PROTEINS: Structure, Function, and Genetics 5:322–336 (1989)

Evolution of CuZn Superoxide Dismutase and the Greek Key β-Barrel Structural Motif

Elizabeth D. Getzoff,[1] John A. Tainer,[1] Michelle M. Stempien,[2] Graeme I. Bell,[2] and Robert A. Hallewell[2]
[1]Department of Molecular Biology, Scripps Clinic and Research Foundation, La Jolla, California 92037 and
[2]Chiron Corporation, 4560 Horton Street, Emeryville, California 94608

ABSTRACT **Detailed analysis of the CuZn superoxide dismutase (SOD) structure provides new results concerning the significance and molecular basis for sequence conservation, intron–exon boundary locations, gene duplication, and Greek key β-barrel evolution. Using 15 aligned sequences, including a new mouse sequence, specific roles have been assigned to all 23 invariant residues and additional residues exhibiting functional equivalence. Sequence invariance is dominated by 15 residues that form the active site stereochemistry, supporting a primary biological function of superoxide dismutation. The β-strands have no sequence insertions and deletions, whereas insertions occur within the loops connecting the β-strands and at both termini. Thus, the β-barrel with only four invariant residues is apparently overdetermined, but dependent on multiple cooperative side chain interactions. The regions encoded by exon I, a proposed nucleation site for protein folding, and exon III, the Zn loop involved in stability and catalysis, are the major structural subdomains not included in the internal twofold axis of symmetry passing near the catalytic Cu ion. This provides strong confirmatory evidence for gene evolution by duplication and fusion followed by the addition of these two exons. The proposed evolutionary pathway explains the structural versatility of the Greek key β-barrel through functional specialization and subdomain insertions in new loop connections, and provides a rationale for the size of the present day enzyme.**

Key words: sequence conservation, exon, gene duplication, protein folding, structure-function, X-ray structure

INTRODUCTION

CuZn SODs are cytoplasmic enzymes, predominantly found in eukaryotes, that protect cells against the toxic effects of the superoxide radical produced as a by-product of aerobic metabolism. Each molecule is a dimer of identical subunits and each subunit contains approximately 153 amino acids and one Cu and one Zn ion. The Cu ion, an essential cofactor in catalysis, is cyclically oxidized and reduced during successive encounters with superoxide. Thus Cu^{2+} accepts an electron from one superoxide radical to produce molecular oxygen and Cu^{+}, which then donates an electron to a second superoxide radical to produce, together with two protons, hydrogen peroxide.[1] Both reactions have rate constants of $2 \times 10^9\ M^{-1}\ sec^{-1}$ and both rates are probably limited by the availability of substrate.[2] Precollision electrostatic guidance, indicated by experimental studies of the ionic strength dependence of the reaction rate[3] and by computational modeling of the molecular electrostatic field,[4] could increase the availability of substrate, making the reaction rate faster than by diffusion alone.

The bovine CuZn SOD structure, refined to a resolution of 2 Å, has as its structural core a flattened, eight-stranded, Greek key β-barrel. This topological motif, named for a design common in Grecian art,[5] is characterized by a set of antiparallel β-strands forming the staves of a barrel; two strands are connected across the barrel at the top and two more at the bottom, while the remainder of the connections occur between adjacent strands. The Cu ion is liganded at the surface of the barrel at the base of a channel formed by two functionally important loops.[6] One loop contains most of the residues important in precollision electrostatic guidance of substrate.[4] The second loop can be divided into two structurally important subdomains, both of which contribute to the high stability of the enzyme: a Zn^{2+} ion-binding region[7] and a region that forms one side of the active site channel, contributes to the dimer contact, and is stabilized by a disulfide bridge to the β-barrel.[6] Close packing in the hydrophobic interfaces both between the subunits and between the two halves of the flattened β-barrel may also contribute to the unusually high stability of the enzyme. The bovine enzyme remains active in 8 M urea[8] and 4% sodium dodecyl sulfate[7] and has a conformational melting temperature of 96°C.[9,10]

Superoxide dismutases are thought to have arisen early in evolution at the time when aerobic life was

Received February 15, 1989; revision accepted May 2, 1989.

Address reprint requests to Elizabeth D. Getzoff, Department of Molecular Biology, Scripps Clinic and Research Foundation, La Jolla, CA 92037.

G.I. Bell's present address: Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637.

developing.[11] Indeed, the existence of CuZn SOD in some bacteria, which were present before the appearance of the oxygen-producing, photosynthetic, blue-green algae, suggests that CuZn SOD activity may have been beneficial even at low atmospheric oxygen tension. Because of their ancient origin and the extensive amino acid sequence data available, sequence alignments of CuZn SODs are used to test methods of constructing phylogenetic trees.[12,13]

CuZn SODs may also contribute to our understanding of the evolution of the Greek key β-barrel, one of the most versatile structural motifs found in biological systems.[14–16] It is used for recognition and binding (immunoglobulins, prealbumin), long-lived and stable structural assemblies (lens crystallins, icosahedral virus capsid proteins), electron transfer (plastocyanin), and catalysis (serine proteases, CuZn SOD). The amino acid sequences and Greek key β-barrel structures of β and γ lens crystallins show evidence for evolutionary gene duplication: the crystallin domains exhibit a twofold axis of symmetry that corresponds to a quite significant level of amino acid duplication,[17] and the proposed ancestral crystallin gene corresponds to one exon in the present day gene.[18] Other more ancient genes, such as those for immunoglobulin and CuZn SOD, encode structural symmetry with less significant sequence duplications.[19] Here we discuss the roles of sequence-conserved residues in the function and three-dimensional structure of CuZn SOD, analyze the exon structure of the CuZn SOD gene in relation to the structural symmetry in the protein, and relate these to the evolution and versatility of the Greek key β-barrel motif.

EXPERIMENTAL PROCEDURES

cDNA Cloning and Sequencing

A mouse placental cDNA library in λgt10[20] consisting of 1.25×10^5 phage plaques was screened by cross-hybridization with the human CuZn SOD (HSOD) cDNA[21] and three positives were obtained. The largest cDNA (see Fig. 2) was subcloned in bacteriophage M13 and sequenced by the chain termination method.[22,23] DNA manipulations were carried out as described previously.[24]

X-Ray Crystallographic Refinement with Hydrogen Atoms

The refined atomic coordinates for bovine CuZn SOD (BSOD) at 2 Å resolution[6] were refit to fragment difference electron density maps as previously described.[25] After regions of bovine SOD with high temperature factors and imperfect geometry were refit, explicit hydrogen atoms were geometrically added to one dimer (two dimers form the asymmetric unit) to aid in the accurate definition of both intermolecular and intramolecular interactions.[26] Hydrogen atoms attached to terminal carbon or nitrogen atoms were placed in the minimal energy or staggered conformation around the preceding single bond. Hydroxyl and sulfhydryl hydrogens were placed in one of the three positions of the staggered conformation. Following the addition of hydrogen atoms, all potentially hydrogen-bonding hydrogen atoms were manually refitted to form hydrogen bonds where appropriate, and changes were also made in some areas of the maps where the addition of hydrogen atoms suggested refitting to allow improved close packing of the hydrophobic side chains. The resulting model was then refined to a crystallographic residual error (R-value) of 19% without solvent[25] using reciprocal space refinement with stereochemical restraints.[27]

To determine the reliability of the stereochemical description for the conserved residues, the agreement and variation among the four equivalent subunits in the crystallographic asymmetric unit were determined by least-squares superposition of the independent subunits. The backbone conformation is well defined in the refined model, being almost entirely consistent among the four subunits. The average deviation is 0.24 ± 0.10 Å for the 604 main chain atoms, 0.69 ± 0.30 Å for the 493 side chain atoms, and 0.44 ± 0.19 Å for all nonhydrogen atoms. In general, all of the interactions involving conserved amino acid residues appear reliably consistent in the four subunits. Assuming that the stereochemical restraints act to reduce the number of independent positional parameters to approximately one-half of the number of atoms,[27] about 6600 variable parameters used in the first refinement of SOD[6] were increased by the addition of hydrogen atoms to about 9800 variable parameters in the second refinement. Using these numbers, the R-value decrease from 25.5% to 19% is significant at the confidence level of 99.9% by the Hamilton test.

The addition of hydrogen atoms is unusual in X-ray structures of proteins and represents the major modification of normal procedures for the SOD refinement. Since the single electron of hydrogen atoms makes only a minor contribution to the scattering of X-rays, its position is not generally distinguishable in the resultant electron density map; exceptions occur for hydrogen atoms participating in strong hydrogen bonds and for electron density maps of small proteins at very high resolution. Judging from comparison of the dimers with and without hydrogens, the inclusion of hydrogen atoms at 2 Å resolution appeared to be neutral or positive for the progress of the crystallographic refinement. However, the use of explicit hydrogens resulted in superior internal side chain packing as judged by the size and number of internal cavities (inconsistent packing defects among the subunits) found by molecular surface calculations with a 1.4 Å probe radius.[28] Thus the addition of hydrogen atoms at this resolution avoided those packing faults that

are caused by the shape inaccuracies resulting from the application of spherical approximations (united atom radii that include implicit hydrogens) to asymmetric hydrogen arrangements on heavy atom groups.

Computational and Computer Graphics Analysis

The BSOD dimer with hydrogen atoms was used for stereochemical analysis and for construction of a human SOD model. Characterization of the role of conserved amino acid residues in the BSOD structure included determination of backbone conformational angles, hydrogen bonds, salt bridges, tight turns, nonbonded interactions, and specific side chain contacts. Torsional angles (ϕ, ψ) of the polypeptide backbone were determined from the refinement program.[27] Hydrogen bonds were identified as distances less than 3.5 Å and an angle of greater than 90° from the hydrogen donor through the hydrogen position to the acceptor atom. Turns were identified by calculation of a separation of less than 7 Å between CA_n and CA_{n+3} and a change in direction of greater than 135°. Solvent accessibility of the molecular surface of the enzyme dimer and nonbonded interactions of surfaces buried between structural elements were determined with the program MS[28] by using a probe radius of 1.40 Å and individual atomic radii with explicit hydrogen atoms: oxygen 1.40 Å, nitrogen 1.54 Å, carbon 1.74 Å, sulfur 1.80 Å, hydrogen 1.20 Å, Cu ion 1.40 Å, and Zn ion 1.40 Å. Intrasubunit interactions were further identified in terms of close contacts between amino acid side chains. The close contacts between side chain atom pairs, identified as distances less than or equal to the sum of the van der Waals radii plus 0.5 Å, were determined in all four subunits and grouped together by their individual residue pairs. Residue pairs with close contacts in less than three subunits were excluded.

The refined atomic coordinates for BSOD at 2 Å resolution were compared to a model for HSOD developed by molecular dynamics and energy minimization using the program AMBER[29] on an initial model derived from the BSOD coordinates with appropriately substituted atoms and side chains. Substituted amino acids were built in by matching atomic positions out along the side chains and placing additional atoms in an extended conformation. Based upon the shifts occurring during complete energy minimization of the solvated bovine CuZn SOD (about 1.1 Å rms), the deviations that occur between the main chain atoms of the bovine SOD crystal structure and human SOD model are minor (usually less than 0.5 Å) except in the area of the two-residue insertion at position 26, and most side chains superimpose. In particular, the dimer contact region is almost identical, with highly conserved sequenced regions accounting for most of the surface area buried by the contact. This HSOD model is also consistent with a preliminary 2.5 Å X-ray structure for the human enzyme (H.E. Parge, R.A. Hallewell, E.D. Getzoff, and J.A. Tainer, unpublished results).

Computer graphics analysis of the BSOD structure and the HSOD model was done using the graphics modeling language GRAMPS[30] and the molecular modeling program GRANNY.[31]

RESULTS AND DISCUSSION

Amino Acid Homology

The sequences of 15 CuZn SODs, including the new mouse sequence, are aligned and numbered against the human sequence in Figure 1A. A schematic of the β-strand and loop secondary structure labeled to show how the loops and exons from the sequence alignment in Figure 1A map onto the structure elements is shown in Figure 1B. The species in the sequence alignment are ordered approximately according to their evolutionary relatedness, but with the human and bovine sequences at the top. Structure–function features are identified above the sequence alignment, while intron positions, structural symmetry and amino acid duplications are shown below the alignment. Although sequence changes between different CuZn SODs cannot be used as a precise evolutionary clock,[12] species relatedness is reflected in the level of amino acid homology. Thus, among the mammals, rat and mouse show almost complete amino acid identity (97%). The human enzyme has an average of 82% amino acid identity with the other mammals and this falls to 67%, 61%, 56%, 54%, and 18% when HSOD is compared with swordfish, fruit fly, plants, fungi, and photobacterium, respectively (insertions and deletions counted as mismatches). A comparison of nucleotide homology between mouse and human cDNAs (Fig. 2) indicates little homology in the noncoding region (37%), whereas the coding sequence has the same level of homology at the nucleotide and amino acid level (84%).

Global Features of the Sequence Alignment

The alignment shown in Figure 1A results in insertions or deletions in six of the seven loop and turn connections between the β-strands (see Figure 3C for a display of the β-strands and loop regions in the three-dimensional structure). Although some insertions and deletions may overlap the ends of β-strands, none is found within the body of these structural elements. Thus from these results with SOD, β-strands are more permissive of sequence variation than of shifts in strand alignment. The absence of single residue insertions is somewhat surprising, since they could presumably be accommodated by β-bulge formation.[51] Larger insertions or deletions within β-strands may disrupt side chain packing interactions.

The two most sequence-variable regions in the

CuZn SOD molecule are between residues 19–36 and 88–105. Both of these segments consist predominantly of antiparallel β-strand hairpins that are solvent exposed because they are not buried underneath the active site channel or in the dimer interface (see Figure 1 for the location of these residues). It is probable that alterations to these solvent exposed residues are less likely to reduce enzyme activity, whereas alterations to residues participating in the dimer contact or the active site are restricted because they are much more likely to inactivate the enzyme.

Invariant Residues

There are 23 invariant amino acids in the 15 species aligned in Figure 1 and these are displayed on the α carbon backbone of human CuZn SOD in Figure 3A. The invariant residues fall into three structure–function categories: active site (15 residues), dimer interface (4 residues) and β-barrel (4 residues). A few of the invariant residues probably have dual functions. The clearest examples of this are Phe-45 and the cysteines forming the disulfide bridge. Phe-45 is categorized above as a β-barrel residue, although it also appears to be intimately involved in orienting the Cu liganding residues His-46 and His-48 of the active site (Fig. 3B). The disulfide bridge (invariant Cys-57 and Cys-146), categorized above as dimer interface, stabilizes a subdomain of loop IV that forms part of the active site channel as well as participating in dimer contact. Invariant residues can also be classified according to amino acid side chain: a hybrid categorization has been used in Figure 3A to emphasize the common structural role of some residues. For example, glycines are the most frequent (9 out of 23) invariant residues because often no other residue can satisfy the spatial or conformational constraints at a given position in the three-dimensional structure (see below.)

Invariant Active Site Residues

The metal liganding residues account for 7 of the invariant active site residues (His-46, 48, 63, 71, 80, 120 and Asp-83). Arg-143 forms part of the superoxide binding pocket above the catalytic Cu ion, where the positive charge of the Arg guanidinium group stabilizes the negatively charged superoxide substrate.[25] Arg-143 also appears to participate in short-range precollision electrostatic guidance of substrate.[4,52] Asp-124 is an invariant residue involved in correctly orienting the Cu and Zn liganding residues: by analogy with the bovine structure, Asp-124 forms strong, charged, hydrogen bonds to both Cu ligand His-46 and Zn ligand His-71.[25]

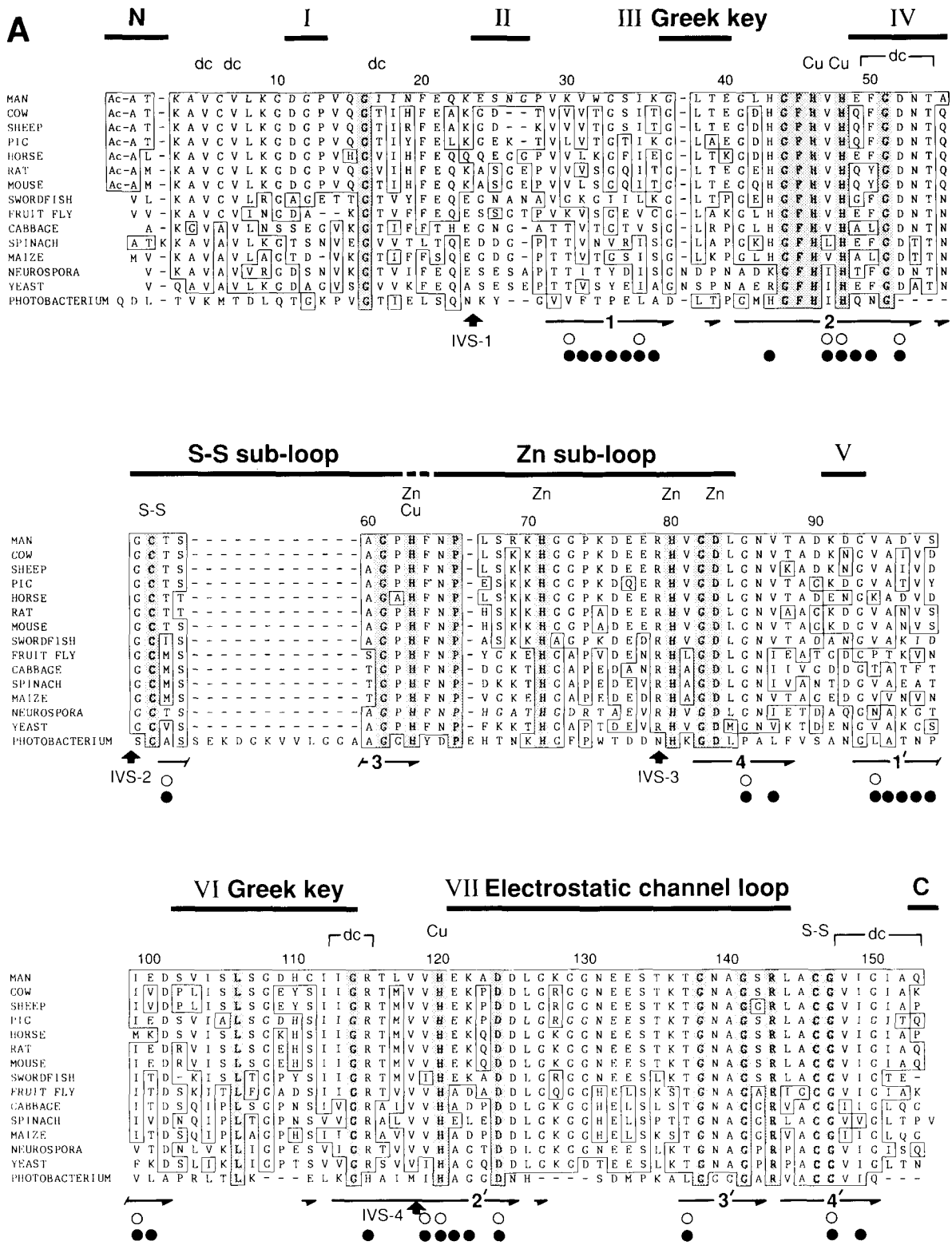
Five invariant glycine residues are involved in maintaining the structure of the active site. Invariant Gly-44 and Gly-138 residues are probably conserved because the larger side chains of any other residue would perturb the short, charged, hydrogen bonds of invariant buried Asp-124, which help position Cu ligand His-46 and Zn ligand His-71. Additionally, Gly-138 has a left-handed α-helical conformation (ϕ and ψ angles forbidden to other amino acids) and forms position 2 of a type 1′ tight turn, where glycines are favored.[14] These features help to position the electrostatic channel loop. Both Gly-61 and Gly-141 form β-bulges[51] at opposite ends of the active site channel. β-bulge (Gly-61, residue 62, residue 49) joins the two ends of the disulfide region of loop IV and β-bulge (Gly-141, residue 142, residue 121) joins the two antiparallel sides of the electrostatic channel loop VII. The strategic placement of these two bulges allows loops IV and VII to peel away from the β-barrel to enclose the active site channel.[6] Invariant Gly-82 is restricted in size by the adjacent positioning of the main chain of residue position 47 (amino acids Val, Leu, or Ile) that covalently joins Cu ligands His-46 and His-48. Mutation of Gly-82 to a residue with a larger side chain would perturb the bonding to the catalytic Cu ion by these histidines.

Pro-66 is unusual because it is the only invariant, bulky hydrophobic residue in the active site category, and it is also the invariant active site residue most distal from the essential, catalytic Cu ion. Pro-66 participates in the tight turn of the Zn liganding region of loop IV that directs the ligands His-63 and His-71 toward the Zn ion. Proline has stronger stereochemical constraints than any other residue, and therefore the replacement of this proline might alter binding of the Zn ion and the conformation of the Zn loop.

Invariant Dimer Contact and β-Barrel Residues

Invariant Gly-51 and Gly-114 both form tight contacts and main chain hydrogen bonds across the dimer interface: the main chain nitrogen atom of Gly-51 and the carbonyl oxygen atom of Gly-114 hydrogen bond to the carbonyl oxygen and main chain nitrogen, respectively, of the nonconserved Ile-151 (Leu, Thr). Together, these three residues of each subunit form all four of the hydrogen bonds across the dimer interface. Mutation of either Gly-51 or Gly-114 could cause collisions of the new side chain with the other subunit of the dimeric enzyme. The two invariant cysteines (Cys-57 and Cys-146) form a disulfide bridge stabilizing a region of loop IV (see Fig. 1) involved in dimer contact by covalently joining it to the β-barrel. Reduction of the disulfide bridge results in a large loss of protein stability.[53]

Two glycines (Gly-16, Gly-147) and two bulky hydrophobic residues (Phe-45, Leu-106) are the invariant residues that appear to be involved in maintaining the stable Greek key β-barrel fold. Due to the β-strand twist, and perhaps constraints from adjacent disulfide-forming Cys-146, a side chain at position 147 would be pointed toward the side chain of


A
N
I
II
III Greek key
IV
S-S sub-loop
Zn sub-loop
V
VI Greek key
VII Electrostatic channel loop
C
dc
Cu
S-S
Zn
IVS-1
IVS-2
IVS-3
IVS-4
1
2
3
4
1′
2′
3′
4′
MAN
COW
SHEEP
PIC
HORSE
RAT
MOUSE
SWORDFISH
FRUIT FLY
CABBAGE
SPINACH
MAIZE
NEUROSPORA
YEAST
PHOTOBACTERIUM

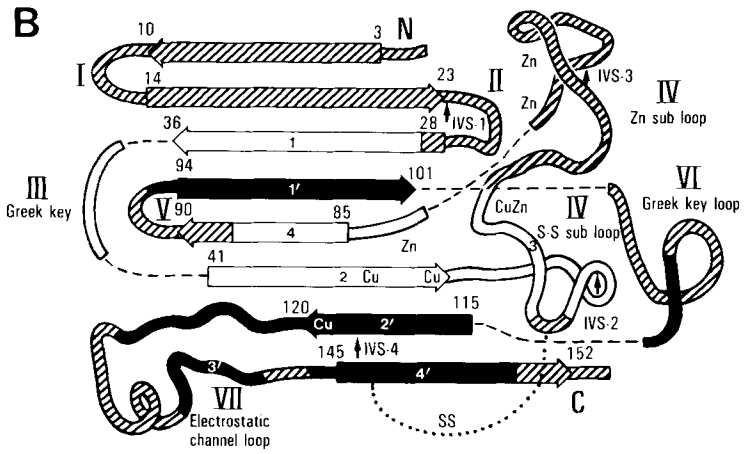


Fig. 1. **A:** Amino acid sequence alignment of CuZn SODs. Residues are numbered according to the human sequence and those identical with human SOD are boxed. Residues invariant among the 15 sequences are shaded grey. Roman numerals and solid lines above the alignment correspond to the seven loops or turns separating the eight β-strands and the N-terminal (N) and C-terminal (C) sequences not involved in β-strands. Residues that disulfide bridge (S–S), ligate the metals (Cu or Zn), or are involved in dimer contact (dc) are marked above the alignment. The locations of the four intervening sequences (introns) of the human gene with respect to the coding sequence are shown below the alignment (IVS-1 through IVS-4) and separate the coding sequence into five exons. Regions exhibiting twofold symmetry in the bovine structure (1 and 1′, 2 and 2′, 3 and 3′, 4 and 4′) are marked with numbered horizontal arrows below the alignment. Identical symmetry-related residues in the bovine sequence (open circles) or among the eukaryotic sequences (closed circles) are indicated below the structural symmetry. All structural assignments are taken from the bovine crystal structure. SOD amino acid sequences were taken from the following sources: man,[21,32] cow,[33] sheep,[34] pig,[35] horse,[36] rat,[37] mouse, ref. 38 and this work, swordfish,[39] fruit fly,[12,40] cabbage,[41] maize,[42] spinach,[43] *Neurospora*,[44] yeast,[45,46] *Photobacterium*.[11,47] **B:** Schematic of the β-strand and loop secondary structure labeled to show how the loops, symmetrical regions, and exons from the sequence alignment in **A** map onto the structural elements. The positions of the N- and C-termini, Cu and Zn ligands, disulfide bond, and of the four intervening sequences are labeled. β-strands are shown as arrows and numbered at their N- and C-terminal residues and corresponding residue numbers are provided in **A.** Loops are labeled by Roman numerals. Greek key loops III and VI are so named because they represent the +3 connections (named by how many strands they move over in the β-sheet) that form the classic Greek key design. Loop IV is subdivided into the S–S and Zn subloops at metal bridging ligand His-63 (labeled CuZn). Loop VII includes 1.5 turns of α-helix (lower left) composed of residues 132–137. β-strand and loop assignments are similar to that previously published,[6] but are modified to include some residues at the ends of β-strands as turns, to reflect the results of further crystallographic refinement (see Methods). The twofold symmetry-related regions, labeled and numbered in **A** and described later in the Results and Discussion, are coded here in white (with black labels 1, 2, 3, and 4) and black (with white labels 1′, 2′, 3′, and 4′), with asymmetric additions striped.

```
                                                                                                                    Met-1
    HUMAN SOD                                                  CAGCGTCTGGGGTTTCCGTTGCAGTCCTCGGAACCAGGACCTCGGCGTGGCCTAGCGAGTTATG
                                                               • *  •  * *  •  **    *  ****    •  *  ****  • • **  ***
    MOUSE SOD                                                  CTTCTCGTCTTGCTCTCTCTGGTCCCTCCGGAGGAGGCCGCCGCGCGTCTTCCGGGGAAGCATG
   Ala1
 1 GCGACGAAGGCCGTGTGCGTGCTGAAGGGCGACGGCCCAGTGCAGGGCATCATCAATTTCGAGCAGAAGGAAAGTAATGGACCAGTGAAGGTGTGGGGAAGCATTAAAGGACTGACTGAA
   ***• *** ** ********************** ** ******** * **** * ************ ***   ** ******   • ***  ***   **** **** * ******
 1 GCGATGAAAGCGGTGTGCGTGCTGAAGGGCGACGGTCCGGTGCAGGGAACCATCCACTTCGAGCAGAAGGCAAGCGGTGAACCAGTTGTGTTGTCAGGACAAATTACAGGATTAACTGAA

121 GGCCTGCATGGATTCCATGTTCATGAGTTTGGAGATAATACAGCAGGCTGTACCAGTGCAGGTCCTCACTTTAATCCTCTATCCAGAAAACACGGTGGGCCAAAGGATGAAGAGAGGCAT
    **** ****** ***** ** *** *** *** ** ******  ********************* ***** **********  ** *  ***** ***** **   ***************
121 GGCCAGCATGGGTTCCACGTCCATCAGTATGGGGACAATACACAAGGCTGTACCAGTGCAGGACCTCATTTTAATCCTCACTCTAAGAAACATGGTGGCCCGGCGGATGAAGAGAGGCCAT

241 GTTGGAGACTTGGGCAATGTGACTGCTGACAAAGATGGTGTGGCCGATGTGTCTATTGAAGATTCTGTGATCTCACTCTCAGGAGACCATTGCATCATTGGCCGCACACTGGTGGTCCAT
    ********* ******************  ** ** ********* ******* *********  ********************** **** ************ *** ***********
241 GTTGGAGACCTGGGCAATGTGACTGCTGGAAAGGACGGTGTGGCCAATGTGTCCATTGAAGATCGTGTGATCTCACTCTCAGGAGAGCATTCCATCATTGGCCGTACAATGGTGGTCCAT
                                                                                                          Gln153
361 GAAAAAGCAGATGACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACAGGAAACGCTGGAAGTCGTTTGGCTTGTGGTGTAATTGGGATCGCCCAATAAACATTCCCTTGGATGTAG
    ** ***  ************************************************ ***** ***** ** ** ***** ***** ** ******** ** ** ******** ** **  **
361 GAGAAACAAGATGACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACTGGAAATGCTGGGAGCCGCTTGGCCTGTGGAGTGATTGGGATTGCGCAGTAAACATTGCCCTGTGTGGTC

481 TCTGAGGCCCCTTAACTCATCTGTTATCCTGCTAGCTGTAGAAATGTA
    * *  *   * *                 •       * *** *
481 TGAGTCTCAGACTCATCTGCTACCCTCAAACCATTAAACTGTAATCTGAAAAAAAAAA
```

Fig. 2. Nucleotide sequence alignment between human and mouse CuZn SOD cDNAs. The initiator methionine codon, N-terminal alanine codon and C-terminal glutamine codon are boxed and labeled. Nucleotide identities are starred between the sequences. A 5′AUUAAA3′ sequence, associated with poly(A) addition,[48] occurs eleven nucleotides 5′ of the mouse poly(A) addition site. The homology comparison was terminated at the mouse poly(A) addition site; the human mRNA polyadenylates at different positions from mouse to give at least two mRNAs.[49,50]

Val-7 (Thr in *Photobacterium*), thus restricting mutation of Gly-147. Similarly, a side chain at position 16 would be pointed into the β-barrel to collide with the large hydrophobic residue at position 35 (Ile, Leu, Val), thus restricting mutation of Gly-16. Phe-45 forms one of only two cross-barrel connections, connecting β-strands directly across the β-barrel and tying the active site to the more regular side of the β-barrel (Fig. 3B). Phe-45 interacts with Ile-18 (or Val) across the barrel and anchors adjacent Cu liganding residues His-46 and His-48. Leu-106 lies in Greek key loop VI, which crosses one end of the barrel. The side chain of Leu-106 points into the barrel creating a "cork" at the end (Fig. 3C) that stabilizes the β-barrel by interacting with buried hydrophobic residues from the β-strands and with residues 112 and 113 (both usually Ile) from the other end of the Greek key loop.

Significance of Invariant Residues

How significant is sequence invariance among the 15 species that have been aligned with the human sequence? Clearly, it has been useful to include distantly related species such as fungi and bacteria in the comparison. For example, the inclusion of the *Photobacterium* sequence removed 25 residues that would otherwise have been invariant. This suggests that some of the current set of invariant residues will be removed as other distantly related CuZn SOD sequences become available. The recently published sequence of human extracellular SOD (EC-SOD; ref. 54) was also aligned with the 15 intracellular SOD sequences (data not shown). This CuZn SOD has additional unique sequence subdomains at the amino and carboxyl termini, which presumably function in tetramer formation and in binding to cell surfaces. When the ECSOD sequence is added to the alignment, 20 residues remain absolutely conserved. Gly-44 (active site) has changed to Ala, which must be a short enough side chain to allow packing with Cu ligand His-46 and the invariant buried Asp-124. Adjacent Phe-45 (β-barrel), which participates in one of the two cross-barrel contacts, has changed to Ile, which is slightly less bulky. Invariant Gly-147 (β-barrel) has become Cys, but close packing between residues 147 and 7 is presumably no longer restrictive, due to extensive changes in the N-terminal sequence.

Overall, it can be concluded that sequence invariance in CuZn SODs is dominated by residues that are directly connected with the active site and the catalytic mechanism (up to 18 residues). From the complete invariance of the metal liganding residues and the importance of their spatial configuration (indicated by the number of invariant residues involved in orienting them) we speculate that alterations to any of the invariant active site residues will reduce enzyme activity.

Alterations to sequence invariant residues in recombinant human CuZn SOD using in vitro mutagenesis provide a direct test of whether a given residue is essential for superoxide dismutase activity. Thus, the single mutations Asp 124 to Gly 124 do inactivate the enzyme. However, mutation of Gly-114 to Asp-114 does not appear to affect enzyme activity and substituting Lys-143 or Ile-143 for the invariant Arg-143 resulted in enzymes that were 50% and 12% active, respectively (ref. 52, and R.A.H. and B.T. Lee, unpublished results). These results emphasize that sequence invariance, even among widely differing species, does not prove that a given residue is essential for enzyme activity. At the same time, sequence alignments are of considerable benefit when considering the effects of changing residues for protein engineering studies. CuZn SOD has been used for protein engineering studies investigating electrostatic attraction of substrate,[52] thermostability (R.A. Hallewell, unpublished results), and the design properties of the Greek key β-barrel structural motif.[55] Such studies are facilitated by the high-resolution crystal structure of the bovine enzyme[6,25] (that of the human enzyme is in progress[56]), high level expression (and rapid purification) of the recombinant protein in microorganisms,[20,52,57] a genetic selection for activity in *E. coli*,[58] and an extensive background body of work on the structural and biochemical properties (reviewed in ref. 15).

Functional Equivalence

Another level of amino acid conservation in proteins reflects the maintenance of an invariant function but does not require amino acid invariance. Such conserved residues or groups of residues show functional equivalence or stereochemical similarity. Functionally equivalent residues usually display limited variability that conserves a specific function at a given three-dimensional location, which can be, but is not necessarily, at the same position of the sequence alignment. Functionally equivalent sites can be recognized by means of the apparent conservation of one or more stereochemical features, such as size and shape, charge and polarity, hydrophobicity, hydrogen bonding, and secondary structure preference.

Perhaps the most common examples of functional equivalence in SODs are the selected replacements allowing large hydrophobic residues at the same position in the sequence alignment. The residues Ile, Leu, and Val are functionally equivalent at nine positions (five of Ile, Leu, or Val at positions 35, 47, 87, 104, 112, and four of Ile or Val at posi-tions 18, 119, 148, 149). As might be expected, all of these side chains are buried from solvent in the interior of the β-barrel, between Greek key loop VI and the β-barrel, or in the dimer interface (see Fig. 1). Although most cases of functional equivalence follow this simple pattern, occasionally more subtle examples

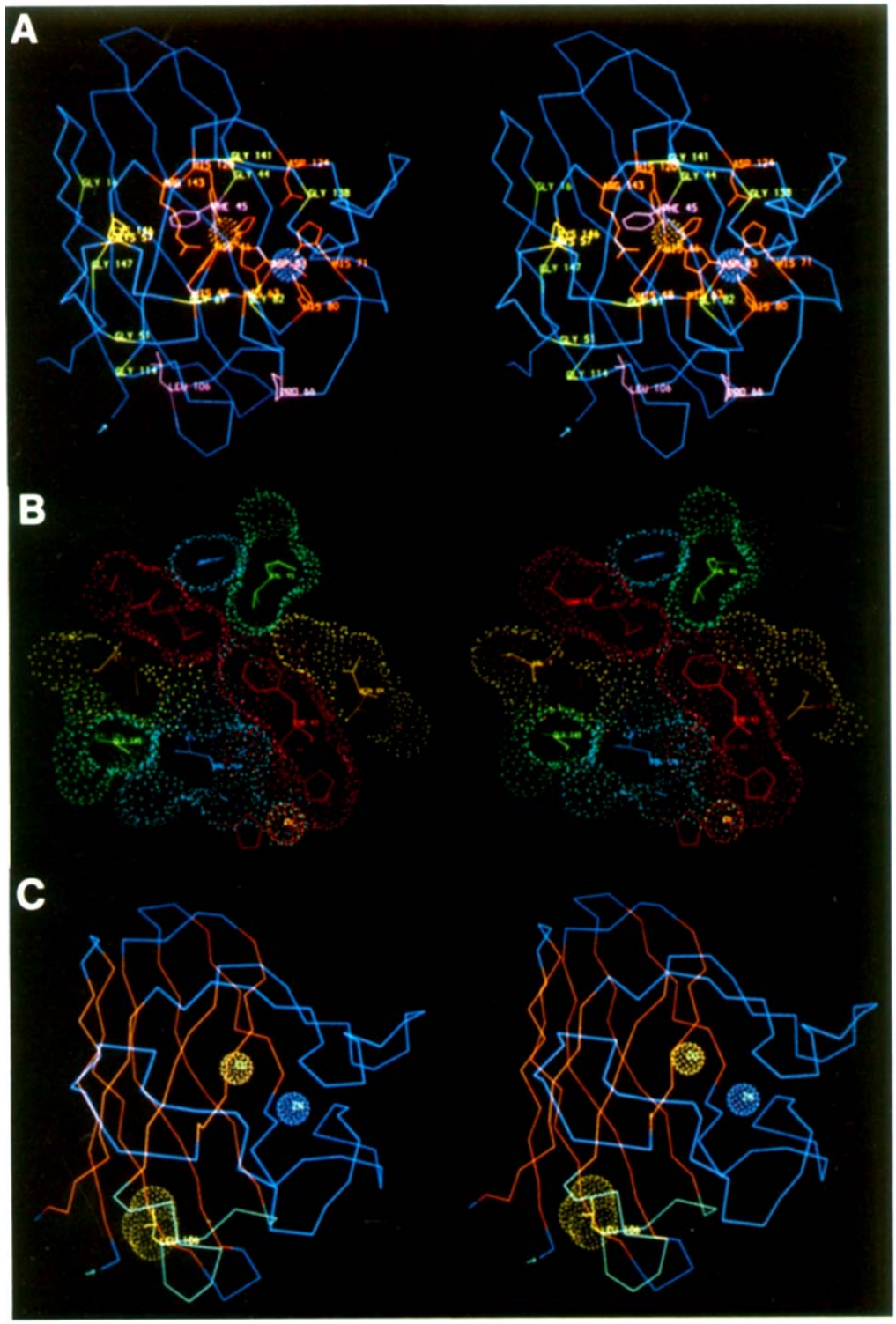


Figure 3.

can be detected. One example is the conserved (except in fungi) bulky hydrophobic residue Leu-38 that forms a "cork" at one end of the β-barrel, analogous to the invariant Leu-106 at the opposite end of the β-barrel (see Fig. 3C). In *Neurospora* and yeast Leu-38 is replaced with two amino acids by the insertion of an additional amino acid in the fungal sequences (see Fig. 1). A particularly interesting example of functional equivalence, where chemical similarity in three-dimensional placement does not require identical sequence alignment, occurs in the electrostatic channel loop. Lys-136, which is important in electrostatic interactions with the substrate,[4] can apparently be replaced in fruit fly, cabbage, spinach, and maize SOD by His-131 (see Fig. 1). Like Lys-136, His-131 can also place a positive charge in a similar three-dimensional position by putting its shorter side chain one turn earlier in the short α-helix of the electrostatic channel loop. The His residue should maintain a positive charge at physiological pH due to the adjacent negatively charged Glu-132. SODs containing His-131 appear to have a requirement for Leu-133; its hydrophobic side chain may be preferentially buried by the more polar His-131 such that the His is displaced outward above the active site channel. Similarly, the SOD of *Photobacterium* contains a deletion in the electrostatic channel loop, but still maintains the potential for the coupled charged side chains (positively charged His-126 and negatively charged Asp-132).

Sequence Conservation and Biological Function

The biological function of CuZn SODs has been the subject of debate both because the superoxide radical appears to have relatively poor reactivity with biological molecules and because some low-molecular-weight copper complexes can catalyze superoxide dismutation at comparable rates to the enzyme. Thus, it has been suggested that the superoxide radical is not particularly toxic and that superoxide dismutation may be an incidental function of the enzyme.[59,60] Since conserved residues in the sequence alignment reflect biological selection for function, it is possible to make inferences about the biological function(s) of a protein by considering all the possible functions of conserved residues in the three-dimensional structure. The predominance of invariant active site residues, and the assignment of the remaining invariant residues to dimer contact and β-barrel categories, suggests that any additional function of SOD must involve the Cu or Zn ions and be restricted to the active site region. The large number of invariant residues involved in the positioning or stabilization of the metal ligating residues strongly suggests that the function of the metals is catalytic. Also the ligand and ligand hydrogen bonding interactions of the Cu network involve a larger number of different structural elements than those of the Zn network, which is compatible with the catalytic role of the Cu site,[61] and may also account for the larger effect of Cu site occupancy on SOD thermal stability.[7] Structural and sequence conservation considerations therefore support a catalytic function for SOD. Furthermore, there is good evidence that several invariant or functionally equivalent residues in electrostatic channel loop VII (Fig. 1) are involved in precollision electrostatic guidance of the negatively charged superoxide[3,4] (see last section). This sequence conservation, to-

Fig. 3. Stereo pairs showing the structural location and apparent function of conserved residues. Where present, the Cu (orange) and Zn (blue) ions are shown as spheres and the N-terminus identified by a green arrow. **A:** Sequence-invariant residue side chain positions and residue labels mapped onto an α carbon model structure for HSOD. Four general classes of invariant residues observed are labeled and color coded onto the α carbon backbone (blue): active site residues (red), glycines (green), the disulfide cysteines (yellow), and hydrophobic residues (purple). The four β-strands forming the more regular side of the β-barrel (at the back in this view) contain only a single invariant residue (Gly-16), although some others are conserved through functional equivalence (see text). The other four β-strands, forming a more twisted sheet on the active site face of the β-barrel, have seven invariant residues. The active site channel is identified (center) by the spheres for the copper (orange) and zinc (blue) ion positions. **B:** Packing in the BSOD β-barrel suggests a critical role for the side chain interactions of invariant Phe-45 with a bulky hydrophobic side chain at position 18. This view (rotated about 90° forward about a horizontal axis from that in **A**) shows a thin cross-sectional slice of the BSOD β-barrel with the eight β-strands surfaced individually to show the packing interactions. Residues are numbered from the BSOD sequence used for this refined 2 Å BSOD structure, which aligns with the HSOD sequence such that two must be subtracted from HSOD residues after position 25. Each strand interacts primarily with its adjacent neighbors, but has some contacts with more distal β-strands. One of two cross-barrel interactions between directly opposing β-strands across the β-barrel is shown here. In order around the β-barrel (as shown in **A**), the more regular β-strands on the back side of the barrel (shown in yellow, red, blue, and green from center left to top center) form a major cross-barrel interaction from Ile-18 (red, upper left) to Phe-45 (red, lower right and numbered 43 in BSOD) of the more irregular strands underlying the active site (shown in green, blue, red, and yellow from lower left to center right). This cross-barrel interaction braces active site Cu ion (orange sphere, bottom right) ligands His-46 (BSOD 44) and His-48 (BSOD 46) through the invariant Phe to the functionally conserved Ile at position 18 at the other side of the β-barrel. The conservation of a bulky hydrophobic side chain at position 18 is an example of a position that shows sequence conservation through functional equivalence. **C:** The critical structural location of sequence invariant Leu-106 shown in HSOD. The hydrophobic side chain of Leu-106 (yellow side chain and dot surface) is located on a loop region (blue backbone) that forms a Greek key connection between β-strands on the two opposite sides of the flattened eight stranded β-barrel (red backbone). The most extensive packing and hydrogen bond interactions in β-structure are between adjacent β-strands thereby forming a strong structural fold around the strands forming the sides of the barrel and weaker interactions across the ends of the barrel. The barrel ends are strengthened by the two covalent Greek key loop connections. The first Greek key connection (loop III) is relatively short with the β-strand hydrogen bonds being interrupted only for two residues. In the second, longer, Greek key connection (loop VI), Leu-106 appears to form a critical "cork" for this end of the β-barrel and has packing interactions with six other side chains (Ala-4, Phe-20, Ala-22, Val-29, Ile-112, Ile-113), which are largely conserved residue positions.

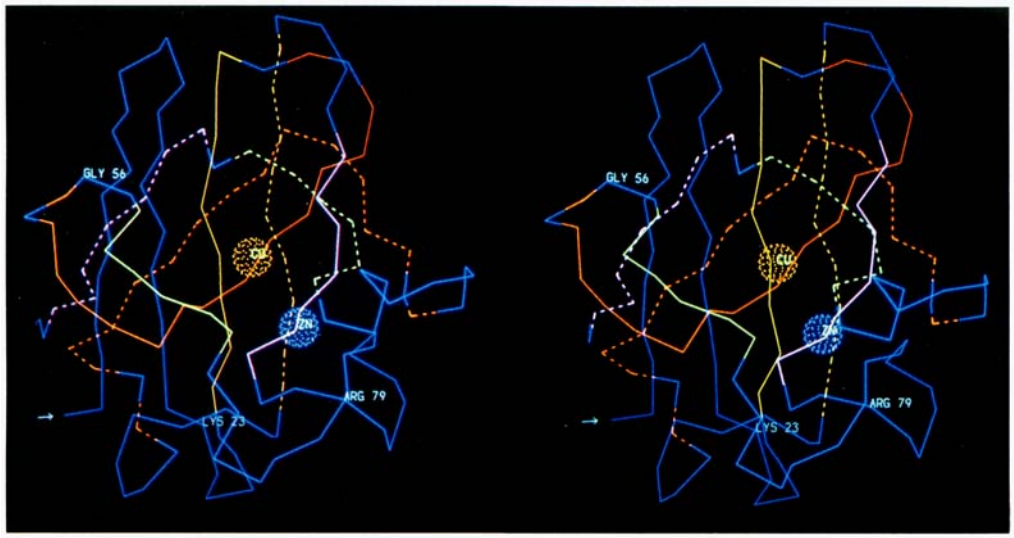


Fig. 4. Stereo pair showing the structural location of the internal twofold symmetry in HSOD. As shown in this view, the internal twofold axis is perpendicular to the plane of the paper and passes though the active site near the copper (orange sphere). The twofold symmetry relates six β-strands and two loop regions, with symmetry pairs identified as solid and dashed α carbon backbone lines in the same color: yellow for symmetry 1 (residues 29–37, solid) and symmetry 1′ (residues 93–101, dashed); red for symmetry 2 (residues 39, 41–53, 55, solid) and symmetry 2′ (residues 111, 113–125, 127, dashed); green for symmetry 3 (residues 58–63, solid) and symmetry 3′ (residues 137–142, dashed); and light purple for symmetry 4 (residues 82–88, solid) and symmetry 4′ (residues 144–150, dashed). Residues that are not symmetrical are shown as solid blue lines. Two exons code for the regions not included within the twofold symmetry: exon I (arrowed Ala-1–Lys-23) and exon III (Gly-56–Arg-79). Exon III also includes part of a loop region included in the twofold symmetry. (See also Figure 1b.)

gether with the evidence that superoxide dismutases protect cells against superoxide-mediated oxidative stress (provided by the study of *E. coli* and yeast strains deficient in superoxide dismutase activity[58,62,63]), strongly suggests that the primary biological function of the CuZn SODs is to remove superoxide.

Structural Symmetry and Gene Duplication

Each CuZn SOD subunit has an internal twofold axis of symmetry relating its α carbon positions. Figure 4 illustrates this three-dimensional structural symmetry in the HSOD model. The symmetrical regions are also mapped to the amino acid sequence alignment and schematic diagram in Figure 1. The Greek key β-barrels of immunoglobulins[19] and β and γ lens crystallins[17] also exhibit internal twofold structural symmetry but the twofold symmetries of the three proteins are distinct. The structural symmetry in SOD led to the proposal that both halves of the primordial SOD were originally identical and encoded by one gene.[19] The primordial gene is believed to have duplicated and fused to form a single gene twice the size of the primordial gene that, after quite extensive further evolution, resulted in the present day gene (see Fig. 6).

The lack of significant amino acid duplication in the present day gene is perhaps evidence against the gene duplication hypothesis. However, three-dimensional structure is usually more conserved than amino acid sequence during evolution, and the CuZn SOD β-barrel may be more overdetermined than most structures. This means that the Greek key β-barrel structures may be able to tolerate considerable sequence variation, provided local packing and the alternation of hydrophobic and hydrophilic side chains in the β-strands are retained. The structurally symmetrical amino acid duplications present in the bovine SOD sequence are shown in Figure 1 together with the more extensive duplications found among the eukaryotic SOD sequences. These additional duplications occur primarily in a highly variable region of the sequence between residues 31 and 36, suggesting that they may not be very significant. Interestingly, the only sequence-invariant duplicated residues are His-48 and His-120, which ligate the activity essential Cu ion, which is also at the center of the axis of symmetry (see Figs. 1 and 3A). This led to the hypothesis that the other histidine ligands and the zinc domain were later evolutionary additions.[19] The more recently evolved lens crystallin β-barrel structures probably evolved by gene duplication since there are several symmetrically duplicated amino acids and each symmetrical region is encoded by one exon in the lens crystallin gene.[18]

Gene Duplication and Exons

Exons often correspond to structural or functional domains or subdomains of proteins[64] and introns probably direct the recombination of exons and their corresponding protein domains to facilitate protein evolution.[65] Intron–exon junctions frequently map to the protein surface, where insertions and deletions are least likely to disrupt the protein structure or

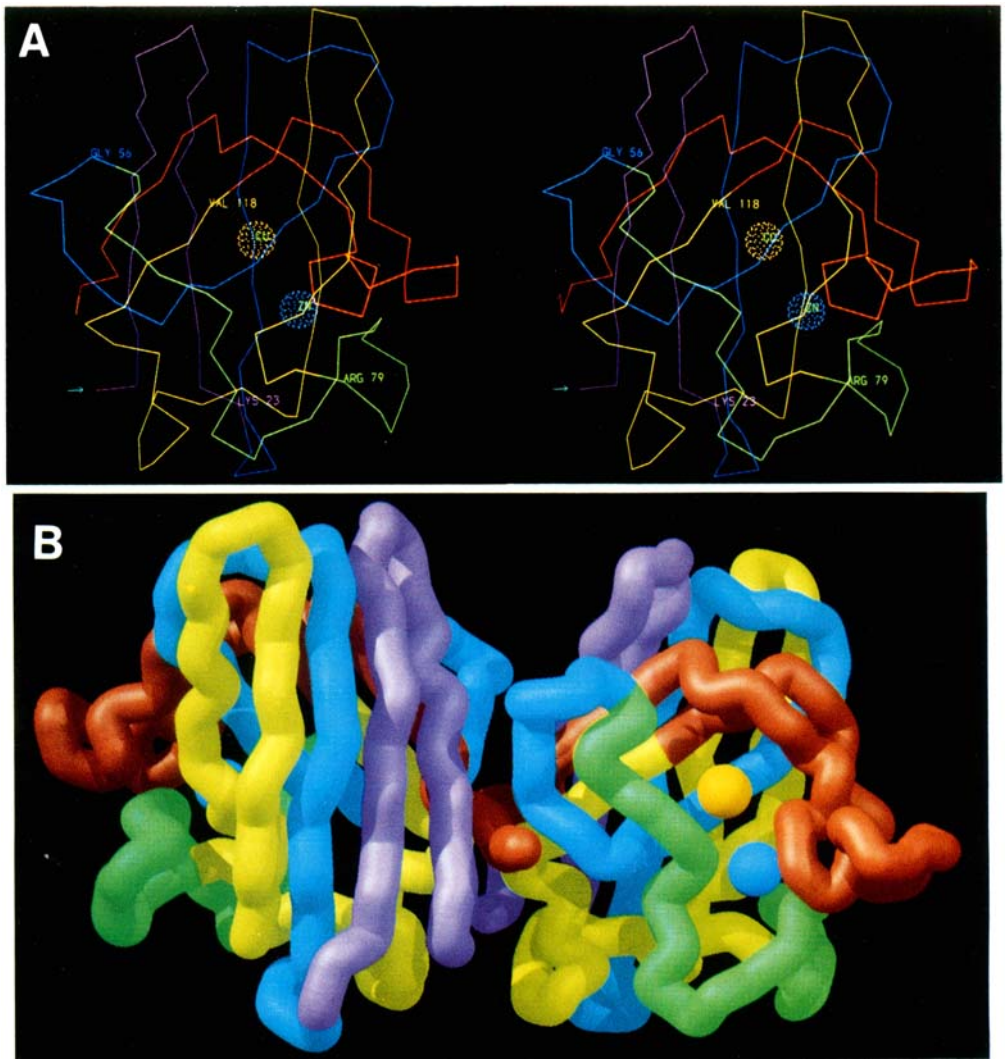


Fig. 5. The five HSOD exons mapped onto the three-dimensional structure of the monomer and native dimer. Amino acid residues are shown color coded for exon I (1–23, purple), exon II (24–55, blue), exon III (56–79, green), exon IV (80–118, yellow), and exon V (119–153, red). The active site channel is identified by spheres representing the Cu (orange) and Zn (blue) ions. All exons except exon III (green), which encodes part of active site loop IV, contribute residues to β-strands of the β-barrel fold. In **a**, the stereo pair shows the exons labeled at their ends and mapped onto the α carbon backbone of a monomer. The N-terminus is indicated with a light blue arrow (lower left). In **b**, solid tubes tracing the α carbon backbones map the relationship of the exons in the native dimer. The view is similar to that in **a** and the N-terminus (purple) and C-terminus (red) of the left subunit are at the front of the molecule and close together. As shown here, of the five exons, only exon III (green) is not involved in any dimer contacts. All intron–exon junctions, except that between exons IV (yellow) and V (red), map to sites on the surface of the protein, suggesting that this buried junction (center) is a relic of the ancestral gene encoding the three β-strand structure (see Fig. 6). This raster picture of 10,000 spheres representing the tracing of the molecule's α carbon backbone atoms was made by Michael Pique, computed on an Ardent Titan using software from Ray-Tracing Corporation, and photographed from a Sun TAAC-1.

function.[66] When the five exons of the human CuZn SOD gene[50,67] are mapped to the protein structure (Fig. 5), all but the last intron–exon junction (at Val-118) map to the protein surface.[50] Exons I and III of the human CuZn SOD gene correspond to structural subdomains consisting of the first antiparallel β-hairpin (exon I) and the Zn subloop (exon III).

Exon I lies outside the structural symmetry (see Figs. 1, 4, 5, and 6), supporting the hypothesis that it was added after gene duplication. The functional significance of exon I is not obvious; it contains a single invariant residue (Gly-16) involved in maintaining the β-barrel structure (see Figs. 1 and 3A) and three residues with minor contributions to the dimer contact (see Fig. 1). An interesting possibility, consistent with the conservation of Gly, Asn, and Asp residues favored in β-hairpin turns[68] at positions 11–13, is that exon I forms a nucleation site to facilitate protein folding. Before the addition of exon I, the first N-terminal β-strand of the six-stranded β-barrel was unable to pair with an adjacent β-strand until the fourth β-strand had been synthesized (see Fig. 6, strands 1 and 1′ of the six-stranded monomer). In contrast, the first β-strands of both halves of the

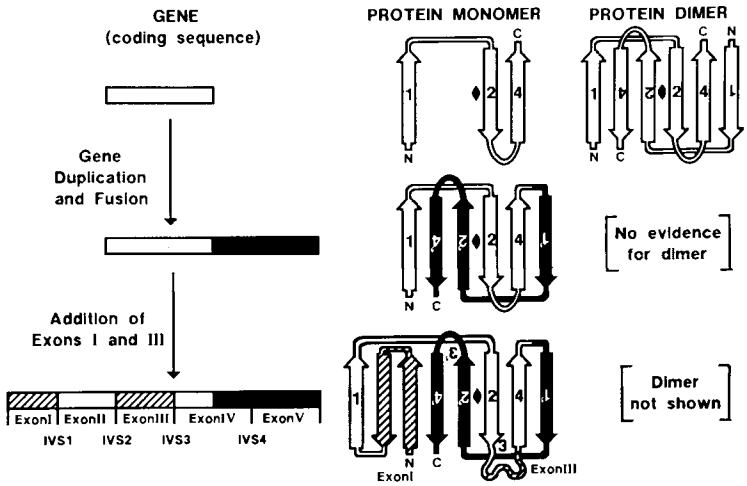


Fig. 6. Proposed evolution of CuZn SOD. The primordial SOD gene and protein structures are shown schematically on the top line and the present day structures on the bottom line. The structurally symmetrical regions 1 and 1′, 2 and 2′, and 4 and 4′ (see Fig. 1) are represented here entirely as β-strands (numbered arrows) connected by loop regions. The first loop in the primordial structure is the Greek key loop (connecting symmetry 1 and 2) that joins nonadjacent β-strands and forms a cross-barrel connection. The structurally symmetrical regions 3 and 3′ (see Fig. 1) are shown only on the present day protein structure, contain no β-strands, and may have been modified from those in the primordial structure (see text). The twofold axis of symmetry (containing the Cu ion) is indicated at the center (♦). Exons I and III are shown as more recent additions to the structure and both were added to the N-terminal symmetrical region. For simplicity, IVS-4 is shown only on the present day gene structure, although it may date from the very early evolution of the gene (see text).

primordial dimer can together form a β-strand pair immediately and the N-terminal β-hairpin of the present day eight-stranded β-barrel is able to form a β-strand pair immediately upon synthesis (see Fig. 6). The ability to rapidly form a β-sheet nucleation site may reduce the likelihood of incorrect folding after sequence changes or the insertion of new subdomains. Furthermore, most Greek key β-barrels (including trypsin, pyruvate kinase, plastocyanin, immunoglobulins, staphylococcal nuclease, γ crystallin, and the icosahedral virus capsid proteins) have an N-terminal β-hairpin.[14] Prealbumin, which does not have an N-terminal β-hairpin, is a more complex case, because the β-sheets of the prealbumin dimer contain strands from both polypeptide chains.

Exon III, which contains most of the zinc-binding subdomain and two of the liganding histidines (Fig. 1 and 5), also includes structurally symmetrical region 3 (Fig. 1 and 4, residues 58–63), which appears to be inconsistent with the addition of exon III after gene duplication. The 3, 3′ symmetry could result from the structural constraints for forming the superoxide channel, rather than from the gene duplication. In support of this possibility, the 3, 3′ symmetry has unique attributes: it does not encode any β-sheet (presumably the core of the primordial structure), it contains a 12-residue insertion in the *Photobacterium* sequence relative to all other known sequences (which probably removes the symmetry in this SOD), and it contains only one duplicated residue pair (Thr-58, Thr-137). Interestingly, exon III is the only exon that does not contribute any residues to the dimer interface (Fig. 5b), suggesting that addition of this Zn-binding loop may have occurred after the formation of the current dimer interface. The concept of exclusion of symmetry 3 from the primordial gene is not entirely satisfactory, however, since 3 and 3′ connect parts of loops IV and VII (2 and 2′ symmetry) back to the β-barrel (Figs. 1 and 4). Furthermore, it seems to make sense to enclose the primordial Cu ion in two such symmetrical loops to form a binding pocket for the superoxide. It is possible that symmetry 3 (Fig. 1, residues 58–63) did not exist in the primordial SOD structure (which contained a different symmetry 3 that has been lost) and, as suggested above, may have evolved due to the structural constraints of forming the superoxide channel and binding pocket. The determination of the exon pattern of SOD genes from plants or other distantly related species, or the solution of the three-dimensional structure of photobacterium SOD may help to unravel the origin of symmetry 3.

Implications of Exon Structure and Gene Duplication

As outlined above, one of the earliest structures in CuZn SOD evolution was probably a symmetrical six-stranded β-barrel (Fig. 6), made from two identical polypeptide chains and containing the catalytic Cu ion at the axis of symmetry. One implication of this hypothesis is that the precursor molecule of the six-stranded barrel was a three β-strand monomer. At present, we are unable to trace the evolutionary pathway that resulted in the formation of this three β-strand structure. The fourth intervening sequence (IVS-4, see Figs. 1 and 5) is the only intron that

maps to a buried position or interrupts a β-strand. IVS-4 may be a relic from the early evolution of the three β-strand structure that mapped to the surface of the primordial β-barrel and was buried by subsequent additions to the loops (including exon III). Thus, buried intron–exon junctions (which will tend to occur in core structures) may reflect relatively ancient events in protein evolution. If IVS-4 was present in the primordial gene it must have been lost from the 5′ (N-terminal) duplicated gene.

The present day exon II was formed from part of one primordial duplicated gene by the introduction of exon III and its flanking intervening sequences (see Fig. 6). Many genes with several introns will have exons similar to exon II: such exons are not necessarily structural or functional subdomains but are a reflection of useful sites in the protein structure able to accept new exons.

Another implication of the gene duplication and internal symmetry hypothesis is that the N and C termini of the different chains in the primordial dimer must have been close together so that gene fusion would not have caused structural distortion in the new monomer (see Fig. 6). Thus, due to the internal twofold symmetry, the N and C termini following gene fusion should also be in close proximity, as is found in many proteins.[69] However, the close proximity of the N and C termini in the present day CuZn SOD is somewhat fortuitous because exon I was added to the N terminus after gene duplication and could have moved the termini apart, as appears to have occurred in the structurally related immunoglobulin Greek key β-barrel.[70]

What evolutionary advantage is there in duplication and fusion of the gene encoding the three β-strand structure? The versatility of the β-barrel design is that it is sufficiently stable to allow a large variety of insertions (subdomains) at the loops between its β-strands, and sufficiently compact to enable such insertions to modify the function in a desirable way. The main limitation of the primordial β-barrel, composed of two three-stranded monomers (Fig. 6), is that its gene encodes only two loop connections in which to insert new subdomains or in which new functions could gradually evolve. Alterations to the existing functions of these loops would have a high probability of reducing enzyme activity and being selected against. Thus, the first turn between β-strands is the Greek key loop, which functions to stabilize the β-barrel by forming a cross barrel connection and by contributing hydrophobic "cork" residue(s) (see Fig. 3C) to plug the ends. The second loop of the primordial barrel corresponds to the present day electrostatic channel loop (see Figs. 1 and 6) and may have contained residues important for local electrostatic orientation or stabilization of the superoxide radical in the binding pocket.[4,52] Thus, alterations to the spatial organization of these residues by the insertion of new sequences would probably have inactivated the enzyme. An important advantage of gene duplication is that the symmetrically duplicated loops are encoded separately, allowing loops that were previously identical to diverge. Gradual functional specialization of one or both loops from a previously identical symmetrical pair seems to have occurred in loop IV (dimer contact/disulfide/Zn binding), Greek key loop VI (which contributes to the dimer contact), and the electrostatic channel loop VII (see Fig. 1). Specialization of loop IV also evolved by the addition of exon III containing the zinc-binding site.

Evolutionary Restrictions on Enzyme Size

Most enzymes are disproportionately large in relation to the size of the region thought to be essential for catalysis.[71] Analysis of the structural arrangement of the invariant and functionally equivalent residues essential for the catalytic function of SOD shows that these active site residues comprise only about 40% of the total molecular structure evaluated both in terms of sequence and volume (see Fig. 3A). Two continuous stretches of the sequence (Gly-44 to Asp-83, and His-120 to Cys-146) fold into the two loops and part of the underlying β-structure that form the active center and active site channel. The remainder of the protein is structural framework consisting of part of the Greek key β-barrel. It is now possible to suggest a reasonably precise account of the evolutionary basis for the size of the SOD molecule, which may have general relevance to other enzyme structures.

The proposed primordial SOD dimer, a six-stranded Greek key β-barrel, appears to have already determined the general dimensions of the current monomer, perhaps imposed by inherent size constraints for β-barrel stability. Since the β-barrel structure provided not only the framework for the active site but also contributed two liganding histidines for the catalytic Cu ion, it could not easily be altered without inactivating the enzyme. In fact, there appear to have been considerable evolutionary advantages in maintaining a framework structure sufficiently stable to tolerate alterations ranging from single residue changes to the insertion of new subdomains. The addition of these new subdomains (exons I and III) represents the most significant increase in size of SOD since the formation of the primordial six-stranded molecule. The history of the CuZn SOD molecule therefore appears to have involved a large number of single amino acid changes, reflected in the lack of amino acid duplication in the present day structure, and a modest increase in size from the primordial six-stranded structure due primarily to the addition of exons I and III. These changes probably contributed to increased efficiency of catalysis and structural stability; the present day enzyme has a reaction rate of $2 \times 10^9\ M^{-1}\ sec^{-1}$ (ref.

2) and is one of the most stable characterized globular proteins.[8–10]

ACKNOWLEDGMENTS

This work was supported in part by Chiron Corporation. Computer graphics and structural analysis studies were supported in part under NIH Grant GM39345 to J.A.T., E.D.G., and R.A.H. We thank Noel Fong for expert technical assistance.

REFERENCES

1. Fridovich, I. Superoxide and superoxide dismutases. Adv. Inorg. Biochem. 1:67–90, 1979.
2. Klug, D., Rabani, J., Fridovich, I. A direct demonstration of the catalytic action of superoxide dismutase through the use of pulse radiolysis. J. Biol. Chem. 247:4839–4842, 1972.
3. Cudd, A., Fridovich, I. Electrostatic interactions in the reaction mechanism of bovine erythrocyte superoxide dismutase. J. Biol. Chem. 257:11443–11447, 1982.
4. Getzoff, E.D., Tainer, J.A., Weiner, P.K., Kollman, P.A., Richardson, J.S., Richardson, D.C. Electrostatic interaction between superoxide and copper, zinc superoxide dismutase. Nature (London) 306:287–290, 1983.
5. Richardson, J.S. β-sheet topology and the relatedness of proteins. Nature (London) 268:495–500, 1977.
6. Tainer, J.A., Getzoff, E.D., Beem, K.M., Richardson, J.S., Richardson, D.C. Determination and analysis of the 2Å structure of bovine copper, zinc superoxide dismutase. J. Mol. Biol. 160:181–217, 1982.
7. Forman, H.J., Fridovich, I. On the stability of bovine superoxide dismutase. The effects of metals. J. Biol Chem. 248:2645–2649, 1973.
8. Malinowski, D.P., Fridovich, I. Subunit association and side-chain reactivities of bovine erythrocyte superoxide dismutase in denaturing solvents. Biochemistry 18:5055–5060, 1979.
9. Lepock, J.R., Arnold, L.D., Torrie, B.H., Andrews, B., Kruuv, J. Structural analyses of various Cu^{2+}, and Zn^{2+}-superoxide dismutases by differential scanning calorimetry and Raman spectroscopy. Arch. Biochem. Biophys. 241:243–251, 1985.
10. Roe, J.A., Butler, A., Scholler, D.M., Valentine, J.S., Marky, L., Breslauer, K.J. Differential scanning calorimetry of Cu,Zn-superoxide dismutase, the apoprotein, and its zinc-substituted derivatives. Biochemistry 27:950–958, 1988.
11. Steffens, G.J., Bannister, J.V., Bannister, W.H., Flohe, L., Gunzler, W.A., Kim, S-M.A., Otting, F. The primary structure of Cu-Zn superoxide dismutase from *Photobacterium leiognathi*: Evidence for a separate evolution of Cu-Zn superoxide dismutase in bacteria. Hoppe-Seyler's Z. Physiol. Chem. 364:675–690, 1983.
12. Lee, Y.M., Friedman, D.J., Ayala, F.J. Superoxide dismutase: An evolutionary puzzle. Proc. Natl. Acad. Sci. U. S. A. 82:824–828, 1985.
13. Feng, D-F., Doolittle, R.F. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25:351–360, 1987.
14. Richardson, J.S. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34:167–339, 1981.
15. Getzoff, E.D., Hallewell, R.A., Tainer, J.A. Structural implications for macromolecular recognition and redesign. In: "Protein Engineering: Applications in Science, Industry and Medicine." Inouye, M. (ed). New York: Academic Press, 1986: 41–69.
16. Getzoff, E.D., Tainer, J.A., Olson, A.J. Recognition and interactions controlling the assemblies of β-barrel domains. Biophys. J. 49:191–206, 1986.
17. Blundell, T., Lindley, P., Miller, L., Moss, D., Slingsby, C., Tickle, I., Turnell, B., Wistow, G. The molecular structure and stability of the eye lens: X-ray analysis of gamma crystallin II. Nature (London) 289:771–777, 1981.
18. Inana, G., Piatigorsky, J., Norman B., Slingsby, C., Blundell, T. Gene and protein structure of a β-crystallin polypeptide in murine lens: Relationship of exons and structure motifs. Nature (London) 302:310–315, 1983.
19. McLachan, A.D. Repeated folding pattern in copper-zinc superoxide dismutase. Nature (London) 285:267–270, 1980.
20. Hallewell, R.A., Mullenbach, G.T., Stempien, M.M., Bell, G.I. Sequence of a cDNA coding for mouse manganese superoxide dismutase. Nucleic Acids Res. 14:9539, 1986.
21. 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., Steimer, K.S., Mullenbach, G.T. Human Cu/Zn superoxide dismutase cDNA: Isolation of clones synthesizing high levels of active or inactive enzyme from an expression library. Nucleic Acids Res. 13: 2017–2034, 1985.
22. Sanger, F., Nicklen, S., Coulson, A.R. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74:5463–5467, 1977.
23. Messing, J. New M13 vectors for cloning. Methods Enzymol. 101:20–78, 1983.
24. Maniatis, T., Fritsch, E.F., Sambrook, J. "Molecular Cloning," Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982.
25. Tainer, J.A., Getzoff, E.D., Richardson, J.S., Richardson, D.C. Structure and mechanism of copper, zinc superoxide dismutase. Nature (London) 306:284–287, 1983.
26. Getzoff, E.D. The refined 2 Å structure of copper, zinc superoxide dismutase: Implications for stability and catalysis. Ph.D. dissertation, Duke University, 1982.
27. Hendrickson, W.A. Stereochemically restrained refinement of macromolecular structures. Methods Enzymol. 115:252–270, 1985.
28. Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. Science 221:709–713, 1983.
29. Weiner, P.K., Kollman, P.A. AMBER: Assisted model building with energy refinement. A general program for modelling molecules and their interactions. J. Comput. Chem. 2:287–303, 1981.
30. O'Donnell, T.J., Olson, A.J. A graphics language interpreter for realtime, interactive, three-dimensional picture editing and animation. Comput. Graphics 15:133–142, 1981.
31. Connolly, M.L., Olson, A.J. GRANNY, a companion to GRAMPS for the real-time manipulation of macromolecular models. Comput. Chem. 9:1–6, 1985.
32. Sherman, L., Dafni, N., Lieman-Hurwitz, J., Groner, Y. Nucleotide sequence and expression of human chromosome 21-encoded superoxide dismutase mRNA. Proc. Natl. Acad. Sci. U.S.A. 80:5465–5469, 1983.
33. Steinman, H.M., Naik, V.R., Abernethey, J.L., Hill, R.L. Bovine erythrocyte superoxide dismutase. Complete amino acid sequence. J. Biol. Chem. 249:7326–7338, 1974.
34. Schinina, M.E., Barra, D., Gentilomo, S., Bossa, F., Capo, C., Rotilio, G., Calabrese, L. Primary structure of a cationic Cu,Zn superoxide dismutase. The sheep enzyme. FEBS Lett. 207:7–10, 1986.
35. Schinina, M.E., Barra, D., Simmaco, M., Bossa, F., Rotilio, G. Primary structure of porcine Cu,Zn superoxide dismutase. FEBS Lett. 186:267–270, 1985.
36. Lerch, K., Ammer, D. Amino acid sequence of copper-zinc superoxide dismutase from horse liver. J. Biol. Chem. 256: 11545–11551, 1981.
37. Steffens, G.J., Michelson, A.M., Puget, K., Flohe, L. The amino-acid sequence of rat Cu-Zn superoxide dismutase. Hoppe-Seyler's Z. Physiol. Chem. 367:1017–1024, 1986.
38. Bewley, G.C. cDNA and deduced amino acid sequence of murine Cu-Zn superoxide dismutase. Nucleic Acids Res. 16:2728, 1988.
39. Rocha, H.A., Bannister, W.H., Bannister, J.V. The amino-acid sequence of copper/zinc superoxide dismutase from swordfish liver. Comparison of copper/zinc superoxide dismutase sequences. Eur. J. Biochem. 145:477–484, 1984.
40. Seto, N.O.L., Hayashi, S., Tener, G.M. The sequence of the Cu-Zn superoxide dismutase gene of Drosophila. Nucleic Acids Res. 15:10601, 1987.
41. Steffens, G.J., Michelson, A.M., Otting, F., Puget, K., Strassburger, W., Flohe, L. Primary structure of Cu-Zn superoxide dismutase of *Brassica oleracea* proves homology with corresponding enzymes of animals, fungi and pro-

karyotes. Hoppe-Seyler's Z. Physiol. Chem. 367:1007–1016, 1986.
42. Cannon, R.E., White, J.A., Scandalios, J.G. Cloning of cDNA for maize superoxide dismutase 2 (SOD2). Proc. Natl. Acad. Sci. U.S.A. 84:179–183, 1987.
43. Kitagawa, Y., Tsunasawa, S., Tanaka, N., Katsube, Y., Sakiyama, F., Asada, K. Amino acid sequence of copper, zinc-superoxide dismutase from spinach leaves. J. Biochem. 99:1289–1298, 1986.
44. Lerch, K., Schenk, E. Primary structure of copper-zinc superoxide dismutase from *Neurospora crassa*. J. Biol. Chem. 260:9559–9566, 1985.
45. Johansen, J.T., Overballe-Peterson, C., Martin, B., Hasemann, V., Svendsen, I. The complete amino acid sequence of copper, zinc, superoxide dismutase from *Saccharomyces cerevisiae*. Carlsberg Res. Commun. 44:201–217, 1979.
46. Steinman, H.M. The amino acid sequence of copper-zinc superoxide dismutase from bakers' yeast. J. Biol. Chem. 255:6758–6765, 1980.
47. Steinman, H.M. Bacteriocuprein superoxide dismutase of *Photobacterium leiognathi*. Isolation and sequence of the gene and evidence for a precursor form. J. Biol. Chem. 262:1882–1887, 1987.
48. Berget, S.M. Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature (London) 309:179–182, 1984.
49. Sherman, L., Levanon, D., Lieman-Hurwitz, J., Dafni, N., Groner, Y. Human Cu/Zn superoxide dismutase gene: molecular characterization of its two mRNA species. Nucleic Acids Res. 12:9349–9365, 1984.
50. Hallewell, R.A., Puma, J.P., Mullenbach, G.T., Najarian, R.C. Structure of the Human Cu/Zn SOD Gene. In: "Superoxide and Superoxide Dismutase in Chemistry, Biology and Medicine." Rotilio, G. (ed). Amsterdam: Elsevier, 1986:249–256.
51. Richardson, J.S., Getzoff, E.D., Richardson, D.C. The β-bulge: A common small unit of nonrepetitive protein structure. Proc. Natl. Acad. Sci. U.S.A. 75:2574–2578, 1978.
52. Beyer, W.F., Fridovich, I. Mullenbach, G.T., Hallewell, R.A. Examination of the role of arginine-143 in the human copper and zinc superoxide dismutase by site specific mutagenesis. J. Biol. Chem. 262:11182–11187, 1987.
53. Abernethy, J.L., Steinman, H.M., Hill, R.L. Bovine erythrocyte superoxide dismutase. Subunit structure and sequence location of the intrasubunit disulfide bond. J. Biol. Chem. 249:7339–7344, 1974.
54. Hjalmarsson, K., Marklund, S.L., Engstrom, A., Edlund, T. Isolation and sequence of complementary DNA encoding human extracellular superoxide dismutase. Proc. Natl. Acad. Sci. U.S.A. 84:6340–6344, 1987.
55. Hallewell, R.A., Laria, I., Tabrizi, A., Carlin, G., Getzoff, E.D., Tainer, J.A., Cousens, L.S., Mullenbach, G.T. Genetically engineered polymers of human CuZn superoxide dismutase. Biochemistry and serum half-lives. J. Biol. Chem. 264:5260–5268, 1989.
56. Parge, H.E., Getzoff, E.D., Scandella, C.S., Hallewell, R.A., Tainer, J.A. Crystallographic characterization of recombinant human CuZn superoxide dismutase. J. Biol. Chem. 261:16215–16218, 1986.
57. Hallewell, R.A., Mills, R., Tekamp-Olson, P., Blacher, R., Rosenberg, S., Otting, F., Masiarz, F.R., Scandella, C.J. Amino terminal acetylation of authentic human Cu,Zn superoxide dismutase produced in yeast. Biotechnology 5: 363–366, 1987.
58. Natvig, D.O., Imlay, K., Touati, D., Hallewell, R.A. Human copper-zinc superoxide dismutase complements superoxide dismutase-deficient *E. coli* mutants. J. Biol. Chem. 262:14697–14701, 1987.
59. Fee, J.A. Is superoxide important in oxygen poisoning? Trends Biochem. Sci. 7:84–86, 1982.
60. Sawyer, D.T., Valentine, J.S. How super is superoxide? Accts. Chem. Res. 14:393–400, 1981.
61. Lipscomb, W.N. In: "Methods for Determining Metal Ion Environments in Proteins: Structure and Function of Metalloproteins." Darnall, D.W. and Wilkins, R.G. (eds). New York: Elsevier North-Holland, 1980:265–302.
62. Bilinski, T., Krawiec, Z., Liczmanski, A., Litwinska, J. Is hydroxyl radical generated by the Fenton reaction in vivo? Biochem. Biophys. Res. Commun. 130:533–539, 1985.
63. Carlioz, A., Touati, D. Isolation of superoxide dismutase mutants in *Escherichia coli*: Is superoxide dismutase necessary for aerobic life? EMBO J 5:623–630, 1986.
64. Gilbert, W. Why genes in pieces? Nature (London) 271:501, 1978.
65. Gilbert, W. Genes-in-pieces revisited. Science 228:823–824, 1985.
66. Craik, C.S., Rutter, W.J., Fletterick, R. Splice junctions: Association with variation in protein structure. Science 220:1125–1129, 1983.
67. Levanon, D., Lieman-Hurwitz, J., Dafni, N., Wigderson, M., Sherman, L., Bernstein, Y., Laver-Rudich, Z., Danciger, E., Stein, O., Groner, Y. Architecture and anatomy of the chromosomal locus in human chromosome 21 encoding the Cu/Zn superoxide dismutase. EMBO J 4:77–84, 1985.
68. Sibanda, B.L., Thornton, J.L. β-hairpin families in globular proteins. Nature (London) 316:170–174, 1985.
69. Thornton, J.M., Sibanda, B.L. Amino and carboxyl-terminal regions in globular proteins. J. Mol. Biol 167:443–460, 1983.
70. Richardson, J.S., Richardson, D.C., Thomas, K.A., Silverton, E.W., Davies, D.R. Similarity of three-dimensional structure between the immunoglobulin domain and the copper, zinc superoxide dismutase subunit. J. Mol. Biol. 102:221–235, 1976.
71. Traut, T.W. What determines the size of enzymes? Mol. Cell Biochem. 70:3–10, 1986.