Chapter 04

In silico Analysis and Homology Modelling of Antioxidant Proteins of Spinach

Archna Sahay[1]* and Madhvi Shakya[2]

[1]Department of Bioinformatics, MANIT, India
[2]Department of Mathematics, MANIT, India

*Corresponding Author: Archna Sahay, Department of Bioinformatics, MANIT, Bhopal-462051 M.P., India, Tel: +91-0755-2670327-1622; Fax: +91-0755-2670562; Email: sahayarchna@gmail.com, archnasahay@yahoo.com

Abstract

Spinach is an important dietary vegetable often associated with beneficial health effects. In this paper, a bioinformatics and molecular modeling approach was adopted to explore properties and structure of spinach antioxidant proteins. The antioxidant proteins selected for this study are ascorbate peroxidase (APX), dehydro ascorbate reductase (DHAR), phospholipid hydroperoxide glutathione peroxidase-like protein (PHGPX) and 2-Cys peroxiredoxin BAS1 (2-CPs). Physico-chemical characterization interprets properties such as pI, EC, AI, GRAVY and instability index and provides data about these proteins and their properties. Prediction of motifs, patterns, disulfide bridges and secondary structure were performed for functional characterization. Three dimensional structures for these proteins were not available as yet at PDB. Therefore, homology models for these antioxidant proteins were developed. The modelling of the three dimensional structure of these proteins shows that models generated by Modeller were more acceptable in comparison to that by Geno3D and Swiss Model. The models were validated using protein structure checking tools PROCHECK and WHAT IF. These structures will provide a good foundation for functional analysis of experimentally derived crystal structures.

Keywords

Spinach; Computational Tools; Isoelectric Point; Disulphide Bridge; Homology Model

Introduction

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 [3] and antioxidative properties [4] in biological systems. A detailed analysis of the antioxidant protein sequences, their probable structures and mode of action has yet to be accomplished.

Reactive oxygen species (ROS) such as superoxide radicals (O_2^-), hydroxyl radical (OH^-), H_2O_2, and hydroperoxides (ROOH) are generated by exogenous sources, including prooxidant allelochemicals. Stress/starvation is an important endogenous source that generates ROS [5]. Antioxidants and antioxidant enzymes interrupt the cascades of uncontrolled oxidation in some organelles [6]. The antioxidant defense is primarily constituted by the actions of glutathione peroxidase (GPX), superoxide dismutase, catalase and ascorbate peroxidase [7]. In this study the antioxidant proteins of spinach have been selected for which three dimensional structures were not available at the protein data bank (PDB). These proteins are ascorbate peroxidase (APX), dehydro ascorbate reductase (DHAR), phopholipid hydroperoxide glutathione peroxidase-like protein (PHGPX) and 2-Cys peroxiredoxin BAS1 (2-CPs).

APX exist as isoenzyme and plays an important role in the metabolism of H_2O_2 in higher plants. APX isoenzymes are distributed in four distinct cellular compartments: stromal APX (sAPX), thylakoid membrane-bound APX (tAPX) in chloroplast, microbody (including glyoxisome and peroxisome) membrane bound APX (mAPX), and cytosolic APX (cAPX) [8]. DHAR is responsible for regenerating ascorbic acid from an oxidized state. It regulates the cellular ascorbic acid redox state, which in turn affects cell responsiveness and tolerance to environmental ROS. Ascorbic acid is a major antioxidant that serves many functions in plants including involvement in the detoxification of ROS, which are produced during aerobic metabolic processes such as photosynthesis or respiration [9]. Ascorbic acid also participates in the regeneration of α-tocopherol (vitamin E) from the tocopheroxyl radical [10].

2-Cysteine peroxiredoxins (2-CPs) constitute a ubiquitous group of peroxidases that reduce cell-toxic alkyl hydroperoxides to their corresponding alcohols [11]. Despite the presence of elaborate en-

zymatic and nonenzymatic antioxidative defense mechanisms, ROS escape from detoxification and oxidize organic compounds such as proteins, nucleic acids, terpenoids and fatty acids to the respective peroxides [12]. In addition, alkyl hydroperoxides are formed by enzymatic reactions in chloroplasts, e.g. lipoxygenase catalyzes peroxidation of fatty acids and other desaturated organic biomolecules, such as carotenoids [13]. Detoxification of alkyl hydroperoxides is important because they can act as long-distance mediators of oxidative damage by oxidizing other biomolecules and initiating radical chain reactions [14]. Another plant protein, PHGPX has been identified in several plant species, including tomato [15]. PHGPX forms reduce phospholipid and cholesterol hydroperoxides and thereby play an important role in protecting biological membranes against oxygen toxicity. The function of this protein in the removal of phospholipid hydroperoxides, which are generated as products of lipoxygenase catalyzed oxygenation of fatty acids has been reported [16].

Computational tools provide researchers to understand physico-chemical and structural properties of proteins. 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 that have been 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. The amino acid sequence provides most of the information required for determining and characterizing the molecule's function, physical and chemical properties. 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 paper the in silico analysis and homology modelling studies on antioxidant proteins of spinach was reported. Three dimensional structures for these proteins were yet not available. Hence to describe its structural features and to understand molecular function, the model structures for these proteins were constructed.

Materials and Methods

Sequences of antioxidant proteins of spinach were retrieved from the SWISSPROT, a public domain protein database [17]. Table I shows the protein sequences considered in this study. The antioxidant proteins sequences were retrieved in FASTA format and used for further analysis.

Physico-Chemical Characterization

For physico-chemical characterization, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient [18], instability index [19], aliphatic index [20] and grand average hydropathy (GRAVY) [21] were computed using the Expasy's ProtParam server [22] (http://us.expasy.org/tools/protparam.html). The results were shown in Table II

Functional Characterization

The SOSUI server performed the identification of transmembrane regions. Table III represents the transmembrane region identified for these antioxidant proteins. Disulphide bonds are important in determining the functional linkages. Table IV shows prediction of "SS" bonds using the primary structure (protein sequence data) by the tool CYS_REC (http://sunI.softberry.com/berry.phtml?topic). CYS_REC identifies the position of cysteins, total number of cysteins present and pattern, if present, of pairs in the protein sequence. Prosite is a database of protein families and domains [23]. Table V 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 [24] was employed for calculating the secondary structural features of the antioxidant protein sequences considered for this study. The results were presented in Table VI.

Model Building and Evaluation

The modeling of the three dimensional structure of the protein was performed by three homology modeling programs, Geno 3D [25], Swissmodel [26] and Modeller [27]. The constructed 3D models were energy minimized in GROMACS force field using steepest descent minimization Algorithms [28]. The overall stereochemical property of the protein was assessed by Ramchandran plot analysis [29]. The validation for structure models obtained from the three software tools was performed by using PROCHECK [30]. The models were further checked with WHAT IF [31]. The results of PROCHECK and WHAT IF analysis was shown in Table VII and Table VIII respectively. Structural analysis was performed and figures representations were generated with Swiss PDB Viewer [32].

Results and Discussion

Table I shows antioxidant proteins of spinach considered in this study. These antioxidant protein sequences were retrieved from the SWISSPROT, a public domain protein database. These protein sequences were retrieved in FASTA format and used for further analysis. Parameters computed using Expasy's ProtParam tool was represented in Table II. The calculated isoelectric point (pI) will be useful because at pI, solubility is least and mobility in an electro focusing system is zero. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of protein is zero. At pI proteins are stable and compact. The computed pI value of APX (Q42459, Q41371) and PHGPX were less than 7 (pI<7) indicates that these antioxidant proteins were considered as acidic. The pI of APX (Q7DN63, O46921, Q7DN73, Q7GDV4), DHAR and 2-CPs are greater than 7 (pI > 7) reveals that these proteins were basic in character. The computed isoelctric point (pI) will be useful for developing buffer system for purification by isoelectric focusing method. Although Expasy's ProtParam computes the extinction coefficient for 276, 278, 279, 280 and 282 nm

wavelengths, 280 nm is favored because proteins absorb light strongly there while other substances commonly in protein solutions do not. Extinction coefficient of AFPs at 280 nm is ranging from 15992.5 to 51465 M^{-1} cm^{-1} with respect to the concentration of Cys, Trp and Tyr. The high extinction coefficient of APX (Q7DN63, O46921, Q7DN73, and Q7GDV4) indicates presence of high concentration of Cys, Trp and Tyr. The computed extinction coefficients help in the quantitative study of protein–protein and protein–ligand interactions in solution. The instability index provides an estimate of the stability of protein in a test tube. There are certain dipeptides, the occurrence of which is significantly different in the unstable proteins compared with those in the stable ones. This method assigns a weight value of instability. Using these weight values it is possible to compute an instability index (II). A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable (Guruprasad et al., 1990). The instability index value for the spinach antioxidant proteins were found to be ranging from 21.90 to 47.14. The result classified APX (Q42459), DHAR (Q9FVE4) and PHGPX (O23814) as stable protein (Table II).

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 antioxidant protein sequences ranged from 67.26 – 84.98. The very high aliphatic index of all antioxidant protein sequences indicates that these antioxidant proteins may be stable for a wide temperature range. The lower thermal stability of APX (Q7DN63 and Q7GDV4) was indicative of a more flexible structure when compared to other antioxidant protein. The Grand Average hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. GRAVY indices of APX are ranging from - 0.095 to -0.519. This low range of value indicates the possibility of better interaction with water.

Functional analysis of these proteins includes prediction of transmembrane region, disulfide bond and identification of important motifs. SOSUI distinguishes between membrane and soluble proteins from amino acid sequences, and predicts the transmembrane helices for the former. The Transmembrane regions and their length were tabulated in Table III. The server SOSUI classifies APX (O46921 and Q7DN73) as membrane protein and other spinach antioxidant proteins as soluble proteins. SOSUI server has identified one transmembrane region in these proteins. The transmembrane regions are rich in hydrophobic amino acids. As disulphide bridges play an important role in determining the thermostability of these proteins. CYS_REC was used to determine the Cysteine residues and disulphide bonds. Possible pairing and pattern with probability were presented in Table IV. Result shows that except APX (Q42459), PHGPX (O23814) and 2-CPs (O24364) all proteins contain disulphide linkages.

The functions of antioxidant proteins of spinach were analyzed by submitting the amino acid sequence to Prosite server. 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 [33]. Prosite analysis suggested the functionality of these proteins with profiles and patterns identified for characteristic functionality were represented in Table V.

The secondary structure of spinach antioxidant proteins were predicted by SOPMA (Self Optimized Prediction Method with Alignment) which correctly predicts 69.5% of amino acids for a state description of the secondary structure prediction [24]. The secondary structure indicates whether a given amino acid lies in a helix, strand or coil. Secondary structure features as predicted using SOPMA were represented in Table VI. The results revealed that random coils dominated among secondary structure elements followed by alpha helix,

extended strand and beta turns for all sequences. The secondary structure were predicted by using default parameters (Window width: 17, similarity threshold: 8 and number of states: 4).

Three dimensional structures are predicted for proteins where such data is unavailable. There is lack of experimental structures for these proteins considered. Out of six ascorbate peroxidase isoenzymes sequences, three dimensional structure was modeled for only Q42459 since it has been reported that the steady-state transcript level of cytosolic APX altered in stress condition [4]. The other three proteins for which the three dimensional structures were modeled includes DHAR, PHGPX and 2-CPs. The modeling of the three dimensional structure of the protein was performed by three homology modeling programs, Geno 3D, Swiss Model and Modeller. The constructed three dimensional models were energy minimized in GROMACS force field using steepest descent minimization Algorithms. The Φ and Ψ distribution of the Ramachandran Map generated by of non glycine, non proline residues were summarized in table VII. A comparison of the results obtained from the Geno 3D, Swiss Model and Modeller, three different software tools in table VII shows that the models generated by Modeller was more acceptable in comparison to that by Geno3D and Swiss Model. The final modeled structures were visualized by Swiss PDB Viewer that was shown in Figure 1.

The stereochemical quality of the predicted models 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 models generated by modeller was shown in Figure 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 over-all 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 APX, DHAR, PHGPX and 2CPS has 96.2%, 80.0%, 95.0% and 91.5% residue respectively in allowed region. The distribution of the main chain bond lengths and bond angles were found to be within the limits for these proteins. Such figures assigned by Ramachandran plot represent a good quality of the predicted models.

The modeled structures of spinach antioxidant proteins were also validated by other structure verification servers WHAT IF. Standard bond angles of the four models are determined using WHAT IF. The results were shown in Table VIII. The analysis revealed RMS Z-scores were almost equal to 1 suggesting high model quality. The predicted structures conformed well to the stereochemistry indicating reasonably good quality.

1: A   1: B

1: C   1: D
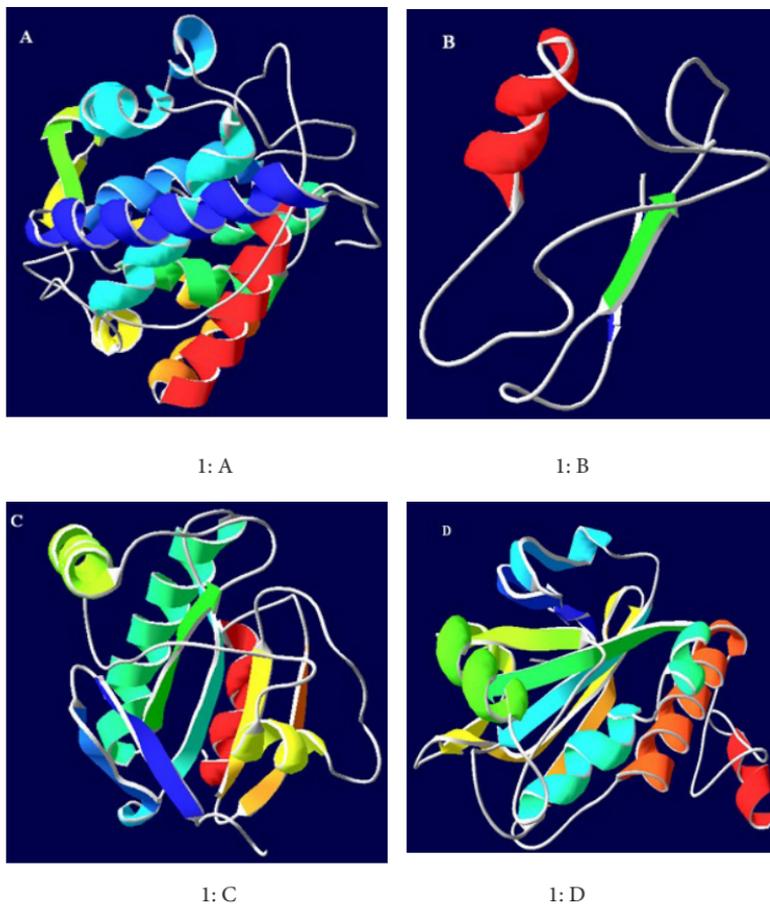
Figure 1: Modeled Structure of spinach proteins (A) Q42459 (B) Q9FVE4 (C) O23814 (D) O24364.

2: A

2: B

2: C

2: D
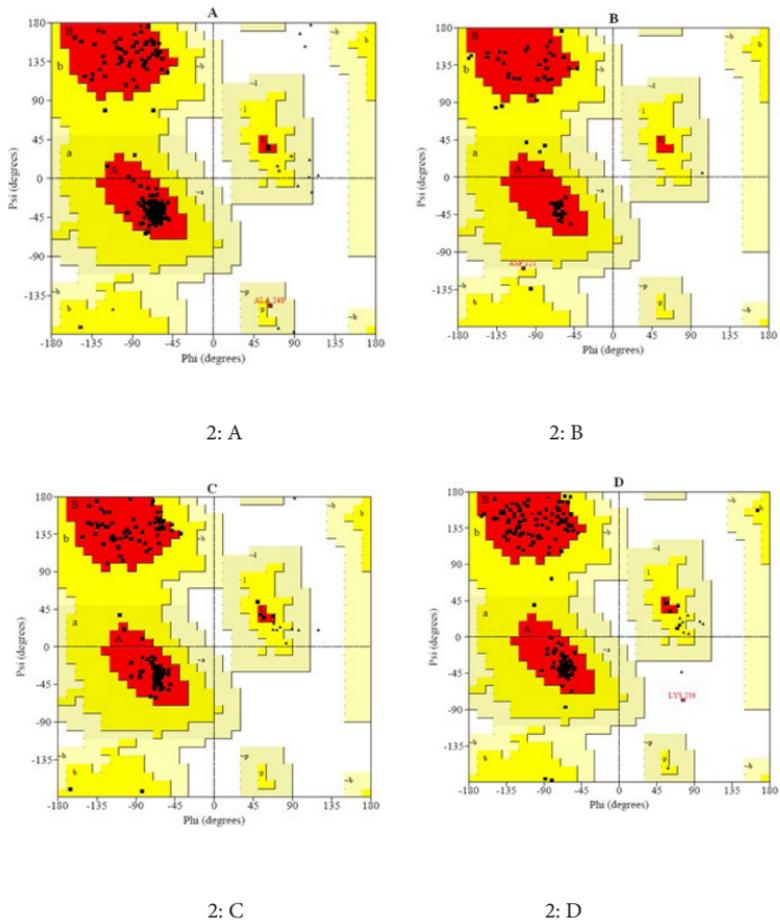
Figure 2: Ramachandran's Map of spinach proteins. (A) Q42459 (B) Q9FVE4 (C) O23814 (D) O24364.

Conclusion

In this study antioxidant protein of spinach were selected. Physico-chemical characterization were 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 these proteins was performed by SOSUI server. For these proteins disulphide linkages, motifs and profiles were predicted. Secondary structure analysis revealed that random coils dominated among secondary structure elements followed by alpha helix, extended strand and beta turns for all sequences. The modelling of the three dimensional structure of the proteins were performed by three automated homology programs, Geno 3D, Swiss model and Modeller. The models were validated using protein structure checking tools PROCHECK and WHAT IF. These structures will provide a good foundation for functional analysis of experimentally derived crystal structures.

Acknowledgement

References

1.	Wang Y, Chang CF, Chou J, Chen HL, Deng X, et al. Dietary supplementation with blueberries, spinach, or spirulina reduces ischemic brain damage. Exp Neurol. 2005; 193: 75-84.

2.	Lomnitski L, Foley J, Ben-Shaul V, Grossman S, Maronpot RR, et al. Effects of apocynin and natural antioxidant from spinach on inducible nitric oxide synthase and cyclooxyge-

nase-2 induction in lipopolysaccharide-induced hepatic injury in rat. Pharmacol Toxicol. 2000; 87: 18–25.

3. Nyska A, Suttie A, Bakshi S, Lomnitski L, Grossman S, et al. Slowing tumorigenic progression in TRAMP mice and prostatic carcinoma cell lines using natural antioxidant from spinach, NAO-a comparative study of three antioxidants. Toxicol Pathol. 2003; 31: 1–13.

4. Bergman M, Varshavsky L, Gottlieb HE, Grossman S. The antioxidant activity of aqueous spinach extract: chemical identification of active fractions. Phytochemistry. 2001; 58: 143–152.

5. Ahmad S, Pardini RS. Mechanisms for regulation of oxygen toxicity in phytophagus insects. Free Radical Biol Med. 1990; 8: 401–413.

6. Shigeoka S, Ishikawa T, Tamoi M, Miyagawa Y, Takeda T, et al. Regulation and function of ascorbate peroxidase isoenzymes. J Exp Bot. 2002; 53: 1305–1319.

7. Barbehenn RV. Gut-based antioxidant enzymes in a polyphagous and graminivorous grosshopper. J Chem Ecol. 2002; 28: 1329–1347.

8. Ishikawa T, Yoshimura K, Sakai K, Tamoi M, Takeda T, et al. Molecular characterization and physiological role of a glyoxysome-bound ascorbate peroxidase from spinach. Plant Cell Physiol. 1998; 39: 23–34.

9. Asada K, Takahashi M. Production and scavenging of active oxygen in photosynthesis. In: DJ Kyle, CB Osmond, CJ Arntzen, editors. Photoinhibition: Topics of Photosynthesis. Elsevier Amsterdam. 1987; 9: 227–287.

10. Asada K. Mechanisms for scavenging reactive molecules generated in chloroplasts under light stress. In: NR Baker, JR Bowyer, editor. Photoinhibition of Photosynthesis: From

Molecular Mechanisms to the Field. Oxford: Bios Scientific Publishers. 1994; 129–142.

11. Baier M, Dietz K-J. Protective Function of Chloroplast 2-Cysteine Peroxiredoxin in Photosynthesis. Evidence from Transgenic Arabidopsis. Plant Physiology. 1999; 119: 1407–1414.

12. Baier M, Dietz K-J. The costs and benefits of oxygen for photosynthesizing plant cells. Prog Bot. 1998; 60: 282–314.

13. Canfield LM, Forage JW, Valenzuela JG. Carotenoids as cellular antioxidants. Proc Soc Exp Biol Med. 1992; 200: 260–265.

14. Elstner EF. Der Sauerstoff. Bl-Wissenschafts-Verlag. Mannheim. 1990; 418–422.

15. Herbette S, Lenne C, Leblanc N, Julien JL, Drevet JR, et al. Two GPX-like proteins from Lycopersicon esculentum and Helianthus annus are antioxidant enzymes with phospholipid hydroperoxide glutathione peroxidase and thioredoxin peroxidase activities. Eur J Biochem. 2002; 269: 2414–2420.

16. Ursini F, Maiorino M, Gregolin C. The selenoenzyme phospholipid hydroperoxide glutathione peroxidase. Biochim Biophys Acta. 1985; 839: 62–70.

17. Bairoch, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 2000; 28: 45-48.

18. Gill SC, Von Hippel PH. Extinction coefficient. Anal Biochem. 1989; 182: 319-328.

19. Guruprasad K, Reddy BVP, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Prot Eng. 1990; 4: 155-164.

20. Ikai AJ. Thermo stability and aliphatic index of globular proteins. J Biochem. 1980; 88: 1895-1898.

21. Kyte J, Doolottle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982; 157: 105- 132.

22. Gasteiger E. Protein Identification and Analysis Tools on the ExPASy Server. In: John M Walker, editor. The Proteomics Protocols Handbook. Humana Press. 2005; 571-607.

23. Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJA, et al. The PROSITE database, its status in 2002. Nucl Acids Res. 2002; 30: 235-238.

24. Geourjon, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995; 11: 681-684.

25. Combet C, Jambon M, Deleage G, Geourjon C. Geno3D: Automatic comparative molecular modelling of protein. Bioinformatics. 2002; 18: 213-214.

26. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006; 22: 195-201.

27. Sali A, Blundelll TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol. 1993; 234: 779-815.

28. Van Der, Spoel D, Lindahl E, Hess B, Groenhof G, et al. GROMACS: Fast, Flexible and Free. J Comp Chem. 2005; 26: 1701-1718.

29. Ramachandran GN, Ramakrishnan C, Sasisekhran V. Stereochemistry of polypeptide chain configuarations. J Mol Biol. 1963; 7: 95-99.

30. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996; 8:477-486.

31. Vriend G. WHAT IF: A molecular modeling and drug design program. J Mol Graph. 1990; 8: 52-56.

32. Guex N, CP Manuel. Data Modeling, Analysis and Classification SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis. 1997; 18: 2714-2723.

33. Christian JAS, Cerutti L, Hulo N, Gattiker A, Falquet L, et al. PROSITE: A documented database using patterns and profiles as motif descriptors. Briefings in Bioinformatics. 2002; 3: 265–274.