Homology Modeling and Comparative Profiling of Superoxide Dismutase Among Extremophiles: *Exiguobacterium* as a Model Organism

Rajiv Pathak · Pankaj Narang · Muktesh Chandra ·
Raj Kumar · P. K. Sharma · Hemant K. Gautam

Received: 27 March 2014 / Accepted: 6 June 2014 / Published online: 20 June 2014
© Association of Microbiologists of India 2014

Abstract Superoxide dismutase (SOD), a well known antioxidant enzyme, is known to exert its presence across bacteria to humans. Apart from their well-known antioxidant defense mechanisms, their association with various extremophiles in response to various stress conditions is poorly understood. Here, we have discussed the conservation and the prevalence of SODs among 21 representative extremophiles. A systematic investigation of aligned amino acid sequences of SOD from all the selected extremophiles revealed a consensus motif D-[VLE]-[FW]-E-H-[AS]-Y-[YM]. To computationally predict the correlation of SOD with the various stress conditions encountered by these extremophiles, *Exiguobacterium* was selected as a model organism which is known to survive under various adverse extremophilic conditions. Interestingly, our phylogenetic study based on SOD homology revealed that *Exiguobacterium sibiricum* was one of the closest neighbors of

Deinococcus radiodurans and *Thermus thermophilus*. Next, we sought to predict 3-D model structure of SOD for *E. sibiricum* (PMDB ID: 0078260), which showed >95 % similarity with *D. radiodurans* R1 SOD. The reliability of the predicted SOD model was checked by using various validation metrics, including Ramachandran plot, Z-score and normalized qualitative model energy analysis score. Further, various physicochemical properties of *E. sibiricum* SOD were calculated using different prominent resources.

Keywords Extremophiles · Superoxide dismutase · *Exiguobacterium* sp. · Homology modeling

Introduction

Extremophiles are microorganisms, which can withstand severe environmental conditions like extreme levels of gamma radiation, temperature, salt-stress, acidic or alkaline conditions. Based on their forbearance, they are categorized as radioresistant, psychrophiles, thermophiles, halophiles, acidophiles and alkaliphiles, respectively [1]. Such type of various stress conditions encountered by these extremophiles result in the formation of reactive oxygen species (ROS) like superoxide, hydroxyl and singlet oxygen, which ultimately cause cell death. These ROS are quenched naturally by some antioxidant enzymes (AOE) like superoxide dismutase (SOD), catalase, glutathione peroxidase, metal ligands (Fe, Mn, Cu, Zn, and Ni) and ubiquinone [2]. Superoxide dismutase (SOD; EC 1.15.1.1) is a class of closely related metallo-antioxidant enzymes that catalyzes the breakdown of the superoxide anions into oxygen and hydrogen peroxide. They act as the first line of defence to ROS and have been categorized on the basis of their metal ligands like Cu/ZnSOD, NiSOD, FeSOD and MnSOD,

Electronic supplementary material The online version of this article (doi:10.1007/s12088-014-0482-8) contains supplementary material, which is available to authorized users.

R. Pathak · M. Chandra · H. K. Gautam (✉)
CSIR- Institute of Genomics and Integrative Biology, Sukhdev
Vihar, Mathura Road, Delhi 110020, India
e-mail: hemant@igib.res.in

P. Narang
School of Computational and Integrative Sciences, Jawaharlal
Nehru University, Delhi 110067, India

R. Kumar
Division of Radiation Biotechnology, Institute of Nuclear
Medicine and Allied Sciences, Timarpur, Delhi 110007, India

P. K. Sharma
Department of Microbiology, Ch. Charan Singh University,
Meerut 250004, India

commonly found in both prokaryotes and eukaryotes [3]. Based on 261 aligned sequences and 12 X-ray structures, it has been coherently shown that dimeric forms of MnSOD or FeSOD are emblematic for bacterial systems, but a tetrameric form of Mn- or FeSOD has also been shown in some prokaryotes, especially in hyperthermophiles [4].

Although investigators have been studying the role of SOD in quenching of ROS in a variety of microorganisms, comparatively little is known about the possible role of SODs in extremophiles, tolerating various stress conditions. In case of extremophiles, some existing reports demonstrate another prominent role of SODs in *Deinococcus radiodurans*, a very well known bacterium which is extremely resistant to both oxidative stress and ionizing radiation [5]. Superoxide dismutase (sodA) mutant of *D. radiodurans* was shown to exhibit a radiosensitive phenotype than the wild type, suggesting the potential role of SODs in regulating radioresistance properties [6]. In a recent study, up-regulated expression of several enzymes related to oxidative stress conditions like catalase (DR1998) and Mn⁺-dependent superoxide dismutase (DR1279) were observed in the case of *D. radiodurans* in response to 6 kGy dose of gamma irradiation, which further strengthens their possible role in radiation resistance [7].

Up-regulation of SOD has also been reported in *Caulobacter crescentus* in response to heavy-metal toxicity, which shows the possible role of SODs in another stressful condition [8]. Evidences from the mentioned reports bring to light the immense pivotal role that SODs may play in rescuing the effects arising under various stress conditions.

To gain a more comprehensive understanding about the role of SODs in various stress-responses, we selected *Exiguobacterium* sp., a gram positive extremophile which is tolerant towards high doses of gamma radiation, high salt stress conditions and has the inherent ability to grow within a temperature range of −2.5 to 40 °C [9]. *Exiguobacterium*, like other extremophiles possesses a free radical scavenging activity due to the presence of antioxidants like glutathione, catalase and SODs [10]. However, unlike most extremophilic microorganisms, biology of *Exiguobacterium* sp. and its adaptability to survive in such extreme environmental conditions is poorly understood. Accumulating evidences from different molecular studies and unusual behavior of *Exiguobacterium*, prompted us to further study the role of SOD in rescuing such type of stress conditions and their conservation among various extremophiles.

Materials and Methods

Sequences Retrieval and Phylogenetic Analysis

The amino acid sequences of SOD from 21 extremophiles were retrieved from NCBI Protein sequence database.

These sequences were aligned using ClustalW with default parameters [11] and conservation of amino acids based on their chemical nature was plotted as a circos plot using Circos [12]. The phylogenetic analysis of 16S rRNA and SOD sequences were performed by neighbor-joining method using Mega5 [13].

Comparative Modeling of SOD and its Evaluation

A 3D model of *Exiguobacterium sibiricum* SOD was built using Modeller 9.10 with default parameters [14]. The predicted structure was first evaluated by Ramachandran plot [15] using the PROCHECK server [16]. To check the reliability of predicted model, Z-score was computed using PROSA-web [17] and secondary structures were predicted using PHYRE [18]. The 3D structure of the template and target were aligned and their root mean square deviation (RMSD) value was calculated using matchmaker tool of the chimera package [19]. The energy of predicted model was compared with other PDB structures of similar sizes using qualitative model energy analysis (QMEAN) server [20]. The different physicochemical properties of the predicted SOD model were calculated using ProtParam [21]. Next, subcellular localization was predicted using CELLO v.2.5 and PSORTb [22].

Functional Analysis and Structure Visualization

The conserved patterns and family of protein based on sequence were searched extensively using different bioinformatics databases such as PROSITE [23] and Pfam [24]. All the structural analysis such as superimposition and visualization were carried out using chimera package [19]. The metal ion specificity and oligomerization mode of the predicted SOD model was identified using SODa webtool [25].

Results and Discussion

Phylogenetic Analysis of SOD Reveals *D. radiodurans* as a Close Neighbor of *E. sibiricum*

Amino acid sequences of SOD were retrieved from 21 extremophilic microorganisms, belonging to different categories of extremophilic environments. These 21 extremophiles comprised; 1 radioresistant, 2 thermophiles, 3 thermoacidophiles, 5 acidophiles, 3 halophiles, 2 haloalkaliphiles, 3 lithotrophs and 1 metallotolerant extremophilic microorganisms, randomly selected from archaea to eubacteria (Supplementary Table 1). To investigate the

Fig. 1 Visualization of conserved amino acid sequences among 21 extremophiles; Bacterial circos plot showing conserved amino acids in all the 21 extremophilic microorganisms based on chemical properties of amino acids (upper panel); Representative view of multiple sequence alignment of superoxide dismutase (SOD)

sequences from position 209 to 285. The rectangles shows the conserved residues at different sites of the SOD (middle panel); The lower panel represents the frequency plot for conserved consensus motif D-[VLE]-[FW]-E-H-[AS]-Y-[YM] from position 253 to 260 positions of SOD

Fig. 2 **a** Heatmap representing the pairwise distance score obtained using pairwise alignment of superoxide dismutase (SODs) sequences from 21 extremophiles. The red to green squares shows the highest to

lowest similarity of SODs, respectively; **b** Phylogenetic tree based on SODs sequences of 21 extremophiles using neighbor-joining method. (Color figure online)

common distribution patterns between SOD sequences from different extremophiles, a comparative analysis was performed by aligning their amino acid sequences using Clustal-W. The multiple sequence alignment showed that the SOD sequences of these extremophiles are highly similar, suggesting extremophiles have evolved through common ancestors during evolution (Supplementary Fig. 1). Amino acid residues with similar chemical properties were also found to be conserved at many positions in different extremophiles, as illustrated by bacterial circos plot (Fig. 1; upper panel). This indicates that these residues were crucial in maintaining the structure of protein, since structure remains more conserved during the course of evolution. Next, we searched for conserved motifs found in amino acid sequences of SOD from different extremophiles to predict key residues of the protein (Supplementary Fig. 1). Interestingly, it was observed that SOD sequences of 11 extremophiles; *Thermus thermophilus*, *N. europaea*, *H. neapolitanus*, *E. sibiricum*, *D. radiodurans*, *C. salexigens*, *C. metallidurans*, *C. aurantiacus*, *B. subtilis*, *A. capsulatus* and *A. acidocaldarius*, had a conserved pattern of a metal binding motif DVWEHAYY with 100 % identity (Fig. 1; middle panel), as reported earlier in the case of cyanobacterial SODs [4]. On an average, it was found that SOD of all the selected extremophiles shared a consensus

pattern D-[VLE]-[FW]-E-H-[AS]-Y-[YM] suggesting that these residues might play a role in the maintenance of structural and functional activity of the protein (Fig. 1; lower panel).

Furthermore, the distance score was calculated between each pair of extremophiles to examine their evolutionary divergence (Supplementary Table 2), which revealed that seven of the extremophiles had higher similarity forming a cluster, as shown in heatmap diagram (Fig. 2a). Based on our analysis, SOD of *E. sibiricum* showed a high degree of conservation score with *D. radiodurans*, *Allicyclobacillus acidocaldarius* and *T. thermophilus* suggesting that structure and function of *E. sibiricum*'s SOD is more similar to SOD of these three extremophiles. Next, a phylogenetic tree based on SOD sequences was constructed, which revealed that *E. sibiricum* and *D. radiodurans* belong to the same cluster and follow the same evolutionary path (Fig. 2b). To strongly build up the relationship, a phylogenetic tree based on nucleotide sequences of 16S rRNA was constructed (Fig. 3). In spite of an evolutionary distant relationship between *E. sibiricum*, *D. radiodurans* and *T. thermophilus* which was based on 16S rRNA, all the three microbes showed the highest similarity of SODs. This further strengthens our prediction about the important role of SOD with special

Fig. 3 Evolutionary phylogenetic tree of 22 extremophiles based on 16S rRNA sequences

reference to tolerance of high dose of gamma radiation and high temperature.

Homology Modeling of SOD and its Evaluation

Homology modeling is one of the methods used for in silico structure modeling of proteins, provided the structure of homologous sequence has been resolved using X-ray crystallography or nuclear magnetic resonance (NMR). Since structure of SOD of *E. sibiricum* has not been resolved yet, we searched for its homologous sequences using Blastp against Protein Data Bank to build its 3D-model. The results showed several homologs of *E. sibiricum* SOD sequence, but SOD of *D. radiodurans* (PDB ID:

1Y67) was selected as a template based on its higher sequence identity, coverage and lower e-value. The pairwise alignment of *E. sibiricum* and *D. radiodurans* sequences showed 63 % identity with 75 % similarity (Fig. 4). Using pairwise alignment five models of target were designed which took into consideration of the various restrictions such as, bond lengths, bond angles and dihedral angles for structure modeling. Out of five, the model with higher number of residues in allowed regions of Ramachandran plot was finally chosen. To improve the modeled structure, the energy was minimized using YASARA server [26]. The final model consists of three strands, twelve helices and eleven turns (Fig. 5a–c) that was superimposed with a template structure which showed the RMSD value

Fig. 4 Identity and similarity (%) graph of SOD from various extremophiles w.r.t. *E. sibiricum*

of 0.44, signifying a good structure alignment of target with SOD of *D. radiodurans* (Fig. 5d).

Several metrics have been actively developed using computational methods to check the reliability of the predicted model. To validate the model, secondary structure of SOD was predicted by using PHYRE, which fully supported its predicted 3-dimensional model. Further, the Ramachandran plot of target structure showed 93 % of residues in most favoured regions, 5 % in additional allowed regions, 1 % in generously allowed regions and remaining 1 % in disallowed regions (Fig. 6a). In order to check the quality of the predicted model, Z-score was calculated which signifies the energy of the predicted models in comparison to distribution of random conformations. The Z-score was found to be -8.36 which was within the range usually seen for similar sizes of proteins (Fig. 6b). The local model quality was checked by plotting energies as a function of residue position using PROSA-web. The negative energy value of most of the residues further indicates the good reliability of predicted structure (Fig. 6c). Next, QMEAN was employed to assess the quality of the model. The QMEAN z-score >-3 indicated that correct folds have been predicted for SOD model (Fig. 6d). Moreover, we calculated different physico-chemical properties of SOD which showed that SOD had a molecular weight of 22.638 kDa and 5.27 isoelectric point (pI). A pI <7 indicates a negatively charged protein with

202 amino acids with an instability index of 32.54. The Grand average of hydropathicity (GRAVY) score was predicted to be -0.461 , indicating its hydrophilic nature. All these measures confirmed the good quality of the predicted model and it was finally submitted to the protein model database (PMDB ID: PM0078260).

Structural and Functional Annotation of SOD

Sub-cellular localization analysis of protein helps to uncover the component of cells, where the protein is found in its active state. In line with the earlier studies, SOD was predicted as an extra-cellular protein. Next, the Pfam and PROSITE database were scanned using Scanprosite, to find out protein families and patterns in SODs. SOD belongs to two protein families; PF00081 and PF02777, both representing SODs family. The pattern PS00088 (D-X-[WF]-E-H-[STA]-[FY]) (where; X is any residue) from PROSITE was predicted to be present in SOD of *E. sibiricum* from residue 164 to residue 171. The superposition of tertiary structure showed >95 % structural overlap and the comparative modeling also showed that the homologs indeed shared two longest common structural motifs of 19 and 14 amino acid residues; YAYDALEPHIDARTMEIHH (at 8th position) and DVWEHAYYLNQNR (at 171 position), respectively (Fig. 5a). The metal ion specificity and oligomerization state (dimer

Fig. 5 **a** Residue-wise comparison of secondary structures of superoxide dismutase (SODs) of *E. sibiricum* and *D. radiodurans*; **b** Secondary structure view of SOD model showing alpha-helices, beta-sheets and turns along with conserved motif in green, orange,

white and blue colour, respectively; **c** A three-dimensional view of SOD model; **d** Superimposition view of secondary structures of SOD from *E. sibiricum* and *D. radiodurans*. (Color figure online)

or tetramer) of all the SODs were identified using SODA webtool. Out of 21 extremophilic SOD sequences, 7 Mn-SOD (dimer), 5 Fe-SOD (dimer), 8 Fe-SOD (tetramer) and 1 Fe/Mn-SOD (dimer) have been categorized (Supplementary Table 1). Interestingly, existence of SODs from *D. radiodurans* and *T. thermophilus* were found to be the same as was in *E. sibiricum* (Mn-dimer) with 100 % fingerprint score. Hence, it might be possible that bacterium with radioresistance and thermophilic properties could acquire the Mn-SOD.

These studies reiterate that *E. sibiricum* shows a similar environmental tolerance as that of *D. radiodurans*. Further

experiments on the SOD structure of *Exiguobacterium sp.* would help to resolve the unusual behaviour of this extremophile and its mode of action in the various metabolic pathways, and its evolutionary hierarchy will provide more insight into this extremophilic antioxidant. It was tempting to speculate that SOD might play an important role in rescuing stress conditions thus enabling *E. sibiricum* to grow at high doses of gamma radiation, salt-stress and at a wide range of temperature. In the near future, *E. sibiricum* may be used as an extremophilic model for investigating the stabilization of some new biomolecules, when exposed to these extreme conditions.

Fig. 6 **a** Ramachandran plot of predicted model of *E. sibiricum* superoxide dismutase (SOD) representing 166 residues in the most favoured region (red), 9 in additional allowed regions (yellow) and 2 in generously allowed regions (light yellow), 2 in disallowed regions (white) with 13 glycine (shown in triangle) and 8 proline residues, respectively; **b** A Z-score of -8.36 for predicted model with 202 residues calculated using PROSA-web is shown as black rectangle

Acknowledgments The authors are grateful to Dr. Rajesh Gokhle, Director of CSIR-IGIB for providing infrastructural facilities and financial assistance from CSIR (BSC0302). R Pathak would also like to acknowledge the University Grants Commission (UGC) India, for providing Senior Research Fellowship. P Narang acknowledges the Department of Biotechnology for providing BINC fellowship.

References

- Rothschild LJ, Mancinelli RL (2001) Life in extreme environments. *Nature* 409:1092–1101
- Le Roes-Hill M, Khan N, Burton S (2011) Actinobacterial peroxidases: an unexplored resource for biocatalysis. *Appl Biochem Biotechnol* 164:681–713
- Priya B, Premanandh J, Dhanalakshmi RT, Seethalakshmi T, Uma L, Prabakaran D, Subramanian G (2007) Comparative analysis of cyanobacterial superoxide dismutases to discriminate canonical forms. *BMC Genom* 8:435–444
- Wintjens R, Noël C, May AC, Gerbod D, Dufernez F, Capron M, Viscogliosi E, Rooman M (2004) Specificity and phenetic relationships of iron- and manganese-containing superoxide dismutases on the basis of structure and sequence comparisons. *J Biol Chem* 279:9248–9254
- Beaume N, Pathak R, Yadav VK, Kota S, Misra HS, Gautam HK, Chowdhury S (2013) Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of *D. radiodurans* involves G4 DNA-mediated regulation. *Nucleic Acids Res* 41:76–89
- Markillie LM, Varnum SM, Hradecky P, Wong KK (1999) Targeted mutagenesis by duplication insertion in the radioresistant bacterium *Deinococcus radiodurans*: radiation sensitivities of catalase (katA) and superoxide dismutase (sodA) mutants. *J Bacteriol* 181:666–669

7. Basu B, Apte SK (2012) Gamma radiation induced proteome of *Deinococcus radiodurans* primarily targets DNA repair and oxidative stress alleviation. *Mol Cell Proteomics* 11:M111.011734
8. Hu P, Brodie EL, Suzuki Y, McAdams HH, Andersen GL (2005) Whole-genome transcriptional analysis of heavy metal stresses in *Caulobacter crescentus*. *J Bacteriol* 187:8437–8449
9. Pathak R, Kumar R, Gautam HK (2013) Cross-species induction and enhancement of antimicrobial properties in response to gamma irradiation in *Exiguobacterium* sp. HKG 126. *Indian J Microbiol* 53:130–136
10. Lee DH, Oh KH, Kahng HY (2009) Molecular analysis of anti-oxidant genes in the extremohalophile marine bacterium *Exiguobacterium* sp. CNU020. *Biotechnol Lett* 31:1245–1251
11. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948
12. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645
13. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
14. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815
15. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7:95–99
16. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486
17. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:W407–W410
18. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4:363–371
19. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
20. Benkert P, Künzli M, Schwede T (2009) QMEAN server for protein model quality estimation. *Nucleic Acids Res* 37:W510–W514
21. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531–552
22. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26:1608–1615
23. Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I (2013) New and continuing developments at PROSITE. *Nucleic Acids Res* 41:D344–D347
24. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301
25. Kwasigroch JM, Wintjens R, Gilis D, Rooman M (2008) SODA: an Mn/Fe superoxide dismutase prediction and design server. *BMC Bioinformatics* 9:257–263
26. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K (2009) Improving physical realism, stereochemistry and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins* 77:114–122