UC Irvine

UC Irvine Previously Published Works

Title

Complete amino acid sequence of copper-zinc superoxide dismutase from *Drosophila melanogaster*.

Permalink

<https://escholarship.org/uc/item/9s31w0gf>

Journal

Archives of biochemistry and biophysics, 241(2)

ISSN

0003-9861

Authors

Lee, YM
Friedman, DJ
Ayala, FJ

Publication Date

1985-09-01

DOI

10.1016/0003-9861(85)90583-1

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at

<https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Complete Amino Acid Sequence of Copper-Zinc Superoxide Dismutase from *Drosophila melanogaster*¹

YOUNG MOO LEE,² DAVID J. FRIEDMAN, AND FRANCISCO J. AYALA³

Department of Genetics, University of California, Davis, California 95616

Received February 21, 1985, and in revised form May 10, 1985

The complete amino acid sequence of the *Drosophila melanogaster* Cu,Zn superoxide dismutase subunit has been determined by automated Edman degradation. Sequence analyses were performed on the intact *S*-carboxymethylated protein, two fragments derived from CNBr cleavage, and three peptides recovered from mouse submaxillary protease digestion of the reduced and *S*-carboxymethylated enzyme. The peptides were aligned by characterizing peptides yielded by trypsin and *Staphylococcus aureus* V8 protease. All the peptides studied were purified exclusively by reverse-phase columns of HPLC and were analyzed with an improved liquid-phase sequencer. A molecular weight of 15,750 (subunit) was calculated from the 151 residues sequenced. The amino acid sequence of the *Drosophila* superoxide dismutase subunit is compared with that of four other eucaryotes: man, horse, cow, and yeast. Comparison of the five primary structures reveals very different rates of evolution at different times. Copper-zinc superoxide dismutase appears to be a very erratic evolutionary clock.

Val-Val-Lys-Ala-Val-Cys-Val-Ile-Asn-Gly-Asp-Ala-Lys-Gly-Thr-Val-Phe-Phe-
20 30
Glu-Gln-Glu-Ser-Ser-Gly-Thr-Pro-Val-Lys-Val-Ser-Gly-Glu-Val-Cys-Gly-Leu-
40 50
Ala-Lys-Gly-Leu-His-Gly-Phe-His-Val-His-Glu-Phe-Gly-Asp-Asn-Thr-Asn-Gly-
60 70
Cys-Met-Ser-Ser-Gly-Pro-His-Phe-Asn-Pro-Tyr-Gly-Lys-Glu-His-Gly-Ala-Pro-
80 90
Val-Asp-Glu-Asn-Arg-His-Leu-Gly-Asp-Leu-Gly-Asn-Ile-Glu-Ala-Thr-Gly-Asp-
100
Cys-Pro-Thr-Lys-Val-Asn-Ile-Thr-Asp-Ser-Lys-Ile-Thr-Leu-Phe-Gly-Ala-Asp-
110 120
Ser-Ile-Ile-Gly-Arg-Thr-Val-Val-Val-His-Ala-Asp-Ala-Asp-Asp-Leu-Gly-Gln-
130 140
Gly-Gly-His-Glu-Leu-Ser-Lys-Ser-Thr-Gly-Asn-Ala-Gly-Ala-Arg-Ile-Gly-Cys-
150
Gly-Val-Ile-Gly-Ile-Ala-Lys.

© 1985 Academic Press, Inc.

¹ This research was supported by the National Institutes of Health Grant GM22221 and Department of Energy Contract PA200-14 Mod #4.

² To whom correspondence should be addressed, at Department of Biochemistry, Michigan State University, East Lansing, Mich. 48824.

³ To whom reprint requests should be addressed.

The Cu,Zn superoxide dismutase (SOD,⁴ EC 1.15.1.1) is an enzyme of considerable

⁴ Abbreviations used: TFA, trifluoroacetic acid; PTH, 3 phenyl-2-thiohydantoin; SDS, sodium dodecyl sulfate; PAM, accepted point mutation.

evolutionary interest. It is polymorphic in natural populations of various organisms. In some species of *Drosophila*, the frequencies of the alleles coding for different "electromorphs" (electrophoretically different forms of the enzyme encoded by the same gene locus) are highly variable from one to another locality (1, 2). This is unusual given that for most enzymes the alleles coding for different electromorphs tend to have very similar frequencies in different populations of the same species, even when the distribution of the species extends over much of a continent. In *D. melanogaster* two naturally occurring allelic forms of the enzyme differ in various biochemical properties, such as specific activity and thermostability (3, 4), and modify the fitness of the individuals carrying them.⁵

The complete amino acid sequence of the enzyme is known in only a few eucaryotes: bovine erythrocyte (5), human erythrocyte (6, 7), horse liver (8), and the yeast, *Saccharomyces cerevisiae* (9, 10). [The complete sequence of this enzyme has apparently been also obtained for the swordfish; see Ref. (11).] Comparison of the yeast amino acid sequence with that of the mammals has led to the suggestion that the enzyme may be evolving very slowly over long evolutionary spans (6). Yet, the degree of variation among the three mammal sequences known would indicate a fairly high rate of evolution.

As part of an extensive program directed towards understanding the evolution and functional significance of the Cu,Zn SOD polymorphisms in nature, we have obtained the complete amino acid sequence of the Cu,Zn dismutase in the fruit fly, *D. melanogaster*. The sequence is reported herein. The rate of evolution of the Cu,Zn dismutases appears to be highly variable at different times and in different lineages.

EXPERIMENTAL PROCEDURES

Materials. An isogenic strain (completely homozygous for the third chromosome where the gene

coding for Cu,Zn dismutase is located) was produced using *D. melanogaster* flies collected in Tunisia, Africa, by crosses to "balanced" laboratory stocks, following described procedures (12). Hundreds of thousands of flies homozygous for the superoxide dismutase locus were raised in laboratory cultures in order to provide the biomass necessary for enzyme purification.

Mouse submaxillaris protease and *Staphylococcus aureus* protease were purchased from Pierce; Tos-Phe-CH₂-Cl-trypsin was from Worthington Products; cyanogen bromide was from Aldrich. Iodoacetic acid (Sigma) was recrystallized from warm carbon tetrachloride. Acetonitrile and 2-propanol used for high-performance liquid chromatography were from Burdick & Jackson. Double-distilled water was filtered through Millipore (HA, 0.45 μ m) before application to HPLC columns. Reverse-phase columns (C-18) were purchased from Altex (Ultrasphere-ODS, 0.46 $\times$ 25 cm) and from Varian (MCH10 and protein C-18, 0.4 $\times$ 30 cm). Trifluoroacetic acid (TFA) and guanidine hydrochloride were Sequenal grade from Pierce.

Carboxymethylation. The protein sample (0.1 to 1 mg) in a presiliconized Pierce Reactivial was reduced with 2-mercaptoethanol and carboxymethylated with iodoacetic acid essentially as described by Gracy (13). The alkylated protein was dialyzed in the dark against water to remove excess buffer and salts.

Cyanogen bromide cleavage reaction. The alkylated protein sample (1 to 2 mg) was suspended in 1 ml of 70% formic acid using a small 14 $\times$ 3.5-mm spin bar in 1.5-ml-capacity Eppendorf tubes. CNBr crystals were added directly into the reaction vessel to give a concentration of 0.1 M. The cleavage reaction was allowed to continue for 24 h at room temperature. The acid and CNBr were evaporated at the completion of the reaction.

Proteolytic enzyme digestion. Arginyl bonds were selectively cleaved using a submaxillaris protease to protein ratio (w/w) of 1:50 in 0.2 M ammonium bicarbonate buffer, pH 7.9, for 16 h at 37°C. Tryptic digestion of derivatized protein was carried out in 0.2 M ammonium bicarbonate buffer, pH 7.9, at 37°C for 6 h with enzyme to substrate ratio (w/w) of 1:100. Cleavage at glutamic acids was accomplished by digestion with an *S. aureus* protease to protein ratio (w/w) of 1:50 in 0.1 M ammonium bicarbonate buffer, pH 7.8, at 37°C for 12 to 16 h. All enzymatic digestions were carried out in tightly sealed Eppendorf microtubes (1.5 ml) in order to avoid pH changes during the reaction periods. The buffer was exhaustively removed after digestion by repeated lyophilization. Freeze-dried enzymatic digests were kept at -20°C and dissolved in 0.1% TFA prior to application to the reverse-phase columns for fractionation.

Amino acid analysis. Acid hydrolysis of peptides was carried out in 5.7 N HCl in sealed, evacuated ignition tubes at 110°C for 24 h (14). Analysis of the

⁵ Unpublished experimental results from our laboratory.

hydrolyzates was made in a Durrum D-500 amino acid analyzer using a single-column, three-buffer elution program.

Sequence determination of purified peptides. Sequence analysis was performed by automated degradation on a Beckman Model 890 sequencer with program no. 031281. The PTH derivatives of the amino acids were identified by two independent methods including HPLC. Gas chromatography (15) as well as thin-layer chromatography (16) were used to confirm HPLC identification of PTH-amino acids, with the exception of PTH-His, which was only identified by HPLC.

Peptide purification on reverse-phase column of HPLC. All peptides generated by the various cleavage methods were exclusively purified by HPLC techniques. For fractionation of relatively large peptides, i.e., CNBr-cleaved peptides, a Varian MCH 10 reverse-phase column (C-18 bonded to 10- μ m silica gel) was used. When shorter peptides were expected (tryptic or Staphylococcal protease digests), an Altex Ultrasphere ODS column (C-18 bonded to 5- μ m silica gel) was adapted in order to achieve higher efficiency in fractionation. Lyophilized enzymatic digests in Eppendorf tubes were dissolved in 0.1% TFA (Solvent A); the tube was placed in an ultrasonic bath for a few seconds to increase the solubility of sparingly soluble peptides and to make a suspension of any insoluble peptides. The sample was then centrifuged before application to the column at 15,600g for 3 min in an Eppendorf microcentrifuge. A Varian Model 5000 and a Waters HPLC were used for peptide analysis and purification. The Varian HPLC was composed of a Rheodyne Model 7125 sample injector with a 10- or 100- μ l sample loop, a Perkin-Elmer Model LC-75 spectrophotometric detector, and a Hewlett-Packard Model 3390A reporting integrator. The Waters system included Models 6000A and M-45 solvent-delivery systems, a Model 660 solvent programmer, and a Model U6K universal injector. The peptide elution programs and C-18 columns used are described in the figure legends.

RESULTS

Strategy. Because the amounts of enzyme available for sequence determination were limited, we had to modify in some respects the conventional approaches. Efforts were made to conserve protein and peptides from loss during the multiple steps involved in the experimental process. Arginine protease was preferred to trypsin in restricting the cleavages at arginine residues of the protein in order to avoid a step for lysine modification. In cleaving the *S*-carboxymethylated protein, we generally avoided any chemical reaction re-

quiring multiple steps or giving incomplete fragmentations as much as possible. In peptide-fractionation procedures, HPLC was the only technical method we adopted for the purification of peptides generated by the various cleavage methods. Reverse-phase columns (C-18) of HPLC allowed us fractionation of nanomole quantities of peptides. Less than 1.5 μ mol of the native enzyme (subunit) was consumed for the complete amino acid sequencing of the *Drosophila* protein due to the efforts described.

N-Terminal residue. The *Drosophila* enzyme (3 nmol) was subjected to five cycles of Edman degradation in order to identify the amino-terminal residue. Quantitative recovery of the PTH derivative of Val was obtained as the terminal residue, followed by Val, Lys, Ala, and Val in that order, indicating that the amino terminus of the enzyme is not blocked. With this information, the first 50 residues of the *S*-carboxymethylated protein (50 nmol) were sequenced by automated Edman degradation (Fig. 1).

Cleavage at Met. A small quantity of the protein was tested to ascertain the extent of cleavage in 0.1 M CNBr in 70% formic acid. It is well known that the cleavage at Met by CNBr is incomplete in certain cases if the sequence is -Met-Thr- or -Met-Ser- (17). The degree of cleavage was monitored with 15% polyacrylamide SDS gels. The gels showed the presence of three bands, corresponding respectively to 16K (uncleaved), 10K, and to a smaller peptide, thus indicating incomplete cleavage [the *Drosophila* protein has only one Met residue as shown previously (3, 4)]. Gel scanning showed less than 50% cleavage by CNBr at the Met residue. The CNBr mixtures (65 nmol) were fractionated on a reverse-phase column as described in Fig. S-1. About 20 nmol of CNBr-1 (~6K) and 5 nmol of CNBr-2 (~10K) were recovered. The amino acid composition of the resolved peptides is shown in Table S-I.

The sequence of the CNBr-1 peptide analyzed up to 40 residues was identical to the sequence obtained from the intact enzyme. Therefore, Met was assumed to be located about one-third of the way

FIG. 1. Fragments of importance generated for the sequence of the Cu,Zn superoxide dismutase from *D. melanogaster*. The position of the residues of importance in fragmentation (m, for methionine; and r for arginine) is indicated in the top bar. The length of the segments is proportional to the number of amino acid residues comprised. Hatched sections represent portions of the sequence actually determined by automated Edman degradation. CNBr denotes peptides cleaved by CNBr; AP, peptides generated by arginine protease; T, tryptic peptides; and SP, staphylococcal peptides. T-SP was generated as a result of resistance to tryptic digestion at Lys 94. nr represents a nonrecovered peptide.

along the molecule from the N-terminal amino acid (Fig. 1). Sequence determination of CNBr-2 up to 30 residues showed a Ser-Ser sequence at the N-terminal side of this peptide, indicating that the sequence in the intact protein is -Met-Ser-Ser-, which accounts for the poor yield of CNBr-cleaved peptides. Due to the incomplete cleavage of Met by CNBr in the *Drosophila* protein, no experiments were designed for further subdigestion of CNBr fragments.

Cleavage at Arg. The enzyme contains three residues of Arg (3, 4); hence, we expected that four peptides or three (if Arg is at the C-terminal) would be obtained by arginine-specific proteolytic enzyme. The arginine protease digested sample was fractionated on the reverse-phase column by HPLC, as described under Cleavage at Met, and three peptide peaks were obtained (Fig. S-2). Identification of the peaks by amino acid analysis (Table S-II) and sequence determination showed a large peptide (AP-1, 1-77), a

medium-size peptide (AP-3, 114-141), and a small peptide (AP-4, 142-151). Amino acid composition analysis of AP-1 manifested that it contained one each of Met, Tyr, Ile, and Arg. Sequence analysis of the first 10 cycles of the peptides agreed with the CNBr-1 sequence, indicating that AP-1 is composed of the entire CNBr-1 and a part of CNBr-2 up to the Arg 77 residue. Since no Arg was found in AP-4, and Lys was at the C-terminal instead, this peptide was assigned to the C-terminus of the molecule. From sequence analysis of CNBr-2 and AP-1, we expected an AP-2 peptide (N-terminal residue is His), but peptide AP-2 was never recovered in our experiment perhaps due to the hydrophobic nature of the peptide. The alignment of the AP-2 peptide is illustrated in Fig. 1.

Cleavage at Arg and Lys. Given that the AP-2 peptide was not recovered under our experimental conditions, we attempted to generate a set of smaller soluble peptides by trypsin digestion of the peptide

AP-2 region. We also hoped to obtain by the same cleavage method a peptide overlapping between CNBr-1 and CNBr-2. We used HPLC fractionation on a reverse-phase column of Ultrasphere ODS (Altex) or a protein C-18 column (Varian) with a solvent gradient from 0.1% TFA to 0.075% TFA in acetonitrile. We recovered to homogeneity 10 peptides out of 12 expected fragments from the trypsin digests of carboxymethylated enzyme (Fig. S-3). Amino acid composition analysis of the recovered peptides (Table S-III) showed that peptide T-12 (residues 142-151) was identical to AP-4; that peptides T-10 (residues 114-133) and T-11 (residues 134-141) were subfragments derived from the peptide AP-3 region; that peptides T-1 (residues 1-3), T-2 (residues 4-13), and T-4 (residues 29-38) corresponded to the CNBr-1 peptide sequenced; and that peptide T-6 (68-77) was a part of the CNBr-2 peptide sequenced. Since T-5 (residues 39-67) contained one residue each of methionine and tyrosine, and part of this peptide had been sequenced from the intact carboxymethylated protein, it was clear that T-5 would overlap between the two CNBr peptides. Amino acid composition data of the remaining three peptides showed that none of them was the T-3 peptide. Therefore, we inferred that the three peptides, T-7 (residues 78-94), T-8 (residues 95-101), and T-9 (residues 102-113), are subfragments of the peptide AP-2. Direct sequence determination of five tryptic peptides (T-5, T-6, T-7, T-8, and T-9) was performed by the automated procedure. The sequenced peptides were aligned as illustrated in Fig. 1 based on the observation that T-9 has Arg at the C-terminal residue and because part of the sequence of T-7 can be deduced from the CNBr-2 sequencing data.

Cleavage at Glu. At this point, we still needed peptides overlapping between peptides AP-2, AP-3, and AP-4. In addition, the peptide T-5, which spans the two CNBr fragments, had been sequenced incompletely (residues 39-54), leaving a gap between residues 55-67 when about 10 nmol of this peptide was assigned for the sequence determination. The problem was caused by poor repetitive yield for each step but was overcome by replacing a

Staphylococcal peptide (SP-5, residues 48-68) for the T-5. Seven peptides out of nine expected SP fragments were recovered to homogeneity by HPLC fractionation (see Fig. S-4 and Table S-IV). In addition to SP-5, peptides SP-8 (residues 87-130) and SP-9 (residues 131-151) were sequenced by automated Edman degradation. We were able to sequence the entire SP-5 (21 residues) with 20 nmol of the material, thereby achieving our goal of overlapping between the two CNBr peptides. The SP-8 and SP-9 peptides provided the overlapping sequence between the three AP peptides from the C-terminal side of the protein.

A trypsin-resistant site. One tryptic peptide peak (T-3 + T-7; see Fig. S-3) separated by HPLC fractionation was not homogeneous, judging from amino acid analysis. Subdigestion with glutamic acid specific protease was done in order to characterize this peak. HPLC fractionation (Fig. S-5) and amino acid analysis of the subdigested peptides (Table S-V) revealed that the nonhomogeneous tryptic peptide peak contained two peptides (one is T-3 and the other comprises residues 78 to 101), indicating that Lys 94 was only partially digested by trypsin. It is not clear why Lys 94 is trypsin resistant. The peptide from residues 78 to 101 yields two smaller peptides (T-SP4 and T-SP5 in Fig. 1) by cleaving at Glu-86 with Staphylococcal protease. T-SP5 was sequenced since it also overlaps between T-7 and T-8.

The amino acid composition obtained from the sequenced *D. melanogaster* dismutase subunit is Asp₁₀, Asn₈, Thr₉, Ser₉, Glu₈, Gln₂, Pro₅, Gly₂₅, Ala₁₁, Cys₅, Val₁₅, Met₁, Ile₉, Leu₇, Tyr₁, Phe₆, His₈, Lys₉, and Arg₃, for a total of 151 residues. Molecular weights calculated from this composition are 15,750 for the subunit and 31,500 for the dimer, values that agree well with those obtained by gel filtration (32,000) and SDS-gel electrophoresis (16,000) (3).

DISCUSSION

The *Drosophila* dismutase was available to us only in small amounts. This is because the fraction of dismutase protein per fly is very small and the wet weight

of an individual fly is about 1 mg. About 1.5 μ mol of the protein (subunit) was sufficient to obtain the complete amino acid sequence of the protein. The process of sequencing was greatly facilitated by using HPLC in the purification of the peptides. One to twenty nanomoles of peptides were easily fractionated in reverse-phase columns that yielded well-compressed and resolved peptide peaks in most cases. The sequencer with an improved design Beckman 890 made it possible to sequence amounts as small as 1 nmol by automated Edman degradation.

A large fraction of the total material was consumed in the sequencing of two peptides, T-5 (residues 39-67) and T-7 (residues 78-94). T-5 has in the middle the sequence -Thr-Asn-Gly-Cys-Met-Ser-. The poor repetitive yield was caused by -Asn-Gly- and the surrounding sequence (9). In the yeast sequence, the problem was solved by purifying the peptide from the Staphylococcal digest of the apo-enzyme, followed by performic acid

oxidation in order to produce cysteic acid with improved repetitive yield. The fruit fly apo-enzyme, however, was not digested to any significant extent by proteolytic enzymes, including the Staphylococcal enzyme. It is known that -Asn-Gly- (or -Asn-Ala-) can undergo cyclization (18), the extent of which is affected by the surrounding sequence (9). The yeast and the *Drosophila* enzyme share the identical segment -Thr-Asn-Gly-Cys-. We solved the problem for the *Drosophila* enzyme by replacing T-5 (residues 39-67) with SP-5 (residues 48-68) in order to resolve the overlap between CNBr-1 and CNBr-2.

The poor yield of T-7 was due to the occurrence of a trypsin-resistant site at Lys 94. This difficulty was eased by obtaining the peptides SP-8 (residues 87-130) and T-SP5 (residues 87-101), as discussed under Results.

Figure 2 lists the complete amino acid sequence of Cu,Zn dismutase in *D. melanogaster* and four other eucaryotes. The sequences have been aligned so as to max-

		1		10		20		30		40																																
Yeast		V	Q	A	V	A	V	L	K	G	D	A	G	V	S	G	V	V	K	F	E	Q	A	S	E	S	E	P	T	T	V	S	Y	E	I	A	G	N	S	P		
Fruit Fly		V	V	K	A	V	C	V	I	N	G	D	A	K	G	T	V	F	F	E	Q	E	S		G	T	P	V	K	V	S	G	E	V	C	G		L	A			
Cow	Ac-A	T	K	A	V	C	V	L	K	G	D	G	P	V	Q	G	T	I	H	F	E	A	K		G	D	T	V	V	V	T	G	S	I	T	G		L	T			
Horse	Ac-A	L	K	A	V	C	V	L	K	G	D	G	P	V	H	G	V	I	H	F	E	Q	Q	Q	E	G	G	P	V	V	L	K	G	F	I	E	G		L	T		
Human	Ac-A	T	K	A	V	C	V	L	K	G	D	G	P	V	Q	G	S	I	N	F	E	Q	K	E	S	D	G	P	V	K	V	W	G	S	I	K	G		L	T		
		41		50		60		70		80																																
		N	A	E	R	G	F	H	I	H	E	F	G	D	A	T	N	G	C	V	S	A	G	P	H	F	N	P	F	K	K	T	H	G	A	P	T	D	E	N	V	R
		K	C	L	H	G	F	H	V	H	E	F	G	D	N	T	N	G	C	M	S	S	G	P	H	F	N	P	Y	G	K	E	H	G	A	P	V	D	E	N	R	
		E	G	D	H	G	F	H	V	H	E	F	G	D	N	T	Q	G	C	T	S	A	G	P	H	F	N	P	L	S	K	K	H	G	C	P	K	D	E	E	R	
		K	G	D	H	G	F	H	V	H	E	F	G	D	N	T	Q	G	C	T	T	A	G	A	H	F	N	P	L	S	K	K	H	G	C	P	K	D	E	E	R	
		E	G	L	H	G	F	H	V	H	E	F	G	D	N	T	A	G	C	T	S	A	G	P	H	F	N	P	L	S	R	K	H	G	C	P	K	D	E	E	R	
		81		90		100		110		120																																
		H	V	G	D	L	G	N	V	K	T	D	E	N	G	V	A	K	G	S	F	K	D	S	L	T	K	L	I	G	P	T	S	V	V	G	R	S	V	V	I	
		H	L	G	D	L	G	N	I	E	A	T	G	D	C	P	T	K	V	N	I	T	D	S	K	I	T	L	F	G	A	D	S	I	I	G	R	T	V	V	V	
		H	V	G	D	L	G	N	V	T	A	D	K	N	G	V	A	I	V	D	I	V	D	P	L	I	S	L	S	G	E	Y	S	I	I	G	R	T	M	V	V	
		H	V	G	D	L	G	N	V	T	A	D	E	N	G	K	A	D	V	D	M	K	D	S	V	T	S	L	S	G	K	H	S	I	I	G	R	T	M	V	V	
		H	V	G	D	L	G	N	V	T	A	D	K	D	G	V	A	D	V	I	E	D	S	V	I	S	L	S	G	D	H	C	I	I	G	R	T	L	V	V		
		121		130		140		150		154																																
		H	A	G	Q	D	D	L	G	K	G	D	T	E	E	S	L	K	T	G	N	A	G	P	R	P	A	C	G	V	I	G	L	T	N							
		H	A	D	A	D	D	L	G	Q	C	G	H	E	L	S	K	S	T	G	N	A	G	A	R	I	G	C	G	V	I	G	I	A	K							
		H	E	K	P	D	D	L	G	R	G	G	N	E	E	S	T	K	T	G	N	A	G	S	R	L	A	C	G	V	I	G	I	A	K							
		H	E	K	Q	D	D	L	G	K	G	G	N	E	E	S	T	K	T	G	N	A	G	S	R	L	A	C	G	V	I	G	I	A	P							
		H	E	K	A	D	D	L	G	K	G	G	N	E	E	S	T	K	T	G	N	A	G	S	R	L	A	C	G	V	I	G	I	A	Q							

FIG. 2. The amino acid sequences of Cu,Zn dismutase in five eucaryotes, aligned to maximize the number of homology sites. The sites at which all five enzymes share the same amino acid are boxed. The yeast dismutase is assumed to have a deletion at position 1 and an insertion at position 37. The fruit fly and cow enzymes have a two-amino acid deletion at 24-25. The NH₂ terminal residues are acetylated (Ac-) in the three mammalian enzymes, but not in the fruit fly or yeast. The sources of the sequences are yeast (9), cow (5), horse (8), human (7).

imize the number of homologous sites among them. The initial sequence of yeast from residues 1 to 37 has been shifted one residue position to the right. With this shift there are seven identical residues for all five organisms in this segment, plus seven others identical between the yeast and three of the four other organisms. Virtually no homologies exist between yeast and the other organisms without this shift, which had been suggested earlier by others (6-8).

The amino acid sequence of the fruit fly and the cow dismutases comprises 151 residues, two fewer than the horse and human sequences. In order to maximize the homology among the four animals, we have placed a two-residue gap at positions 24-25 in the fruit fly and cow sequences. This gap has been placed for the cow at positions 27-28 by Lerch and Ammer (8) and at positions 25-26 by Barra *et al.* (6). Inspection of the sequences indicates that greater homology is achieved by placing this deletion at positions 24-25 as we have done it in Fig. 2.

All critical amino acid residues for function and structure are essentially conserved in all five Cu,Zn SODs: the six histidine residues at positions 47, 49, 64, 72, 81, and 121 (Fig. 2), which contribute the two metal ligands, as well as Asp 84, which is also involved in metal liganding (25, 26). The glycine residues known to be essential to the formation of the characteristic β -strand barrel structure in Cu,Zn SOD are also largely conserved; 17 glycine residues (out of a total of 22 to 26 found

in the various sequences) are in identical positions in all five sequences.

We have evaluated the degree of differentiation among the five eucaryotic dismutases in various ways. First, we have counted the number of amino acid differences between each pair of organisms; and, using the genetic code, we have calculated the minimum number of nucleotide substitutions required to account for the observed amino acid differences. During the long period of time elapsed in the evolution of the lineages leading to the five organisms, some amino acid (or nucleotide) replacements may have been reversed; others may have been mediated by more than one substitution. There are several methods to account for these reversed and superimposed substitutions. The simplest method is to use Tables 36 and 37 from Ref. (19). These tables give the increments (per 100 residues) required to transform the observed residue differences into PAMs ("accepted point mutations"). One way of estimating the number of nucleotide substitutions per codon that separate the DNA sequences coding for two proteins is the statistic, D_E , proposed by Kimura and Ohta (20). Holmquist (21) has used somewhat different assumptions to estimate the total number of fixed nucleotide substitutions, or "random evolutionary hits" (REH), between two DNA sequences coding for a protein.

The results obtained with these various methods are summarized in Table I. For comparison purposes, this table also gives the average number of amino acid differ-

TABLE I
EVOLUTIONARY DIVERGENCE AMONG THE Cu,Zn SUPEROXIDE DISMUTASES OF FIVE EUCARYOTIC ORGANISMS

	Superoxide dismutase				Cytochrome c: amino acids	Time since divergence (million years)
	Amino acids	Nucleotides	D_E	REH		
Horse-cow	20.4 (1.0)	8.0 (1.0)	0.245 (1.0)	82.2 (1.0)	2.9 (1.0)	63 (1.0)
Human-horse/cow	22.6 (1.1)	8.8 (1.1)	0.265 (1.1)	86.3 (1.1)	10.0 (3.4)	75 (1.2)
Fruit fly-mammals	63.7 (3.1)	23.8 (3.0)	0.720 (2.9)	287.6 (3.5)	23.2 (8.0)	600 (9.5)
Yeast-animals	70.0 (3.4)	24.5 (3.1)	0.738 (3.0)	275.1 (3.3)	53.8 (18.6)	1200 (19.0)

Note. The amino acid replacements and minimum number of nucleotide substitutions are the incremented ("PAM") values per 100 residues. The amino acid replacements in cytochrome c are given for comparison. The normalized values obtained by dividing each figure by the horse-cow difference are given in parentheses.

is only somewhat greater than expected if it is assumed that the probability of such substitutions is constant (24). This is the case, for example, for cytochrome *c*, one of the proteins best studied in this respect. Table I shows the average number of amino acid replacements (PAM values) in the cytochromes *c* of the five eucaryotes; the agreement between the number of substitutions and the time elapsed is quite acceptable in this case.

What circumstances may account for the apparent variation in the evolutionary rate of the dismutase? One possible explanation is that there are a number of sites that are fixed; that is, that if function is to be preserved, there are sites at each of which only a given amino acid may be present. The number of amino acid replacements observed would then increase with evolutionary time, asymptotically approaching the maximum allowed. But, the maximum number of amino acid differences has not been reached between any pair of the organisms studied. When all five organisms are compared, the total number of variable sites is 92, although the largest number of amino acid differences between any two is only 74 (between *Drosophila* and yeast). In addition, it seems unlikely that 92 is the maximum number of variable sites in this enzyme. When the horse and the cow are compared, the number of variable sites is 27; when the human sequence is added, the number of variable sites increases by 12. Addition of the fruit fly sequence brings about 26 more variable sites; and the yeast sequence increases them by 17 more. It seems probable to us that when the sequences of other fungi and animal phyla are known, additional variable sites will appear.

One more anomaly seems to have occurred in the evolution of Cu,Zn SOD. The sequence alignment shown in Fig. 2 indicates that a deletion of two amino acids has occurred at positions 24 and 25 (or very near them) in the cow and in the *Drosophila* sequences. Because this double deletion is not present in the other two mammals, we must postulate that the deletion has occurred twice, independently in the cow and in the *Drosophila* lineages (see the branches marked with a cross in

Fig. 3). (Alternatively, we would need to postulate that an insertion of two amino acids at these two sites would have occurred thrice independently: in the human, the horse, and the yeast lineage.)

The independent occurrence of identical events of this kind in separate lineages is an unlikely evolutionary phenomenon. An explanation that would not require two independent evolutionary events to account for the same deletion in the cow and *Drosophila* lineage would be to postulate that some of the dismutase comparisons involve paralogous rather than orthologous genes. Paralogous genes are those that result from a gene duplication and thus reflect the time when the duplication occurred. Orthologous genes are those that have differentiated in different species derived from a common ancestral species, and thus reflect the time since the divergence of the species. We could assume in the present case that the cow and the fruit fly dismutases are orthologous to each other but paralogous to the human, horse, and yeast dismutases. The gene duplication and the 24-25 deletion in one of the duplicates would have occurred before the split of the *Drosophila* and the mammal lineages. But, in this case the cow dismutase should be as different as the *Drosophila* enzyme from the dismutase of the other two mammals. This is not the case and, therefore, this explanation is unlikely.

ACKNOWLEDGMENTS

We thank Mr. A. J. Smith and his staff in the Protein Structure Laboratory for their invaluable aid in the operation of the Beckman 890M sequencer and the Durrum D-500 analyzer.

REFERENCES

1. AYALA, F. J., POWELL, J. R., AND TRACY, M. L. (1972) *Genet. Res.* **20**, 19-42.
2. AYALA, F. J., POWELL, J. R., TRACY, M. L., MOURAO, C., AND PEREZ-SALAS, S. (1972) *Genetics* **70**, 113-139.
3. LEE, Y. M., AYALA, F. J., AND MISRA, H. P. (1981) *J. Biol. Chem.* **256**, 8506-8509.
4. LEE, Y. M., MISRA, H. P., AND AYALA, F. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7052-7055.

5. STEINMAN, H. M., NAIK, V. R., AVERNETHY, J. L., AND HILL, R. L. (1974) *J. Biol. Chem.* **249**, 7326-7338.
6. BARRA, D., MARTINI, F., BANNISTER, J. V., SCHININA, M. E., ROTILIO, G., BANNISTER, N. H., AND BOSSA, F. (1980) *FEBS Lett.* **120**, 53-56.
7. JABUSCH, J. R., FARB, D. L., KERSCHENSTEINER, D. A., AND DEUTSCH, H. F. (1980) *Biochemistry* **19**, 2310-2316.
8. LERCH, K., AND AMMER, D. (1981) *J. Biol. Chem.* **256**, 11545-11551.
9. JOHANSEN, J. T., OVERBALLE-PETERSON, C., MARTIN, B., HASEMANN, V., AND SVENDSEN, I. (1979) *Carlsberg Res. Commun.* **44**, 201-217.
10. STEINMAN, H. M. (1980) *J. Biol. Chem.* **255**, 6758-6762.
11. BARRA, D., SCHININA, M. E., SIMMACO, M., BANNISTER, J. V., BANNISTER, W., ROTILIO, G., AND BOSSA, F. (1984) *J. Biol. Chem.* **259**, 12595-12601.
12. McDONALD, J. F., AND AYALA, F. J. (1978) *Genetics* **89**, 371-388.
13. GRACY, R. W. (1977) in *Methods in Enzymology* (Hirs, C. H. W., and Timasheff, S. N., eds.), Vol. 47, pp. 195-204, Academic Press, New York.
14. MOORE, S., AND STEIN, W. H. (1963) in *Methods in Enzymology* (Colowick, S. P., and Kaplan, N. O., eds.), Vol. 6, pp. 819-831, Academic Press, New York.
15. PISANO, J. J., AND BRONZERT, T. J. (1969) *J. Biol. Chem.* **244**, 5597-5607.
16. KULBE, K. D. (1975) *Anal. Biochem.* **59**, 564-573.
17. SCHROEDER, W. A., SHELTON, J. B., AND SHELTON, J. R. (1969) *Arch. Biochem. Biophys.* **130**, 551-554.
18. HERMODSON, M. A., ERICSSON, L. H., TITANI, K., NEURATH, H., AND WALSH, K. A. (1972) *Biochemistry* **11**, 4493-4502.
19. DAYHOFF, M. O. (1978) in *Atlas of Protein Sequence and Structure*, Vol. 5, p. 375, Natl. Biomed. Res. Found., Silver Spring, Maryland.
20. KIMURA, M., AND OHTA, T. (1972) *J. Mol. Evol.* **2**, 87-90.
21. HOLMQUIST, R. (1978) *J. Mol. Evol.* **12**, 17-24.
22. DOBZHANSKY, TH., AYALA, F. J., STEBBINS, G. L., AND VALENTINE, J. W. (1977) *Evolution*. Freeman, San Francisco.
23. KIMURA, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge Univ. Press, London.
24. FITCH, W. M. (1976) in *Molecular Evolution* (Ayala, F. J., ed.), pp. 160-178, Sinauer Assoc., Sunderland, Massachusetts.
25. RICHARDSON, J. S., THOMAS, K. A., RUBIN, B. H., AND RICHARDSON, D. C. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1349-1353.
26. TAINER, J. A., GETZOFF, E. D., BEEM, K. M., RICHARDSON, J. S., AND RICHARDSON, D. C. (1982) *J. Mol. Biol.* **160**, 181-217.

MINIPRINT SUPPLEMENT

Table S-I. Amino acid composition of the cyanogen bromide peptides.

Peptides:	CNBr-1 ^a	CNBr-2 ^a
Residues:	1 - 56	57 - 151
CN-Cys	2.4 (3)	1.2 (2)
Asx	4.8 (5)	12.7 (13)
Thr	2.8 (3)	5.9 (6)
Ser	3.4 (3)	5.4 (6)
Glx	5.2 (5)	6.1 (5)
Pro	0.9 (1)	3.8 (4)
Gly	8.6 (9)	15.9 (16)
Ala	3.2 (3)	7.6 (8)
Val	7.9 (9)	7.3 (6)
Met	-- (1)	-- (0)
Ile	0.8 (1)	6.4 (8)
Leu	2.5 (2)	5.0 (5)
Tyr	(0)	0.9 (1)
Phe	3.3 (4)	2.4 (2)
His	2.6 (3)	4.9 (5)
Lys	4.0 (4)	5.0 (5)
Arg	(0)	3.0 (3)
Total	(56)	(95)

^a Cyanogen bromide-cleaved peptides were purified by HPLC fractionation as illustrated in Fig. S-1. The integral numbers in parentheses are from the complete amino sequence determined in our laboratory.

Table S-II. Amino acid composition of the arginine protease peptides^a.

Peptides:	AP-1	AP-2 ^b	AP-3	AP-4
Residues:	1-77	78-113	114-141	142-151
CN-Cys	3.2 (3)	1		0.7 (1)
Asx	7.9 (8)	6	3.8 (4)	
Thr	3.9 (3)	4	1.6 (2)	
Ser	4.2 (5)	2	2.0 (2)	
Glx	6.6 (7)	1	1.9 (2)	
Pro	3.8 (4)	1		
Gly	11.2 (12)	5	5.0 (5)	2.8 (3)
Ala	3.8 (4)	2	3.7 (4)	1.0 (1)
Val	8.4 (10)	1	2.2 (3)	1.3 (1)
Met	0.7 (1)			
Ile	1.2 (1)	5		2.1 (3)
Leu	2.5 (2)	3	1.7 (2)	
Tyr	1.0 (1)			
Phe	4.6 (5)	1		
His	4.9 (5)	1	1.5 (2)	
Lys	5.0 (5)	2	1.0 (1)	1.0 (1)
Arg	1.0 (1)	1	1.2 (1)	
Total	(77)	36	(28)	(10)

^a Peptides were purified on reverse phase column by HPLC as illustrated in Fig. S-2.

^b This peptide was not recovered under the HPLC elution program described in Fig. S-2. Integral numbers are obtained from the sequence determined in our laboratory.

Table S-III. Amino acid composition of tryptic peptides of carboxymethylated superoxide dismutase.

Tryptic	T-1	T-2	T-3 ^a	T-4	T-5	T-6	T-7	T-8	T-9	T-10	T-11	T-12 ^b
Residues:	1-3	4-13	14-28	29-38	39-67	68-77	78-94	95-101	102-113	114-133	134-141	142-151
CN-Cys		0.9 (1)		0.7 (1)	1.2 (1)		-- (1)					0.6 (1)
Asx		1.9 (2)			3.0 (4)	2.0 (2)	2.6 (3)	1.9 (2)	1.0 (1)	3.0 (3)	1.0 (1)	
Thr			2		0.9 (1)		1.4 (2)	0.9 (1)	0.9 (1)	0.8 (1)	0.9 (1)	
Ser			2	1.0 (1)	1.3 (2)			1.1 (1)	0.9 (1)	0.9 (1)	0.9 (1)	
Glx			3	1.3 (1)	1.1 (1)	2.0 (2)	1.2 (1)			2.0 (2)		
Pro			1		2.5 (2)	1.0 (1)	0.9 (1)					
Gly		1.1 (1)	2	2.3 (2)	5.2 (6)	1.0 (1)	2.5 (3)		2.0 (2)	3.0 (3)	1.9 (2)	2.8 (3)
Ala		1.7 (2)		0.9 (1)		1.1 (1)	1.2 (1)		1.0 (1)	2.1 (2)	1.9 (2)	1.0 (1)
Val	1.8 (2)	1.8 (2)	2	1.9 (2)	1.0 (1)	1.1 (1)	0.8 (0)	1.0 (1)		2.6 (3)		1.3 (1)
Met					0.8 (1)							
Ile		0.6 (1)					0.9 (1)	0.8 (1)	2.4 (3)			2.4 (3)
Leu				1.2 (1)	0.9 (1)		1.6 (2)		1.0 (1)	1.9 (2)		
Tyr					1.0 (1)							
Phe			2		3.2 (3)				1.0 (1)			
His					4.0 (4)	1.0 (1)	0.7 (1)			1.8 (2)		
Lys	1.0 (1)	1.0 (1)	1	1.0 (1)	1.0 (1)		1.0 (1)	1.0 (1)		1.0 (1)		1.0 (1)
Arg						0.9 (1)			0.7 (1)		0.9 (1)	

^a From the determined amino acid sequence of *Drosophila* dismutase. T-3 was coeluted with a peptide comprising all the residues 78-101 (due to tryptic resistant Lys 94). See Text. All integral numbers in parentheses and residue numbers are also from the sequence determined in our laboratory.

^b T-12 was not eluted by the water-acetonitrile solvent gradient system, but rather with the water-isopropanol system (result is not shown).

Table S-IV. Amino acid composition of the *Staphylococcal* V8 protease peptides of carboxymethylated dismutase.

Peptides:	SP-1 ^a	SP-2 ^a	SP-3	SP-4	SP-5	SP-6	SP-7	SP-8	SP-9
Residues:	1-19	20-21	22-32	33-47	48-68	69-75	76-86	87-130	131-151
CN-Cys	1			0.7 (1)	0.7 (1)			-- (1)	0.7 (1)
Asx	2				3.9 (4)	1.0 (1)	2.7 (3)	6.9 (7)	1.3 (1)
Thr	1		0.9 (1)		1.1 (1)			4.1 (5)	1.0 (1)
Ser			2.4 (3)		1.7 (2)			1.6 (2)	1.4 (2)
Glx	1	2	1.1 (1)	1.1 (1)	1.2 (1)	1.1 (1)	1.0 (1)	2.2 (2)	
Pro			1.0 (1)		1.9 (2)	0.9 (1)		0.9 (1)	
Gly	2		1.9 (2)	3.0 (3)	3.8 (4)	0.9 (1)	1.9 (2)	6.0 (6)	4.6 (5)
Ala	2			1.2 (1)		0.9 (1)		3.4 (4)	2.8 (3)
Val	5		1.9 (2)	1.9 (2)		0.9 (1)		4.3 (4)	0.8 (1)
Met					0.3 (1)				
Ile	1						0.9 (1)	3.9 (4)	2.6 (3)
Leu				1.7 (2)			1.7 (2)	2.2 (2)	0.9 (1)
Tyr					0.9 (1)				
Phe	2			0.9 (1)	1.8 (2)			1.0 (1)	
His				2.4 (3)	1.0 (1)	0.8 (1)	0.9 (1)	1.9 (2)	
Lys	2		1.0 (1)	1.0 (1)	1.0 (1)			2.0 (2)	2.0 (2)
Arg							0.9 (1)	1.1 (1)	0.8 (1)

^a From the determined primary structure of *Drosophila* dismutase. SP-1 and SP-2 were not characterized by amino acid composition analysis. Residue number and integral numbers in parentheses are also from the known amino acid sequence.

Table S-V. Amino acid composition of *Staphylococcal* protease-cleaved peptides of inhomogeneous tryptic fragments^a.

Peptides:	T-SP1	T-SP2 ^b	T-SP3	T-SP4	T-SP5
Residues:	14-19	20-21	22-28	78-86	87-101
CN-Cys					0.2 (1)
Asx				2.1 (2)	3.2 (3)
Thr	0.9 (1)		0.9 (1)		2.9 (3)
Ser			1.8 (2)		0.9 (1)
Glx	0.9 (1)	2		1.0 (1)	
Pro			1.0 (1)		0.9 (1)
Gly	1.0 (1)		1.2 (1)	2.6 (2)	1.1 (1)
Ala					1.0 (1)
Val	1.0 (1)		1.0 (1)		1.0 (1)
Met					
Ile				1.2 (1)	0.9 (1)
Leu				1.8 (2)	
Tyr					
Phe	1.7 (2)				
His				0.9 (1)	
Lys			1.0 (1)		2.0 (2)
Arg					

^a Unresolved peptides peak in Fig. S-3. Mixture of T-3 and a tryptic peptide. Residues 78-101, due to trypsin resistant Lys 94.

^b From the determined sequence of the *Drosophila* protein. Residue number and integral values in parentheses are also from the known amino acid sequence.

Fig. S-1. Fractionation of CNBr-cleaved peptides of the *Drosophila* β -carboxymethylated protein using HPLC. The column used was a reverse phase column C-18 (MCN-10, 0.45x30 cm from Varian). Ten to 65 nmole of the CNBr peptides were applied to the column and a gradient elution was performed from 0.1% TFA in water to 60% propanol-2 (0.07% TFA included) using the Varian HPLC system described in Experimental Procedures. Flow rate was 0.7 ml/min (1.42/min increase in solvent B) at room temperature. Peptide peaks were collected manually and concentrated by evaporation under vacuum for amino acid composition and sequence analyses. The two large peaks at retention times about 5 and 21 min are not peptide peaks since no detectable amino acids were found after 6N HCl hydrolysis. No uncleaved polypeptide was recovered under the experimental conditions.

Fig. S-2. Fractionation of arginine protease digests of the *Drosophila* S-carboxymethylated protein by HPLC. Ten to 50 nmol of peptides were applied to the column. Experimental conditions are the same as in Fig. S-1. Peaks with retention time less than 5 min. and retention times of 10 and 32.5 min. are not peptide peaks since the peaks were seen in blank samples (buffer and arginine protease) as well, and no detectable amino acids were found after 6 N HCl hydrolysis of the peaks. The two AP-1 peaks at about 38 and 40 min showed identical amino acid composition; they were not fully sequenced to confirm identity (most of the sequence information in AP-1 was already known from the data in Fig. S-1, CN8r-1 and CN8r-2).

Fig. S-3. Fractionation of tryptic peptides of the *Drosophila* S-carboxymethylated protein by HPLC. About 50 nmol of tryptic peptides were applied to the protein C-18 reverse phase column (large pore size from Varian) and washed for 5 min. with 0.1% TFA and eluted with an acetonitrile gradient system (0 to 50%). Flow rate was maintained at 1 ml/min using the Waters solvent delivery system described in Experimental Procedures. T-7' denotes a peptide, residues 78 to 101 (due to a trypsin resistant Lys 94). See Fig. S-5 for identification and characterization of this peptide.

Fig. S-4. Fractionation of the peptides cleaved by *Staphylococcus aureus* V-8 protease from S-carboxymethylated *Drosophila* dismutase on a reverse phase column. The C-18 column used was Altex, Ultrasphere ODS (0.4 x 25 cm). About 20 nmol of V-8 protease digest were applied to the column and washed with 0.1% TFA for 5 min. before gradient elution (described in Fig. S-3) started. Major peaks were collected and concentrated for further structural analysis. Amino acid composition analysis of the peaks at retention time 60.20 and 61.93 min. showed that they are one and the same peptide (SP-8), which was confirmed by sequence analysis.

Fig. S-5. HPLC fractionation of *S. aureus* V-8 protease subdigest of unresolved tryptic peptides (T-3 and T-7' in Fig. S-3). About 30 nmol of peptides were applied and fractionated as described in Fig. S-3; UD: undetermined peak.