Isolation and sequence of complementary DNA encoding human extracellular superoxide dismutase

(placental cDNA expression library/nucleotide sequence/amino acid sequence/molecular evolution/oxygen radicals)

KARIN HJALMARSSON*, STEFAN L. MARKLUND^{†‡}, ÅKE ENGSTRÖM[§], AND THOMAS EDLUND[¶]

*Department of Molecular Genetics and Cell Biology, SYN-TEK AB, Box 1451, S-901 24 Umeå, Sweden; [†]Department of Clinical Chemistry, Umeå University Hospital, S-901 85 Umeå, Sweden; [§]Department of Immunology, Biomedical Center, Box 582, S-751 23 Uppsala, Sweden; and [§]Department of Microbiology, Umeå University, S-901 87 Umeå, Sweden

Communicated by Irwin Fridovich, May 8, 1987

ABSTRACT A complementary DNA (cDNA) clone from a human placenta cDNA library encoding extracellular superoxide dismutase (EC-SOD: superoxide:superoxide oxidoreductase, EC 1.15.1.1) has been isolated and the nucleotide sequence determined. The cDNA has a very high G+C content. EC-SOD is synthesized with a putative 18-amino acid signal peptide, preceding the 222 amino acids in the mature enzyme, indicating that the enzyme is a secretory protein. The first 95 amino acids of the mature enzyme show no sequence homology with other sequenced proteins and there is one possible N-glycosylation site (Asn-89). The amino acid sequence from residues 96–193 shows strong homology (\approx 50%) with the final two-thirds of the sequences of all known eukaryotic CuZn SODs, whereas the homology with the P. leiognathi CuZn SOD is clearly lower. The ligands to Cu and Zn, the cysteines forming the intrasubunit disulfide bridge in the CuZn SODs, and the arginine found in all CuZn SODs in the entrance to the active site can all be identified in EC-SOD. A comparison with bovine CuZn SOD, the three-dimensional structure of which is known, reveals that the homologies occur in the active site and the divergencies are in the part constituting the subunit contact area in CuZn SOD. Amino acid sequence 194-222 in the carboxyl-terminal end of EC-SOD is strongly hydrophilic and contains nine amino acids with a positive charge. This sequence probably confers the affinity of EC-SOD for heparin and heparan sulfate. An analysis of the amino acid sequence homologies with CuZn SODs from various species indicates that the EC-SODs may have evolved from the CuZn SODs before the evolution of fungi and plants.

Extracellular superoxide dismutase (EC-SOD; superoxide: superoxide oxidoreductase, EC 1.15.1.1) is the major SOD isoenzyme in extracellular fluids such as plasma, lymph (1), and synovial fluid (2) but occurs also in tissues (3, 4). EC-SOD is heterogenous with regard to binding to heparin-Sepharose and can be separated into three fractions: A, without affinity; B, with weak affinity; and C, with relatively high affinity (5). Most EC-SOD in the vascular system is apparently bound to endothelial cell surfaces (6). Membranebound heparan sulfate is the likely receptor for the enzyme (K. Karlsson and S.L.M., unpublished data) and EC-SOD fraction C is the form that binds (6).

EC-SOD is a tetrameric glycoprotein with an apparent subunit molecular weight of $\approx 30,000$ (5). Like the CuZn SODs, EC-SOD contains one Cu and one Zn atom per subunit (5, 7) and is inhibited by cyanide, azide, diethyldithiocarbamate, and H₂O₂ (8). As for the CuZn SODs, H₂O₂ is the product of the catalyzed reaction (9). Still, despite the similarities, the amino acid compositions of human CuZn SOD and EC-SOD are clearly different (5) and no crossreactions between rabbit polyclonal antibodies directed toward the enzymes have been observed (8). The present paper presents the isolation and characterization of a complementary DNA (cDNA) clone from human placenta encoding EC-SOD. The deduced amino acid sequence is compared with those of sequenced CuZn SODs.^{||}

MATERIALS AND METHODS

Chemicals. Restriction endonucleases were obtained from Boehringer Mannheim. ${}^{35}S$ - and ${}^{32}P$ -labeled nucleotides and nitrocellulose filters were purchased from Amersham. DuPont Cronex 4 x-ray film was obtained from E. I. DuPont de Nemours.

Strains, Phages, and Plasmids. Escherichia coli Y 1090 (10) was used for plating of the λ gt11 placenta cDNA library, *E. coli* HB101 (11) was used for propagation of plasmids, and *E. coli* strain JM 103 (12) was for propagation of M13 vector mp9 derivatives. The human placenta cDNA library prepared in the vector λ gt11 was obtained from Clontech Laboratories (Palo Alto, CA) (catalogue no. HL 1008 lot 1205). For determination of the nucleotide sequence of cDNA clones, the M13 vector mp9 (13) was used. Subcloning of cDNA fragments was carried out in plasmid pUC18 (14).

Amino Acid Sequence Analysis. EC-SOD type C, isolated from human umbilical cords (7), was desalted by gel filtration in 50 mM NH₄HCO₃. It was then reduced with dithiothreitol (2 mM) in 6 M guanidinium hydrochloride/0.5 M Tris-HCl, pH 8.0, for 1 hr at 37°C and alkylated with iodo[³H]acetic acid (2.2 mM) for 1 hr at room temperature in the dark. Digestion of the reduced and alkylated EC-SOD with trypsin 1:50 (wt/wt) was done in 0.1 M NH₄HCO₃ containing 0.1 mM CaCl₂ at 37°C for 1.5 hr. Separation of tryptic peptides was done by reversed-phase chromatography on a μ Bondapak C₁₈ column (Waters Associates) eluted with a linear gradient of acetonitrile in 0.1% trifluoroacetic acid. The eluate was monitored at 220 nm and liquid scintillation counting identified cysteine-containing fractions. The intact protein and isolated tryptic peptides were sequenced by automated Edman degradation (15) on a Beckman 890 C sequencer as described (16). In the intact protein, the sequence of the first 33 amino-terminal amino acids was determined (cf. Fig. 1).

Cloning and Sequencing of Human EC-SOD. *Preparation of a DNA probe.* On the basis of the amino-terminal amino acid sequence of human EC-SOD (see Fig. 1) and the postulated codon usage for eukaryotic proteins (17), a synthetic 48-mer deoxyoligonucleotide (5' CAGGGACATGTATGCCAAG-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SOD, superoxide dismutase; EC-SOD, extracellular SOD.

[‡]To whom reprint requests should be addressed.

The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02947).

		-18	-10	
		MetLe	uAlaLeuLeuCysSerCysLeuLeuLeu	AlaAlaGlyAlaSerAsp
CTGGGTGCAGCTCTCTTTC	AGGAGAGAAAAGCTCTCTTGGAGGAG	CTGGAAAGGTGCCCGACTCCAGCCATGCT	GGCGCTACTGTGTTCCTGCCTGCTCCTG	GCAGCCGGTGCCTCGGAC
1				120
-1 +1	10	20	30	
AlaTrpThrGlyGluAspSe	rAlaGluProAsnSerAspSerAla	GluTrpIleArgAspMetTyrAlaLysVa	lThrGluIleTrpGlnGluValMetGlr	ArgArgAspAspAspGly
GCCTGGACGGGGGGGGGGGGGCGAGGACTCO	GGCGGAGCCCAACTCTGACTCGGCG	GAGTGGATCCGAGACATGTACGCCAAGGT	CACGGAGATCTGGCAGGAGGTCATGCAG	CGGCGGGGACGACGACGGC
10	50	6		240
The Lougin Michle Concel	50 - Vel Cla Das Caultantaria	60	70	
ACCCTCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	nvalginproseralarnrLeuAsp	AlaAlaGInProArgvalfhrGlyvalva	<u>ILeupheArgGinLeuAlaproArgAla</u>	LysLeuAspAlaPhePhe
ACCELCACOCCOCCIOCCA	GOIGCAGCCGICGGCCACGCIGGAC	GCCGCGCAGCCCCGGGTGACCGGCGTCGT	CUTUTTCCGGCAGCTTGCGCCCCGCGCC	AAGCTCGACGCCTTCTTC
80	90	100	110	360
AlaLeuGluGlvPheProTh	rGluProAsnSerSerGerArgAla	TleHieValHieGlnPheGluAcpTeuSe	relacivereciuserThrelyProHic	TURNERBROLOUNIAU
GCCCTGGAGGGCTTCCCGAC	CGAGCCGAACAGCTCCAGCCGCGCC	ATCCACGTGCACCAGTTCGGGGACCTGAG	CCAGGGCTGCGAGTCCACCGGGCCCCAC	TACAACCCGCTGGCCGTG
				480
120	130	140	150	
ProHisProGlnHisProGl	yAspPheGlyAsnPheAlaValArg	AspGlySerLeuTrpArqTyrArqAlaGl	vLeuAlaAlaSerLeuAlaGlvProHig	SerIleValGlyArgAla
CCGCACCCGCAGCACCCGGG	CGACTTCGGCAACTTCGCGGTCCGC	GACGGCAGCCTCTGGAGGTACCGCGCCGG	CCTGGCCGCCTCGCTCGCGGGCCCGCAC	TCCATCGTGGGCCGGGCC
				600
160	170	180	190	
ValValValHisAlaGlyGl	uAspAspLeuGlyArgGlyGlyAsn	GlnAlaSerValGluAsnGlyAsnAlaGl	yArgArgLeuAlaCysCysValValGly	ValCysGlyProGlyLeu
GTGGTCGTCCACGCTGGCGA	GGACGACCTGGGCCGCGGCGGCAAC	CAGGCCAGCGTGGAGAACGGGAACGCGGG	CCGGCGGCTGGCCTGCTGCGTGGTGGGG	GTGTGCGGGCCCGGGCTC
300				720
200	210	220		
TrpGluArgGInAlaArgGI	UHISSerGluArgLysLysArgArg	ArgGluSerGluCysLysAlaAla***		
IGGGAGCGCCAGGCGCGGGGA	GCACTCAGAGCGCAAGAAGCGGCGG	CGCGAGAGCGAGTGCAAGGCCGCCTGAGC	GCGGCCCCCACCCGGCGGCGGCCAGGGA	ACCCCCGAGGCCCCCCTCT
GCCTTTGAGCTTCTCCTCTC	~~~~	COMPANY COMPANY CONTRACTOR NO CONCENCIÓN CONCENCIÓN CONTRACTOR CONCENCIÓN CONTRACTOR CONCENCIÓN CONCENCENCIÓN CONCENCIÓN CONCENCENCENCENCENCENCENCENCENCENCENCENCEN		840
deciritanderrerecterd	CICCARCAGACACCIICCACICIGA	GGICICACCIICGCCICIGCIGAAGICIC	CUCGUAGECUTUTULAUCUAGAGGTUTU	LUTATACUGAGACCCACC
ATCOTTCOATCOTGAGGACCO	GCCCCAACCCTCGGAGCCCCCCACT		G & & & C & C & C & C & C & C & C & C &	980
		CAGINGGICIGANGGCCICCATTIGIACC	GARACACCCCGCICACGCIGACAGCCIC	1080
CTTTCCACCCAGACCCTCCT	TCCCCACCCCATAAGCCCTGAGACT	CCCGCCTTTGACCTGACGATCTTCCCCCT	TECEGEETTEAGGTTEETAGGEGE	
				1200
GTTGCCTCGAGTCCCCCCAC	CCCTCCCCACCACCACCGCTCCCG	CGGCAAGCCAGCCCGTGCAACGGAAGCCA	GGCCAACTGCCCCGCGTCTTCAGCTGT"	TTCGCATCCACCGCCACCC
				1320
CACTGAGAGCTGCTCCTTTG	GGGGAATGTTTGGCAACCTTTGTGT	TACAGATTAAAAATTCAGCAATTCAAAAA	AA	
		13	96	

FIG. 1. cDNA sequence and deduced amino acid sequence of EC-SOD. Underlined amino acid sequences have been confirmed by amino acid sequence analysis of intact EC-SOD and of tryptic peptides.

GTGACTGAGATCTGGCAGGAGGTGATGCA 3') complementary to the coding strand of the EC-SOD gene was synthesized (18).

Screening of a human placenta cDNA library. The recombinant phages of the $\lambda gt11$ human placenta cDNA library were screened for human EC-SOD cDNA sequences by plating the phages on the indicator strain E. coli Y 1090. Transfer of plaques to and treatment of nitrocellulose filters were done essentially as described by Maniatis et al. (19). Eight nitrocellulose filters, to which 20,000 plaques had been transferred per filter, were presoaked in 0.75 M NaCl/75 mM Na₃ citrate, and then prehybridized for 1 hr at 41°C in 40 ml of 20% formamide/750 mM NaCl/75 mM Na₃ citrate/0.1% (wt/vol) bovine serum albumin/0.1% (wt/vol) Ficoll/0.1% (wt/vol) polyvinylpyrrolidone/50 mM sodium phosphate, pH 6.8/50 μ g of denatured sonicated calf thymus DNA per ml. The hybridization was performed in the prehybridization solution supplemented with 100 μ M ATP (final concentration) and 7 × 10⁵ cpm of the [γ -³²P]ATP end-labeled 48-mer probe per ml, described above, for 18 hr at 41°C. After incubation, the filters were washed once in 30 mM NaCl/3 mM Na₃ citrate at 37°C followed by four washes in the same solution containing 0.1% NaDodSO4 at 37°C and allowed to air dry. The filters were exposed to DuPont Cronex 4 x-ray film overnight. Each filter contained about six positive plaques. Phages from plaques showing a positive hybridization reaction were isolated and purified and DNA was extracted (20). The length of the cDNA inserts was determined by agarose gel electrophoresis after cleavage of the phage DNA with EcoRI. The recombinant phage carrying the largest cDNA insert was designated λ sp3 and was chosen for further studies.

Subcloning and sequencing of cDNA encoding human EC-SOD. About 30 μ g of λ sp3 was cleaved with EcoRI and the cDNA insert was separated from λ DNA by electrophoresis in a 6% polyacrylamide gel. The cDNA fragment was isolated from the gel by electroelution, phenol and chloroform extraction, and ethanol precipitation (19). Then, 0.05 μ g of the isolated cDNA fragment was ligated to 1 μ g of EcoRI-digested alkaline phosphatase-treated pUC18 DNA and then transformed into strain E. coli HB101. Transform

ants were selected on plates containing ampicillin. A recombinant plasmid carrying the cDNA insert was identified and designated pLS3. Plasmid pLS3 DNA was subjected to restriction endonuclease analysis. The complete nucleotide sequence of both DNA strands of the cDNA insert of phage λ sp3 was determined (12, 21–23) after subcloning into the *Eco*RI site of the M13 vector mp9 and generating a sequential series of overlapping clones by the method of Dale *et al.* (24).

Computer Programs and Analysis. All nucleotide sequencing data were compiled and analyzed using the GENEUS computer system (25). Homologies with other proteins were searched for using the Protein Identification Resource, containing 3557 sequences.**

RESULTS

Nucleotide Sequence of Human EC-SOD cDNA. The nucleotide sequence and deduced amino acid sequence of the cDNA insert of λ sp3 are shown in Fig. 1. The insert was found to be 1396 base pairs (bp) long and contained an open reading frame of 240 amino acids surrounded by 69-bp and 607-bp untranslated regions of the 5' and 3' ends, respectively. Underlined amino acid sequences have been confirmed by amino acid sequence analysis of intact EC-SOD and of tryptic peptides. The start of mature EC-SOD (cf. Materials and Methods) corresponds to amino acid 19 of the deduced amino acid sequence. The preceding 18 amino acids apparently represent a signal peptide. Like known signal peptides, this sequence of amino acids is rich in hydrophobic amino acids and the last residue is alanine, which is one of the amino acids found in this position in known signal peptides (26). The DNA sequence found at the translation initiating codon, -CAGCCAUGC-, is homologous to the postulated consensus sequence for eukaryotic initiation sites, -CCACC-AUG(G)- (27). A possible polyadenylylation signal with the sequence -ATTAAA- homologous to the postulated consen-

^{**}Protein Identification Resource (1986) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 8.0.

sus sequence AATAAA is found 14 bp upstream of the polyadenylylation tail (28).

The codon usage of the gene encoding EC-SOD is shown in Table 1. There is a strong preference for codons containing a maximal number of G+C. The usage of G+C at the third position of codons is one of the highest (96%) found in vertebrate genes (29).

Deduced Amino Acid Sequence. The deduced amino acid sequence of EC-SOD is presented in Figs. 1 and 2. In Fig. 2, the sequences of the human, pig, cow, horse, swordfish, fruit fly, spinach, yeast, and *P. leiognathi* CuZn SODs are aligned along the EC-SOD sequence. The sequence starts with an 18-amino acid signal peptide preceding the amino terminus of mature EC-SOD as discussed above. The mature protein contains 222 amino acids and the calculated molecular weight is 24,174. The amino acid composition of the deduced sequence is very similar to the amino acid compositions of human native EC-SOD isolated from lung (5) and umbilical cord (7) and recombinant EC-SOD produced by expression of the cDNA in Chinese hamster ovary cells (7).

The carboxyl-terminal end of EC-SOD (residues 194–222) is very hydrophilic and the most distinctive feature is the high content (nine) of positively charged amino acids. The first 95 amino acids of EC-SOD, among which the only N-glycosylation site is found (Asn-89), show no sequence homologies with the CuZn SOD family (Fig. 2). Nor was it possible to find any significant homologies with other proteins in the Protein Identification Resource.

In the region His-96 to Gly-193, the EC-SOD sequence shows strong homologies with the CuZn SOD family. Of the 76 positions in this region, in which at least six of the nine CuZn SODs share amino acids (enclosed in Fig. 2), EC-SOD has an identical amino acid in 49 positions. EC-SOD shares amino acids in 22 of the 23 positions in which all the CuZn

 Table 1.
 Codon usage for the gene encoding human EC-SOD including signal peptide

Amino		No. of	Amino		No. of
acid	Codon	codons	acid	Codon	codons
Leu	UUA	0	Ala	GCU	1
	UUG	0		GCC	21
	CUU	1		GCA	1
	CUC	7		GCG	9
	CUA	1	Gly	GGU	1
	CUG	11	-	GGC	17
Ser	UCU	1		GGA	0
	UCC	4		GGG	5
	UCA	1	Ile	AUU	0
	UCG	5		AUC	4
	AGU	0		AUA	0
	AGC	6	Phe	UUU	0
Arg	CGU	0		UUC	7
	CGC	8	Tyr	UAU	0
	CGA	1	-	UAC	3
	CGG	10	His	CAU	0
	AGA	0		CAC	9
	AGG	1	Gln	CAA	0
Val	GUU	0		CAG	11
	GUC	7	Asn	AAU	0
	GUA	0		AAC	7
	GUG	10	Lys	AAA	0
Pro	CCU	0		AAG	5
	CCC	5	Asp	GAU	0
	CCA	0		GAC	14
	CCG	8	Glu	GAA	0
Thr	ACU	0		GAG	15
	ACC	3	Cys	UGU	1
	ACA	0		UGC	7
	ACG	4			

SODs contain identical amino acids in the region. If EC-SOD is compared with bovine CuZn SOD, the three-dimensional structure of which is known (40), it is found that all ligands to Cu (His-96, His-98, His-113, and His-163) and Zn (His-113, His-121, His-124, and Asp-127) can be identified in EC-SOD. The cysteines forming the intrasubunit disulfide bridge in the CuZn SODs also have their counterparts in EC-SOD; Cys-107 and Cys-189. The arginine in the active site entrance of the CuZn SODs (40), the chemical modification of which leads to inactivation of the enzyme (41), is also found in EC-SOD, Arg-186. The homology is lower from residues 117-123, where some amino acids are deleted in EC-SOD. In bovine CuZn SOD, this corresponds to a part of the loop in the "zinc ligand region" (40). Another less homologous sequence is in the region 137-147. This part is also less conserved among the CuZn SODs. The last significant difference is seen at residues 175-180, corresponding to a part of the "active site lid loop" in bovine CuZn SOD (40).

DISCUSSION

The amino acid sequence of human EC-SOD, as deduced from the cDNA, contains a signal peptide, which indicates that EC-SOD is a secretory protein as postulated (1, 3-5). The sequence of the central portion of the mature EC-SOD is highly homologous with the final two-thirds of the sequences of the CuZn SODs. The carboxyl-terminal end of EC-SOD is very hydrophilic and contains many positively charged amino acid residues. It is therefore likely to extend out into the solvent. We propose that this part of the protein confers the affinity of EC-SOD for heparin (5) and analogues such as heparan sulfate leading *in vivo* to binding to cellular surfaces (ref. 6; K. Karlsson and S.L.M., unpublished data).

The region of the EC-SOD sequence, residues 96–193, homologous to the CuZn SOD family corresponds to residues 44–148 in bovine CuZn SOD, the three-dimensional structure of which is known in detail (40). The first 43 residues of bovine CuZn SOD constitute B-strands "1a, 2b, 3c," and the first half of "6d." EC-SOD was also deviant from residues 137–147 (92–102 in bovine CuZn SOD), which correspond to B-strand "4f." These parts of bovine CuZn SOD are positioned in the outer surface of the "B-barrel," which constitutes the main part of the subunit contact area in the enzyme. Since EC-SOD is a tetramer, its subunit contact area should be different and larger. It is possible that the deviant sequences in the homologous region and parts of the first 95 amino acids of EC-SOD are responsible for the tetramer subunit contact.

The parts of bovine CuZn SOD that are homologous to EC-SOD form the environment of the Cu and Zn atoms. All ligands to the Cu and Zn atoms can be identified in EC-SOD, as can the cysteines forming the intramolecular disulfide bond. Arg-186 in EC-SOD corresponds to Arg-141 in the entrance to the active site of bovine CuZn SOD. The conformation of EC-SOD around the active site is therefore expected to be very similar to that of the CuZn SODs. This would explain the similarities of EC-SOD and CuZn SODs with respect to specific activity (5) and sensitivity to inhibitors like cyanide, azide, diethyldithiocarbamate, and H_2O_2 (8).

Comparison of EC-SOD with sequences of CuZn SODs from different species allows some speculations on the evolutionary history of the EC-SODs (Fig. 2). If the 98 amino acids of EC-SOD in sequence positions 96–193 are compared with homologous positions in the CuZn SODs, the following numbers of identical amino acids are found: human, 53; pig, 56; cow, 55; horse, 52; swordfish, 52; fruit fly, 49; spinach, 54; bakers' yeast, 48; *P. leiognathi*, 30. Except for *P. leiognathi*, the identities with the CuZn SODs from various species cannot be judged to differ significantly and follow any order of evolution. For comparison, a reasonable order

Biochemistry: Hjalmarsson et al.

1	1 WTE	G	D :	s A	N E	P	10) I S	D	S	A	E	W	I	R	2 D	0 M	Y	A	ĸ	۷	T	E	I	W	Q Q	0 E '	V I	M	QI	RF	₹ C) (0) G	
	40 T L H Human Pig Cow Horse Sword Fruit Spina Baker P. le	A fi fi ch s jog	A uZ ish y yea yna	c (n st th) \ Sor " " "	(()) F # # # # ! !	50 SA T A T A L V L V N A T V N	A			D V V V V V V V V T	A C C C C C C C C C C V	A V V V V V V K		PKKKKRNKKT	6 R G G G G G G G G G G G G G G G G G G		T G G G G A S A O	G P P P P E K N G T	V V V V G V G G	VQQQHTTESK		F I T T V T F V V V	R I I V F V G	Q N Y H H Y E T K T			PI QI QQ QQ L	RKKKQESEAS	A I G I G I G I D I S Q	(. 0 6 - 0 6 0 7 0 7 0 7 0 7 0 7 1 7 1 7 1 7 1 7 1 7 1 7 1 7 1 7 1 7 1	D A FI		
	80 FALW VVT VVLK GKVS VS VS FT VF	F. G. G. G. G. G. V. Y. P	G S T S F I E R E E	F I I I I I I I I	PKKTEKCSA	T G G G G G G G G G G G G G G G G G G G	E - - - - -		90 N 90 T P T P T P		SLDDDELKEM		A G G G G G G G G G G G G G G G G G G G	I F F F F F F F F F F F	****	V V V V V V V V I I I		1 Q E Q E G E E E Q	OC FFFFFFFF N	000000000000000000000000000000000000000					666666666	000000000000000000000000000000000000000	E T T T M M V A	1 S S S S S S S S S S S S S S S S S S S	10 T A A A A A A A A A A A A	6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	P P P P P P P P G	н н н н н н н	Y F F F F F F F	N N N N N N N N	P P P P P P P P P P P P P	
	L A S R E S S K K K K K K K K K K K K K K K K K		0 H H H H H H H H H H H H H H H H H H H	- <u> </u>	- G G G G G G A A F	P P P P P P P P P	- K K K K V E T W	D D D D D D D D T					6 6 6 6 6 6 6 6 6 6 6		FLLLLLML	4000000000	13(N N N N N N N N N N N		F T T T T T E V K F		V D G D D D T N D S	R K K K E A G T E A	D D N N D D N N		S V V V V V V V V V V V V V V V V V V V	L A A A A A A A A A A	1 W D T I D K K E K T	40 R V V V V I V A G N	Y S Y D D N T S P	R I I M I I F V	A E V K T T V K L	GDDDDDDDA	L S S P S - S N S P	A V V L V K K Q L -		
	150 L A C L S C L S C L S C L T C L T C L T C L T C		H H H Y H Y D N T -	S S S S S S S S S S L	I I I I V V L	V I I I V K	666666666666		1 A T T T T T T T A S A	60 VL M M M V L V I I				G K K K K K D L G G	E A P P Q A A E Q G				17 GGGGGGGGG		6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6					S S S S S S S P	V T T T L K P L K	1 ΕΚΚΚΚΚΟΤΚΑ	80 N T T T T T T T T T	000000000000000000000000000000000000000	N N N N N N N N N N N N N N	A A A A A A A A G	666666666666666666666666666666666666666	R S S S S S S S S S A G P A	R R R R R R R R R R R R R R	
	A C A C A C A C A C A C A C A C A C A C			000000000000	V I I I I I L L	C A T A A E A T T	G Q Q K P K P N	P V	G	2 L H P C H S F S B P	00 W ig ow or vo ru ak	E F an se rd it nac er:	fi fi fi sy) A ish iy yea	R	: E	н	S C	EuZ	21 R	0 80	R D	(R	R	R	ε	S	E	c	221 K	A	A				

FIG. 2. Comparison of the deduced amino acid sequence (designated by the single-letter code) of mature human EC-SOD with those of CuZn SODs from human (30, 31), pig (32), cow (33), horse (34), swordfish (35), fruit fly (36), spinach (37), bakers' yeast (38), and *P. leiognathi* (39). Homologous amino acids are enclosed in boxes in positions where six or more are shared by the nine CuZn SODs. *, Sequence (SEKDGKVVLGGA) in *P. leiognathi* not presented.

emerges if the same 98 amino acid positions are compared among the CuZn SODs with the human enzyme as the basis: pig, 86; cow, 83; horse, 84; swordfish, 77; fruit fly, 64; spinach, 63; bakers' yeast, 62; P. leiognathi, 30. This analysis indicates that the EC-SODs have evolved from the CuZn SODs, after the bacteriocupreins but before the occurrence of fungi and plants. The need for an extracellular and cell-surface-associated SOD must have occurred early in evolutionary history. In spite of the proposed long evolutionary separation between the EC-SODs and the eukaryotic CuZn SODs, the homologous sequence is highly conserved and must be important for the function of the enzymes. In accordance with this view is the fact that only 12 of the 53 amino acids identical in EC-SOD and human CuZn SOD are identical on the nucleotide level (Fig. 1; refs. 30 and 31). The EC-SODs might be widely distributed among higher phyla. Almost nothing is known about this distribution; the EC-SODs have so far only been looked for and found in mammals (4), birds, and fish (S.L.M., unpublished data).

The skillful technical assistance of Ms. Stina Olofsson and assistance with computer analysis of Dr. Robert Harr is gratefully acknowledged. Dr. Lena Tibell and Dr. Gunnar Skogman are thanked for valuable discussions. The study was supported by Swedish Medical Research Council Grants 04761, 07149, and 03556; Swedish Natural Science Council Grant 4875-103; and Swedish National Board for Technical Development Grant 85-3412.

- Marklund, S. L., Holme, E. & Hellner, L. (1982) Chim. Acta 126, 41-51.
- Marklund, S. L., Bjelle, A. & Elmqvist, L. G. (1986) Ann. Rheum. Dis. 45, 847-851.
- 3. Marklund, S. L. (1984) J. Clin. Invest. 74, 1398-1403.
- 4. Marklund, S. L. (1984) Biochem. J. 222, 649-655.
- 5. Marklund, S. L. (1982) Proc. Natl. Acad. Sci. USA 79, 7634-7638.
- Karlsson, K. & Marklund, S. L. (1987) Biochem. J. 242, 55-59.
- Tibell, L., Hjalmarsson, K., Edlund, T., Skogman, G., Engström, Å. & Marklund, S. L. (1987) Proc. Natl. Acad. Sci. USA 84, in press.

- 8. Marklund, S. L. (1984) Biochem. J. 220, 269-272.
- 9. Marklund, S. L. (1985) FEBS Lett. 184, 237-239.
- Huynh, T. V., Young, R. A. & Davis, R. W. (1984) in DNA 10. Cloning Techniques: A Practical Approach, ed. Glover, D. (IRL, Oxford), pp. 49-78.
- 11. Boyer, H. W. & Roulland-Dussoix, D. (1969) J. Mol. Biol. 41, 459-472.
- 12. Messing, J., Crea, R. & Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309-321.
- 13. Messing, J. & Vieira, J. (1982) Gene 19, 269-276.
- 14. Norrander, J., Kempe, T. & Messing, J. (1983) Gene 26, 101-106.
- 15 Edman, P. & Begg, G. (1967) Eur. J. Biochem. 1, 80-91.
- Engström, Å., Engström, P., Tao, Z.-J., Carlsson, A. & 16. Bennich, H. (1984) EMBO J. 3, 2065-2070.
- 17. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981) Nucleic Acids Res. 9, r43-r74.
- 18. Matthes, H. W. D., Zenke, W. M., Grundström, T., Staub, A., Wintzerith, M. & Chambon, P. (1984) EMBO J. 3, 801-805.
- 19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- 20. Davis, R. W., Botstein, D. & Roth, J. R. (1980) Advanced Bacterial Genetics (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- 21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
- Sanger, F. & Coulson, A. R. (1978) FEBS Lett. 87, 107-110. 22.
- Schreier, P. H. & Cortese, R. (1979) J. Mol. Biol. 129, 23. 169-172
- Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) 24. Plasmid 13, 31-40.
- 25. Harr, R., Fällman, P., Häggström, M., Wahlström, L. & Gustafsson, P. (1986) Nucleic Acids Res. 14, 273-284.

- Proc. Natl. Acad. Sci. USA 84 (1987)
- 26. Von Heijne, G. (1983) Eur. J. Biochem. 133, 17-21.
- Kozak, M. (1984) Nucleic Acids Res. 12, 857-872. 27.
- 28. Proudfoot, N. J. & Brownlee, G. G. (1976) Nature (London) 263, 211-214.
- 29. Aota, S.-I. & Ikemura, T. (1986) Nucleic Acids Res. 14, 6345-6355.
- 30. Sherman, L., Dafni, N., Lieman-Hurwitz, J. & Groner, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 5465-5469.
- 31. Hallewell, R. A., Masiarz, F. R., Najarian, R. C., Puma, J. P., Quiroga, M. R., Randolph, A., Sanchez-Pescador, R., Scandella, C. J., Smith, B., Steiner, K. S. & Mullenbach, G. T. (1985) Nucleic Acids Res. 13, 2017-2034.
- 32. Hering, K., Kim, S.-M. A., Michelson, A. M., Ötting, F., Puget, K., Steffens, G. J. & Flohé, L. (1985) Biol. Chem. Hoppe-Seyler 366, 435-455.
- 33. Steinman, H. M., Naik, V. R., Abernethy, J. L. & Hill, R. L. (1974) J. Biol. Chem. 249, 7326-7338.
- 34. Lerch, K. & Ammer, D. (1981) J. Biol. Chem. 256, 11545-11551.
- 35. Rocha, H. A., Bannister, W. H. & Bannister, J. V. (1984) Eur. J. Biochem. 145, 477-484.
- 36. Lee, Y. M., Friedman, D. J. & Ayala, F. J. (1985) Proc. Natl. Acad. Sci. USA 82, 824-828.
- 37 Kitagawa, Y., Tsunasawa, S., Tanaka, N., Katsube, Y., Sakiyama, F. & Asada, K. (1986) J. Biochem. 99, 1289-1298. 38
- Steinman, H. M. (1980) J. Biol. Chem. 255, 6758-6765.
- Steffens, G. J., Bannister, J. V., Bannister, W. H., Flohé, L., 39. Günzler, W. A., Kim, S.-M. A. & Ötting, F. (1983) Hoppe-Seylers Z. Physiol. Chem. 364, 675-690.
- 40. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S. & Richardson, D. C. (1982) J. Mol. Biol. 160, 181-217.
- 41. Borders, C. L., Saunders, J. E., Blech, D. M. & Fridovich, I. (1985) Biochem. J. 230, 771-776.