THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 10, Issue of March 5, pp. 9248–9254, 2004

Printed in U.S.A.

Specificity and Phenetic Relationships of Iron- and Manganese-containing Superoxide Dismutases on the Basis of Structure and Sequence Comparisons*[S]

Received for publication, November 11, 2003, and in revised form, December 10, 2003
Published, JBC Papers in Press, December 12, 2003, DOI 10.1074/jbc.M312329200

René Wintjens‡, Christophe Noël§¶, Alex C. W. May‖, Delphine Gerbod§, Fabienne Dufernez§, Monique Capron§, Eric Viscogliosi§, and Marianne Rooman‡‡§§**

*From the ‡Université Libre de Bruxelles, Institut de Pharmacie, Chimie Générale, CP 206/04, Campus de la Plaine, Boulevard du Triomphe, B-1050 Bruxelles, Belgium, the §Institut Pasteur de Lille, INSERM Unité 547, 1 Rue du Professeur Calmette, BP 245, 59019 Lille cedex, France, the **Institut de Pathologie Cellulaire Christian de Duve (ICP-TROP) 74.39, Avenue Hippocrate 74-75, 1200 Bruxelles, Belgium, the ‖National Institute for Medical Research, Division of Mathematical Biology, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom, and the ‡‡Université Libre de Bruxelles, Bioinformatique Génomique et Structurale, CP 165/61, 50 Avenue Roosevelt, B-1050 Bruxelles, Belgium*

The iron- and manganese-containing superoxide dismutases (Fe/Mn-SOD) share the same chemical function and spatial structure but can be distinguished according to their modes of oligomerization and their metal ion specificity. They appear as homodimers or homotetramers and usually require a specific metal for activity. On the basis of 261 aligned SOD sequences and 12 superimposed x-ray structures, two phenetic trees were constructed, one sequence-based and the other structure-based. Their comparison reveals the imperfect correlation of sequence and structural changes; hyperthermophilicity requires the largest sequence alterations, whereas dimer/tetramer and manganese/iron specificities are induced by the most sizable structural differences within the monomers. A systematic investigation of sequence and structure characteristics conserved in all aligned SOD sequences or in subsets sharing common oligomeric and/or metal specificities was performed. Several residues were identified as guaranteeing the common function and dimeric conformation, others as determining the tetramer formation, and yet others as potentially responsible for metal specificity. Some form cation-π interactions between an aromatic ring and a fully or partially positively charged group, suggesting that these interactions play a significant role in the structure and function of SOD enzymes. Dimer/tetramer- and iron/manganese-specific fingerprints were derived from the set of conserved residues; they can be used to propose selected residue substitutions in view of the experimental validation of our *in silico* derived hypotheses.

Aerobic organisms have developed mechanisms to protect against reactive oxygen intermediates arising from oxidative processes. The superoxide dismutase (SOD)[1] metalloenzymes (EC 1.15.1.1) constitute an example of such a defense against oxidative damage (1, 2). They catalyze the degradation of toxic superoxide radicals to oxygen and hydrogen peroxide (3–7). They are subdivided into two structurally distinct families: (i) copper/zinc-containing SODs (Cu/Zn-SODs) that use copper and zinc simultaneously in their active sites and are found in eukaryotes and bacteria; and (ii) iron/manganese-containing SODs (Fe/Mn-SODs) that bind specifically either Fe or Mn (8). Fe-SOD is found in prokaryotes, chloroplasts, and protozoans, and Mn-SOD is in both prokaryotes and mitochondrial matrices.

Iron and manganese SODs exhibit a high degree of sequence and structure similarity, strongly suggesting that these enzymes originate from a common ancestry. Each monomer adopts a similar α/β-fold, which combine to form a dimeric or tetrameric structure in solution (9, 10). Despite the similarity of the three-dimensional molecular environment around the metal (7, 11, 12), these proteins generally require a specific metal ion for activity (13–15); only a small number of SODs, called cambialistic, are functional with both iron or manganese. Despite several attempts to explain the strict metal specificity (10, 12, 16–21), its precise sequence and structural basis remain to be identified.

In the present paper, we concentrate on the known Fe/Mn-SOD structures. A phenetic analysis of the Fe/Mn-SOD family is achieved on the basis of three-dimensional structure superimpositions and compared with sequence-based phylogenetic investigations. The conservation of structure and sequence patterns within subsets of SOD sequences is systematically explored with the aim of identifying the reasons underlying the oligomerization characteristics and metal specificity. The results of our research directly apply to function identification in the context of genome sequencing and to knowledge-based modeling of Fe/Mn-SOD enzymes in the framework of structural genomics.

MATERIALS AND METHODS

Sequence and Structure Data—Our analysis is based on 261 Fe/Mn-SOD sequences listed in Table II in the supplemental material (available in the on-version of this article). The oligomerization state and the type of catalytic metal ion of each SOD were assigned according to the

* This work was supported by Communauté Française de Belgique Action de Recherche Concertée Grant 02/07-289 and European Commission Concerted Action Quality of Life Grant 2001-3-8.4. (to R. W. and M. R.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental material in the form of Fig. 4 (a depiction of the r.m.s. deviation of heavy main chain atoms after coordinate superimposition of the 66 pairs of SOD structures) and Table II (a listing of 261 SOD proteins and the assignment of their specific metal ions and their oligomeric states in solution).

¶ Present address: Molecular Biology Unit, Microbiology Group, Dept. of Zoology, The Natural History Museum, Cromwell Rd., London SW7 5BD, UK. E-mail: chrn@nhm.ac.uk.

§§ Research Director of the Belgian Fund for Scientific Research and to whom correspondence should be addressed. E-mail: mrooman@ulb.ac.be.

[1] The abbreviations used are: SOD, superoxide dismutase; Fe-SOD, iron-containing SOD; Mn-SOD, manganese-containing SOD; r.m.s., root mean square; PDB, Protein Data Bank.

TABLE I

Distance scoring matrix of sequence similarities (right upper triangle, amino acid change score) and structure similarities (left lower triangle, r.m.s. of heavy main chain atoms after coordinate superimposition)

	1isa	1dt0	1qnn	1b06	1sss	1coj	1ids	1bsm	1n0j	1kkc	1d5n	1mng
1isa		0.44	0.85	1.52	1.53	2.11	1.43	1.42	1.34	1.47	1.20	1.11
1dt0	0.73		0.91	1.59	1.50	2.14	1.41	1.44	1.24	1.35	1.19	1.24
1qnn	1.02	0.76		1.55	1.54	2.07	1.51	1.56	1.29	1.46	1.31	1.20
1b06	4.37	4.91	4.56		0.15	1.83	1.26	1.18	1.30	1.31	1.47	1.54
1sss	4.47	4.71	4.66	0.44		1.78	1.26	1.18	1.34	1.26	1.44	1.55
1coj	4.44	5.04	4.61	2.72	2.71		1.99	1.83	1.82	1.98	2.18	1.78
1ids	4.22	4.51	4.37	1.44	1.44	2.31		0.53	0.83	1.04	1.27	1.44
1bsm	4.22	4.60	4.34	1.64	1.61	2.29	0.76		0.95	1.03	1.13	1.25
1n0j	3.86	4.18	3.99	1.93	1.98	2.52	1.31	1.40		0.82	1.03	0.90
1kkc	4.22	4.36	3.53	2.25	2.03	2.84	2.05	1.92	1.83		0.96	1.03
1d5n	1.43	1.52	1.26	5.16	5.02	5.05	3.66	3.66	3.15	3.43		0.86
1mng	1.26	1.32	1.84	4.77	4.90	4.91	3.56	3.52	3.06	3.53	1.14	

data base annotations and the literature. We identified 156 dimers (71 iron, 72 manganese, and 13 cambialistic SODs) and 105 tetramers (20 iron, 82 manganese, and 3 cambialistic SODs). We would like to stress that this assignment is quite a delicate point. Indeed, neither the metal specificity nor the oligomeric state exhibits all-or-none behavior. For example, some SOD enzymes are dimeric or tetrameric, depending on the experimental conditions (22). Other SODs can bind either iron or manganese according to the metal available in the culture medium (23–27). The existence of such SODs, named cambialistic, render unambiguous metal ion attribution difficult, especially considering that they may be cambialistic to various degrees. As a matter of fact, only a few SODs have had their preferred metal experimentally identified. Most attributions were made on the basis of sequence similarity using BLAST (28) and pairwise sequence comparisons.

For 12 of these SOD sequences, a high-resolution x-ray structure is available, which we retrieved from the Protein Quaternary Structure server (http://pqs.ebi.ac.uk). These are SODs from *Escherichia coli* (Protein Data Bank (PDB) accession numbers 1isa and 1d5n; x-ray resolution 1.80Å and 1.55Å, respectively), *Homo sapiens* mitochondria (1n0j; 2.20Å), *Sulfolobus acidocaldarius* (1b06; 2.20Å), *Propionium freudenreichii* subsp. *shermanii* (1bsm; 1.35 Å), *Pseudomonas ovalis/putida* (1dt0; 2.10Å), *Mycobacterium tuberculosis/vaccae* (1ids; 2.00Å), *Thermus thermophilus* (1mng; 1.80Å), *Porphyromonas gingivalis* (1qnn; 1.80Å), *Sulfolobus solfataricus* (1sss; 2.30Å), *Aquifex pyrophilus* (1coj; 1.90Å), and *Aspergillus fumigatus* (1kkc; 2.00Å).

Sequence- and Structure-based Similarity Trees—Two distance matrices were computed on the basis of the sequence and structure similarities, respectively, of the 12 known SOD structures. The former was obtained with the ProtDist program of the PHYLIP package (29) using the JTT substitution model (30). For the latter, we performed pairwise superimpositions of three-dimensional structures of apoenzymes with the SoFiSt program (31). This program is designed to yield optimal superimpositions with respect to the root mean square (r.m.s.) deviation of heavy main chain atoms after coordinate superimposition, generally considered to be the most reliable measure for comparing closely related structures. It was applied in two steps. First, the subset of secondary structure elements that are the most similar in the two structures, *i.e.* that superimpose with, at most, 2-Å r.m.s. deviation, was identified. The aligned segments were then extended in such a way that all residues except those corresponding to insertions and deletions are aligned. The r.m.s. deviations of the pairwise superimpositions so obtained were used as entries of the second distance matrix.

A hierarchical classification (tree) was obtained from each distance matrix using the unweighted pair-group method with arithmetic averages (UPGMA), the most widely utilized algorithm for generating hierarchical classifications (32). The reliability of this classification was assessed on the basis of a jackknife test (33) so as to identify the reliable nodes (34, 35). The trees were drawn using the TreeView program (36).

Sequence Conservation—From the multiple sequence alignment of 261 SOD proteins, we identified residues that were conserved at a given position in at least 90% of all SOD protein sequences. Simultaneously, we searched for residues that are specific to the oligomeric state (dimer or tetramer) and/or type of metal (iron or manganese). More precisely, we defined four subgroups (iron dimer, iron tetramer, manganese dimer, and manganese tetramer) and looked for residues that occur in at least 80% of the proteins of one or several subgroups and in, at most, 20% of the proteins of each of the remaining subgroups. A gap in the sequence was considered to be a 21st residue. We also searched for conserved residue properties, *e.g.* aromatic, aliphatic, and charged, with a stricter threshold of 90%.

Analysis of Structure Features—Secondary structure assignments were defined according to the DSSP program (37), and hydrogen bonds were determined using HBPLUS (38). Cation-π and amino-π interactions were defined geometrically by a distance and an angle criterion (39). For simplicity, both interactions are here referred to as cation-π. Amino acids in the vicinity of the metal ion were identified as having a side chain atom at less than 6.5 Å from the metal. Buried residues were defined as having <10% side chain solvent accessibility in the dimeric form. Solvent accessibilities were computed with NACCESS (40). Residues at the dimeric or tetrameric interface were defined as having a side chain solvent accessibility of 25% at least in the monomeric or dimeric form, and losing >60% of their solvent accessibility upon complex formation.

RESULTS

Phenetic Classifications—The 12 SOD structures present a pairwise structural similarity ranging from 0.4 to 5.2 Å r.m.s. deviation, with a sequence identity between 25.4% and 85.4%. On the basis of the distance matrices defined from pairwise sequence and structure similarities (Table I), two phenetic trees were inferred and are depicted in Fig. 1, *a* and *b*. Note that the sequence-based tree can be viewed as phylogenetic, as its construction involves an evolutionary model (41). To test the reliability of the trees, a jackknife procedure was applied (see "Materials and Methods"), which showed that all of the nodes of the three-dimensional structure-based tree (Fig. 1*b*) are reliable, whereas there are two unreliable nodes in the sequence-based tree (Fig. 1*a*).

The connectivities of the two trees show marked differences, which can be considered as significant despite the limited number of SOD structures. First, the structure-based tree makes a clear-cut distinction between dimer and tetramer SODs. Only the Mn-SOD of *T. thermophilus* (1mng) is grouped with dimers, but it is quite a special kind of tetramer. Indeed, structure superimpositions of complete tetramer units reveal three types of tetramer organization characterized by a solvent accessible surface area buried by the dimer-dimer interface of ~10,500 Å^2 (in the iron or cambialistic SODs 1b06, 1sss, 1ids, 1coj and 1bsm), 4500 Å^2 (in the Mn-SODs 1n0j and 1kkc), and <3000 Å^2 (in 1mng only). In fact, 1mng exhibits very few dimer-dimer contacts (Fig. 2*a*) and is a totally atypical tetramer that will be shown to possess all the sequence characteristics of the dimers.

The dimer/tetramer distinction is less clear-cut in the sequence-based tree, where the dimeric Mn-SOD 1d5n appears merged with the tetramers. Some dimer/tetramer distinction is nevertheless perceptible, considering that 1b06 and 1sss proteins are dimeric at room temperature (22) and that the node linking them to the main tree is unreliable according to the jackknife test.

The comparison of the two trees reveals the existence of significant structural variations between dimers and tetramers induced by less marked sequence alterations. These variations involve a shorter H1 helix in the dimers and a H2 helix sepa-

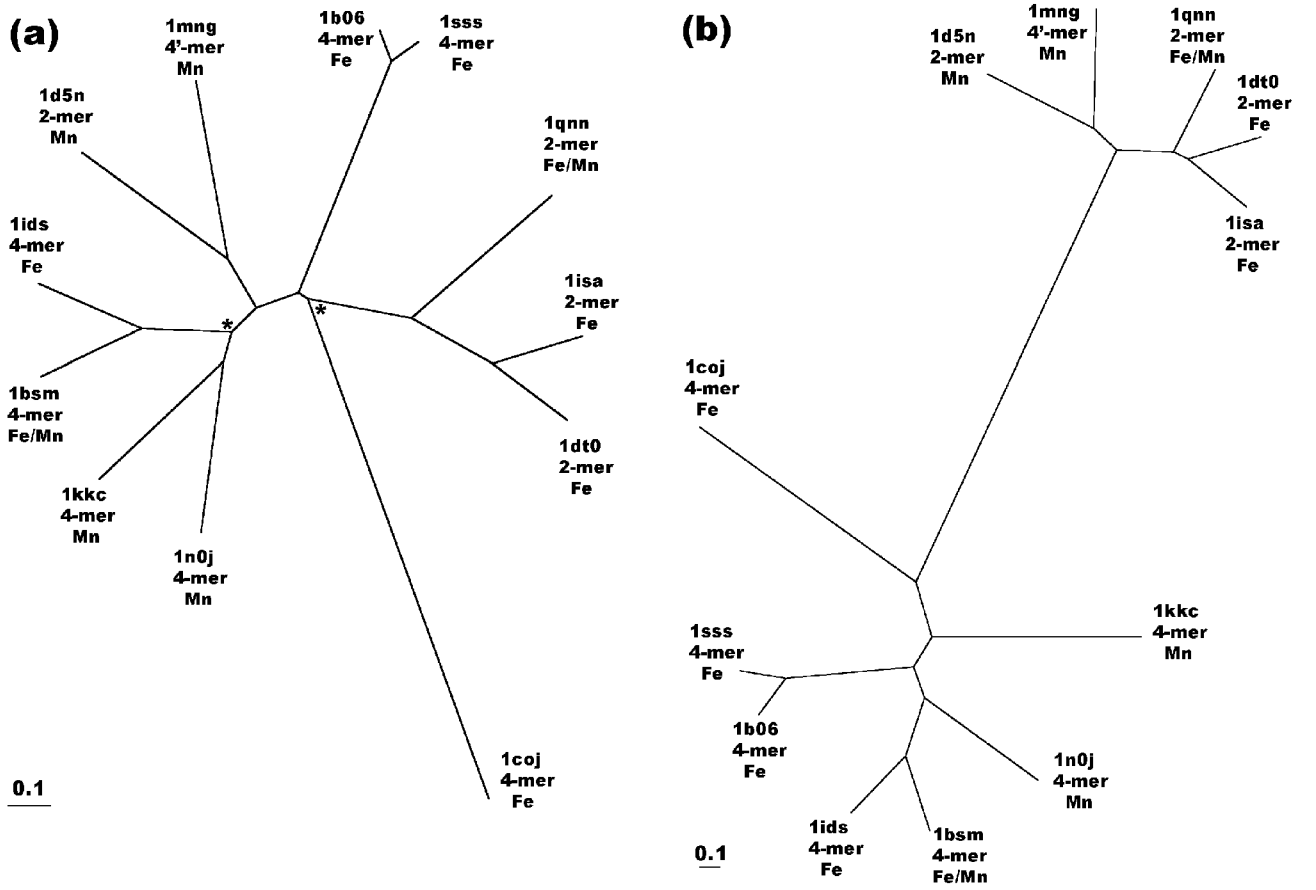

FIG. 1. **Phenetic trees based on the amino acid sequence (*a*) or the spatial structure (*b*) of the 12 known x-ray structures of SOD enzymes.** Each protein is labeled by its protein code; the metal ion type and the oligomeric state of the active form are also indicated. An *asterisk* indicates non-reliable nodes according to the jackknife test (see "Materials and Methods"). The *scales* below the trees indicate a length of 0.1 in the distance measures used (*i.e.* amino acid change score for *panel a*, Å for *panel b*).

rated into two pieces (H2a and H2b) by a loop (Fig. 2*a*). They are accompanied by a modification of the solvent-accessible surface area buried upon dimer formation, which is <8000 Å2 in dimeric SODs and >8000 Å2 in tetrameric forms.

A second noticeable feature of the structure-based tree is the segregation between manganese and iron enzymes among the dimeric SODs, with the cambialistic SOD (1qnn) located in between. A structural difference explaining this segregation implicates the H2a-H2b loop, which is longer in manganese than in iron dimers. The manganese/iron differentiation is less manifest in the tetramer subgroup. The sequence-based tree also distinguishes between Fe- and Mn-SODs to a certain extent, but the dimer/tetramer and iron/manganese classifications overlap.

In the sequence-based tree, the iron tetramer from *A. pyrophilus* (1coj), an extremely heat-stable enzyme occurring in a hyperthermophilic bacterium (42), appears very distant from all other SODs. This suggests that the sequence has evolved to ensure optimal heat resistance while maintaining the functional SOD structure.

The sequence- and structure-based trees thus yield distinct phenetic views of SOD enzymes, which is not surprising considering the limited correlation between sequence and structure similarity scores (as monitored by a correlation coefficient of 0.4; see Fig. 4 in the supplemental material that can be found in the on-line version of this article). Comparison of these trees allows us to identify features, such as the oligomeric state or the metal ion specificity in dimers, that are obtained by means of a minimum of sequence modifications but that induce significant structural changes. It also allows us to detect properties, such as hyperthermophilicity, that require, on the contrary, few structural changes but drastic sequence alterations.

Conserved Sequence and Structure Characteristics—Residues conserved in at least 90% of all 261 available SOD sequences are indicated in Fig. 2*b*. Six are perfectly conserved, among which four ligand the metal ion (His^{26}, His^{73}, Asp^{156}, and His^{160}; we use the PDB numbering of 1isa) and three occur near the dimer interface (His^{160}, Glu^{159}, and Tyr^{163}). The latter three residues ensure the conservation of two interchain interactions at the dimer interface, namely a 100% conserved double salt bridge between Glu^{159} and His^{160} (43) and an almost conserved double H-bond between His^{30} and Tyr^{163} (44).

Some other highly conserved residues are in the immediate environment of the metal cofactor (His^{30}, His^{31}, Tyr^{34}, Trp^{77}, Trp^{122}, and Trp^{158}) or occur near the dimer interface (His^{30}, Tyr^{34}, Ser^{120}, Trp^{158}, Ala^{161}, Tyr^{162}, and Asn^{168}). The channel leading to the active site is situated at this interface, with the entrance gated by the conserved residues His^{30} and Tyr^{34} (45). Still other conserved residues are part of the protein core or can be expected to be structurally important. In particular, Asn^{72}-Trp^{122} form a cation-π interaction, and so do Lys^{107}-Trp^{178}, whose level of conservation is, however, slightly lower than 90%. Residues Ile^{96}, Leu^{125}, Leu^{133}, and Trp^{183} contribute to hydrophobic packing. The conserved proline Pro^{16}, always in *cis*-conformation, is located at the end of the N-terminal arm. This arm adopts, in all SODs, the same extended structure packed against the helices H1 and H2, ensured by the conserved residues Leu^{7}, Ala^{13}, Leu^{14}, and Asn^{39}.

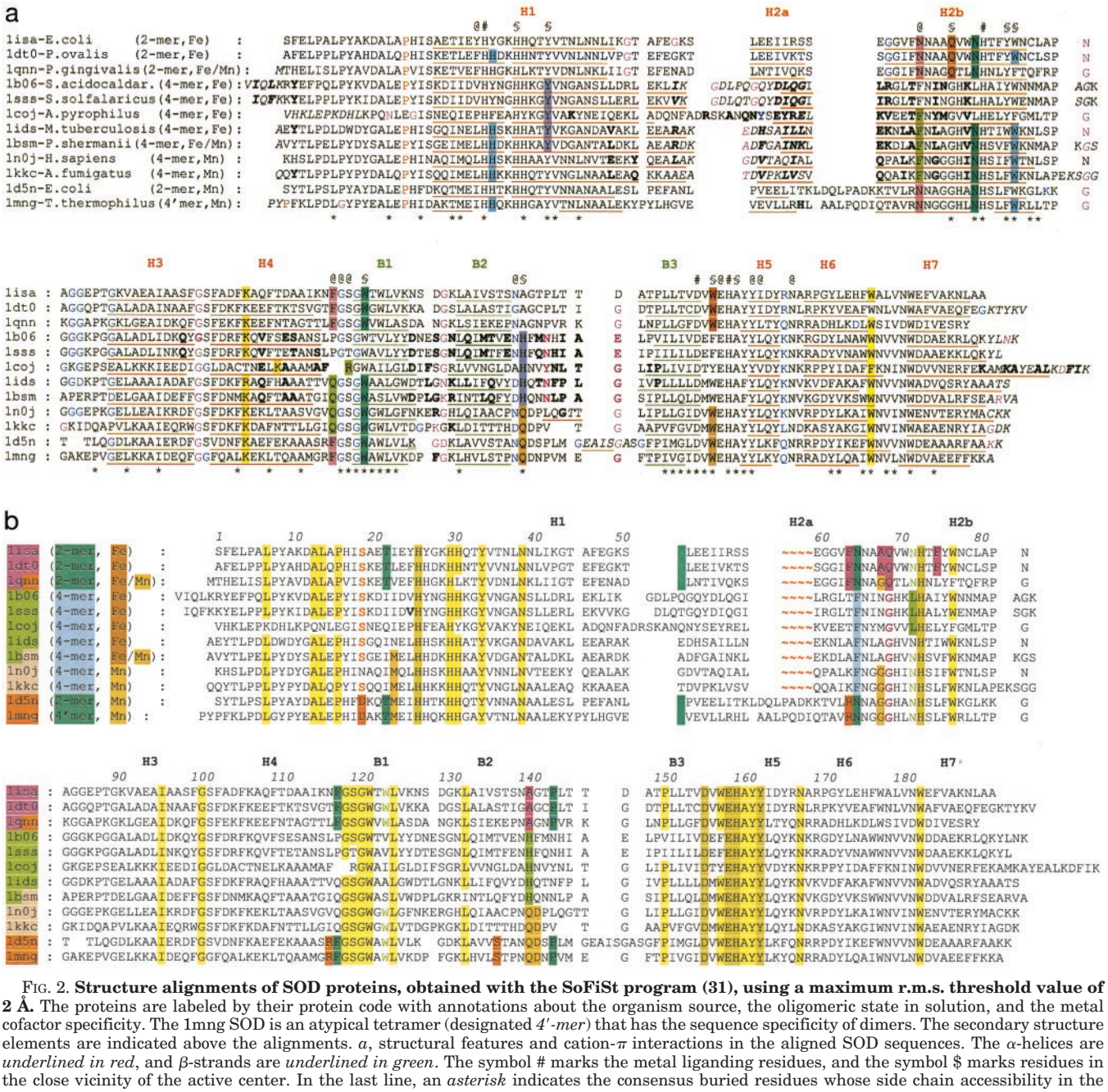

FIG. 2. **Structure alignments of SOD proteins, obtained with the SoFiSt program (31), using a maximum r.m.s. threshold value of 2 Å.** The proteins are labeled by their protein code with annotations about the organism source, the oligomeric state in solution, and the metal cofactor specificity. The 1mng SOD is an atypical tetramer (designated *4′-mer*) that has the sequence specificity of dimers. The secondary structure elements are indicated above the alignments. *a*, structural features and cation-π interactions in the aligned SOD sequences. The α-helices are *underlined in red*, and β-strands are *underlined in green*. The symbol # marks the metal liganding residues, and the symbol $ marks residues in the close vicinity of the active center. In the last line, an *asterisk* indicates the consensus buried residues whose side chain accessibility in the dimeric form is <10%. Positions that are not structurally aligned with the 1isa structure are indicated in *italics*, residues involved in the dimer-dimer interface are in *boldface*, prolines in *cis* conformation are in *red*, residues with a positive ϕ left-handed helical conformation are in *purple*, and residues in extended conformation with a positive ϕ dihedral angle are in *blue*. Residues on colored backgrounds are involved in cation-π interactions. *Yellow background*, Lys^{107}-Trp^{178} cation-π interaction in 1isa numbering; *pink*, interchain Asn^{65}-Phe^{118} interaction; *green*, interchain Gln^{118}-Phe^{65}; *mauve*: His^{141}-Tyr^{34}; *turquoise blue*, Asn^{72}-Trp^{122}; *sky blue*, His^{27}-Trp^{77}; *light brown*, Gln^{141}-Trp^{158}; *dark brown*, Gln^{69}-Trp^{158}. *b*, specific residue conservation in SOD sequences. The first line contains the PDB sequence numbering from the 1isa PDB file. *Dark yellow background*, residues conserved in 100% of SOD sequences; *light yellow background*, residues conserved in at least 90% of SOD sequences; *blue background*, residues specific to dimers (Thr^{22} in 1isa numbering, the gap of at least seven residues following residue 51, and Asn^{65}, Phe^{118}, and Pro^{144}); *purple background*, tetramer-specific (Phe^{65}); *orange background*, manganese-specific (Met^{23}, Gly^{68}, Gln^{141}, and Asp^{142}); *violet background*, iron dimer-specific (Phe^{64}, Ala^{68}, Gln69, Phe^{75}, and Ala^{141}); *violet letters*, specific for all but iron dimers (Gly^{69}); *red background*, manganese dimer-specific (Asp^{19}, Arg^{64}, Arg^{117}, and Ser^{137}); *red letters*, specific for all but manganese dimers (Ser^{19}, the gaps of at least 11 residues following residue 59, and the other proteins having gaps of, at most, four residues); *green background*, iron tetramer-specific (Leu^{72} and His^{141}); *green letters*, specific for all but iron tetramers (Asn^{72} and aromatic position 124). Note that the results are derived from the 261 SOD sequences and represented on the 12 SOD structures; therefore, some residues appearing as conserved in the structures are not colored, as they are not conserved enough among all sequences.

Some residues are conserved to maintain particular types of turns between successive secondary structures. Gly^{101} has a positive ϕ value and is situated in the two-residue H3-H4 turn. In general, this turn is of type αGBα (G denotes left-handed helical conformation, B is an extended β-type conformation, and α is an α-helix) with the sequence GS. In the thermophilic proteins 1mng and 1coj, however, it is an αGEα turn (E denotes a positive ϕ extended conformation) (46) with sequence GG. The conserved glycines Gly^{119} and Gly^{121} adopt a positive ϕ E-type extended conformation and are required for the particular αBBEBEβ turn (47) between helix H4 and β-strand B1. Finally, the αEABα turn linking the α-helices H5 and H6

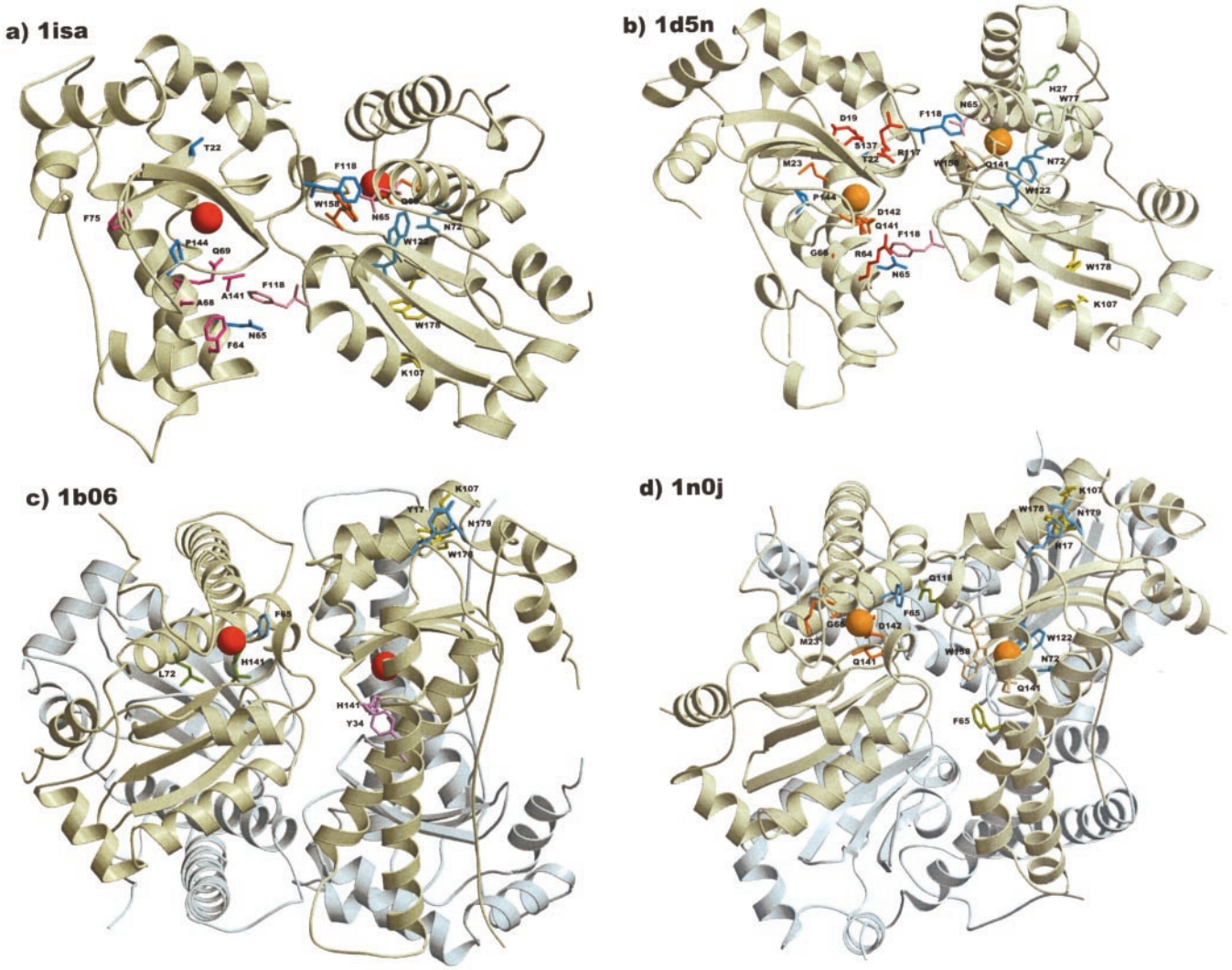

FIG. 3. **Ribbon views of typical SOD structures.** Dimer Fe-SOD 1isa (*a*), dimer Mn-SOD 1d5n (*b*), tetrameric Fe-SOD 1b06 (*c*), and tetrameric Mn-SOD 1n0j (*d*) are depicted. Iron and manganese atoms are represented in *red* and *orange*, respectively. Ribbons in the foreground dimeric unit are in *sable color*, whereas, for the tetrameric protein the background ribbons are *gray*. In the *right part* of the dimeric unit, the residues involved in conserved cation-π interactions are labeled and depicted according the color convention of Fig. 2*a*; in the *left part*, the specific residue conservations in SOD sequences are shown and colored as in Fig. 2*b*. See the Fig. 2 legend for more details.

contains two conserved residues, namely Asn[168] close to the dimer interface and, at position 170, a positively charged residue located near the substrate funnel entry (45). This particular loop seems thus to be necessary for the enzymatic function.

Dimer/Tetramer-specific Sequence and Structure Characteristics—Several residues are conserved in subsets of the 261 SOD sequences, characterized by specific oligomeric and/or metal binding properties. They are listed in Fig. 2*b* and depicted in Fig. 3.

Residues Thr[22], Asn[65], Phe[118], and Pro[144] are systematically encountered in dimers and never in tetramers, whereas Phe[65] is conserved among tetramers. Tetramers also exhibit a shorter loop, by one residue at least, at the end of helix H1. The larger number of dimer-specific residues compared with tetramer-specific residues is not surprising, as several kinds of tetramers are merged in the tetramer subset (see above).

In dimers, Asn[65] and Phe[118] are linked by an interchain cation-π interaction (Fig. 2*a*) across the dimer interface situated not far from the entrance of the main substrate channel. In tetramers, these two positions are mutated simultaneously, with Asn[65] always substituted by Phe and Phe[118] often by Gln; in the latter case, a cation-π interaction is formed again. The conserved Pro[144] in dimers introduces a short 3_{10} helix located in the loop between the β-strands B2 and B3. Thr[22] is situated between the metal-liganding residue His[26] and residue Ser[19] or Asp[19], which has been suggested as forming the entrance of an alternative pathway, allowing the substrate to reach or leave the functional site (44).

Manganese-specific Sequence and Structure Characteristics—Four manganese-specific residues are conserved in both dimers and tetramers. Met[23] is near the entry of the potential alternative channel to the functional site; Gly[68], Gln[141], and Asp[142] are spatially close and situated near the dimer interface. Note that Asp[142] frequently forms a salt bridge with residue 64, which is usually positively charged, especially in manganese dimers. Gly[68] is flanked by Gly[69] (also present in iron tetramers) and often by Gly[67]. This GGG pattern occurs in the middle of helix H2, an unusual feature that can be expected to locally weaken the structure and is required by packing constraints. The last manganese-specific residue, Gln[141], forms a cation-π interaction along the dimer interface with Trp[158], which is conserved in almost all SODs and situated close to the metal ion.

No residues are specific to manganese tetramers, whereas Asp[19], Arg[64], Arg[117], and Ser[137] are specific to manganese

dimers. Arg^{64}, in a salt bridge with the manganese-specific residue Asp^{142}, faces Arg^{117} across the dimer interface and flanks the dimer-specific Asn^{65}-Phe^{118} cation-π interaction. Ser^{137} and Arg^{117} are spatially close but do not seem to be involved in structure or function. Asp^{19} is situated at the entry of the possibly alternative funnel toward the active site (44); all other SODs possess a serine at that position. Finally, manganese dimers display a large insertion between helices H2a and H2b, which is located in the vicinity of the tetramer interface in the tetrameric SODs.

Iron-specific Sequence and Structure Characteristics—Fe-SODs have no conserved features, but iron dimers and iron tetramers do. The iron dimer-specific residues are Phe^{64}, Ala^{68}, Gln^{69}, Phe^{75}, and Ala^{141}. Phe^{64} is next to Asn^{65}, which forms the dimer-specific interchain cation-π interaction Asn^{65}-Phe^{118}; it occupies the same position as Arg^{64} in manganese dimers. Residue Phe^{75} is in helix H2b and points in the direction opposite to the metal; it is stacked against an aromatic residue at position 71, located one helix turn ahead. Gln^{69} is near the metal ion and forms a cation-π interaction with Trp^{158}, which is known to be involved in the reactivity (51); note that, in manganese enzymes, Trp^{158} forms an alternative cation-π interaction with Gln^{141}, whose side chain fills the place occupied by the Gln^{69} side chain in iron dimers. Steric reasons impose the presence of small residues at positions 68 and 141 when the cation-π interaction Gln^{69}-Trp^{158} is formed; this explains the conservation of Ala^{68} and Ala^{141}.

There are only two conserved residues in iron tetramers, Leu^{72} and His^{141}; moreover, these proteins never present an aromatic residue at position 124, probably due to structural constraints. Leu^{72} is positioned just before one of the metal-liganding His residues; in all but the iron tetramers this position is occupied by an Asn residue. His^{141} forms a cation-π interaction with the conserved Tyr^{34} located at the entry of the main substrate access funnel. In addition, His^{141} is hydrogen bonded with the OH of Tyr^{34} and with a metal-bound water molecule (H_2O or OH^-, depending on the oxidation state of the metal ion). The same two H^- bonds are observed for Gln^{141} in Mn-SODs and Gln^{69} in iron dimers. Not surprisingly, therefore, residues at positions 69 and 141 are largely described as influencing the metal specificity (11, 20, 48–50). The nature of the residues at these positions is thought to play an important role in the catalytic fine tuning of the enzyme, and their mutation induces large effects on the catalytic activity.

DISCUSSION

Several explanations have been proposed to understand the strict metal ion specificity of SOD proteins (10, 12, 16–20). The most convincing is from Vance and Miller (19, 21), who showed that the active site environment of *E. coli* SOD induces redox potential tuning that is appropriate for reactions with one type of metal ion only. However, the chemical groups responsible for the redox tuning effects have not yet been fully identified. This task is quite complicated, given that metal selectivity and specificity is not overruled by a single residue but by the synergetic effects of several key groups.

The systematic investigation of conserved sequence and structure characteristics of SOD enzymes performed in this paper led us to recover and specify known features, but it also revealed new ones. The relevance of these characteristics in explaining oligomeric states or metal specificities should first, of course, be confirmed computationally using semi-empirical approaches and, ultimately, by experimental means through, for instance, site-directed mutagenesis. Our results lead us to propose a concrete list of single-site or concerted mutations to be tested, which can be deduced from Fig. 2*b* and involve residues that either appear to guarantee the common SOD structure or function or to modulate the quaternary structure or metal specificity.

This list of residues is not meant to be exhaustive; other residues also influence structure or function. For example, position 154, located ~10 Å away from the active site, has recently been shown to affect metal specificity (52). It is occupied by a Gly in 74% of the Mn-SODs and by a Thr in 72% of the Fe-SODs. According to our criteria, which require 80% conservation at least, these residues are not considered as representing typical manganese/iron characteristics. The fact that they, nevertheless, play a role demonstrates that specificity is achieved by an ensemble of residues of which some are situated far from the active site. The latter observation supports our prediction of specificity influencing residues not situated in the neighborhood of the metal ion.

Seven cation-π interactions between aromatic and (partially) charged groups were identified as being particularly well conserved (Fig. 2*a*) and are thus suspected to play important structural and functional roles. Besides their obvious structural role, it can be argued that these cation-π interactions play a functional role by fixing the exact positions of residue side chains in the vicinity of the metal ion and probably by tuning the redox potential of the metal ion by exploiting the electronic properties of aromatic amino acids. For example, the His^{141}-Tyr^{34} pair, which only forms when the histidine is protonated and increases the pK_a (53), is especially likely to play an active role in the function. Indeed, it is situated at the entry of the main pathway to the active site, where it can be suggested to play the role of a gate (17, 45, 49, 54, 55) and even to be involved in the catalysis as a proton donor.

The dimer interface generates two symmetrical substrate funnels that lead from the bulk solvent to the metal ions (45). There are several conserved side chains found to line this funnel; others close its bottom and seem to prevent substrates from reaching the metal ion of the other monomer. Note that the metal-specific residues are situated near the metal and along the channel entrance, not at its bottom, which is consistent with their role in tuning specificity.

The existence of an alternative substrate access channel has been proposed with the aim of explaining the rapid turnover, which is incompatible with a single pathway toward the active site (44). This alternative channel has been suggested as being situated at the interface between the N-terminal helical domain and the C-terminal α/β domain of each monomer. The conserved residues Asp^{19} in manganese dimers, Ser^{19} in all but manganese dimers, Thr^{22} in dimers, and Met^{23} in manganese enzymes are situated along this alternative pathway and, hence, tend to support its very existence.

The present analysis also reveals that SOD dimers, whether binding iron or manganese, have common sequence and structure characteristics that differentiate them from tetramers. In contrast, tetramers only present common features after subdivision into iron- and manganese-specific enzymes. This can be taken to mean that iron and manganese dimers have recently evolved from a common ancestor, whereas the common ancestor of iron and manganese tetramers dates from the very remote past.

The *T. thermophilus* SOD (1mng) presents quite an atypical behavior. Despite the experimental evidence indicating its tetrameric behavior in solution (56–61), it has all the sequence and structural features of typical dimers (see Figs. 1 and 2). The few contacts across the tetramer interface, moreover, indicate quite a loose tetrameric packing. This apparent contradiction probably reflects the possibility that the oligomeric state of 1mng depends on the experimental conditions such as temperature, pH, protein concentration, or ionic strength.

More generally, the ensemble of residues that we identified as ensuring the metal specificity and/or oligomeric state of SOD enzymes, summarized in Fig. 2*b*, can be used to define manganese/iron- and dimer/tetramer-specific fingerprints. If a given sequence presents some deviations from the typical fingerprints, it can be thought to adopt alternative oligomeric states or to have cambialistic tendencies. Large deviations can even be taken to indicate possible misannotations in the databases.

REFERENCES

1. McCord, J. M., and Fridovich, I. (1988) *Free Radic. Biol. Med.* **5,** 363–369
2. Fridovich, I. (1995) *Annu. Rev. Biochem.* **64,** 97–112
3. Klug-Roth, D., Fridovich, I., and Rabani, J. (1973) *J. Am. Chem. Soc.* **95,** 2786–2790
4. McAdam, M. E., Fox, R. A., Lavelle, F., and Fielden, E. M. (1977) *Biochem. J.* **165,** 71–79
5. Bull, C., and Fee, J. A. (1985) *J. Am. Chem. Soc.* **107,** 3295–3304
6. Bull, C., Niederhoffer, E. C., Yoshida, T., and Fee, J. A. (1991) *J. Am. Chem. Soc.* **113,** 4069–4076
7. Stallings, W. C., Metzger, A. L., Pattridge, K. A., Fee, J. A., and Ludwig, M. L. (1991) *Free Radic. Res. Commun.* **12–13,** 259–268
8. Bannister, J. V., Bannister, W. H., and Rotilio, G. (1987) *Crit. Rev. Biochem.* **22,** 111–180
9. Stallings, W. C., Pattridge, K. A., Strong, R. K., and Ludwig, M. L. (1984) *J. Biol. Chem.* **259,** 1095–10699
10. Jackson, S. M., and Cooper, J. B. (1998) *Biometals* **11,** 159–173
11. Parker, M. W., Blake, C. F., Barra, D., Bossa, F., Schinina, M. E., Bannister, W. H., and Bannister, J. V. (1987) *Protein Eng.* **1,** 393–397
12. Parker, M. W., and Blake C. F. (1988) *FEBS Lett.* **229,** 377–382
13. Brock, C. J., and Harris, J. L. (1977) *Biochem. Soc. Trans.* **5,** 1533–1539
14. Ose, D. E., and Fridovich I. (1979) *Arch. Biochem. Biophys.* **194,** 360–364
15. Beyer, W. F., and Fridovich, I. (1991) *J. Biol. Chem.* **266,** 303–308
16. Yamakura, F., Kobayashi, K., Ue, H., and Konno, M. (1995) *Eur. J. Biochem.* **227,** 700–706
17. Whittaker, M. M., and Whittaker, J. W. (1997) *Biochemistry* **36,** 8923–8931
18. Edwards, R. A., Whittaker, M. M., Whittaker, J. W., Jameson, G. B., and Baker, E. N. (1998) *J. Am. Chem. Soc.* **120,** 9684–9685
19. Vance, C. K., and Miller, A.-F. (1998) *J. Am. Chem. Soc.* **120,** 461–467
20. Hiraoka, B. Y., Yamakura, F., Sugio, S., and Nakayama, K. (2000) *Biochem. J.* **345,** 345–350
21. Vance, C. K., and Miller, A.-F. (2001) *Biochemistry* **40,** 13079–13087
22. Kardinahl, S., Schmidt, C. L., Petersen, A., and Schäfer, G. (1996) *FEMS Microbiol. Lett.* **138,** 65–70
23. Martin, M. E., Byers, B. R., Olson, M. O. J., Salin, M. L., Arceneaux, J. E. L., and Tolbert, C. (1986) *J. Biol. Chem.* **261,** 9361–9367
24. Yamakura, F., Kobayashi, K., Tagawa, S., Morita, A., Imai, T., Ohmori, D., and Matsumoto, T. (1995) *Biochem. Mol. Biol. Int.* **36,** 233–240
25. Santos, R., Bocquet, S., Puppo, A., and Touati, D. (1999) *J. Bacteriol.* **181,** 4509–4516
26. Chen, H.-Y., Hu, R.-G., Wang, B.-Z., Chen, W.-F., Liu, W.-Y., Schröder, W., Frank, P., and Ulbrich, N. (2002) *Arch. Biochem. Biophys.* **404,** 218–226
27. Tabares, L. C., Bittel, C., Carrillo, N., Bortolotti, A., and Cortez N. (2003) *J. Bacteriol.* **185,** 3223–3227
28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410
29. Felsenstein, J. (1995) *PHYLIP* (*Phylogeny Inference Package*) Version 3.57c, Dept. of Genetics, University of Washington, Seattle, WA.
30. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) *Comput. Appl. Biosci.* **8,** 275–282
31. Boutonnet, N. S., Rooman, M. J., Ochagavia, M. E., Richelle, J., and Wodak, S. J. (1995) *Protein Eng.* **8,** 647–662
32. Romesburg, H. C. (1984) *Cluster Analysis for Researchers*, Lifetime Learning Publications, Belmont, CA
33. Lanyon, S. M. (1985) *Syst. Zool.* **34,** 397–403
34. May, A. C. W. (1999) *Proteins* **37,** 20–29
35. May, A. C. W. (1999) *Structure Fold. Des.* **7,** R213
36. Page, R. D. M. (1996) *Comput. Appl. Sci.* **12,** 357–358
37. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637
38. McDonald, I. K., and Thornton, J. M. (1994) *J. Mol. Biol.* **238,** 777–793
39. Wintjens, R., Liévin, J., Rooman, M., and Buisine, E. (2000) *J. Mol. Biol.* **302,** 395–410
40. Hubbard, S. J. (1992) *NACCESS Program*, University College, London
41. Ridley, M. (1996) *Evolution*, 2nd Ed., Blackwell Science, Cambridge, MA
42. Lim, J. H., Yu, Y. G., Han, Y. S., Cho, S., Ahn, B. Y., Kim, S. H., and Cho, Y. (1997) *J. Mol. Biol.* **270,** 259–274
43. Whittaker, M. M., and Whittaker, J. W. (1998) *J. Biol. Chem.* **273,** 22188–22193
44. Edwards, R. A., Whittaker, M. M., Whittaker, J. W., Baker, E. N., and Jameson, G. B. (2001) *Biochemistry* **40,** 4622–4632
45. Lah, M. S., Dixon, M. M., Pattridge, K. A., Stallings, W. C., Fee, J. A., and Ludwig, M. L. (1995) *Biochemistry* **34,** 1646–1660
46. Wintjens, R. T., Rooman, M. J., and Wodak, S. J. (1996) *J. Mol. Biol.* **255,** 235–253
47. Wintjens, R., Wodak, S. J., and Rooman, M. (1998) *Protein Eng.* **11,** 505–522
48. Lévêque, V. J.-P., Stroupe, M. E., Lepock, J. R., Cabelli, D. E., Tainer, J. A., Nick, H. S., and Silverman, D. N. (2000) *Biochemistry* **39,** 7131–7137
49. Edwards, R. A., Whittaker, M. M., Whittaker, J. W., Baker, E. N., and Jameson, G. B. (2001) *Biochemistry* **40,** 15–27
50. Hunter, T., Bannister, J. V., and Hunter, G. J. (2002) *Eur. J. Biochem.* **269,** 5137–5148
51. Cabelli, D. E., Guan, Y., Lévêque, V., Hearn, A. S., Tainer, J. A., Nick, H. S., and Silverman, D. N. (1999) *Biochemistry* **38,** 11686–11692
52. Yamakura, F., Sugio, S., Hiraoka, B. Y., Ohmori, D., and Yokota, T. (2003) *Biochemistry* **42,** 10790–10799
53. Loewenthal, R., Sancho, J., and Fersht, A. R. (1992) *J. Mol. Biol.* **224,** 759–770
54. Sines, J., Allison, S., Wierzbicki, A., and McCammon, J. A. (1990) *J. Phys. Chem.* **94,** 959–961
55. Hunter, T., Ikebukuro, K., Bannister, W. H., Bannister, J. V., and Hunter, G. J. (1997) *Biochemistry* **36,** 4925–4933
56. Sato, S., and Harris, J. I. (1977) *Eur. J. Biochem.* **73,** 373–381
57. Sato, S., and Nakazawa, K. (1978) *J. Biochem.* **83,** 1165–1171
58. Stallings, W. C., Pattridge, K. A., Powers, T. B., Fee, J. A., and Ludwig, M. L. (1981) *J. Biol. Chem.* **256,** 5857–5859
59. Stallings, W. C., Pattridge, K. A., Strong, R. K., and Ludwig, M. L. (1985) *J. Biol. Chem.* **260,** 16424–16432
60. Ludwig, M. L., Metzger, A. L., Pattridge, K. A., and Stallings, W. C. (1991) *J. Mol. Biol.* **219,** 335–358
61. Whittaker, M. M., and Whittaker, J. W. (1999) *J. Biol. Chem.* **274,** 34751–34757