

The extended environment of mononuclear metal centers in protein structures

SAMUEL KARLIN*[†], ZHAN-YANG ZHU*, AND KENNETH D. KARLIN[‡]

*Department of Mathematics, Stanford University, Stanford, CA 94305-2125; and [‡]Department of Chemistry, Remsen Hall, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218

Contributed by Samuel Karlin, October 3, 1997

ABSTRACT The objectives of this and the following paper are to identify commonalities and disparities of the extended environment of mononuclear metal sites centering on Cu, Fe, Mn, and Zn. The extended environment of a metal site within a protein embodies at least three layers: the metal core, the ligand group, and the second shell, which is defined here to consist of all residues distant less than 3.5 Å from some ligand of the metal core. The ligands and second-shell residues can be characterized in terms of polarity, hydrophobicity, secondary structures, solvent accessibility, hydrogen-bonding interactions, and membership in statistically significant residue clusters of different kinds. Findings include the following: (i) Both histidine ligands of type I copper ions *exclusively* attach the N^{δ1} nitrogen of the histidine imidazole ring to the metal, whereas histidine ligands for *all* mononuclear iron ions and nearly all type II copper ions are ligated via the N^{ε2} nitrogen. By contrast, multinuclear copper centers are coordinated predominantly by histidine N^{ε2}, whereas diiron histidine contacts are predominantly N^{δ1}. Explanations in terms of steric differences between N^{δ1} and N^{ε2} are considered. (ii) Except for blue copper (type I), the second-shell composition favors polar residues. (iii) For blue copper, the second shell generally contains multiple methionine residues, which are elements of a statistically significant histidine–cysteine–methionine cluster. Almost half of the second shell of blue copper consists of solvent-accessible residues, putatively facilitating electron transfer. (iv) Mononuclear copper atoms are never found with acidic carboxylate ligands, whereas single Mn²⁺ ion ligands are predominantly acidic and the second shell tends to be mostly buried. (v) The extended environment of mononuclear Fe sites often is associated with histidine–tyrosine or histidine–acidic clusters.

Metals in protein structures serve structural, catalytic, and regulatory purposes ranging from electron transfer, substrate oxidation–reduction, transport processes, and active chemistry in regulation of enzymatic activity. For recent comprehensive reviews, see refs. 1 and 2. In this and the following paper, we define and analyze the extended metal environment centering on the ligands coupled to the “second shell.” The concept of the protein second shell of a metal center has been widely appreciated. For example, in certain proteins the second shell of metal sites is stated to be mainly hydrophobic (3, 4). Others distinguish *direct* vs. *indirect* ligands, the indirect ligands connected through hydrogen bonds to one of the direct ligands (5). Our analysis of the residues of the second shell emphasizes similarities and contrasts in residue polarity vs. hydrophobicity, the relative distribution of these residues into α -helices, β -strands, and coils, their degree of solvent accessibility, their hydrogen-bonding networks, and relationships to various sta-

tistically significant residue clusters. It is of interest to compare situations where it is clear that second-shell residues are “preorganizing” direct ligands via hydrogen-bonding and other specific interactions, thereby controlling metal-binding affinity, as well as situations in which the second shell appears to have little directing function on the direct ligands. Residues in the second shell may influence a variety of structural and chemical facets of the protein, including hydrogen-bonding interactions to the ligands, pK_a values of ligand groups, and the control of the metal center oxidation state and redox potential. In addition to influencing the polarity of the metal environment, neighboring side chains also exert steric and chemical controls over the ability of the metal ion to bind or discriminate substrate and to accommodate conformational changes.

The metal environment of the protein structure emphasizes four facets. (i) *Metal core*. This term refers to the metal type(s) in mononuclear or multinuclear sites (6).

(ii) *Ligand group*. Residues of the protein structure that directly bond to the metal(s). Foremost ligands of Cu, Fe, Mn, and Zn ions include an imidazole nitrogen(s) of histidine (H; one-letter code used), carboxylates (D/E), sulfur (C/M), solvent, and carbonyl oxygens. There are differences depending on the metal and function. Thus, Cu ions are never ligated by an acidic residue. Fe ions are predominantly ligated by multiple H residues in conjunction with one or two acidic residues and occasionally Y. The primary ligands of Mn²⁺ are often only acidic residues. The principal ligands coordinating Zn²⁺ ions consist of combinations of H, D, E, C, and sometimes Y, N, S, and T.

The two nitrogens in the histidine imidazole ring have the ability to donate ligand electrons to metals via either N^{ε2} or N^{δ1} (7). There are bridging bidentate residues that simultaneously coordinate two metal ions or a unibidentate with two bonds (e.g., O^{δ1} and O^{δ2} of aspartate or side chain plus backbone) to a single metal. Copper coordination rarely includes unibidentates, whereas iron coordination often involves one or more unibidentate residues or molecules. Coordination patterns and metal–ligand bond lengths reflect the flexibility of the coordination geometry, steric constraints, stability in the coordination process, alternative electrostatics, and redox potential.

(iii) *Second shell*. We define the second shell to consist of all residues and non-amino-acid molecules whose “distance” to any non-hydrogen atom among the ligand group is within 3.5 Å. The distance between a residue pair in the three-dimensional (3D) protein structure is calculated as the minimum distance between side-chain atoms or between side-chain and main-chain atoms (cf. ref. 8).

(iv) *Significant residue clusters*. Another useful descriptor of metal environments is in terms of a distinctive 3D residue cluster that may highlight the metal ligands and additional neighboring residues (9). Types of 3D residue clusters include the following: concentrations of residues with high net positive

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

© 1997 by The National Academy of Sciences 0027-8424/97/9414225-6\$2.00/0 PNAS is available online at <http://www.pnas.org>.

Abbreviation: 3D, three-dimensional.

[†]To whom reprint requests should be addressed. e-mail: fd.zgg@forsythe.stanford.edu.

charge (positive-charge clusters) designated {KR}, and analogously negative-charge clusters {ED}, or mixed-charge clusters {KRED}; 3D environments rich in histidine (histidine clusters) {H}, or cysteine–histidine {CH}, or histidine–acidic {HED}, or cysteine–histidine–methionine {CHM}. For other types of residue clusters, see refs. 9 and 10.

Methods for Identifying Significant Residue Clusters

Statistically significant 3D residue clusters are identified by appropriately representing a protein structure by a matrix of N (N the length of the primary sequence) linear amino acid sequences (each of length N) and evaluating these sequences for unusual residue sequence clusters (9). Specifically, from each residue a_k in a structure, a sequence S_k is generated with respect to a distance measure between residues (9, 10), as follows. Method 1 (M1): in the sequence $S_k = \{s_{k,1}, s_{k,2}, s_{k,3}, \dots\}$, $s_{k,1}$ is a_k , and the j th residues, $s_{k,j}$, of S_k ($j > 1$) is the next closest residue to any of the residues $s_{k,1}, s_{k,2}, \dots, s_{k,j-1}$. Thus, M1 accumulates residues in S_k , reflecting the shape and density of the structure surrounding the starting residue $s_{k,1}$. Method 2 (M2): Again $s_{k,1}$ is a_k and the j th residue $s_{k,j}$ ($j > 1$) is the closest in the 3D structure to the previous residue $s_{k,j-1}$ that is distinct from $s_{k,1}, s_{k,2}, \dots, s_{k,j-1}$. M2 in generating S_k proceeds along a pathway that often reveals channels relevant for transporting suitable molecules or substrates. Method 3 (M3) incorporates residues in the order of increasing distances from a_k . Using scoring theory (9), we ascertain whether or not a statistically significant high-scoring segment at 1% level occurs at the start of the sequence S_k generated from $s_{k,1}$. The test is applied to every sequence S_k of the protein structures, using all three methods. Two significant residue clusters are considered the same if the smaller cluster overlaps the larger cluster in at least 50% of its residues. For precision and elaborations, see ref. 9. Of special interest are distinctive residue clusters overlapping the metal environment. However, other residue clusters separated from the metal may contribute to protein–protein interactions, to substrate channeling, to enhancing stability and/or catalysis at the active site, and to quaternary structure formation (9, 10).

Data sources of the present study are described in the legend to Table 1.

Patterns and Contrasts of the Extended Environment for Mononuclear Copper, Iron, Manganese, and Zinc Sites

Table 1 describes the ligand, second-shell composition, and characteristics for representative Cu sites of type I (blue copper motif, electron transfer function) and type II; for general references, see refs. 6 and 11. Table 2 reviews examples of mononuclear Fe (nonheme) extended metal environments; for general references, see refs. 4, 6, and 12. The corresponding listings and descriptions for single Zn metal sites are presented in the following paper (13).

Tautomeric Histidine–Metal Preferences. Both histidine ligands of type I Cu ions adopt strictly $N^{\delta 1}$ nitrogen contacts, whereas histidine ligands for all mononuclear Fe ions and nearly all type II Cu ions adopt the $N^{\epsilon 2}$ tautomeric conformation (see Tables 1 and 2). By contrast, histidine diiron tautomeric contacts are predominantly $N^{\delta 1}$. In fact, from the five available diiron structures, methane monooxygenase (Protein Data Bank identification number 1mmo), ribonucleotide reductase (1rib), rubrerythrin, Δ^9 -desaturase (dsa), and hemerythrin (2hmz), the first two and dsa are coordinated by two H and four acidic residues, rubrerythrin is coordinated by one H and four E, and hemerythrin is coordinated with five H and two acidic bridging bidentates. Hemerythrin ligates via $N^{\epsilon 2}$ for all its five histidine ligands, whereas the other diiron associations ligate strictly via $N^{\delta 1}$. Histidine tautomeric metal bonding patterns for mononuclear zinc ions are variable (see ref. 13). The bonding of the trinuclear copper complexes of ceruloplasmin (hCP), ascorbate oxidase (1aoz), and the binuclear Cu ions of hemocyanin (1oxy) are ligated in all but one contact (in 1aoz) by histidines via the $N^{\epsilon 2}$ geometry (14, 15).

The foregoing contrasts in histidine–metal tautomeric conformations raise challenging questions. There may be electronic and/or chemical differences of $N^{\epsilon 2}$ vs. $N^{\delta 1}$ metal contacts—e.g., in the ability to deprotonate the N–H group. Yet, steric differences between $N^{\delta 1}$ and $N^{\epsilon 2}$ may be decisive. The $N^{\epsilon 2}$ –metal interaction has the histidine main-chain displaced radially away from the metal site, whereas for the $N^{\delta 1}$ contact the tautomeric geometry is displaced mostly in a tangential direction. In particular, two histidine $N^{\delta 1}$ ligation contacts would sandwich the metal ion between them and thereby provide a more stable metal–ligand coordination and proximity to the backbone and second-shell environment. This may be a favorable situation for an electron transfer center such as in type I Cu. On the other hand, $N^{\epsilon 2}$ ligand contacts afford more space to allow for metal–substrate catalytic interactions such as for O_2 transport, O_2 detoxification, or substrate oxidation occurring at mononuclear Fe and type II Cu sites. Zinc metalloproteins, including an abundance of hydrolases and transcription factors, do not engage in redox activity, and they manifest quite variable histidine tautomeric patterns (13). Notably, there are to date no examples of metalloproteins having three or more histidine ligands that involve more than one $N^{\delta 1}$ contact for any metal site. In particular, each iron site in diiron structures (except hemerythrin) tends to have a single histidine ligation invariably via $N^{\delta 1}$. The existence of a single $N^{\delta 1}$ contact may enhance stability of the metal–ligand complex and still leave space for other chemical interactions.

Ligand Group and Second Shell. The familiar type I Cu^{2+} ion ligands consist of H, C, H, M, occurring in that order in the primary sequence and the latter three ligands subscribing to the motif CX_mHX_nM , m and n mostly 2–4. At least three different structural motifs occur for the ligand set. The peptides that contain the latter C, H, and M ligands form a rather similar local 3D conformation in nitrite reductase (2afn, Cu-501), ascorbate oxidase (1aoz, Cu-701), and ceruloplasmin (hCP, Cu-33 and Cu-35) (2, 16), where the cysteine is at the C-cap of a β -strand, followed by a loop, and a short α -helix containing the second histidine and the methionine is in the turn following the helix (Table 1; consult secondary structure designation of ligands). A glycine putatively structurally important to form a tight turn is observed before the methionine in the primary sequences. The corresponding ligands in azurin, pseudoazurin, and plastocyanin are all contained in a single loop. However, in amicyanin the ligands are associated with two β -strands sandwiching the Cu ion.

For type I Cu^{2+} ions, the polar residues of the second shell include generally at least one N (asparagine), often part of an N–S or N–T pair, and one or more acidic residues. The hydrophobic part of the second shell generally involves multiple M residues and/or one or more large aromatic residues. The type I Cu center and its direct ligands are almost always totally buried, whereas the second shell contains about 1/3 solvent-exposed residues. [A residue is considered buried if its side-chain solvent accessibility is $\leq 10\%$ (17).] The exposed residues may contribute to the electron transfer function by providing pathways to an exogenous physiological electron transfer acceptor/donor. The blue motif second-shell residues are mostly in coil elements. The solvent-accessible coil residues suggest an intrinsic relationship between the type I Cu metal environment and the protein surface.

Proteins with type II Cu sites include galactose oxidase and two copper amine oxidase structures: 1oac from *Escherichia coli* and 1amo from pea seedling. An essential organic molecule in these copper amine oxidases is the modified tyrosine cofactor TPQ (2,4,5-trihydroxyphenylalanine quinone), and on this basis these proteins are quinoproteins. The copper ion in chain A is tightly coordinated by H-524, H-526, H-689, and the TPQ residue. Strikingly, the second shell, practically all buried, contains no hydrophobic residues but features eight polar (including two acidic) residues and four water molecules.

Table 1. Single copper ions in protein 3D structures

Ligands		Bond Length & Atom	Second Shell: residues, hydrophobic Φ vs hydrophilic Π , secondary structure, side-chain solvent accessibility	Significant 3D Residue Clusters
1. 2afn: Nitrite Reductase (<i>Alcaligenes faecalis</i>): Homo trimer, 331 aa each chain (A, B & C); protein structural class: all β				
Cu -501	H-95	2.22; N _{δ1}	E-47, N-87, L-93;	Φ =(1F, 1V, 1L, 3M) Π =(1E, 1N, 1H) others=(1P)
	C-136	2.08; S _γ	P-138, V-146;	
	H-145	2.07; N _{δ1}	M-94, M-141;	{CHM} cluster about Cu-501; {HED} cluster traverses both Cu-501 and Cu-502 environments
	M-150	2.59; S _δ	M-62, F-64, H-135;	
Cu -502	H-100	2.01; N _{δ2}	V-133, G-286, H-B255, H-B260, E-B279	Φ =(1I, 2V, 1M, 1A) Π =(1R, 1D, 1E, 2H) Others={2G}
	H-135	2.08; N _{δ2}	D-98, V-146, M-150	
	H-B306	2.13; N _{δ2}	A-248, G-285, G-286, R-B253, H-B255	1 α , 4 β , 7 coil 12 buried, 0 exposed
	MTQ-503	2.09; O	D-98, I-B257; Residues in Chain A if not specified	
2. 2aza: Azurin (<i>Alcaligenes denitrificans</i>): Homo dimer 129 aa each chain (A & B); all β				
Cu A130	H-46	2.08; N _{δ1}	S-9, N-10, H-35, I-87;	Φ =(1I, 2F, 1M), Π =(2N, 1S, 1M, 1 [†] Y), Others=(1G, 1P, 1Wat)
	C-112	2.12; S _γ	Y-15, N-47;	
	H-117	2.01; N _{δ1}	G-45, F-114, P-115, Wat-137;	{CHM} cluster covers copper center in each chain
	M-121	3.12; S _δ	M-13, F-111;	
3. 1paz: Pseudoazurin (<i>Alcaligenes faecalis</i>): Single chain, 123aa; all β				
Cu 124	H-40	2.20; N _{δ1}	N-9, D-37, K-38, Wat-170, Wat-176;	Φ =(1V, 1M), Π =(1D, 1K, 1N); Others=(3Wat);
	C-78	2.14; S _γ	Wat-148	
	H-81	2.27; N _{δ1}	M-16, V-17;	{CHM} cluster about the copper
	M-86	2.66; S _δ		
4. 1plc: Plastocyanin (Poplar Leaves): Single chain, 99aa; all β				
Cu 100	H-37	1.91; N _{δ1}	L-5, N-31, A-33, G-34, F-35;	Φ =(2L, 2F, 2A); Π =(2N); Others=(1G, 1P);
	C-84	2.07; S _γ	N-38;	
	H-87	2.06; N _{δ1}	G-34, P-60, A-90;	1 α , 3 β & 7 coil 4 buried & 6 exposed
	M-92	2.82; S _δ	L-12, F-14;	
5. 2mta: Methylamine Dehydrogenase (<i>Paracoccus denitrificans</i>) (chain H - 373aa and L - 125aa) complexed with Amicyanin (chain A - 105aa) & Cytochrome C5511 (chain C - 147aa); The copper center is contained in Amicyanin (Chain A, all β).				
Cu -A0	H-53	2.22; N _{δ1}	N-47, E-49, M-51;	Φ =(2M), Π =(2E, 2N, 1H, 1 [†] Y) Others=(1Wat)
	C-92	2.35; S _γ	Y-30, N-54;	
	H-95	2.29; N _{δ1}	E-101, Wat-56;	1 α , 3 β & 5 coil 5 buried & 3 exposed
	M-98	2.98; S _δ	M-28, Y-30, H-91;	
6. 1aoz: Ascorbate Oxidase (Zucchini): Homo dimer, 552aa each chain (A & B); all β				
Cu A701	H-445	2.09; N _{δ1}	W-362, Wat-80, Wat-83;	Φ =(I, W); Π =(N); Others=(3Wat)
	C-507	2.13; S _γ	I-509;	
	H-512	2.05; N _{δ1}	I-509, Wat-18, Wat-83;	{HED} cluster extends over Cu-A701 and the trinuclear copper core
	M-517	2.90; S _δ	N-351;	
7. Ceruloplasmin (human, PDB code not assigned): Single chain, 1022aa; α + β (mainly β)				
Cu 33	H-975	2.09; N _{δ1}	D-973, I-999;	Φ =(1L, 3I, 1V, 2A) Π =(1E, 2D, 1T, 1H)
	C-1021	2.05; S _γ	T-976, V-1023, I-1027;	
	H-1026	2.06; N _{δ1}	E-272, I-972, L-974, V-1023, D-1025, A-1029;	3 α , 4 β & 5 coil 7 buried & 5 exposed
	M-1031	3.20; S _γ	H-940, A-941;	
Cu 34	H-276	2.10; N _{δ1}	D-274, N-296, L-297;	Φ =(1L, 2V, 1A, 1Y), Π =(1D, 2N), no others
	C-319	2.03; S _γ	A-277;	
	H-324	2.05; N _{δ1}	Y-241, V-273, V-275, N-321;	8 buried & 4 exposed
Cu 35	H-637	2.08; N _{δ1}	F-563, S-629, D-635, N-657, L-658, T-682;	Φ =(1L, 1M, 1A, 1Y) Π =(1E, 1D, 1N, 2T, 1S) Others=(1P)
	C-680	2.02; S _γ	T-682, Y-686;	
	H-685	2.06; N _{δ1}	A-634, T-682, E-971;	1 α , 5 β & 5 coil 7 buried & 4 exposed
	M-690	3.07; S _γ	T-564, M-604;	
8. 1gof: Galactose Oxidase (<i>Dactylium dendroides</i>): Single chain, 639aa; all β				
Cu 700	Y-272	1.93; O _H	C-228, W-290, N-333, H-334, Y-405;	Φ =(1C, 1W, 2F) Π =(1N, 1H, 1T, 1 [†] Y) Others=(1G, 5Wat)
	Y-495	2.69; O _H	Y-405, Wat-(51, 103, 108, 294);	
	H-496	2.11; N _{δ2}	G-513;	1 α , 1 β & 8 coil 7 buried & 2 exposed
	H-581	2.14; N _{δ2}	H-334, T-580, Wat-22;	
	Acetate Ion	2.26; Oxt	F-194, W-290, F-464, Wat-294;	
9. 1oac: Copper Amine Oxidase (<i>E. coli</i>): Homo dimer, 719aa each chain (A & B); mainly β				
Cu 801	H-524	2.03; N _{δ2}	P-368, T-522, Wat-112, Wat-159;	Φ =(none) Π =(1E, 1D, 1N, 1H, 1T, 1 [†] Y) Others=(1P, 4Wat)
	H-680	1.97; N _{δ2}	Y-468, Wat-184;	
	H-689	2.01; N _{δ1}	E-490, H-682, E-695;	{H} & {HY} clusters about the copper
	TPQ-466	2.05; O ₄	T-462, N-465, D-467, Y-468, T-487, Wat-(112, 326) (TPQ: 2,4,5-trihydroxyphenylalanine)	
10. 1amo: Amine Oxidase (Pea Seedling): Homo dimer, 642aa each chain; mainly β				
Cu 650	H-442	2.20; N _{δ2}	P-285, I-440, Wat-C17	Φ =(1I), Π =(1D) Others=(2P, 2Wat)
	H-444	2.09; N _{δ2}	Wat-C162;	
	H-603	2.16; N _{δ1}	P-605, D-609;	{KRED} cluster distant from the copper center; about a Mn metal site
	Wat-651	2.51; O		
Wat-652	1.98; O			

Copper centers 1–7 are classified type I (blue copper); copper centers 8–10 are type II. A ligand residue or a residue in the second shell is shown by the one-letter amino acid code followed by chain identifier and primary sequence position (Protein Data Bank residue number). A chain identifier of a residue is not designated unless necessary. Bond lengths are in Å. Residues are underlined when the residue is in an α -helix, in italic when the residue is in a β -strand, or in ordinary font when the residue is in a coil location. Bold letters indicate that the residue side-chain atoms are buried (side-chain solvent accessibility is less than 10%). A residue in second shells is designated by the symbol * when one of its side-chain atoms forms a hydrogen bond with a side-chain atom of a ligand; the symbol † when one of its main-chain atoms forms a hydrogen bond with a side-chain atom of a ligand; the symbol ‡ when one of its side-chain atoms forms a hydrogen bond with a main-chain atom of ligand. Exclusive main chain and main chain hydrogen bonds are not indicated. *Data set*: a representative set of protein structures was based on the list of Hobohm and Sander (22), version of December, 1996, with pairwise primary sequence identity less than 25%. Total number of proteins in the list is 443 and those with metal, heme, or iron-sulfur linkage is 129 (29%). The structure data set was augmented with several recent protein structures known to contain metal centers.

Table 2. Single iron ions in protein 3D structures

Ligands	Bond Length & Atom	Second Shell: residues, hydrophobic Φ vs hydrophilic Π , secondary structure, side-chain solvent accessibility	Significant 3D Residue Clusters	
1. Iisc (Fe-A193): Iron Superoxide Dismutase (<i>E. coli</i>); Homodimer, 192aa each chain; protein structural class: all α				
H-26 H-73 D-156 H-160 Wat-194 Azide-195	2.20; N ₂ 2.13; N ₂ 2.04; O ₂ 2.15; N ₂ 2.02; O 2.09; N ₁ 2.81; N ₂	T-220, H-30, H-31, Wat-439; Q-69, Y-76, W-77, W-122; W-122, W-158, A-161, Y-173; W-158, E-159 (B chain); H-30, H-31, Y-34;	Φ =(1A, 3W) Π =(1E, 1Q, 2H, 1T, 3Y) Others=(1Wat) 8 α , 1 β & 3 coil 10 buried & 2 exposed	{H} cluster about the iron center {HY} cluster overlaps the iron center in chain A and involves second shell residues of the corresponding iron in chain B
2. Iao (Fe-1): Aldehyde Ferredoxin Oxidoreductase (<i>Pyrococcus furiosus</i>); two identical chains (A&B), 605aa each chain, Fe-1 at the chain interface; α+β				
E-A332 E-B332 H-A383 H-B383	2.01; O ₁ 2.71; O ₂ 2.05; O ₁ 2.72; O ₂ 2.07; N ₂ 2.05; N ₂	T-A303, A-A328, S-A329, H-A335, Wat-23; T-B303, A-B328, H-B335, Wat-24; S-A329, T-A379, P-B302, D-B326, A-B328, S-B329, Wat-23; D-A326, A-A328, S-A329, S-B329, T-B379;	Φ =(2A); Π =(2D, 2H, 4T, 2S) Others=(1P, 2Wat) 6 α , 0 β & 7 coil 11 buried & 2 exposed	{C} cluster about FS4 (iron/sulfur linkage) {HY} & {KRED} clusters distant from the iron center
3. Ict (Fe-400): Lactoferrin (N-terminal half molecule) (human); Single chain, 324aa, α/β				
D-60 Y-92 Y-192 H-253 Bicarbonate-401	1.99; O ₁ 2.00; O ₁ 1.92; O ₁ 2.15; N ₂ 2.23; O ₁ 2.09; O ₂	G-62, R-121, T-122, Wat-504, Wat-517; R-210, E-211, V-250, P-251, Wat-517; R-121, F-196, R-210, T-213; L-59, E-80, Y-82, Wat-516; T-117, R-121, A-123, G-124;	Φ =(1A, 1V, 1L, 1F) Π =(2E, 2R, 3T, 1Y) Others=(2G, 1P, 3Wat) 3 α , 5 β & 7 coil 12 buried & 3 exposed	{HY} cluster about the iron {CHY} cluster distant from the iron
4. Iprc (Fe-607, at L & M interface) Photosynthetic Reaction Center (<i>Rhodospseudomonas viridis</i>); Four chains: C: 333aa, L: 273aa, M: 323aa, H: 258aa; all α				
To Fe-607: H-L190 H-L230 H-M217 E-M232 H-M264	1.99; N ₂ 2.41; N ₂ 2.10; N ₂ 2.17; O ₁ 2.02; O ₂ 2.04; N ₂	A-L186, I-L194, UQ1 (Ubiquinone-1)-614; A-L226, L-L227, I-L229, Wat-525; L-M213, I-M221, MQ7 (Menaquinone-7)-608 I-L194, Q-M235, I-M236, Wat-525; I-M260, E-M261, W-M268;	Φ =(5I, 2L, 1W, 2A) Π =(1E, 1Q) Others=(UQ1, MQ7, 1Wat) 12 α , 0 β & 0 coil 10 buried & 2 exposed	{HED} cluster about the iron at the interface of chains H, M and L {KR} a rare positive charge cluster, {KRED} cluster, see Zhu & Karlin (1996) {CHM} clusters about heme
5. Iyge (Fe-840): Lipoxigenase-1 (Soybean); Single chain, 839aa; α+β				
H-499 H-504 H-690 N-694 I-839 Wat-842	2.23; N ₂ 2.26; N ₂ 2.21; N ₂ 3.05; O ₁ 2.16; Oxt 2.56; O	H-494, T-503; W-500, T-503, I-538, N-539, I-837; T-503, I-685, Wat-1059, Wat-1085; F-695, Q-697, L-754, Wat-852 L-754	Φ =(2I, 1F, 1L, 1W) Π =(1Q, 1N, 1H, 1T) Others=(3Wat) 5 α , 0 β , 4 coil all buried	{KRED} & {FYW} clusters distant from the iron center, see Figure 1.
6. 2pdc (Fe-M600): Protocatechuate 3,4-Dioxygenase (<i>Pseudomonas aeruginosa</i>); Chain A: 200aa, Chain M: 233aa; α+β (mainly β)				
Y-M408 Y-M447 H-M460 H-M462 Wat-M827	1.81; O _H 1.90; O _H 2.26; N ₂ 2.33; N ₂ 1.91; O	Y-A16, Q-M401, D-M413, Wat-M643 G-M406, P-M448, Wat-M638 P-M444, R-M457, P-M458, Q-M477, L-M478 M-M399, W-M400, Wat-M605, Wat-M606	Φ =(1L, 1W, 1M, 1Y) Π =(1R, 1D, 2Q) Others=(1G, 3P, 4Wat) 0 α , 6 β & 6 coil 10 buried & 2 exposed	{HY} cluster about the iron environment
7. Ihan (Fe-500): 2,3-Dihydroxybiphenyl 1,2-Dioxygenase (<i>Pseudomonas cepacia</i>); Single chain, 288aa; α+β (mainly β)				
H-146 H-210 E-260 Wat-3001 Wat-3012	2.15; N ₂ 2.25; N ₂ 1.96; O ₁ 2.41; O 2.12; O	H-195, T-196, M-212, Wat-1017; F-147, V-148, H-209, E-258; M-212, H-241, S-248, Y-250; H-241, D-244; H-195;	Φ =(1M, 1F, 1V) Π =(1E, 1D, 3H, 1T, 1S, 1Y) Others=(1Wat) 0 α , 8 β & 3 coil 9 buried & 2 exposed	{HED} cluster overlaps the iron environment

See legend to Table 1.

The two acidic residues, E-490 and E-695, are hydrogen-bonded to the ligand H-689, and Y-468 is hydrogen-bonded to ligand H-526. E-695 and H-689 interact, connecting the side-chain oxygen atom O^{e1} of E-695 with N^{e2} of H-689 at the distance about 2.7 Å. Iamo is a dimer homologue of Iao, each unit containing a single Cu²⁺ ion and a separate Mn²⁺ ion not present in Iao. Each chain of copper amine-oxidase (Iao) contains two separated Ca²⁺ ions. The ligands of Ca-A802 are D-533, D-535, D-678, L-534, A-679 (L and A coordinating the metal through the carbonyl oxygen), and one water molecule. The second shell consists of Φ = {2Y}, Π = {2K, R, E, N, *W}, one glycine, and two water molecules. K-133 connects with the ligand D-533 by a salt bridge involving the side-chain nitrogen N^s of K and the side-chain oxygen O^{e2} of D-533 at 2.68 Å distance. The residue W-137 of the second shell hydrogen-bonds with D-535, involving the side-chain atoms (imino nitrogen) N^{e1} of W and O^{e2} of D-535 (2.69 Å apart). The direct environment of Ca-A803 consists of the coordinating residues E-573, Y-677 (carbonyl ligation), D-670, E-672, and two water molecules; second shell Φ = {F}, Π = {R, E, 2N, T, H}, {1G}, 3 water.

Galactose oxidase (Igo) functions in catalysis of the ste-

reospecific oxidation of a broad range of primary alcohol substrates. The active site is the mononuclear copper cofactor. Apart from two H, the other ligands are two Y residues and an exogenous acetate ion presumed to be the site of substrate binding (11). Eight of nine residues of the second shell are in coils and seven are buried. The nearby C-228 forms a covalent link to a ring carbon of Cu-ligand Y-272. This Y-C moiety constitutes an active-site cofactor that stores a protein oxidizing equivalent critical for catalysis. The second-shell W-290 (hydrogen-bonded to Y-272) may also serve an important role via its π -stacking with Y-272 and C-228 (11). Galactose oxidase also contains a separated sodium (Na-702) ion (ligands: K-29, D-32, N-34, T-37, A-141, E-142 mostly exposed, with the bonding atom for these residues generally the carbonyl oxygen). The second shell consists of seven exposed and two buried residues summarized by Φ = {I, F, L}, Π = {R, D, E}, others {2G, 1P} with its polar residues all charged and most of its residues located in coil elements.

Iron placements occur at the interface of heterodimer structures, aldehyde ferredoxin oxidoreductase (Iao) and the photosynthetic reaction center (Iprc), with ligands and sec-

ond-shell residues contributed from both chains. Lactoferrin (1lct), a member of the transferrin family, which transports iron, and the two dioxygenases 1han and 2pcd are regarded to have multiple domains where the iron placement is near the interface of these domains (cf. refs.4, 12, and 18). Iron-ligand centers tend to be buried, but access to solvent may be vital for substrate binding and catalytic processes. A metal center proximal to an interface may provide flexibility in this respect. The structure of 1aor possesses two unidentates, E-332A and E-332B, bonded to the iron atom at the interface. Iron superoxide dismutase (1isc) has an exogenous azide bonding via two nitrogen atoms. E-232 of chain M of 1prc is a unidentate to the Fe ion. In 1lct the bicarbonate-401 molecule bonds to Fe-400 via two oxygen molecules.

Polar residues predominate in the second shell of mononuclear Fe sites. In particular, the second shell contains one or more acidic residues (exception, 1yge). Multiple histidine residues occupy the second shell of 1aor, 1han, and 1isc. A tyrosine hydrogen bonding to an acidic ligand occurs in structures 1han, 2pcd, and 1isc. The second shell of iron superoxide dismutase (1isc) involves about equally "hydrophobic" and polar residues. The hydrophobic component highlights five aromatic residues, three W and two Y of mixed hydrophobic/hydrophilic character but no F residue. Y-173 hydrogen bonds via its hydroxyl group with the carbonyl oxygen of the ligand D-156, and W-158 hydrogen bonds through its main-chain nitrogen atom with the carboxylate of D-156. Aldehyde ferredoxin oxidoreductase (1aor) has two chains, A and B, with a single iron atom at the interface. The second shell of the Fe atom is predominantly polar but strikingly buried.

There are available six examples with a mononuclear manganese metal center (19); for details, see <http://gnomic.stanford.edu/~zhu>. The ligands of both Mn proteins (2chr, 370 aa, and 2mnr, 357 aa) are purely acidic, exhibiting virtually equal spacings, DX₂₅EX₂₄D and DX₂₅EX₂₅E, respectively. The second shell enclosing the acidic residues have the same hydrophobic residues $\Phi = \{V, W, M\}$ and many common polar residues ($\Pi = \{3K, D, 2E, N, H, S\}$ in 2mnr and $\Pi = \{3K, E, N, Q\}$ in 2chr). The three lysines suggest that these may help balance the excess net negative charge about the metal–ligand core.

It is interesting to compare the metal environment of the *E. coli* Fe-sod (1isc), that of the *T. thermophilus* Mn-sod (1mng), and that of human Mn-sod (1amb). All are homodimers of similar sizes with identical ligand groups consisting of 3H, D, and one bound water. Their secondary structure dispositions are almost identical. The complete primary sequences align in the range of (40–50%) identity. The human Mn-sod (1abm) and *E. coli* Fe-sod (1isc) significantly differ in the polar composition of the second shell (see Web site). On the other hand, the extended Mn environments of 1isc and 1mng are very similar.

Numbers of Hydrophobic (Φ) vs. Polar (Π) Residues of the Second Shell and Numbers of the Different Types of Closest Atoms. For each mononuclear metal type, the aggregate residue counts in Φ and Π [glycine (G), proline (P), and water counted separately] are given, and the counts of the closest atoms from second shell residues to ligands [carbon (C), nitrogen (N), oxygen (O), sulfur (S)] are cumulated in Table 3.

These counts (except for type I Cu) show that the second shell in aggregate favors polar residues, the closest atoms are predominantly oxygen, and in total there are many more O, N,

and water contacts than C contacts. The designations in Tables 1 and 2 (see the legend of Table 1) indicate possible hydrogen-bond connections between side-chain atoms and between main-chain and side-chain atoms. The foregoing corroborates the general prevalence of polarity over hydrophobicity of the second shell.

Statistically Significant 3D Residue Clusters in Metalloprotein Structures. Almost all the blue copper motifs are embedded in {CHM} clusters usually carrying at least two methionine and at least three histidine residues. The homotrimer nitrite reductase (2afn) features a {HED} cluster about the type I copper Cu-A501, consisting of (the notation H-A145, for example, signifies histidine at position 145 in chain A) E-A47, H-A95, Cu-A501, H-A145, C-A136, M-A150, I-A114, H-A135, Cu-A502, H-A100, MTO, H-B306, R-B253, P-B254, M-B260, DA-251, D-A98, E-B279, and H-B255 encompassing all the ligands (double underlined) and residues of the second shell (underlined); MTO is a water molecule bound to Cu-A502. Strikingly, the cluster {HED} simultaneously envelops in chain A the blue copper Cu-501 direct ligands together with the direct ligands of Cu-502 of type II, plus some residues of the second shell and other residues (E and D are not ligands). Thus, these two copper ions interrelate and putatively function cooperatively, allowing electron transfer from Cu-501 to Cu-502, where the nitrite substrate binding and reduction occurs (16).

Azurin (2aza), thought to transfer electrons from cytochrome *c*-551 to cytochrome oxidase, contains a highly significant {CHM} cluster embodying residues of both chains (from chain B: six M, three H, two C, two Y, three G, P, F, S, and D and from chain A: four M, three H, C, S, and F) and covering the Cu sites of both chains. The large number of methionine residues and aromatic residues in the cluster with many glycine fillers is striking. Azurin is of all β -structure class, yet most of the second shell are residues of coil elements. Pseudoazurin (2paz) and plastocyanin (1plc) contain a significant {CHM} cluster enveloping the blue copper ion. Among type II copper sites galactose oxidase contains a highly significant {CHY} cluster about the copper ion.

The structure (1aor) contains an iron–sulfur (FeS₄) linkage ligated by four cysteines and further possesses a {HY} cluster distant from the Fe site. In this structure, there is a significant mixed charge cluster {KRED} with residues contributed from chains A and B. This mixed-charge cluster is part of the cleft formed by chains A and B, however, with no overlap to the iron environment. Lactoferrin (1lct) possesses a {HY} cluster about the Fe-400. The protein also possesses a cysteine {C} cluster distant from the iron center juxtaposing three disulfide bridges: C-157—C-173, Y-189, P-159, L-172, C-198—C-115, E-187, and C-181—C-170. This may have a role in nucleating the protein conformation and in structural stability (9). The photosynthetic reaction center (1prc) features a {HED} cluster about the iron site. Distant from the Fe center of lipoxigenase (1yge), there is a significant exposed mixed charge cluster {KRED} (Fig. 1). Is it possible that this mixed-charge cluster is a region of protein–protein interaction or a help in mediating the movement of the lipophilic substrate toward the active metal center? An aromatic cluster {FWY} (Fig. 1) may facilitate formation of a hydrophobic cavity or channel for moving the lipid substrate to the active site (6, 20). The

Table 3. Numbers of residues and closest atoms in the metal environment

Metal	Aggregate no. of			Aggregate residue counts					No. of closest atoms				
	Structures	Ligands	2nd-shell residues	Φ	Π	G	P	HOH	C	N	O	HOH	S
Type I Cu	9	35	86	38	34	2	4	8	25	7	52	9	0
Type II Cu	4	18	47	10	19	3	3	12	14	3	24	14	1
Single Fe	7	37	101	30	46	3	5	15	43	17	48	18	0
Single Mn	7	38	79	18	43	3	1	14	27	28	31	15	0

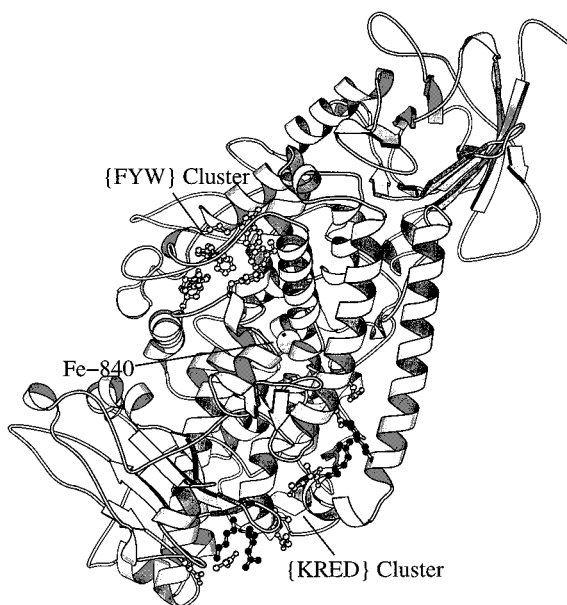


FIG. 1. Lipooxygenase-1 contains an aromatic {FYW} cluster and a mixed charge {KRED} cluster. Both clusters are distant from the iron center.

dioxygenase structures 2pcd and 1han feature an {HY} and {HED} cluster, respectively, near the iron site.

Despite the similar metal-ligand environments in {2chr} and {2mnr}, these manganese proteins contrast sharply away from the Mn site. For example, 2chr carries two distinct exposed mixed-charge clusters. Both {KRED} clusters intriguingly consist of exactly the same charged residues two R, one K, two E, and three D in separate regions of the protein surface. By contrast, 2mnr possesses no significant residue cluster of any kind.

Dispelling a Dogma. There is a commonly held belief that the second shell of metal binding sites is predominantly hydrophobic (e.g., see refs. 3 and 4). However, our analysis shows that the second-shell composition depends on the metal type and the metal multiplicity. A hydrophobic shell putatively inhibits solvent ions and/or water from reaching the active site while accommodating nonpolar groups, whereas a channel coated with polar residues would facilitate movement of ions and/or polar substrates to or from a metal active site. A hydrophobic second shell apparently offers stronger contrasts in the immediate vicinity of the metal by causing changes in electric polarizability and by allowing tuning of the metal center redox potential and electron transfer rate. The detailed structure is important. In some cases polar molecules could concentrate at the surface, but aliphatic portions of these same residues could contribute to a hydrophobic shielding environment. For type I Cu^{2+} ion environments, we find that nitrite reductase and plastocyanin show a hydrophobic prevalence in the second shell but three others do not. Among the type II copper centers, only the second shell of nitrite reductase (2afn) (Cu-502) at the interface is predominantly hydrophobic. Strikingly, the second shell of the *E. coli* copper amine oxidase (1oac), although completely buried, is devoid of hydrophobic residues. Of single-iron centers only the photosynthetic reaction center (1prc) has a large majority of hydrophobic residues in the second shell. The mononuclear iron placements of 1aor and of 1prc are both at an interface of two protein units, but these second shells drastically differ in the degree of polarity. More than 80% of zinc-ligand surroundings, including carbonic anhydrase II (2caa), enolase (4enl), and thermolysin (8tn), are

predominantly hydrophilic. They are often associated with histidine-acidic clusters and charge residues (13). The hydrophobic component of the second shell of Cu,Zn superoxide dismutase (2sod) is virtually all aromatic, highlighting Y and W residues (data not shown). The second shells of the diiron carboxylate protein structures methane monooxygenase (1mmo), R2 ribonucleotide reductase (1rib), rubrerythrin, Δ^9 -desaturase, and hemerythrin (2hmz) (cf. ref. 21) are very different. Specifically, the second shell of 1mmo is predominantly polar, that of 1dsa is modestly polar, that of 1rib and 1rub are equally polar vs. hydrophobic, and that of 2hmz carries a preponderance of hydrophobic residues (data not shown).

Conclusion

Extensive comparisons of metalloprotein structures from analysis of x-ray data, especially among metal environments, yield new observations and insights. In particular, we ascertained statistically significant residue clusters that envelop the ligand group and identify other residue clusters separated from the metal. Commonalities and contrasts of tautomeric histidine-metal preferences and composition of the second shell are dependent on metal and functional types. Further analysis and experimental approaches should aid the understanding of how these newly observed 3D elements contribute to protein function.

We convey our appreciation for valuable comments on the manuscript by Drs. B. E. Blaisdell, J. Griffin, E. Solomon, and W. Weiss from Stanford University. S.K. was supported in part by National Institutes of Health Grants 5R01GM10452-33 and 5R01HG00335-09 and National Science Foundation Grant DMS9403553-002. K.D.K. was supported in part by National Institutes of Health Grant GM28962.

1. (1991) *Adv. Protein Chem.* **42**.
2. Holm, R. H., Solomon, E. I., eds. (1996) *Chemical Reviews: Bioinorganic Enzymology* (1996), Nov. issue.
3. Yamashita, M. M., Wesson, L., Eisenman, G. & Eisenberg, D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5648–5652.
4. Howard, J. B. & Rees, D. C. (1991) *Adv. Protein Chem.* **42**, 199–273.
5. Christianson, D. W. & Fierke, C. A. (1996) *Acc. Chem. Res.* **29**, 331–339.
6. Holm, R. H., Kennepohl, P. & Solomon, E. I. (1996) *Chem. Rev.* **96**, 2239–2314.
7. Chakrabarti, P. (1990) *Protein Eng.* **4**, 57–63.
8. Karlin, S., Zuker, M. & Brocchieri, L. (1994) *J. Mol. Biol.* **239**, 227–248.
9. Karlin, S. & Zhu, Z.-Y. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8344–8349.
10. Zhu, Z.-Y. & Karlin, S. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8350–8355.
11. Klinman, J. (1996) *Chem. Rev.* **96**, 2541–2561.
12. Que, L., Jr., & Ho, R. Y. N. (1996) *Chem. Rev.* **96**, 2607–2624.
13. Karlin, S. & Zhu, Z.-Y. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 14231–14236.
14. Messerschmidt, A., Ladenstein, R., Huber, R., Bolognesi, M., Avigliano, L., Petruzzelli, R., Rossi, A. & Finazzi-Agro, A. (1992) *J. Mol. Biol.* **224**, 179–205.
15. Messerschmidt, A. (1993) *Adv. Inorg. Chem.* **40**, 121–185.
16. Murphy, M. E. P., Turley, S., Kukimoto, M., Nishiyama, M., Horinouchi, S., Sasaki, H., Tanokura, M. & Adman, E. T. (1995) *Biochemistry* **34**, 12107–12117.
17. Richmond, T. J. & Richards, F. M. (1978) *J. Mol. Biol.* **119**, 537–555.
18. Han, S., Eltis, L. D., Timmis, K. N., Muchmore, S. W. & Bolin, J. T. (1995) *Science* **270**, 976–980.
19. Dismukes, G. C. (1996) *Chem. Rev.* **96**, 2909–2926.
20. Minor, W., Steczko, J., Stec, B., Otwinowski, Z., Bolin, J. T., Walter, R. & Axelrod, B. (1996) *Biochemistry* **35**, 10687–10701.
21. Kurtz, D. M., Jr. (1997) *J. Biol. Inorg. Chem.* **2**, 159–167.
22. Hobohm, U. & Sander, C. (1994) *Protein Sci.* **3**, 522–524.