Sequence to Structure Analysis of DOPA Protein from *Mucuna pruriens*: A Computational Biology Approach

Authors
Laxman Nagar\(^1\)*\(^\ddagger\), Anuj Kumar\(^2\), Vimala. Y\(^1\), MNV Prasad Gajula\(^3\)*

\(^1\)Department of Botany, Ch. Charan Singh University, Meerut 250004, UP, India
\(^2\)Bioinformatics Infrastructure Facility (BIF), CCS University, Meerut 250004, India
\(^3\)Institute of Biotechnology, PJTSAU, Rajendra Nagar, Hyderabad-500030, India
*\ddagger* Equal Contribution
*Corresponding Author

MNV Prasad Gajula
Institute of Biotechnology, PJTSAU, Rajendra Nagar, Hyderabad-500030, India
Email: email2gajula@gmail.com

Abstract

L-DOPA, (L-3, 4-dihydroxyphenylalanine), an anti-nutritional compound is an important intermediate of secondary metabolism in higher plants and is known as a precursor of alkaloids, betalain, melanine, and others. We analyzed the amino acid sequence of DOPA protein from *M. pruriens* by using computational methods. The DOPA protein *M. pruriens* was subjected to sequence, structural and functional annotation. Functional domain prediction through Conserved Domain Database (CDD) suggested that DOPA proteins of *M. pruriens* contain Aspartate aminotransferase (AAT) superfamily associated with various cellular activities. The analysis of predicted secondary structure by PsiPred and SOPMA has shown that high percentage of helices in the protein structure makes DOPA more flexible for folding processes. Sub-cellular localization predictions suggested it to be a cytoplasmic `protein. Homology modeling method was used to deduce the three-dimensional (3D) structure of selected DOPA proteins from *M. pruriens*. From template search results it has been identified that all the hypothetical proteins share more than 50% sequence identity with crystal structure of *Sus scrofa*, indicating proteins are evolutionary conserved. Several quality assessment and validation parameters computed indicated that predicted homology models are in accordance with literature. Current study will help researchers to better understanding of the predicted models and refinement.

Key words: L-DOPA, Homology Modeling, Secondary Structure.

1. Introduction

L-DOPA, (L-3, 4-dihydroxyphenylalanine), an anti-nutritional compound is an important intermediate of secondary metabolism in higher plants and is known as a precursor of alkaloids, betalain, melanine, and others. It is also a precursor of catecholamines in animals and anti-nutritional compounds confer insect and disease resistance to plants \(^1\), \(^2\). L-DOPA is a potent neurotransmitter precursor that is believed, in part, to be responsible for the toxicity of the *Mucuna* seeds \(^3\). L-DOPA plays a very important role in plant physiology and for the treatment of Parkinson's disease \(^4\). Plants synthesize great variety of non-protein amino acids, among which L-DOPA, a compound with strong allelopathic activity, stands out \(^5\). This allelochemical is found in large quantities (1% and 4–7% in the leaves and seeds, respectively) in velvet bean [*Mucuna pruriens* (L.) var. *utiliz*], a legume of the Fabaceae family that has a nutritional quality...
comparable to the soybean [6],[7]. Velvet bean is often used for soil cover and as silage due to the large amount of organic matter with high digestibility it produces. It is estimated that velvet bean can release about 100–450 kg ha−1 of L-DOPA into the soil. Furthermore, its ability to control weeds and nematodes greatly reduces the need to apply artificial chemicals to the crops [6], [8], [9]. In addition, its cultivation in tropical areas is aimed at enriching the soil due to its ability to fix nitrogen [10].

In our study we have considered M. pruriens, which is considered to be one of the most popular Indian medicinal plant, traditionally used in Ayurvedic Indian medicine, for treatment of different diseases including parkinsonism, male infertility, nervous disorders, and also as an aphrodisiac [11]. Due to high protein concentration (23–35%) M. pruriens is considered an important source of various dietary supplements [12], [7]. DOPA from this herbal medicine plant has not been well studied so far at genomic level and the x-ray crystal or NMR structure is also not available in Protein Data Bank (PDB). So we analyzed the amino acid sequence of DOPA protein from M. pruriens by using different bioinformatics algorithms. We predicted the 3D model structure of DOPA from M. pruriens by using homology modeling method. We then studied the stereochemical properties of the modeled structure using Ramachandran plot. We also compared the amino acid sequence with the other DOPA protein sequence reported in different organisms.

2. Methods

2.1 Sequence Retrieval and Functional Domain Analysis

DOPA/tyrosine decarboxylase DC1 from M. pruriens was obtained from the Protein sequence database of NCBI (GI: ABK97629). Functional domain analysis was assessed using Conserved Domain Database (CDD) available at NCBI which is a collection of multiple sequence alignment representing protein domains conserved in molecular evolution. It contains the alignment data from protein families’ databases such as Pfam and SMART [13], [14].

2.2 Characterization of DOPA/tyrosine decarboxylase DC1 Protein

The physicochemical properties such as molecular mass, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) were calculated by ProtParam from EXPASy server (http://www.expasy.org/ ). The TargetP1.1 was used to analyze the subcellular localization of DOPA [15]. In order to know the key residues responsible for catalytic activity of the enzyme, multiple sequence alignment of DOPA with other closely related protein of decarboxylase family having known crystallographic structures was done on ClustalW2 program (http://www.ebi.ac.uk/Tools/msa/clustalw2/ )

2.3 Secondary structure prediction

The secondary structure prediction was done by SOPMA and PsiPred servers [16], [17]. The predicted secondary structural information of the enzyme was considered to improve the target-template alignment and for building conformations for 3D model of the DOPA.

2.4 Model building by Homology Modeling

Homology modeling or comparative modeling is the one of the most robust method for protein structure prediction. This method was supplemented for protein structure prediction based on the experimentally determined structure of another homologous protein [18]. The position specific iterated BLAST (PSI-BLAST) against Protein Data Bank (PDB) was carried out to identify their homologous structures based on the maximum identity with high score and lower e-value [19], while 3D structure was predicted using automated Swiss-Model server [20], [21].The modelled 3D structure of DOPA was visualized using UCSF CHIMERA 1.10 [22].
2.5 Structure validation by SAVES
The Ramachandranplot of the modeled structure was calculated by analyzing phi (Φ) and psi (ψ) torsion angles using PROCHECK available on SAVES server (http://nihserver.mbi.ucla.edu/SAVES/) [23], [24], Molprobity, and PSVS server [25], [26]. Further, the ProSA (https://prosa.services.came.sbg.ac.at/prosa.php) verified the modeled structure (DOPA) structure from X-ray analysis, NMR spectroscopy and other theoretical calculations. The energy criteria of the modeled structures were compared with large set of known protein structures.

3. Results and Discussion
3.1 Sequence Retrieval and Functional Domain Analysis
DOPA/tyrosine decarboxylase DC1 from *M. pruriens* was extracted from the Protein sequence database of NCBI (GI: ABK97629) having 164 amino acids in length. Conserved domains analysis carried by CDD revealed that DOPA protein belongs to Aspartate aminotransferase (AAT) superfamily (Pssm-ID: 276217) with conserved domain length 164, bit score: 224.02 and e-value: 4.89e-71.

3.2 Characterization of DOPA Protein
Physicochemical properties of DOPA by ProtParam tools are presented in (Table 1). Results show that DOPA protein has a molecular weight of 17261.1 Daltons and an isoelectric point of 6.78. The computed isoelectric point will be useful for separating the protein on a polyacrylamide gel by isoelectric focusing. The extinction coefficient can be used to calculate the concentration of a protein in solution. Stability of DOPA was studied by analyzing the values for instability index, aliphatic index and Grand average of hydropathicity (GRAVY) index. The value of instability index was 27.38 hence it could be safely predicted as a stable protein. The aliphatic index refers to the relative volume of a protein that is occupied by aliphatic side chains and contributes to the increased thermo stability of protein. Aliphatic index of DOPA was 106.52. GRAVY index indicates the solubility of proteins, GRAVY index of DOPA was 0.896. GRAVY value for DOPA describes it to be hydrophilic in nature. Subcellular localization was predicted by TargetP1.1 revealed that DOPA found in the cytoplasm.

### Table 1. Physicochemical properties of DOPA/tyrosine decarboxylase DC1 predicted by ProtParam program

<table>
<thead>
<tr>
<th>Sr.No</th>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Molecular Weight</td>
<td>17261.1 Dalton</td>
</tr>
<tr>
<td>2</td>
<td>Theoretical PI</td>
<td>6.78</td>
</tr>
<tr>
<td>3</td>
<td>Extinction coefficient*</td>
<td>15720 at Abs 0.1% 0.896</td>
</tr>
<tr>
<td>4</td>
<td>Instability Index</td>
<td>27.38</td>
</tr>
<tr>
<td>5</td>
<td>Aliphatic index</td>
<td>106.52</td>
</tr>
<tr>
<td>6</td>
<td>Grand average of hydropathicity(GRAVY)</td>
<td>0.392</td>
</tr>
</tbody>
</table>

* Extinction Coefficient units M⁻¹cm⁻¹ at 260 nm

3.3 Multiple Sequence Alignment
The identification of catalytic residues is a key to understanding the function of proteins. With the information from other functionally similar sequences with known crystallographic structures we can identify the key catalytic residues. ClustalW2 server was used for multiple sequence alignment of DOPA with other decarboxylases from human (PDB Id: 3RBF_A), Human aromatic L- amino acid decarboxylase (PDB Id: 3RCH_A) and Human histidine decarboxylase (PDB Id: 4E1O_A) shown in (Fig.1). The compared sequences varied in length but essentially conserved the key catalytic residues which have been highlighted with an asterisk (*) symbol.
3.4 Secondary Structure Prediction

Secondary structure was predicted by PsiPred and SOPMA servers (Fig.2). The results from SOPMA are shown in (Table 2).

<table>
<thead>
<tr>
<th>Secondary structure Element</th>
<th>No of amino acid involved</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha helix</td>
<td>76</td>
<td>46.34%</td>
</tr>
<tr>
<td>Extended strand</td>
<td>29</td>
<td>17.68%</td>
</tr>
<tr>
<td>Beta turn</td>
<td>10</td>
<td>6.10%</td>
</tr>
<tr>
<td>Random coil</td>
<td>49</td>
<td>29.88%</td>
</tr>
</tbody>
</table>

High percentage of helices in the structure makes the protein more flexible for folding, which might increase protein interactions. Moreover the predicted secondary structural information of DOPA was considered to improve the target-template alignment and for building 3D model of the DOPA.

3.5 Homology modeling of DOPA Protein

We selected a template from Protein data bank (PDB) as a result of PSI BLAST which showed 56% similarity (PDB ID: 1JS3, Chain A, Crystal structure of DOPA decarboxylase in complex with the inhibitor carbiDOPA, X-ray, Minimized Mean Structure, Resolution: 2.25Å) to build 3D structure of M. prurients DOPA using automate homology modeling server Swiss-Model (Biasini et al. 2014). Generated model showing the closest Cα RMSD (root-mean-square deviation) with respect to their template upon superposition were selected for further structural characterization. Generated model was visualized by UCSF CHIMERA 1.10. (Fig. 3 (A))

Fig.1. Multiple Sequence Alignment (MSA) of DOPA and templates proteins structure using CLUSTAL 2.1.

Fig.2. Predicted secondary structure of DOPA protein by using PsiPred Server.

Fig.3. Structure to validation: (A): Predicted 3D model of DOPA protein; (B): Ramachandran plot for predicted DOPA 3D model; and (C): ProSA plot of DOPA protein
3.6 Structure Validation

Ramachandran dihedral statistics for DOPA revealed a total of 88.7%, 9.2%, 0.7% and 1.4% residues in most favored, additionally allowed, generously allowed and disallowed regions, respectively. The Ramachandran plot generated by PROCHECK showed the satisfactory geometry of the predicted model (Fig. 3 (B)), (Table. 3). PSVS predicted G-factor score (all dihedral angles) -0.11 revealed that model was acceptable. Verified3D Z-score was -2.85 suggested that generated model have good quality. ProsaII (-ve) and MolProbity predicted. The Z-score -2.15 and -9.27 also suggested a good model. Similar studies were performed by Anuj et al. has shown that the predicted models can be used for further research [27, 28]. ProSA predicted Z-score value was -5.00, evidencing highly reliable structure (Fig.3 (C)). (Table. 3)

Table 3. Summary of structure quality factors of DOPA predicted by using SAVES and PSVS 1.5 server

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Mean score</th>
<th>SD</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procheck G-factor e(phi / psi only)</td>
<td>-0.14</td>
<td>N/A</td>
<td>-0.24</td>
</tr>
<tr>
<td>Procheck G-factor e(all dihedral angles)</td>
<td>-0.11</td>
<td>N/A</td>
<td>-1.65</td>
</tr>
<tr>
<td>Verify3D</td>
<td>0.28</td>
<td>0.0000</td>
<td>-2.89</td>
</tr>
<tr>
<td>ProsaII (-ve)</td>
<td>0.13</td>
<td>0.0000</td>
<td>-2.15</td>
</tr>
<tr>
<td>MolProbity clash score</td>
<td>62.91</td>
<td>0.0000</td>
<td>-9.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran Plot Summary from Procheck</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most favoured regions</td>
</tr>
<tr>
<td>Additionally allowed regions</td>
</tr>
<tr>
<td>Generously allowed regions</td>
</tr>
<tr>
<td>Disallowed regions</td>
</tr>
</tbody>
</table>

Residues selected based on: Dihedral angle order parameter, with S(phi)+S(psi)>=1.8
Selected residue ranges: 1A-163A

With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score

4. Conclusion

In this present study, we have used sequence analysis, secondary structure analysis, functional domain prediction and structure prediction to assign to DOPA protein from M. pruriens. L-DOPA is the drug of choice in the treatment of Parkinsonism disease. This study will lead for further research in optimizing the functionality of DOPA which has large implications in treatment of various human diseases. Experimental validation through X-ray crystallography and / or any other spectroscopic techniques can provide more insight into the actual function of this protein.

Acknowledgment

Authors acknowledge DST for the Ramanujan fellowship award SRN/S2/RJN-22/2011.

References

and C. Devendra, (Eds.) CAB International, Wallingford, UK.


