A Second Superoxide Dismutase Gene in the Medfly, Ceratitis capitata

G. K. Banks,* A. S. Robinson,† J. Kwiatkowski,‡ F. J. Ayala,§ M. J. Scott,** and D. Kriticou†

*Department of Plant Science, University of Arizona, Tucson, Arizona 85721, †International Atomic Energy Agency, Vienna, Austria, ‡Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92717, §Institute of Botany, Warsaw University, 00-478 Warsaw, Poland, †Department of Microbiology and Genetics, Massey University, New Zealand, and **Department of Biology, Division of Genetics, Cell and Developmental Biology, University of Patras, Patras, Greece

ABSTRACT

We report the first case of two Cu/Zn Sod genes (ccSod1 and ccSod2) that have been cloned and sequenced from an insect, the medfly, Ceratitis capitata. Biochemical evidence suggested the presence of two Sod genes in the medfly. The two genes are isolated using different molecular strategies: ccSod1 via cross-hybridization to a genomic library using a heterologous probe and ccSod2 from cDNA using a homologous probe generated by PCR. Sequence analysis shows that ccSod1 and ccSod2 are different genes. The inferred amino sequences show that all essential residues of the active site are strictly conserved, which suggests both genes encode functional Cu/Zn superoxide dismutase (SOD). Phylogenetic analysis by the maximum parsimony method with bootstrap resampling of previously known Cu/Zn SOD reveals two monophyletic groups, vertebrates and insects. The position of ccSod2 in this phylogeny is undefined with respect to dipteran ccSOD1, vertebrate, plant, fungal, and extracellular Cu/Zn SOD, which suggests that the duplication detected in Ceratitis is ancient, perhaps as old as the origins of the arthropod phylum in the Cambrian more than 500 million years ago. In situ hybridization to polytene chromosomes places the genes on different chromosomes, which is consistent with an ancient gene duplication.

THE medfly, Ceratitis capitata, is an important agricultural pest for which molecular (ZACHAROPOULOU et al. 1992) and genetic maps (MALACRIDA et al. 1988) are being constructed. A large number of biochemical loci have been mapped in the medfly (MILANI et al. 1989). Oligonucleotide primers have been designed, based on sequence information from other organisms, to amplify by PCR homologous sequences from the medfly. These PCR products are then used as probes for isolating the genes from genomic and cDNA libraries (SCOTT et al. 1993). The cloned genes have in turn been located on polytene chromosomes by in situ hybridization, thus correlating the genetic with the cytological map.

One gene included in this genome mapping encodes superoxide dismutase (SOD; EC1.15.1.1), which catalyzes the conversion of $O_2^-$ to $H_2O_2 + O_2$, thus protecting the cell from oxidative damage. In most nonplant eukaryotes, SOD is present in two forms: a mitochondrial enzyme containing Mn and a cytoplasmic one containing Cu/Zn (FRIDOVICH 1986). We cloned a Cu/Zn Sod gene, hereafter called ccSod1, from a genomic library by cross-hybridization to a Drosophila melanogaster probe (KIWTOWSKI et al. 1992). Independently, a second Sod gene fragment was isolated by PCR from adult medfly cDNA. This fragment was in turn used as a probe to identify a number of clones from a third-instar larval cDNA library. All these clones contained the identical Sod sequence, and this gene is herein designated ccSod2 (GenBank accession L35494).

Earlier biochemical analysis suggested the presence of two Cu/Zn SOD genes in the medfly (FERNANDEZ-SOUZA and MICHELSON 1976). Comparison of the coding sequences, as well as their relative location on the polytene chromosomes, shows that ccSod1 and ccSod2 are two different Sod genes. We present a comparative analysis of the two Sod coding regions, their cytological location, and discuss their phylogeny.

MATERIALS AND METHODS

General methods: Medfly genomic DNA was isolated from the Benakion strain by the method of HOLMES and BONNER (1973). Total RNA was prepared by the procedure described by ASHBERNER (1989). cDNA was synthesized following the protocol of Stratagene. S. BROGNA provided a third-instar larval Lambda ZapII-cDNA library, which was screened by standard procedures (MANTATIS et al. 1982) using nitrocellulose membranes. The prehybridization and hybridization buffers comprised 50% formamide, 6X SSC, 0.1% SDS and 5X Denhardt’s; the filters were incubated at 42°C for both steps. Two sequential cold washes (5 and 15 min) at room temperature and two hot washes at 65°C were carried out in 2X SSC and 0.1% SDS. DNA was sequenced via the chain termination method (SANGER et al. 1977).

PCR primers and amplification: Four degenerate primers (I and II, 5'-3'; III a and b, 3'-5') were synthesized: primer I, 5'-CA-(CT)GG(ACGT)TT(CA)GT(ACGT)CA-3'; primer II, 5'-GG(ACGT)CG(ACGT)CA(CA)TT(CA)(AA)GT-3'; primer
Figure 1.—Nucleotide sequence of the cDNA for the ccSod2 gene in the medfly, Ceratitis capitata. The inferred amino acid sequence is shown below the corresponding codons. The terminating codon is indicated with an asterisk; putative regulatory sequences are underlined; regions corresponding to the PCR primers are shaded. The sequence has been submitted to GenBank/EMBL under the accession number L35494.

IIIa, 5'-CC(AGT)AT(ACGT)AC(ACGT)CC(AG)CA(ACGT)-CC-3'; and primer IIIb, 5'-CC(AGT)AT(ACGT)AC(ACGT)CC(AG)CA(ACGT)-GC-3'. The primers are based on three conserved regions in known SOD enzymes (see Figure 3); primer I from aa44 to aa49 (HGFHVH), primer II from aa62 to aa67 (GPHFNP), and primer IIIa and B cover the region aa147 to aa152 (GPHFNPVH and ACGVIG, respectively). The reason for the single amino acid difference between primers IIIa and IIIb is that nondipteran SOD enzymes have alanine at position 49, whereas two known species of Drosophila have glycine.

Figure 1 gives the nucleotide sequence of ccSod2 cDNA, as well as the inferred amino acid sequence. The 453 nucleotide-long sequence codes for a 150 amino acid-long polypeptide. Two putative regulatory sequences are identified: a TATA box at position 121 and a polyadenylation signal at 1011.

We used the cDNA ccSod2 to screen an Embl 3 genomic library and isolated the genomic homologue (Sod2 gen) for this gene. Figure 2 shows simplified restriction maps for the genomic clones of the two medfly Sod genes.

**Amino acid comparisons:** Figure 3 displays the alignment of nine Cu/Zn SOD amino acid sequences, six from insects including ccSod1 and ccSod2 and three from baculovirus, human, and fluke. Invariant in all enzymes compared are all essential amino acid residues of the active site (GETZOFF et al. 1989), including metal ligands (His-47, 49, 64, 72, 81, 122 and Asp-84) and two other residues (Arg-145 and Asp-126) involved in catalytic reaction, as well as the residues that determine the stability of the environment of the active site (Gly-45, 62, 83, 140, 143 and Pro-67). The disulfide bridge residues (Cys-58 and 148) are strictly conserved, whereas dimer-contact

**RESULTS**

PCR, library screening and genomic Southern analysis: Neither primer I nor II in combination with IIIa resulted in any PCR amplification product from the adult medfly cDNA template. Amplification with primers I and IIIb resulted in a 330-bp fragment and with primers II and IIIb, in a 260-bp fragment. The 330-bp fragment was isolated and used for probing a third-instar larval Lambda-Zap II cDNA library; a secondary screen was carried out with the 260-bp fragment. Nine secondary positive clones were identified and subcloned into Bluescript SK' via in vitro excision. Partial sequence (using PCR primer I) was obtained for all clones that proved to be identical over 200 bp downstream of primer I.
amino acids located at positions 6, 8, 18, 51–55, 115–117 and 150–155 are conserved either absolutely or functionally in most enzymes. This conservation suggests that both medfly SOD are functional enzymes.

The average amino acid differences between the Drosophilidae Cu/Zn SOD and ccSOD1 is 30.8 ± 0.9, which is 20% of the total 153 amino acids (see Table 1). This degree of divergence (as well as the divergence among the Drosophilid species) corresponds well to the phylogenetic relationships between the medfly and the Drosophilids, thus indicating that ccSod1 is orthologous with the Drosophilid Cu/Zn Sod, that is, the divergence among these SOD reflects speciation events.

A different picture emerges for ccSod2, which differs from the Drosophilid species by 68.4 ± 0.7 (46%) amino acids on the average. This is approximately the degree of differentiation between the Drosophilid species and those of baculovirus (44%), human (39%), and the fluke (38%). This indicates that ccSod2 is paralogous to ccSod1 and to the Drosophilid Cu/Zn Sod, that is, it probably arose by gene duplication in a rather distant ancestor.

**Phylogenetic analysis:** Figure 4 shows a phylogenetic tree obtained by the maximum parsimony method with bootstrap resampling (Felsenstein 1985). The lengths of the branches reflect PAM differences (Dayhoff 1978) between sequences. The tree includes a number of distinct species, including vertebrates, plants, fungi and prokaryotes. Figure 4 shows a deep division between prokaryotes and eukaryotes. Among the eukaryotes a deep branch separates two extracellular enzymes (human and fluke) from all intracellular SOD. Yeast and plants appear as monophyletic groups (with fairly high incidence of bootstrap resampling, 89 and 65%, respectively). Within the animal branch, a fluke (Schistosoma) protein is undifferentiated, but there are two monophyletic groups, vertebrates (71%) and insects, including ccSOD1 (97%). The position of Ceratitis ccSod2 is undefined with respect to dipteran ccSOD1 and the vertebrate, plant, fungal, and extracellular Cu/Zn SOD, which suggests that the duplication of this gene is very ancient.

**Cytological location:** The 1.6-kb cDNA clone for ccSod2 maps to a site on chromosome 4L at position 42B (see Figure 5A). However, there is also a faint signal repeatedly observed at position 36D on chromosome 3L (see Figure 5B). This weak signal is not due to cross-

**TABLE 1**

<table>
<thead>
<tr>
<th>Species</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster</td>
<td>20</td>
<td>19</td>
<td>25</td>
<td>33</td>
<td>71</td>
<td>65</td>
<td>58</td>
</tr>
<tr>
<td>D. virilis</td>
<td>19</td>
<td>22</td>
<td>91</td>
<td>69</td>
<td>67</td>
<td>60</td>
<td>58</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>21</td>
<td>30</td>
<td>57</td>
<td>66</td>
<td>60</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>Chymomyza</td>
<td>29</td>
<td>68</td>
<td>66</td>
<td>60</td>
<td>59</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Ceratitis SD1</td>
<td>67</td>
<td>66</td>
<td>60</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ceratitis SD2</td>
<td>74</td>
<td>69</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baculovirus</td>
<td>72</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td></td>
<td></td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.—Amino acid sequence alignment of Cu/Zn SOD from flies, baculovirus, humans and a fluke. Residues identical with those in D. melanogaster are represented by dots; gaps in sequence alignments by dashes. Sequences are from the following sources: D. melanogaster, D. virilis, Chymomyza amoena, C. capitata SOD1 (KWIATOWSKI et al. 1992); D. willistoni (KWIATOWSKI et al. 1994); Baculovirus, Autographa californica (TOMALSKI et al. 1991); human (SHERMAN et al. 1983); fluke, Schistosoma mansoni (CORDEIRO DA SILVA et al. 1992).
hybridization with ccSod1, because the 4-kb genomic clone of ccSod1 maps exclusively to chromosome 6 at position 99A (Figure 5C).

**DISCUSSION**

This is the first case reported of two Cu/Zn Sod genes that have been cloned and sequenced from an insect. The isolation of two different Sod genes confirms preliminary biochemical evidence suggesting the presence of two Sod genes in the medfly (Fernandez-Souza and Michelson 1976).

It is pertinent to explain how and perhaps why two different gene sequences were isolated. ccSod1 was cloned from a genomic medfly library by cross-hybridization to a D. melanogaster cDNA probe. All positive clones contained the same Sod gene sequence. ccSod2 was isolated from a medfly third-instar cDNA library, using as a probe a PCR fragment amplified from adult cDNA. Thus, the cloning of ccSod1 was dependent on the degree of homology with the D. melanogaster counterpart, whereas the cloning of ccSod2 was dependent on the specificity of the PCR primers, as well as the level of expression of Sod in adults and third-instar larvae.

The two genes ccSod1 and ccSod2 are different to the extent that they do not cross-hybridize even at low to moderate stringency on a library screen, genomic Southern blot, or in situ hybridization to polytene chromosomes (data not shown). This might explain why only ccSod1 sequences were isolated from a genomic library when the homologous gene from D. melanogaster was the probe and, similarly, why only ccSod2 sequences from the cDNA library showed as positive clones when a ccSod2 fragment was used as a probe.

Differences in gene expression through development may have also played a role. To cover the 3' end (aa147–152) of the gene, we designed two primers, IIIa and IIIB, that differ at position 147 in that IIIB assumes alanine (the amino acid present in nondipteran Cu/Zn SOD) and IIIa assumes glycine (the amino acid present in Drosophila SOD). IIIB was effective, in conjunction with either primer I or II, for PCR amplification from adult cDNA; IIIa failed with both 5'-end primers. Because we know that ccSod1 codes for glycine at position 147, it would seem to follow that this gene is likely to be expressed in the adult medfly only at low levels or not at all. Consistent with this tentative inference, preliminary Northern blot analysis indicates that ccSod1 is expressed mainly in early embryo rather than later developmental stages. Graf and Ayala (1986) have shown, on the basis of CRM assays, that in D. melanogaster SOD content is low in young adults but it gradually increases after emergence from the pupa, so that by the seventh day SOD levels are two to three times higher than just after emergence. SOD levels are also high in embryos but low through the larval and pupal stages. Using the same RNA blots, stripped and reprobed, ccSod2 shows hybridization signal from third-instar larvae through adult but not in the early developmental stages.

The inferred amino acid sequences show that the coding region for ccSod1 is three amino acids longer than that of ccSod2. Nevertheless, it would seem that both genes encode functional Cu/Zn SOD because all the essential amino acid residues of the active site are strictly conserved, including those involved in metal ligand binding, catalytic reaction, and stability of the active site. The lack of a signal peptide in either gene suggests both genes may have a cytoplasmic function. The human Cu/Zn SOD is localized in the peroxisomes in fibroblasts and hepatoma cells (Keller et al. 1991), and this locale may be imposed by the putative peroxisome-targeting tripeptide Ser-Arg-Leu (Figure 3, positions 144–146) near the c-terminal of the polypeptide.
ccSOD1 also has the canonical tripeptide Ala-Arg-Leu (Gould et al. 1989) in positions 144–146. But neither of these tripeptides occurs in ccSOD2, which has Gly-Arg-Leu instead, which suggests the subcellular locale may be different for ccSOD1 and ccSOD2.

There is some anecdotal evidence, both cytological and molecular, that the medfly is particularly susceptible to duplication events. For example, tandem duplications are frequently detected on polytene chromosomes (A. Zacharopoulos, personal communication), and at least two molecularly characterized gene systems, vitellogenin (Rina and Savakis 1991) and Adh (C. Savakis, personal communication), include duplications of functional genes. However, the presence of two Cu/Zn Sod genes in Ceratitis is unlikely to be a result from a recent tandem duplication because the two genes are located on different chromosomes. Moreover, ccSod2 has an intron at position aa67/68, whereas ccSod1 has an intron at aa95/96, 28 amino acids away (data not shown). An intron similarly located as in ccSod2 occurs in a gene coding for extracellular SOD in the fluke, Schistosoma mansoni (Simurda et al. 1988).

Evidence against a recent duplication derives also from the degree of divergence between ccSOD1 and ccSOD2 (Table 1). The difference amounts to 67 amino acids, approximately as large as between human SOD and either ccSOD1 (60 amino acids) or ccSOD2 (69 amino acids). In addition, the average number of amino acid differences compared with Drosophila SOD is 30.8 ± 0.9 for ccSOD1 but 68.4 ± 0.7 for ccSOD2. The possibility that ccSod2 may have evolved at a much faster rate than ccSod1 after a recent duplication seems unlikely in view of the large differentiation between ccSod1 and ccSod2. Furthermore, such rapid evolution of ccSod2 (accounting for 38 additional amino acid differences) should have similarly increased the degree of differentiation of ccSod2 relative to human and Schistosoma SOD, which are only slightly or not at all more different from ccSod2 (69 and 63 amino acid differences, respectively) than from ccSod1 (59 and 63 differences, respectively).

These results point out an ancient duplication of ccSod1 and ccSod2. The fact that ccSod2 is as different from ccSod1 as it is from human and Schistosoma SOD suggests that the duplication is about as old as the divergence of the phyla chordata and either the arthropoda or platyhelminthes phyla, perhaps as old as the beginning of the Cambrian period, some 590 million years ago. The inference that the duplication is ancient assumes that SOD has evolved at an approximately constant rate. The assumption of an SOD molecular clock has been challenged (Ayala 1986), but it has been recently shown that, under certain reasonable assumptions, SOD behaves as a fairly accurate molecular clock to the extent that similar numbers of amino acid differences reflect similar spans of time since divergence (Fitch and Ayala 1994). The conclusion is that the duplication of ccSod1 and ccSod2 is ancient and may be as old, or nearly, as the emergence of the arthropod phylum.

If the ccSod1/ccSod2 duplication is ancient, two hypotheses are possible. The first hypothesis is that ccSod2 has been transmitted vertically in the Ceratitis lineage since it first appeared by duplication. If so, it follows that ccSod2 would have been inherited in other dipteran (and even arthropod) lineages. Unless it became lost in all other lineages, ccSod2 should be present in other dipterans as well. The second hypothesis is that ccSod2 became incorporated in the Ceratitis lineage by horizontal gene transfer. There is a baculovirus homolog...
logue of Sod (Tomalski et al. 1991), which the virus may have acquired from an animal species, possibly a lepidopteran host. Horizontal transfer of repetitive DNA sequences (P elements) between Drosophila species has been demonstrated (Daniels et al. 1990). Horizontal transfer of autologous protein-coding genes between eukaryotes has not yet been discovered, but its possibility cannot be convincingly excluded. However, if horizontal transfer from some other species to Ceratitis would have occurred, it still follows that ccSod2 would be present in lineages other than the donor's, unless it would have been lost in all of them.

Evidence exists of other ancient Sod duplications (e.g., Kwiatowski et al. 1991; Smith and Doolittle 1992; Fitch and Ayala 1994). For example, a duplicated gene coding for an extracellular form of the enzyme exists in rats and humans that is ostensibly as old, or older than, the divergence of the animal, plant, and fungi kingdoms. Ancient duplications, although not quite as old, are present in C. elegans and S. mansoni. All higher plants studied (including scots pine, garden pea, spinach, and tomato) share an Sod duplication, yielding a chloroplast and a cytoplasmic form, that preceeds Cu/Zn superoxide dismutase and that the Greek key beta-barrel structural motif. Proteins 5: 322–336.


Communicating editor: W.-H. Li

LITERATURE CITED


