Sequence to Structure Analysis of SOD1 and SOD2 from Fresh Water Turtles

Rituparna Sarma¹, Dhirendra K Sharma²*

¹Bioinformatics Centre, Gauhati University, Guwahati, Assam, India
²School of Biological Sciences, University of Science and Technology, Meghalaya, Techno city, Baridua, India

RS and DKS designed the research work. Manuscript was prepared and verified by both the authors.
RS carried out the experiments.

Corresponding Author: Dhirendra Kumar Sharma, School of Biological Sciences, University of Science and Technology, Meghalaya, Techno city, Kiling Road, Baridua-793101, India. Tel: +91 9435147412; E-mail: dksgu@yahoo.co.uk

Abstract
Superoxide dismutase (SOD) responsible for dismutation of ROS produced in cell controls the aging and longevity of animals. An attempt has been made to report on sequence to structure analysis of the genes and proteins of SOD1 and SOD2 of freshwater turtles, Pelodiscus sinensis and Mauremys reevesii. Analysis of gene and protein sequences of these SODs retrieved from the NCBI database suggested that there were minor variations in their molecular weight of the gene sequences, melting temperature, folding, aliphatic index and isoelectric point. Gene sequences were all AT rich with 5 restriction sites each in SOD1 of both the turtles and SOD2 of Pelodiscus sinensis while 8 restriction sites in SOD2 of Mauremys reevesii were obtained. SOD1 were dominated by β Strands, whereas, SOD2 were by the alpha helices. Homology models were generated by MODELLER 9.12 presented that all the models of SODs within acceptable range. Solvent accessible surface area (SASA) and active sites analysis of refined models of the SOD proteins were acidic and with 5 to 11 number of active sites in all the proteins and high percentage of exposed aliphatic residues. Therefore, it could well be inferred that these models have the potentiality to be used for understanding the aging process.

Keywords: SOD1; SOD2; Structural characterization of Protein; Turtles P. sinensis; M. reevesii

Introduction
Superoxide dismutase (SOD) is the most effective enzyme in the aging process and longevity, regulates the Reactive Oxygen Species (ROS), produced in the metabolic and physiological events of animals¹⁻³. Unbalanced concentration of ROS often contributes to diseases like cancer, diabetes, premature aging, inflammation and hypertension⁴. The SOD1 is found in cytoplasm and outer mitochondrial space, protects the cells against any lethal effects of radiation, drugs or toxicity of ROS⁵ while the SOD2 found in inner mitochondrial space, promotes cellular differentiation, apoptosis, tumorgenesis and hypoxia induced pulmonary disease⁶⁻⁸. Among all the Superoxide dismutase found in organisms, only the SOD1 (Cu, Zn SOD) and SOD2 (Mn SOD) have been sequenced so far in a limited number of chelonians, but without structural information⁹. The group turtles, many of them survived for a prolonged period and that too with very active life has been an interesting model to understand the aging¹¹,¹². Reliable 3-D structural predictions using Homology modeling of these proteins might be significant to understand the aging process¹³. Pelodiscus sinensis Wiegmann, 1835 and Mauremys reevesii Grey, 1831 are the two freshwater turtles¹⁴ have often been used as models for turtle evolution and development studies, where SOD1 and SOD2 have been sequenced recently¹⁵⁻¹⁷. Therefore, an attempt has been made to analyze the SOD gene and protein sequences and their characterization (SOD1 and SOD2) in Pelodiscus sinensis and Mauremys reevesii at structural and functional level. Prediction of the secondary structures, analysis of gene and protein sequence properties, restriction sites, homology modeling and evaluation, Solvent Accessible Surface Area (SASA) and
active site prediction of the SODs of Pelodiscus sinensis and Mauremys reevesii were carried out to understand their structural and functional status.

Methods

Sequence Retrieval and Sequence Analysis

SOD nucleotide and protein sequences of Pelodiscus sinensis (GenBank SOD1: GenBank: JX470524.1, GenPept: AEK80392.1, SOD2: GenBank: JX470525.1, GenPept: AEK22120.1) and Mauremys reevesii (SOD1: GenBank: JX843790.1, GenPept: AFX95918.1, SOD2: GenBank: JX843791.1, GenPept: AFX95919.1) were retrieved from NCBI. Sequence lengths of SOD1 (155 aa) and SOD2 (226 aa) were similar in both the turtle species. Nucleotide lengths of SOD1 were 727 bp and 749 bp, while the SOD2 had 1436 bp and 1687 bp in Pelodiscus sinensis and Mauremys reevesii respectively. Nucleotide and protein sequences were run in CLC workbench package (CLC Bio) to analyze sequence properties, secondary structure prediction as well the restriction sites.

Homology Modeling of Protein and Evaluation

The 3D structures of SOD1 and SOD2 were generated using comparative method in MODELLER 9.12[20]. Validation of the models was done by PROCHECK[21] and RAMPAGE[22]. The minimization of energy and refinement of protein structures were carried out by Discovery Studio package (Accelrys)[23] and Chiron[24]. Refined structures were evaluated using PROSESS[25], PROCHECK and RAMPAGE and ProFunc[26].

Solvent Accessible Surface Area (SASA) and Active Site Prediction

SASA and active site predictions were carried out by using the best refined model structures of the proteins. Get Area[27], Discovery Studio Client 4.0 was used to find out the SASA percentage. Active sites and cleft predictions of protein structures were carried out by Discovery Studio package (Accelrys)[23] and Chiron[24]. Refined structures were evaluated using PROSESS[25], PROCHECK and RAMPAGE and ProFunc[26].

Restriction Enzyme Digestion

The nucleotide SOD sequences were restriction digested in-silico with 8 restriction enzymes namely, BglII, EcoRI, FokI, HindIII, MspI, PstI, Smal and Xbal based on 4-base and 6-base cutters, virtual gene map of the gene sequences (Table 2). Results showed that the Gene sequences of SOD1 and SOD2 of Pelodiscus sinensis have 5 restriction sites, while the SOD2 of Mauremys reevesii presented 8 restriction sites.

Protein Sequence Analysis

Analysis of protein sequences for SOD1 of Pelodiscus sinensis had molecular weight 15.819 kDa, isoelectric point 6.47 with aliphatic index at 76.71 whereas SOD1 of Mauremys reevesii had 16.022 kDa molecular weight, isoelectric point 6.87 and aliphatic index at 82.78. SOD2 in Pelodiscus sinensis demonstrated molecular weight of 25.166 kDa, isoelectric point 9.01 and aliphatic index 83.319 compared to 25.049 kDa, 9.04 and 83.761 respectively for SOD2 of Mauremys reevesii (Table 3).
Model Prediction and Evaluation

MODELLER 9.12 was used to predict the 3D structures of SOD1s and SOD2s for _Pelodiscus sinensis_ and _Mauremys reevesii_. Final templates selected for MODELLER 9.12 were as follows; for SOD1, 3GTV_A (153aa, 78% identical for _P. sinensis_) and 3GTT_A (153aa and 75% identical for _M. reevesii_); and for SOD2, 1PL4_A (198aa, 89% identical for _P. sinensis_) and 1NO-J_A (199aa, 86% identical for _M. reevesii_). In MODELLER, 3D structures were generated comparing target sequences with the help of template sequences and aligned with the maximum similar structure on the basis of DOPE Score and mole PDF Score. In PROCHECK analysis, SOD1 of _P. sinensis_ and _M. reevesii_ scored 91.0% and 89.5% while the SOD2 scored at 92.7% and 96.7% respectively. On the other hand, in RAMPAGE analysis, SOD1 secured 98.0% against 98.7% in SOD2 for _P. sinensis_, compared to 12% of SOD2 in _M. reevesii_. β-strand and coil dominated the SOD1 structures with only 2-3% helix in both the models. In SOD2, helix was noted to be dominating the model with 58% in _M. reevesii_ against the 50% in the model of _P. sinensis_. Covalent bond and packaging bond qualities were within the acceptable range for all the SOD structures.

Discovery studio 4.0 calculated that the refined SOD1 of _P. sinensis_ had 1366nos. of atom at a molecular weight of 15,801.8 kDa with net formal charge (-2) demonstrating the chemical formula as C_{671}H_{1108}N_{206}O_{222}S_{7}, while the SOD2 had 2,177 number of atoms, molecular weight of 25,139.8kDa and net formal charge of 5 presenting the chemical formula C_{1129}H_{1747}N_{316}O_{220}S_{5}. Whereas, SOD1 of _M. reevesii_ had 1,383 atoms, 16,003.1 kDa molecular weight, net formal charge -3 showing the chemical formula as C_{307}H_{1104}N_{208}O_{221}S_{3}. The net formal charge indicated that the SOD1 of _P. sinensis_ as well the SOD1 of _M. reevesii_ had higher anion atoms, whereas, the SOD2 of _P. sinensis_, (MODELLER) had large numbers cationic charge. All the SODs were of single stranded. The size of the exposed protein groups were found to be higher than the size of the buried protein groups in all the SODs. Hydrophilic residue size was higher than the hydrophobic residue size in SOD1s, while vice-versa was detected in SOD2s.

Solvent Accessible Surface Area (SASA) and Active site predictions

Get Area server analysis predicted that the total solvent accessibility of SOD1s were 7508.24 and 7835.29 as well for SOD2 was 13424.35 and 8767.77 respectively for _P. sinensis_ and _M. reevesii_. A polar surface area was found to be dominant in all the proteins. Likewise, in Discovery studio it was found that solvent accessibility for SOD1 and SOD2 was 7,663.26 and 11,653.98 kDa respectively.
13,724.3 for \textit{P. sinensis} models while 7,978.09 and 8,964.37 was achieved for \textit{M. reevesii} respectively.

The pocket properties were tabulated in (Table 4). The largest active site volume (Å$^3$) in SOD1 of \textit{P. sinensis} was 579.50 against 333.06 in \textit{M. reevesii}, whereas, the SOD2 had presented the largest active site volume with 559.17 and 991.94 respectively for both the model of \textit{P. sinensis} and \textit{M. reevesii}. Gly residue was found to be dominant in both the largest active sites of SOD1 models followed by Val residues. Hydrophobicity ratio of 0.38 of SOD1 model in \textit{M. reevesii} was better than the 0.52 of \textit{P. sinensis} in the largest active site. However, in SOD2 model, the largest pockets had the highest number of amino acid residues of Gly and Leu in both the \textit{P. sinensis} and \textit{M. reevesii} models with hydrophobicity ratio of 0.49 and 0.65 respectively. ProFunc results showed that the SOD1 models of \textit{P. sinensis} and \textit{M. reevesii} had the largest cleft size of 2045.67 Å and 2234.67 Å, while SOD2 had 1666.41 Å and 2122.88 Å respectively. Comparison of residue types in the largest clefts suggested that the SOD1 model of \textit{P. sinensis} had dominant equal numbers of positive, neutral and aliphatic residues (10 residues in each) while the model of \textit{M. reevesii} was aliphatic residue dominant (11 residues in each) followed by negative residues (10 residues). The SOD2 was with positive and neutral residues (6 residues each) however, was dominant in the model of \textit{P. sinensis} while aliphatic (11 residues) followed by aromatic residues (10 residues) were dominant in the model of \textit{M. reevesii}.

Table 4: Cavity size and properties of SOD1 and SOD2 of \textit{P. sinensis} and \textit{M. reevesii} analyzed using DogSiteScorer server

<table>
<thead>
<tr>
<th></th>
<th>\textit{P. sinensis}</th>
<th>\textit{M. reevesii}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textbf{P0}</td>
<td>Naume: 597.50</td>
<td>Naume: 333.06</td>
</tr>
<tr>
<td></td>
<td>Vol: 125.53</td>
<td>Vol: 127.37</td>
</tr>
<tr>
<td></td>
<td>Sur: 323</td>
<td>Sur: 7.61</td>
</tr>
<tr>
<td>\textbf{P1}</td>
<td>Naume: 426.61</td>
<td>Naume: 228.67</td>
</tr>
<tr>
<td></td>
<td>Vol: 559.17</td>
<td>Vol: 506.07</td>
</tr>
<tr>
<td></td>
<td>Sur: 492</td>
<td>Sur: 519.95</td>
</tr>
<tr>
<td>\textbf{P2}</td>
<td>Naume: 352.34</td>
<td>Naume: 220.67</td>
</tr>
<tr>
<td></td>
<td>Vol: 281.15</td>
<td>Vol: 483.83</td>
</tr>
<tr>
<td></td>
<td>Sur: 582</td>
<td>Sur: 421.61</td>
</tr>
<tr>
<td>\textbf{P3}</td>
<td>Naume: 271.72</td>
<td>Naume: 178.94</td>
</tr>
<tr>
<td></td>
<td>Vol: 172.74</td>
<td>Vol: 414.18</td>
</tr>
<tr>
<td></td>
<td>Sur: 352</td>
<td>Sur: 123.20</td>
</tr>
<tr>
<td>\textbf{P4}</td>
<td>Naume: 390.76</td>
<td>Naume: 170.62</td>
</tr>
<tr>
<td></td>
<td>Vol: 162.69</td>
<td>Vol: 406.42</td>
</tr>
<tr>
<td></td>
<td>Sur: 294</td>
<td>Sur: 104.96</td>
</tr>
<tr>
<td>\textbf{P5}</td>
<td>Naume: 447.23</td>
<td>Naume: 120.70</td>
</tr>
<tr>
<td></td>
<td>Vol: 137.92</td>
<td>Vol: 143.76</td>
</tr>
<tr>
<td></td>
<td>Sur: 367</td>
<td>Sur: 101.25</td>
</tr>
</tbody>
</table>

Discussion

Nucleotide base composition analysis of SOD1s and SOD2s suggested that all the genes were AT rich where ‘Adenine’ was dominating (Table 1). DNA replication starts at the AT rich regions and these regions are universally the most conserved regions found in replicons\cite{39-41}. In both the SOD1, adenine percentage was followed by almost the same frequency ranges for guanine, thymine and low percentage for cytosine. While in SOD2, adenine percentage is marginally higher than the thymine followed by guanine and low cytosine percentage. Analysis of SOD1 and SOD2 nucleotide sequences showed that there were minor variations in melting temperatures and no coding regions were defined in the sequences. In silico analysis of restriction site variability has been suggestive of their differences may lead to high degree of polymorphism. Further, similarity in the conserved sequence is in fact suggestive predicted their identical longevity.

Aliphatic index of SOD1 of \textit{Mauremys reevesii} were higher, indicative of its stability over different temperatures ranges than the SOD1 of \textit{Pelodiscus sinensis}. Isoelectric point indicated that the pH of both the SOD1 were acidic when not carrying any net electrical charge. Results indicated that SOD2 of \textit{Mauremys reevesii} had more stability than the SOD2 of \textit{Pelodiscus sinensis}. Half-life of all the SODs were > 20 hrs. SOD1 of both the turtles had the similar frequency of hydrophobic residues, but there were variations in frequencies of hydrophilic residue with both the negative and positive charges. Where as in case of SOD2, charged residue frequencies were of same but there were variations in frequencies of hydrophobic and hydrophilic residues. Glycine (G) residue distribution was found to be dominant in both the SOD1, while the distribution frequency of leucine (L) was dominant in SOD2s. Due to dominance of Glycine, helix forming probabilities were low in SOD1, while such helix forming probabilities were evident in the SOD2 with leucine dominance. Moreover, all the four SOD sequences might be highly conserved since the Glycine and Leucine have low mutability and are more frequent in conserved sequence elements\cite{32}.

Prediction of secondary structures locates the positions of the amino acid residues, whether they lie in helixes, strands or in coils\cite{33}. Secondary structure prediction indicated that the SOD1 of both the \textit{Pelodiscus sinensis} and \textit{Mauremys reevesii} had same 12 numbers of \(\beta\) strands with no \(\alpha\) helices. Though the number of beta strands was same, yet variations at the 2nd, 3rd and 4th strand positions of both the structures were noticed. SOD2 of \textit{Pelodiscus sinensis} had 13 \(\alpha\) helices and 4 \(\beta\) strands compared to 12 \(\alpha\) helices and 5 \(\beta\) strands of SOD2 from \textit{Mauremys reevesii}. The structure was suggestive of the dominance of \(\alpha\) helix in both the SOD2. Further, sequence alignment suggested that the SOD1 and SOD2 of both the turtles had 90.32% and 97.78% conserved sequence similarity respectively.

From the ProFunc evaluation, it could be outlined that all the proteins were associated with cellular oxygen and reactive oxygen species and metabolic processes. SASA prediction helps in understanding the probable binding oriented conformational changes that may occur in the protein structures\cite{34}. It could be suggested from the results that the models had greater potentiality of binding to the ligands in solvent.

Active site predictions are essential for prediction of functions, classification and drug binding ability of proteins\cite{35}. It has been found that SOD1 of \textit{P. sinensis} had 6 numbers of pockets against the model of \textit{M. reevesii}, which had 7 pockets; on the other hand, the SOD2 had 11 and 5 numbers of pockets for the models of \textit{P. sinensis} and \textit{M. reevesii} respectively.
The aromatic residues at the active sites stabilize the monomers in a hydrophobic core. Since the binding site affinities and specificities are mainly achieved by hydrogen bond interactions\[50\], the results of active sites and clefts predictions, it could well be inferred that both the SOD1 and SOD2 proteins of the two turtles had acceptable and applicable range of active sites and cavity size.

Conclusion

Analysis of the nucleotide sequences of the SODs suggested that the genes were AT rich and had minor melting temperature differences with good number of restriction sites indicate high degree of polymorphism. Analysis of protein sequences of SOD1 and SOD2 and evaluation of the predicted secondary and tertiary refined structures of the two SODs generated by MODELLER 9.12 indicated that all the 4 SODs structure(s) were within acceptable range. Comparison among the models with their counterparts suggested that although they had the sequence similarity of around 90\%, yet all the SODs had their own individual structural characteristics. SOD structures were submitted to PMDB database under the IDs PMDB ID: PM0079765 (SOD1 of *P. sinensis*), PMDB ID: PM0079766 (SOD2 of *P. sinensis*), PM0079772 (SOD1 of *M. reevesii*) and PM0079773 for SOD2 of *M. reevesii*. Thus it could be assumed that these models and the data have the potentiality to be used as source for further understanding on the aging process and drug binding related attempts.

Conflict of Interest: There was no conflict of interest.

Acknowledgement

All the searches were carried out in Bioinformatics Facility Centre (BIFGU), Gauhati University as well in the Department of Zoology of the University of Science and Technology, Meghalaya. The authors gratefully acknowledge Department of Biotechnology (DBT), Government of India for their help and assistance.

References