ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

Comparative Protein Modeling of Superoxide Dismutase Isoforms in Maize.

Kaliyugam Shiriga[1, 2], Rinku Sharma[1], Krishan Kumar[2], Firoz Hossain[1] and NepoleanThirunavukkarasu[1*]

Division of Genetics, Indian Agricultural Research Institute, New Delhi – 110012, India[1]

School of Life Sciences Singhania University, Rajasthan – 333515, India[2]

[*]Corresponding author

Abstract: Superoxide dismutase (SOD) is one of the major classes of antioxidant enzymes, which protects the cellular and subcellular components against harmful reactive oxygen species. In maize, three types of SODs are present based on their constituent metal ions, namely Cu /Zn-SOD, Mn-SOD and Fe-SOD. In this study we critically assess the phylogenetic relationship and structural models for maize Cu/Zn SOD, Mn SOD and Fe SOD. The phylogenetic analysis showed that Mn-SOD and Fe-SODs in maize have a greater degree of similarity between them as compared to Cu/Zn-SODs. The secondary structure of Mn-SOD and Fe-SODs demonstrated similar characteristics of helices, sheets, turns and coils revealed that Mn-SOD and Fe-SOD were closely related, whereas Cu/Zn-SOD evolved independently. We found Mn-SOD and Fe-SOD had structurally analogous beta sheets. Homology modeling-enabled three-dimensional structure prediction helped to understand the molecular functions of SOD proteins in improving tolerance to oxidative stress in maize.

Keywords: Maize, superoxide dismutase, phylogenetic analysis, Protein modelling, secondary structure.

I. INTRODUCTION

Maize (Zea mays L.) is one of the most important global food crops, and used for various applications, ranging from food and feed to industrial purposes. Maize crop frequently affected by biotic and abiotic stresses that cause injury, limit their growth and adversely affect their productivity. The most common result of such stress is the production of toxic reactive oxygen species (ROS). ROS such as superoxide radicals (O_2-), hydroxyl radical (OH-) and H_2O_2, are crucial for many physiologic processes and usually exist in the cell in a balance with antioxidants. However, excess ROS disrupts normal biological functions. This condition is referred to as "oxidative stress". Superoxide dismutases are ubiquitous enzymes and involved in protection from oxygen toxicity. SODs are metalloproteins that occur in three isoforms: copper zinc SOD (Cu/Zn SOD), manganese SOD (Mn-SOD), iron SOD (Fe-SOD), all of which are highly stable because of the β-barrel structure and low content of α-helix strands.

Protein sequence comparison is the greatest powerful tool in characterizing proteins because of the information stored in conserved protein domains throughout the evolutionary procedure. Sequence comparison is very useful in homologous proteins, because they share common active sites or binding domains. The comparative study of protein structures allows the study of functional relationships between proteins and it bears huge importance in homology search and threading methods in structure prediction [1]. Multiple structure alignment of protein is desirable in order to group proteins into families, which allows a subsequent analysis of evolutionary issues. Detailed computational studies of protein sequence homology are essential for a range of purposes, therefore protein sequence homology study become routine in computational molecular biology and bioinformatics field. It is also seen that protein structure prediction is possible through bioinformatics tools. The functional analysis has been essential to confirm such predictions. Computational tools provide researchers to understand physicochemical
and structural properties of proteins. A large number of computational tools are available from diverse sources for making predictions concerning the identification and structure prediction of proteins. The major disadvantages of experimental methods that have been used to distinguish the proteins of various organisms are the time frame involved, high cost and the fact that these methods are not agreeable to high throughput techniques. In silico approaches offer a viable solution to these problems.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

The present investigation is an effort to analyze the sequence of maize SODs by using computational tools and techniques in order to understand the biological functions and evolutionary relationships among the maize SODs. We performed comparative in-silico structural and phylogenetic analysis of maize SOD isoforms (Cu/Zn, Mn and Fe SOD) with Arabidopsis SODs, to understand the similarity and dissimilarity among them at sequence level as well as at protein structural level.

II. MATERIALS AND METHODS

A. Maize and Arabidopsis SOD sequences collection

Sequences of genes and proteins corresponding to SOD isozymes in *Arabidopsis*, namely, Mn-SOD (AT3G10920), Fe-SOD (AT4G25100) and Cu/Zn-SOD distributed in chloroplasts (AAD10208) and cytosol (AT1G08830), in maize cytosolic SOD9 (NM_001111953) chloroplast Cu/Zn-SOD (EU408345), Mn-SOD (NM 001138523) and Fe-SOD (AB201543) sequences were collected from NCBI database. These sequences were used for phylogenetic analysis and for assessment of homology at the protein level.

B. Phylogenetic analysis of sequences

All sequences were aligned with a gap open penalty of 10 and a gap extension penalty of 0.2 using ClustalW. The result of the sequence alignment was used for constructing an unrooted phylogenetic tree by the Neighbor-Joining method. The confidence level of monophyletic groups was estimated using bootstrap analysis of 1000 replicates. Multiple sequence alignment (MSA) and phylogenetic analysis were performed using MEGA 5.10.

C. Secondary structure prediction of SOD proteins

SOPMA(Self Optimized Prediction Method with Alignment)[2] was employed for calculating the secondary structural features of the SOD protein sequences considered for this study. The secondary structure was predicted by using the default parameters, window width 17, similarity threshold 8 and number of states 4.

D. 3D structure

The 3D structure models for maize and *Arabidopsis* SODs were developed using Phyre2 (Protein Homology/Analog Y Recognition Engine; http://www.sbg.bio.ic.ac.uk/phyre2) for predicting the protein structure by homology modeling under 'intensive' mode.

E. Model evaluation

The dihedral angles Φ vs Ψ of amino acid residues in the protein structures were visualized and analyzed with Ramachandran plots [3]. The evaluation of models predicted in-silico is essential in order to avoid errors resulting from trivial and non-trivial mistakes. To avoid ambiguities and to improve accuracy, the predicted SOD models were evaluated using the ProSA and SAVS (Structural Analysis and Verification Server) web servers. For a specific PDB structure, ProSA calculates the overall quality score and validates a low resolution structure for approximate models using C-alpha atoms of the input structure. The output provides a z-score for the model that indicates the overall model quality; this value was determined from the plot during structure prediction.

III. RESULTS AND DISCUSSION

A. Phylogenetic analysis

The phylogenetic relationships of maize SODs were evaluated with respect to *Arabidopsis* SODs by using the Neighbor-Joining method implemented in MEGA5.10. Phylogenetic analysis of the maize and *Arabidopsis* provided information on the evolutionary ancestry development of all the SOD groups. The protein structure and catalytic residues that determine the substrate specificity are generally conserved within the family. Our phylogenetic analysis showed that SODs are segregated into two major clusters, with cytosolic, chloroplast Cu/Zn in one cluster and Mn-SOD and Fe-SOD in another. Phylogenetic analysis showed an evolutionary relationship between maize and *Arabidopsis* SODs, as maize cytosolic SOD was segregated with *Arabidopsis* cytosolic SOD (AT1G08830), Zmchloroplast Cu/Zn-SOD (EU408345) with AtCu-Zn chloroplast SOD (AAD10208), ZmMn-SOD (NM 001138523) with AtMn-SOD (AT3G10920) and ZmFe-SOD (AB201543) with AtFe-SOD (AT4G25100) (Figure 1).

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

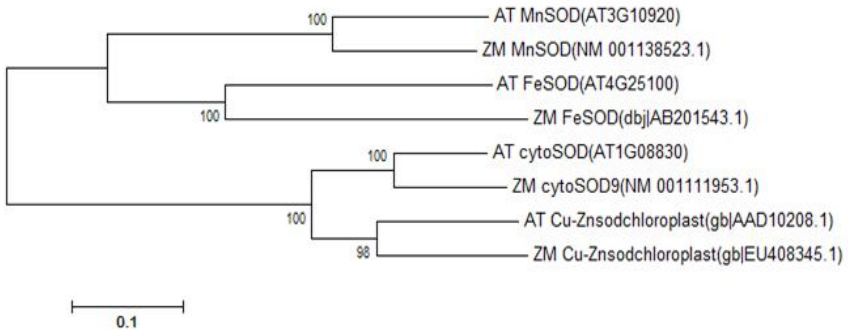

Figure 1. Phylogenetic tree of maize and Arabidopsis constructed by the Neighbor-Joining method and 1000 bootstrap replicates.

Sequence homology in grouping reflected the strong similarities in the genetic patterns of these two maize and *Arabidopsis* plants. However, many sequence-based families are polyspecific, which means they include genes that encode proteins with different functions. Phylogenetic tree reflects gene duplication and evolutionary divergence, with the acquisition of new protein functions [4].

B. Secondary structure analysis

The secondary structure of maize and *Arabidopsis* SOD proteins were predicted by SOPMA. Secondary structural features as predicted using SOPMA are represented in Table I. The results revealed a clear cluster separation between Mn-SOD, Fe-SOD with Cu-Zn, and Cytosolic SOD by distribution of alpha helix, beta sheet, beta turn and coil. Coils dominated among secondary structure elements followed by alpha helix, beta sheet, and beta turns for all sequences in Mn-SOD and Fe-SOD, whereas, in Cu/Zn-SOD, coils dominated followed by beta sheet, alpha helix and beta turns. Cytosolic SOD showed the highest number of residues in coil, followed by beta sheet, beta turn and alpha helix. The percentage of β-sheets in Fe-SOD and Mn-SOD were 12.7%, 12.9% in Arabidopsis and 14.3%, 15% in maize respectively.

C. Three dimensional protein structure modeling

Proteins are complex chemical entities with a large number of variable atoms and a convoluted topology that makes their description complicated [5]. The rapid increase in the number of gene sequences being deposited in databases means that there is a need to identify the functions that form the basis of defining protein groups. Three-dimensional models of maize and Arabidopsis Cu/Zn, Mn and Fe-SOD proteins having higher homology were selected (Figure 2).

TABLE I
SECONDARY STRUCTURE OF ARABIDOPSIS AND MAIZE SOD PROTEINS

Secondary structure	AtCu/Zn-SOD %	ZmCu/Zn-SOD %	AtCyto SOD%	ZmCyto SOD%	AtFe-SOD %	ZmFe-SOD %	AtMn-SOD %	ZmMn-SOD %
Alpha helix	15.28	16.50	6.58	1.97	55.19	44.57	52.81	49.79
Beta sheet	34.26	30.10	34.87	36.18	12.74	14.34	12.99	15.02
Beta Turn	12.50	11.17	12.50	11.18	5.66	4.65	6.49	8.15
Coil	37.96	42.23	46.05	50.66	26.42	36.43	27.71	27.04

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

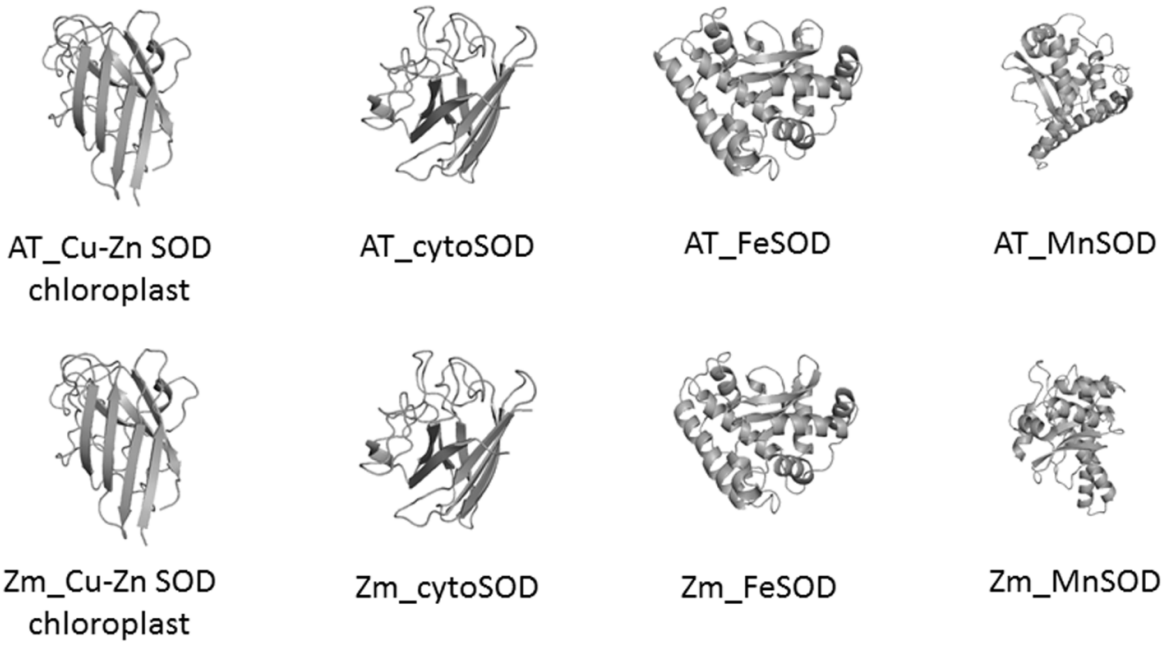

Figure 2. Three dimensional structural representation of maize and Arabidopsis SODs.

D. Model evaluation

The stereo chemical quality and accuracy of the predicted protein model was analyzed using residue-by-residue geometry and the overall geometry of the protein structures was analyzed with Ramachandran plots provided by the SAVS (Structural Analysis and Verification Server) web server. These plots help visualize the dihedral angles Φ vs Ψ of amino acid residues in proteins. Figure (3) shows the quantitative evaluation of maize and Arabidopsis protein structures done using Ramachandran plots. The main chain parameters plotted are Ramachandran plot quality, 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/favored regions, whereas the yellow regions represent allowed regions. Glycine is represented by triangles and other residues are represented by squares.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

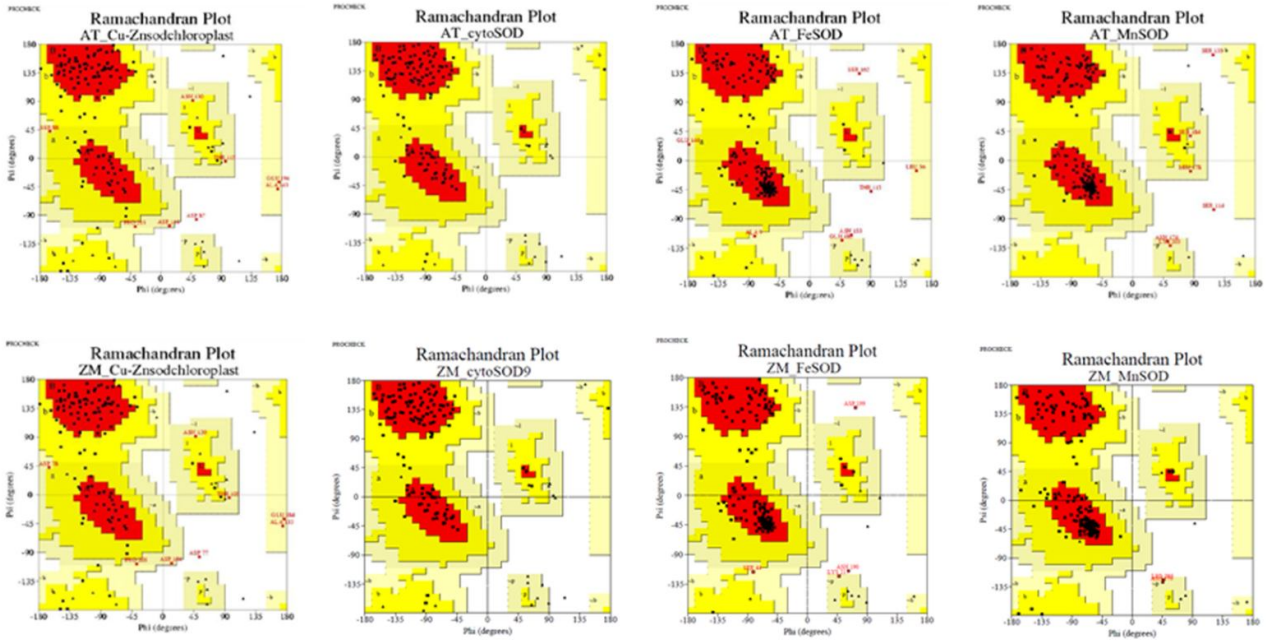

Figure 3.Validation of SOD structures using Ramchandran plots. The Ramachandran plots revealed that > 90% of SOD amino acid residues from the modeled Arabidopsis structure were incorporated in the favoured and allowed regions of the plot.

Chloroplast Cu/Zn-SOD of both plants had almost same percentage of residues in all regions. The absence of residues in the αL region in both structures was an interesting feature of this Cu/Zn protein and indicated that the structure had no left helices. Interestingly, Cytosolic SODs from both the plants had no residues in their generously allowed and outlier regions. All of the Cu/Zn and Cytosolic structures had fewer residues in the αR region compared to the β region. The maize Fe-SOD structure had 97.8% of its residues in expected regions and a negligible proportion (1.1%) in the outlier region; the corresponding values for Arabidopsis were 96.1% and 1.7%, respectively. Maize Mn-SOD showed higher quality compared to the Arabidopsis protein, with no residues in the outlier region and more residues in the allowed region. The Fe-SOD and Mn-SOD structures had good clustering of residues and a greater number of helices. The structural quality of chloroplast Cu/Zn-SOD of both the plants was lower than in the remaining structures. From all of the structures, the αL region was much less populated, i.e., few or no residues. The distribution of the main chain bond lengths and bond angles were found to be within the limits for these proteins. The prominence of residues in the αR and β regions suggested that the structure was rigid with many right helices. Such figures assigned by Ramachandran plot represent a good quality of the predicted models. The result revealed that the modeled structure for maize Cu/Zn, Cytosolic, Mn and Fe SODs has 94.2%, 100.0%, 98.9% and 96.8% residue respectively, in favored and allowed region.

The Z-score value, a measure of model quality that predicts the total energy of the structure [8], was predicted for Maize and Arabidopsis SODs using the ProSA server (Protein Structure Analysis). The ProSA-web gives z-scores of all protein chains in the PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length (Figure 4). The plot shows only chains with less than 1000 residues and a z-score ≤ 10. The Z-score values for chloroplast and Cytosolic SODs were -6.13 and -6.82, respectively; the corresponding values for Mn-SOD and Fe-SOD were -7.18 and -8.48, respectively.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

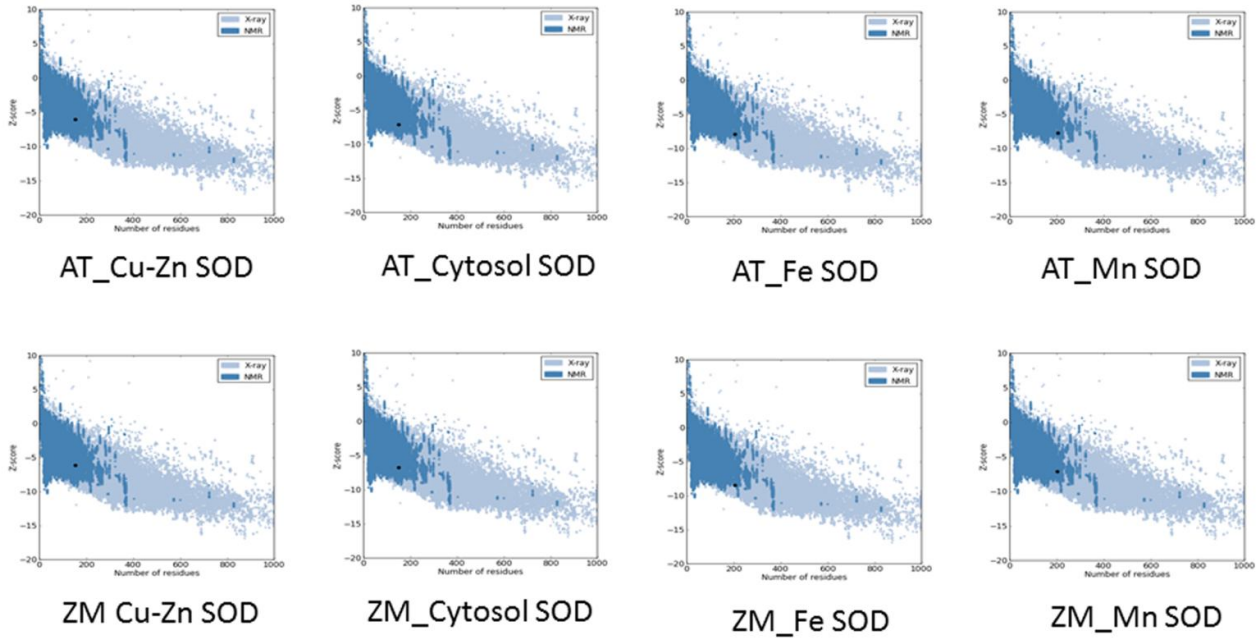

Figure 3. ProSA-web z-score chimeric protein plot. The z-score indicates overall model quality. The ProSA-web z-scores of all protein chains in PDB were determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length. The plot shows results with a z-score < 10. The z-score for SOD is highlighted as a large dot. The value is within the range of native conformations.

IV. CONCLUSION

In our study, we selected maize SODs isoforms consist of metal ion for comparative analyses with Arabidopsis SODs and, a phylogenetic tree was constructed to know their functional relationship. Results showed that Mn-SOD and Fe-SOD were in one cluster while, chloroplast Cu/Zn-SOD and Cytosolic SOD formed separate clusters. The secondary structure of SOD proteins revealed that the distribution of helix, sheets and turns in Mn-SOD and Fe-SOD were same, while they were different in chloroplast Cu/Zn-SOD and Cytosolic SOD. Mn-SOD and Fe-SOD had a similar number of β-sheets in their secondary structure. Cytosolic SOD had different structural pattern as compared to other SODs. These analyses provided insights into the molecular function of SOD isoenzymes with respect to their interactions with different cellular organelles. Further studies can lead to design new proteins having all possible characters of SODs to produce maize cultivars tolerant to oxidative stress.

ACKNOWLEDGEMENT

The authors sincerely acknowledge the ICAR network project on Transgenics in Crops (Maize Functional Genomics Component) for funding the study.

REFERENCES

[1] Balasubramanian, A., Das, S., Bora, A., Sarangi, S. , and Mandal, A. B, "Comparative Analysis of Structure and Sequences of Oryza sativa Superoxide Dismutase", *American Journal of Plant Sciences*,vol. 3, pp.1311–1321, 2012.

[2] Geourjon, C., and Deléage, G.,"SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments", *Computer Applications in Biosciences*, vol.11,pp. 681–684, 1995.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

[3] Ramachandran, G.N., Ramakrishnan, C. , and Sasisekharan, V., "Stereochemistry of polypeptide chain configurations", *Journal of Molecular Biol*ogy, vol. 7, pp. 95–99, 1963.
[4] Emanuele, L ., Yi, L., and McQueen-Mason, S. J., "Phylogenetic analysis of the plant endo--1, 4-glucanase gene family", *Journal of Molecular Evolution*, vol. 58, pp.506–515, 2004.
[5] Ingale, A.G., and Chikhale, N.J., "Prediction of 3D structure of paralytic insecticidal toxin (ITX-1) of Tegenariaagrestis (hobo spider)", *Journal of Data Mining in Genomics and Proteomics*, vol.1, pp.102–104, 2010.
[6] Soding, J., "Protein homology detection by HMM-HMM comparison", *Bioinformatics*, vol.21, pp. 951–960, 2005.
[7] Jefferys, B.R., Kelley, L.A., Sternberg, M.J.E. , "Protein folding requires crowd control in a simulated cell", *Journal of Molecular Biology*,vol. 397, pp. 1329–1338, 2010.
[8] Wiederstein, M., and Sippl, M.J. , "ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins", *Nucleic Acids Res*earch, vol.35, pp. 407–410, 2007.