not yet determined experimentally (X-ray or NMR), and therefore models are built following homology modeling protocol. A large number of computational tools are available from different sources for making predictions regarding the identification and structure prediction of proteins. The major drawbacks of experimental methods, used to characterize the proteins of various organisms are the time frame involved, high cost and the fact that these methods are not amenable to high throughput techniques.

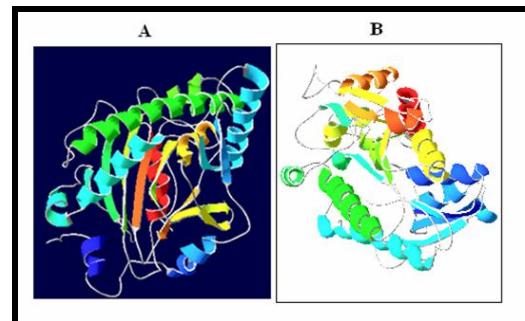
*In silico* approaches provide a viable solution to these problems. Computationally based characterization of the features of the proteins found or predicted in completely sequenced proteomes is an important task in the search for knowledge of protein function. In this study, an effort is

made to generate three-dimensional (3D) models of ANS and DFR based on the available template crystal structure from protein data bank. The built

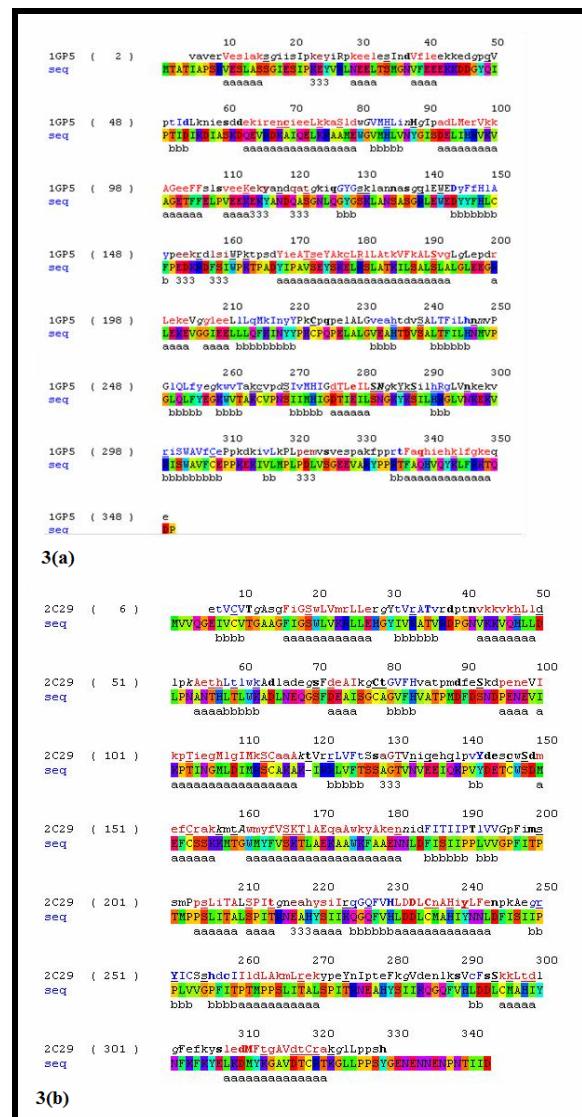
models are validated using widely followed Procheck and WHATCHECK programmes.



**Figure 1(a):** Sequence alignment of ANS from spinach using CLUSTAL W server. **Figure 1(b):** Sequence alignment of DFR from spinach using CLUSTAL W server



**Figure 2:** The final 3D structure of (A) ANS (B) DFR.



3(b)

**Figure 3 (a):** Structure based sequence alignment of template and final structures of the spinach ANS using JOY program. The key to the JOY annotation is as follows: lowercase red letter,  $\alpha$ -helix; lowercase blue letter,  $\beta$ -strand; lowercase maroon letter,  $3_{10}$  helix; uppercase letter, solvent-inaccessible residue; lowercase letter, solvent-accessible residue; italic lowercase letter, positive  $\Phi$ . **Figure 3 (b):** Structure based sequence alignment of template and final structures of the spinach DFR using JOY program. The key to the JOY annotation is as indicated Figure 3(a).

### Methodology:

**Modeling of anthocyanin synthase and dihydroflavonol 4-reductase:** The amino acid sequence of ANS and DFR were retrieved from the sequence database SWISSPROT, a public domain protein database [10]. It is ascertained that the three-dimensional structure of these proteins is not available in Protein Data Bank.

BLAST [11] search was performed against Brookhaven Protein Data Bank (PDB) with the default parameters to find suitable templates for homology modeling. Sequences were aligned and the one that showed the maximum identity with high score and lower e-value and maximum sequence identity is used as a reference structure to build a 3D model for ANS and DFR. The ClustalW program [12] was used for sequence alignment. The secondary structures were analyzed and compared by the JOY [13] server (protein sequence-structure representation and analysis).

The academic version of MODELLER 9v7 [14] was used for model building. Out of 10 models generated by MODELLER for each protein, the one with the best G-score of PROCHECK [15] was selected. The constructed 3D models were energy minimized in GROMOS96 force field using steepest descent minimization Algorithms [16]. The validation for structure models was performed by using WHATCHECK [17] and PROCHECK. The stereo chemical qualities of the three dimensional models were assessed by PROCHECK analysis and Ramachandran plot was drawn [18]. Structural analysis is performed and figures representations were generated with Swiss PDB Viewer [19]. The weighted root mean square deviation (RMSD) of the modeled protein is calculated using combinatorial extension (CE) algorithm [20].

### Results and Discussion:

#### Model Building:

Experimental (X-ray or NMR) structures of the spinach proteins ANS and DFR were not available therefore models are built following homology modeling protocol. BLAST search was performed against Brookhaven Protein Data Bank (PDB) with the default parameters to find suitable templates for homology modeling. Based on the maximum identity with high score and lower e-value in the BLAST search, 1GP5 and 2C29 were used as the structural template for modeling of the spinach ANS and DFR proteins respectively. Sequence alignment between target and template was done by using ClustalW program. The sequence structure alignment used for model building was shown in **Figure 1(A)** and **Figure 1(B)**. The predicted 3-D models will be very useful while studying the real structure of the protein. The academic version of MODELLER 9v7 (Sali et al., 1993) (<http://www.salilab.org/modeler>) was used for model building. Out of 10 models generated by MODELLER, the one with the best G-score of PROCHECK (Laskowski et al., 1993) was selected. The final stable structures of the ANS and DFR obtained were shown in **Figure 2(A)** and **Figure 2(B)**.

The secondary structures were analyzed and compared by the JOY (Mizuguchi K et al, 1998) server (protein sequence-structure representation and analysis). Given their PDB files, secondary structures were analyzed and compared by the JOY program. The secondary structures of templates and final models were highly conserved which showed that final models were highly reliable as shown in **Figure 3(A)** and **Figure 3(B)**.

#### Evaluation and validation of the 3-D structure:

The stereo chemical quality of the predicted models was evaluated using PROCHECK. Validations of the models were carried out after the refinement process using Ramachandran Map calculations computed with the PROCHECK program. The  $\Phi$  and  $\Psi$  distribution of the Ramachandran Map of non glycine, non proline residues were summarized in **Table 1** (See **Supplementary material**). The overall PROCHECK G-factor for ANS and DFR homology modeled structures were -0.04 and -0.23. The score indicates that the modeled structures were acceptable as value is greater than the acceptable value – 0.50. Table I shows that more than 90% of the residues were in allowed regions for ANS and 89.3% residues were in allowed region for DFR, leading to a good validation for the model.

The constructed models of ANS and DFR from spinach were examined for validation using different criteria. Standard bond angles of the models are determined using WHATCHECK (**Table 1**, See **Supplementary material**). The analysis revealed RMS Z-scores of bond angles for ANS and DFR were found to be 1.090 and 1.349, respectively. The values are almost equal to 1 suggesting high model quality. The quality of model was also assessed by comparing predicted structure to experimentally solved structure via superimposition and atoms root mean square deviation (RMSD) assessment. A model can be considered as reliable or accurate model when its RMSD is less than 3 – 4 Å (accurate  $<=2$  Å reliable  $<=4$  Å) [21]. The structural superimposition of C trace of templates and query sequence (ANS and DFR) are estimated by using CE program (<http://cl.sdsc.edu/ce.html>). The weighted root mean square deviation of C trace between the template and final refined model for ANS and DFR were found to be 0.23 Å and 0.38 Å respectively which suggests further that the models were reliable.

The information concerning biological function can be derived from a 3D protein structure. The homology models developed in this study for ANS and DFR will be used for the identification of active site and for docking study. Active sites were explored for determining important residues. The residues that are buried in the core of the molecule or exposed to solvent on its surface can be identified. Protein-ligand complexes carry functional information such as where the ligand is bound and which residues in the active site interact with the ligand can be identified. These models could be further explored for characterizing the protein.

#### Conclusion:

A 3D structure of anthocyanidin synthase and dihydroflavonol 4-reductase from spinach were constructed in this study through homology modeling. The homology modeling was done by using the MODELLER 9v7 software. The energy of models is minimized by applying molecular mechanics method. The final models were assessed by PROCHECK and WHATCHECK which showed that the final refined models were reliable. The molecular models of ANS and DFR will provide a good foundation for functional analysis of experimentally derived crystal structures.

#### Acknowledgements:

Financial assistance from Council of Scientific and Industrial Research (CSIR), Govt. of India, for the Research Associate Fellowship (RA) is gratefully acknowledged. The authors are grateful to the Department of Bioinformatics and Department of Mathematics, MANIT, Bhopal for providing the necessary facilities and encouragement.

#### References:

- [1] Y Wang *et al.* *Exp. Neurol.* **193**: 75 (2005)
- [2] L Lommitski *et al.* *Pharmacol. Toxicol.* **87**: 18 (2000) [PMID: 10987211]
- [3] M Bergman *et al.* *Phytochemistry* **58**: 143 (2001) [PMID: 11524124]
- [4] L H Cazarolli *et al.* *Mini Rev. Med. Chem.* **8**: 1429 (2008)
- [5] JA Ross *et al.* *Annu. Rev. Nutr.* **22**: 19 (2002) [PMID: 12055336]
- [6] JM Kong *et al.* *Phytochemistry* **64**: 923 [PMID: 14561507]
- [7] S Shimada *et al.* *J. Exp. Bot.* **58**: 957 (2007) [PMID: 17185736]
- [8] S Martensa *et al.* *FEBS Lett.* **531**: 453 (2002)
- [9] AM Amorini *et al.* *Free Radic. Res.* **37**: 453 (2003) [PMID: 12747740]
- [10] Bairoch & R Apweiler. *Nucleic Acids Res* **28**: 45 (2000) [PMID: 10592178]
- [11] SF Altschul *et al.* *J. Mol. Biol.* **215**: 403 (1990) [PMID: 2231712]
- [12] JD Thompson *et al.* *Nucleic Acids Res* **22**: 4673 (1994) [PMID: 7984417]
- [13] K Mizuguchi *et al.* *Bioinformatics* **14**: 617 (1998) [PMID: 9730927]
- [14] A Sali & TL Blundell. *J. Mol. Biol.* **234**: 779 (1993) [PMID: 8254673]
- [15] RA Laskowski *et al.* *J. Appl. Cryst.* **26**: 283 (1993) [doi:10.1107/S0021889892009944]
- [16] D Van der Spoel *et al.* *J. Comp. Chem.* **26**: 1701 (2005) [PMID: 16211538]
- [17] RWW Hooft *et al.* *Nature* **381**: 272 (1996)
- [18] GN Ramachandran *et al.* *J. Mol. Biol.* **7**: 95 (1963) [PMID: 13990617]
- [19] N Guex & CP Manuel. *Electrophoresis* **18**: 2714 (1997) [PMID: 9504803]
- [20] IN Shindyalov & PE Bourne. *Protein Engineer* **11**: 739 (1998)
- [21] A Rayan. *Bioinformation* **3**: 263 (2009)

Edited by P. Kangueane

Citation: Sahay & Shakya. Bioinformation 5(6): 259-263(2010)  
License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

**Supplementary material:****Table 1:** Validation of modeled protein structure using Procheck and WHATCHECK server.

Server	Property	ANS	DFR
Procheck	Residues in the most Favored Region	93.5%	89.3%
	Residues in additionally allowed region	6.1%	9.1%
	Residues in generously allowed region	0.3%	1.0%
	Residues in disallowed region	0.0%	0.7%
Whatcheck	RMS Z-score for bond angles	1.090	1.349