In Vitro and In Silico Cloning of Xenopus laevis SOD2 cDNA and Its Phylogenetic Analysis

MICHELE PURRELLO, CINZIA DI PIETRO, MARCO RAGUSA, ALFREDO PULVIRENTI, ROSALBA GIUGNO, VALENTINA DI PIETRO, GIOVANNI EMMANUELE, SALVO TRAVALLI, MARINA SCALIA, DENNIS SHASHA, and ALFREDO FERRO

ABSTRACT

By using the methodology of both wet and dry biology (i.e., RT-PCR and cycle sequencing, and biocomputational technology, respectively) and the data obtained through the Genome Projects, we have cloned Xenopus laevis SOD2 (MnSOD) cDNA and determined its nucleotide sequence. These data and the deduced protein primary structure were compared with all the other SOD2 nucleotide and amino acid sequences from eu-karyotes and prokaryotes, published in public databases. The analysis was performed by using both Clustal W, a well known and widely used program for sequence analysis, and AntiClustAl, a new algorithm recently created and implemented by our group. Our results demonstrate a very high conservation of the enzyme amino acid sequence during evolution, which proves a close structure–function relationship. This is to be expected for very ancient molecules endowed with critical biological functions, performed through a specific structural organization. The nucleotide sequence conservation is less pronounced: this too was foreseeable, due to neutral mutations and to the species-specific codon usage. The data obtained by using Clustal W are comparable with those produced with AntiClustAl, which validates this algorithm as an important new tool for biocomputational analysis. Finally, it is noteworthy that evolutionary trees, drawn by using all the available data on SOD2 nucleotide sequences and amino acid and either Clustal W or AntiClustAl, are comparable to those obtained through phylogenetic analysis based on fossil records.

INTRODUCTION

Molecular oxygen is essential for efficient energy production by aerobic organisms, since it is the final acceptor of electrons carried through the respiratory chain. On the other hand, it is also very toxic; thus, it has to be rapidly and efficiently metabolized together with reactive oxygen species (ROS), its highly reactive and obligatory byproducts. ROS have been shown to be causally involved in aging as well as in several degenerative pathologies (e.g., Parkinson disease) and in cancer (Dreher and Junod, 1996; Melov et al., 1998; Salviodi et al., 2001). Accordingly, it is not surprising that the enzymes scavenging them (i.e., superoxide dismutases, SODs) are very ancient and ubiquitary molecules. Interestingly, SODs are present not only in organisms using oxygen for their energy metabolism, but also in some facultative anaerobes (Ryan Fink and Scandalios, 2002). To date, five different types of SODs have been identified: two forms of Cu/Zn SOD (SOD1 and SOD3), Mn SOD (SOD2), Fe SOD, Ni SOD (Ryan Fink and Scandalios, 2002; Zelko et al., 2002). Eukaryotic SOD2 is encoded by a nuclear gene but is localized within the mitochondrion, where it catalyzes the dismutation of O2− and other oxygen species into H2O2 (Borgstahl et al., 1992). By inactivating ROS at their major site of production and blocking their harmful effects on essential cell molecules (e.g., nucleic acids, proteins, and phospholipids), SOD2 performs a critical role in metabolism. In fact, it allows cells to use oxygen as the final acceptor of electrons (derived from nutrients and carried by NADH and FADH2) and

1 Dipartimento di Scienze Biomediche, Sezione di Biologia e Genetica, Università di Catania, Catania, Italy.
2 Dipartimento di Matematica e di Informatica, Università di Catania, Catania, Italy.
3 Dipartimento di Scienze Biomediche, Sezione di Patologia clinica e di Oncologia molecolare, Università di Catania, Catania, Italy.
4 Courant Institute of Mathematical Sciences, New York University, New York, New York.
to obtain a much greater amount of energy than that produced by glycolysis alone, without the high price of heavy structural damage to be otherwise paid. Since it was missing from the sequences available in public databases, we decided to clone *Xenopus laevis* SOD2 cDNA with the aim of determining its primary structure and performing a phylogenetic analysis.

**MATERIALS AND METHODS**

RT-PCR was performed on total RNA, purified from a *Xenopus laevis* liver, by using methods that have been described (Purrello et al., 2001; Di Pietro et al., 2004). Primers were designed with Omiga 2.0 software (Oxford Molecular). We used *Homo sapiens* SOD2 cDNA and, after its cloning, *Xenopus laevis* SOD2 cDNA as templates (Fig. 1). The sequence of primers and relative PCR conditions are shown in Table 1. Both strands of cDNA fragments were subjected to cycle sequencing with SEQ 4/H11003/4 Personal Sequencing System ([Amersham Pharmacia Biotech, Arlington Heights, IL], by using Cy5.5 dye-labeled ddNTP terminators (Thermo Sequenase Cy5.5 Dye Terminator Cycle Sequencing Kit from Amersham Pharmacia Biotech). Amplified cDNAs were purified of unincorporated dye terminators with spin columns (AutoSeqTM G-50 from Amersham Pharmacia Biotech), following the manufacturer’s protocol. To compare SOD2 protein and nucleic acid sequences and to align

<table>
<thead>
<tr>
<th>Name</th>
<th>Primers</th>
<th>cDNA</th>
<th>RT-PCR protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>BO13</td>
<td>5′GATCGCTTTCTAAATAGTCTGG3′</td>
<td>304 bp</td>
<td>50°C 30 min, 94°C 4 min; 94°C 60 sec, 51°C 120 sec, 72°C 120 sec; 72°C 10 min</td>
</tr>
<tr>
<td></td>
<td>5′CAACCTTTCCTAATGGG3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BO23</td>
<td>5′TGACCCACGGCAAGACCC3′</td>
<td>417 bp</td>
<td>50°C 30 min, 94°C 4 min; 94°C 60 sec, 53°C 120 sec, 72°C 120 sec; 72°C 10 min</td>
</tr>
<tr>
<td></td>
<td>5′CGTGCTCCCACACATAC3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BO25</td>
<td>5′TGGTGTCATATCAATCA 3′</td>
<td>373 bp</td>
<td>50°C 30 min, 94°C 4 min; 94°C 60 sec, 51°C 120 sec, 72°C 120 sec; 72°C 10 min</td>
</tr>
<tr>
<td></td>
<td>5′TCTTTGAGTTAACATTCCCC3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BO26</td>
<td>5′CTTGGTGCAATCAAGACC3′</td>
<td>452 bp</td>
<td>50°C 30 min, 94°C 4 min; 94°C 60 sec, 52°C 120 sec, 72°C 120 sec; 72°C 10 min</td>
</tr>
<tr>
<td></td>
<td>5′TTAACAAGTGGTGCTGCG3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BO30</td>
<td>5′TGACCCACAGCAAGCC3′</td>
<td>486 bp</td>
<td>50°C 30 min, 94°C 4 min; 94°C 60 sec, 55°C 120 sec, 72°C 120 sec; 72°C 10 min</td>
</tr>
<tr>
<td></td>
<td>5′CCGATTTGATAACATTCC3′</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIG. 2. (A) Multiple alignment of SOD2 cDNA sequences from organisms at different levels of the evolutionary scale: *Homo sapiens*, *Gallus gallus*, *Xenopus laevis*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Aspergillus nidulans*, *Escherichia coli*, and *Halobacterium salinarum*. (B) *Xenopus laevis* SOD2 nucleotide and amino acid sequences. (C) Multiple alignment of SOD2 amino acid sequences from *Homo sapiens*, *Gallus gallus*, *Xenopus laevis*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Aspergillus nidulans*, *Escherichia coli*, and *Halobacterium salinarum*. 
FIG. 3. Phylogenetic trees based on SOD2 cDNA and amino acid sequence evolution, drawn by using either Clustal W (A, C) or AntiClustAl (B, D). Their different spatial orientation derives from the different computational procedures used (see, also, Results and Discussion). It is noteworthy that Chlamydomonas (a Chlorophyton) is an out of order species in both trees.
them, we used two software programs, Clustal W and Anti-ClustAl; the latter is a new algorithm recently created and implemented by our group and available at http://alpha.dmi.unict.it/~cnyu/ (Di Pietro et al., 2003; Cantone et al., 2004). Comparative protein modeling was performed with SWISS–MODEL, version 3.5, an Automated Comparative Protein Modelling Server (http://www.expasy.org/swissmed/).

RESULTS AND DISCUSSION

The absence of Xenopus laevis SOD2 cDNA sequence in public databases was noteworthy, because of the peculiar evolutionary position of Amphibia. In fact, they were the first vertebrates to leave the submarine world and to colonize an oxygen-rich environment; accordingly, they were forced to profoundly modify their metabolism (Dobzhansky et al., 1997). Furthermore, Xenopus laevis is an important model organism with a Genome Project in progress (www.xenbase.org), so its further biological characterization is valuable. To reach our aim, we first compared the nucleotide cDNA sequence of Homo sapiens SOD2 to all available eukaryotic and prokaryotic SOD2 sequences from GenBank (www.ncbi.nlm.nih.gov/entrez/) by using BLAST (www.ncbi.nlm.nih.gov/BLAST/) and Clustal W, a widely used tool for solving the Multiple Sequence Alignment (MSA) problem (Higgins et al., 1994). This allowed us to identify the most conserved regions of SOD2, which shared at least 90% homology and also comprised the four Mg$^{2+}$ binding sites (Figs. 1A and 2A). Primers designed on these segments were then used to perform RT-PCR from Xenopus laevis liver RNA (Table 1; Fig. 1A and B). We obtained several overlapping cDNA fragments that were subjected to cycle sequencing on both strands (Fig. 1B). By using the newly amplified sequences as well as uncharacterized ESTs published in NCBI (X1.23883) as templates, we designed primers specific for the evolutionarily poorly conserved segments of Xenopus laevis SOD2 cDNA and we were able to fill the cDNA gaps (Figs. 1B, C, and 2B). The resulting nucleotide sequence (GeneBank No. AY362041) and the deduced primary protein structure were then compared with all the other SOD2 sequences, available in public databases (Fig. 2A and C). This analysis was performed by using two different alignment software, Clustal W and AntiClustAl (see Materials and Methods). The latter exploits a simple technique based on randomized tournaments designed to approximate solutions of optimization problems in metric spaces such as 1-median and diameter computation (Cantone et al., 2004a, 2004b). Winners are those sequences that locally optimize the function, whose global minimum/maximum is being sought (Cantone et al., 2004a, 2004b). AntiClustAl is based on a top down bisector tree clustering algorithm (called Antipole Tree) and a linear approximate 1-median computation, instead of the quadratic exact procedure of ClustalW. Drawing phylogenetic trees with Antipole Tree gives a substantial speed improvement over ClustalW with comparable quality of the data, especially in the case of low-sequence homology. The two programs differ only in the clustering algorithm. ClustalW uses a hierarchical bottom-up clustering, trying to keep similar sequences in the same cluster (Fig. 3A and C), whereas AntiClustAl tries to put dissimilar sequences in different clusters (Fig. 3B and D): the clustering method used by ClustalW is more precise, but at the cost of a quadratic running time computation of all pairwise distances, whereas AntiClustAl produces a lower quality grouping but in linear time. Moreover, the lower height of the AntiClustAl tree reduces the error produced by the profile alignments in the bottom-up alignment. This explains why at the end the two algorithms produce similar results (Fig. 3A–D). It could be reasonable to use ClustalW for small sets of highly similar and not very long sequences, whereas AntiClustAl may be used in connection with large sets

**FIG. 4.** The rate of amino acid changes within SOD2 is comparable to that of Cytochrome c and clearly different from that of very well-conserved proteins (such as histones) or very poorly conserved ones (such as fibrinopeptides).
of sequences endowed with low similarity but quite long. More precisely, AntiClustAl gives a significant improvement in precision and speed when similarity goes below 40% and sequence length exceeds 5000 nucleotides, respectively. Our results demonstrate a high conservation of SOD2 primary structure, comparable to that of cytochrome c (Figs. 2A, C, and 4); this is expected for ancient molecules endowed with important biological functions, performed through a specific structural organization. Accordingly, the tertiary structure of the protein is also well conserved (not shown). On the other hand, the cDNA nucleotide sequence is less conserved, most likely because of neutral mutations and as a consequence of the species—specific codon usage (Fig. 2A). The data obtained by using either ClustalW or AntiClustAl are comparable and both agree with the phylogenetic analysis based on fossil records (Fig. 3A–D).

ACKNOWLEDGMENTS

This project was supported by funds from MIUR to Prof. Michele Purrello (Ministero dell’Istruzione, dell’Università e della Ricerca, Piani di Potenziamento della Rete Scientifica e Tecnologica), and from the Università di Catania to Prof. Michele Purrello and Prof. Alfredo Ferro (Progetti di ricerca di Ateneo).

REFERENCES


Address reprint requests to:
Prof. Michele Purrello
Dipartimento di Scienze Biomediche
Sezione di Biologia e Genetica
Università di Catania
Via Santa Sofia 87
95124 Catania, Italy

E-mail: purrello@unict.it

Received for publication July 8, 2004; accepted July 26, 2004.
PURRELLO

AU1
Location of Oxford Molecular?

AU2
Cantone et al. 2004 a or b?

AU3
Can you update Cantone 2004a?