Extended Metal Environments of Cytochrome *c* Oxidase Structures†

Samuel Karlin,*,‡ Zhan-Yang Zhu,§ and Kenneth D. Karlin||

Department of Mathematics, Stanford University, Stanford, California 94305-2125, and Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218

Received June 12, 1998; Revised Manuscript Received October 12, 1998

ABSTRACT: The metals of the cytochrome *c* oxidase structures of the bovine heart mitochondrion (PDB code 1occ) and of the soil bacterium *Paracoccus denitrificans* (1ar1) include a dicopper center (Cu_A), magnesium, two proximal hemes, a copper (Cu_B) atom, and a calcium. The mitochondrial structure also possesses a bound distant zinc ion. The extended environments of the metal sites are analyzed emphasizing residues of the second shell in terms of polarity, hydrophobicity, secondary structure, solvent accessibility, and H-bonding networks. A significant difference in the Cu_A metal environments concerns D-51 I in 1occ, absent from 1ar1. The D-51 I appears to play an important role in the proton pumping pathway. Our analysis uncovers several statistically significant residue clusters, including a cysteine-histidine-tyrosine cluster overlapping the Cu_A−Mg complex; a histidine-acidic cluster enveloping the environment of Mg, the two hemes, and Cu_B; and on the protein surface a mixed charge cluster, which may help stabilize the quaternary structure and/or mediate docking to cytochrome *c*. These clusters may constitute possible pathways for electron transfer, for O_2 diffusion, and for H_2O movement. Many hydrogen bonding relations along the interface of subunits I and II demarcate this surface as a potential participant in proton pumping.

The extended environment of single or multinuclear metal centers within a protein embodies at least three layers: the metal core, the ligand group, and the "second shell" (second coordination sphere) which is defined to consist of all residues that have some component atom distant less than 3.5 Å from some ligand atom (*1*, *2*). The second-shell residues can be analyzed in terms of polarity, hydrophobicity, secondary structures, solvent accessibility, H-bonding networks, and the presence of statistically significant three-dimensional (3D) residue clusters. In protein structures, identification of 3D residue clusters of various types may help to infer functionally important parts of the structures and, in particular, in 1occ and 1ar1, help to characterize quaternary structural arrangements and protein docking sites (in the case at hand for cytochrome *c*), and suggest pathways for electron transfer and for proton pumping.

Cytochrome *c* oxidase, for example, in most eukaryotic aerobic organisms is the terminal oxidase of cell respiration, a process that reduces molecular oxygen to water, coupled with pumping protons from the matrix side of the mitochondrial membrane to the cytosolic side. Electrons are supplied from cytochrome *c* on the cytosolic outside. The structure 1occ is a dimer (total of 3560 amino acids) in which each monomer contains 13 distinct subunits. Major functional subunits I and II hold all the redox metal sites (*3*−*5*). The role of the large subunit III is unclear. The primary sequences of subunits I from 1occ and 1ar1 show about 60% sequence identity (see Figure 1) (*6*). Intriguingly, this pronounced conservation applies both near (including all ligands and almost all second-shell residues) and far from the metal regions (see below). The degree of sequence identity between subunits II of the two structures is about 40%. The high degree of global sequence conservation for subunits I and II supports the importance of the primary sequence with respect to function and structure. Many of the conserved positions can be interpreted as part of electron transfer or proton pumping pathways (see below).

The structures contain two iron (heme) sites (Fe_a and Fe_{a_3}), two coupled copper atoms (dicopper), Cu_A, linked to a magnesium ion at the interface of subunits I and II, a single copper Cu_B atom adjacent to heme Fe_{a_3} (see Figure 2), and in 1occ a distant isolated zinc ion. The flow of electrons is generally considered (*7*, *8*) to proceed in the following order

$$\text{Cyt-}c \xrightarrow{e^-} Cu_A \xrightarrow{e^-} Fe_a \xrightarrow{e^-} Fe_{a_3}-Cu_B \xrightarrow{e^-} O_2 \rightarrow H_2O$$

where the last arrow summarizes several steps including protonation of reduced oxygen derivatives and proton pumping across the membrane. The proton gradient produced drives ATP synthesis.

METHODS, DATA, AND TERMINOLOGY

Data. Two complete cytochrome *c* oxidase structures are now available: from the bovine heart mitochondrion (1occ) (*3*, *9*) traversing the inner mitochondrial membrane and from the α-proteobacterium *Paracoccus dentifricans* (1ar1) (*10*)

† S.K. supported in part by NIH Grants 5R01GM10452-34 and 5R01HG00335-10, and NSF Grant DMS9704552.

* Corresponding author. Telephone: (650) 723-2204. Fax: (650) 725-2040. E-mail: fd.zgg@forsythe.stanford.edu. K.D.K. supported in part by NIH Grant GM28962.

‡ Stanford University.

§ Present address: Molecular Simulations, Inc., 9685 Scranton Rd., San Diego, CA 92121.

|| Johns Hopkins University.

10.1021/bi981390t CCC: $15.00
Published on Web 12/05/1998

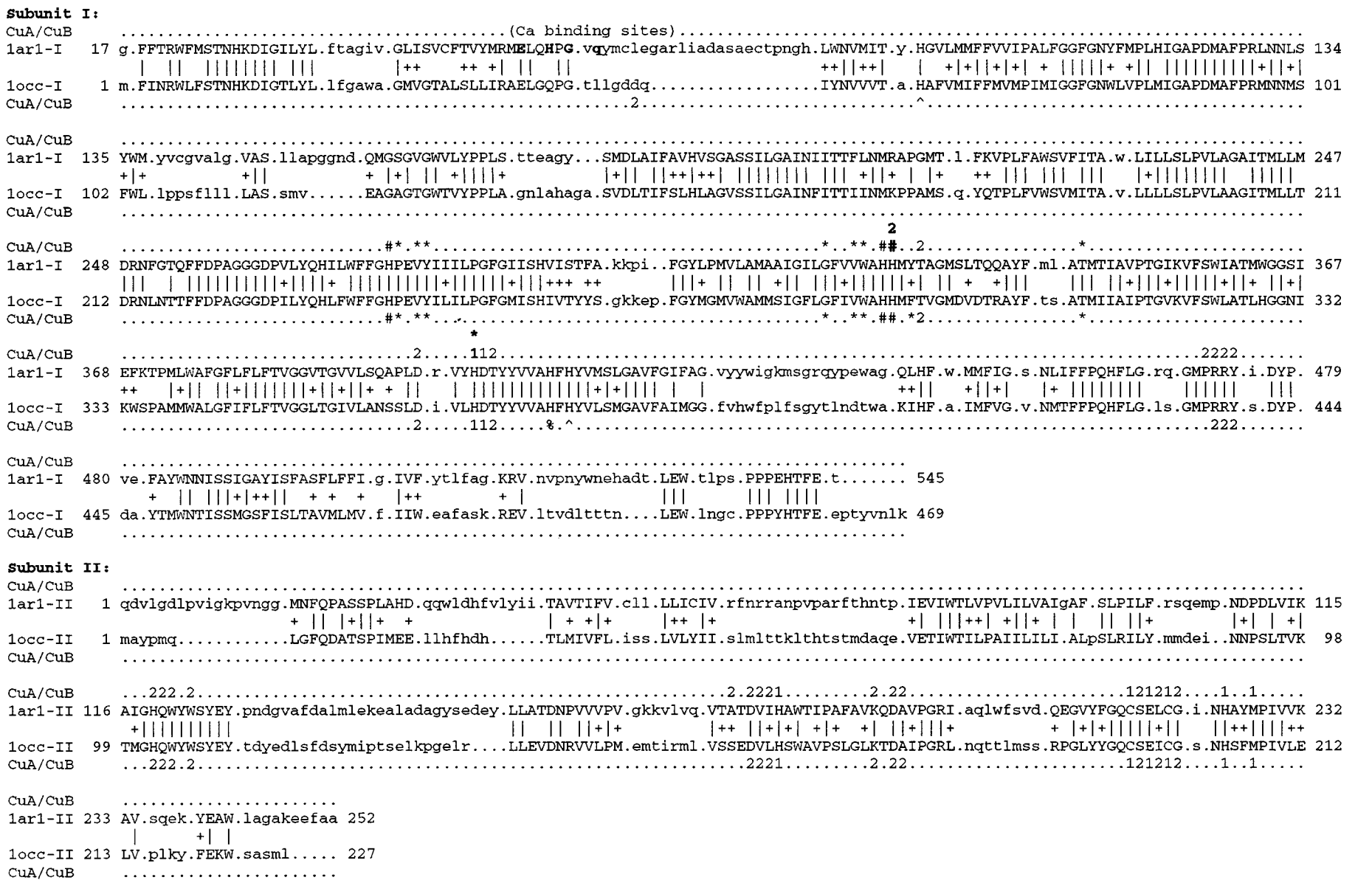

FIGURE 1: Sequence alignment of cytochrome *c* oxidase of *Paracoccus denitrificans* (PDB code 1ar1) with that of bovine heart mitochondrium (1occ). Aligned positions are in uppercase. Nonaligned positions are in lowercase. Identical residues are connected by vertical bars, and the symbol + corresponds to similar residue pairs according to the BLOSUM62 amino acid substitution matrix. The ligands and second-shell residues are indicated by 1 and 2 for the metal center Cu_A-Mg, respectively, and # and * for the metal center Cu_B. These notations appear above the alignment for 1ar1 and below for 1occ, respectively. Note in 1ar1, H-326 I is a ligand residue of Cu_B and also a second-shell residue of Cu_A-Mg; H-403 I is a ligand residue of Cu_A-Mg and a second-shell residue of Cu_B. Axial ligands to heme *a* are indicated by the symbol ∧ and by the symbol % for heme a_3.

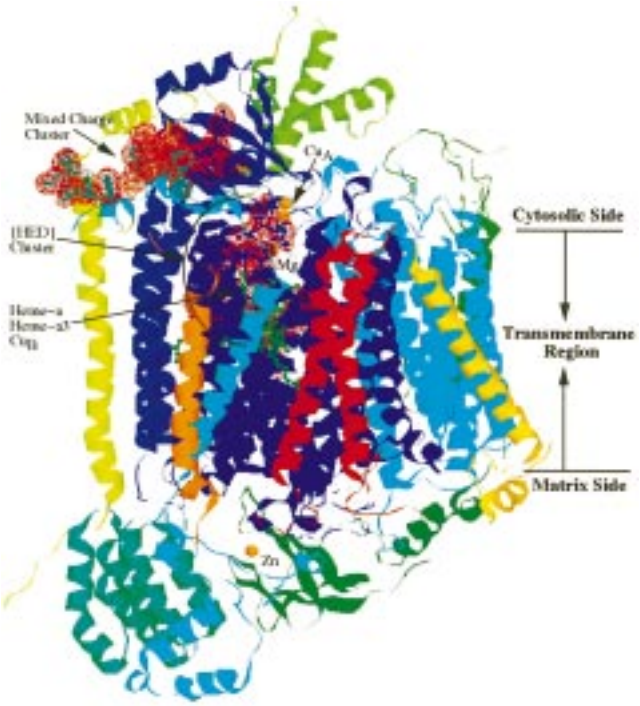

FIGURE 2: Structure of cytochrome *c* oxidase from bovine heart (1occ). A mixed charge cluster and a histidine-acidic residue cluster {HED} are shown as dotted surfaces. The structure is colored by chain units.

crossing the inner periplasmic membrane. The structures were resolved in their entirety.

Ligand Groups. These consist of residues of the protein structure and non-amino acid molecules (e.g., solvent and cofactors) that directly coordinate the metals whose bond lengths are assessed as the distance ($\leq$2.8 Å) of the ligand (side chain or main chain) atoms from the metal. There are also large complexes that sometimes contribute in coordinating a metal group (e.g., ATP, azide, porphyrin, and phosphonic acid). Patterns and distances of the individual coordinating atoms (e.g., preferences between the tautomeric $N_{\epsilon 2}$ vs $N_{\delta 1}$ bonding of the imidazole ring of histidine) can affect the geometry, orientation, electrostatics, stability, and catalytic efficiency. For a detailed discussion of tautomeric perferences, see ref *1*. For example, with respect to histidine coordination, mononuclear copper type I sites are invariably ligated by $N_{\delta 1}$, and non-heme single iron ions are invariably ligated by $N_{\epsilon 2}$ (*1*, *11*).

Second-Shell S (the Extended Environment of the Metal Center). We define S to consist of all residues (and non-amino acid molecules or complexes) whose distance to a ligand or the metal atom(s) is less than 3.5 Å. For our purposes, the distance between a residue pair in the 3D protein strucure is calculated as the minimum distance between side chain atoms or between side chain and main chain atoms (excluding hydrogens). Other definitions of the second shell might be less conservative and reach out to 4.5 Å, and would probably encompass residues often less relevant to the metal complex. Previous discussions in the literature have been more subjective and instinctive (*12*, *13*). Our definition is applied in a consistent way to all metal centers, providing meaningful comparisons across different metalloprotein classes (*1*, *2*, *11*). Each residue of the second shell is referred to close ligands and shows hydrogen bond

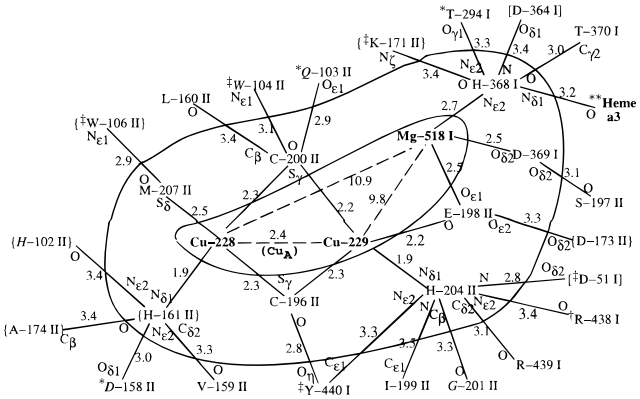

FIGURE 3: Environment of the multinuclear metal center of Cu_A and Mg of 1occ. Residues are enclosed in brackets when the residue is a component of an α-helix and enclosed in braces when the residue is a component of a *β*-strand; otherwise, the residue is part of a coil element. Italic letters indicate the residue side chain atoms are exposed (side chain solvent accessibility greater than 10%). A residue in a second shell has attached the symbol ∗ if one of its side chain atoms forms a hydrogen bond with a side chain atom of a ligand, or the symbol † if one of its main chain atoms forms a hydrogen bond with a side chain atom of a ligand, or the symbol ‡ if one of its side chain atoms forms a hydrogen bond with a main chain atom of a ligand. The second shell of the Cu_A−Mg core contains two residues in α-helices, five in *β*-strands, and 13 in coil elements. Five residues are solvent-exposed, and 15 are buried. Non-main chain···main chain hydrogen bonds between ligand residues and second-shell residues include H-161 II···D-158 II, C-196 II···Y-440 I, H-204 II···R-438 I···H-204 II···D-51 I, H-361 I···heme a_3, H-368 I···T-294 I, H-368 I···K-171 II, C-200 II···Q-103 II, C-200 II···W-104 II, and M-207 II···W-106 II.

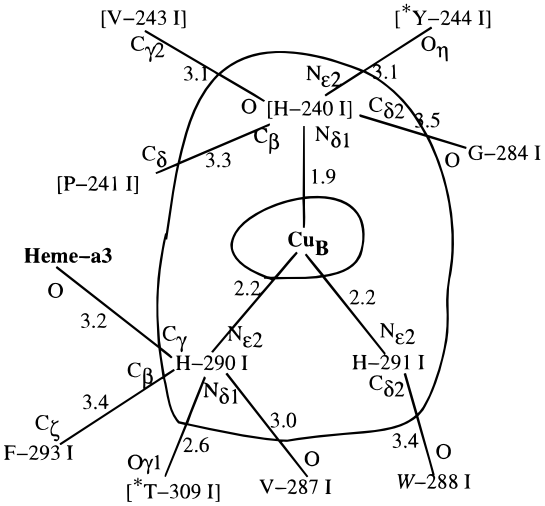

FIGURE 4: Environment of the Cu_B metal site of 1occ. See Figure 3 regarding the meaning of symbols. The second shell of Cu_B possesses four residues in α-helices and four in coil elements. One residue is solvent-exposed, and seven are buried. Non-main chain···main chain hydrogen bonds between ligand residues and second shell residues include H-290 I···T-309 I and H-240 I···Y-244 I.

relationships, degree of solvent accessibility, position in terms of secondary structure, and interconnections to other members of the second shell (see Figures 3 and 4 and Table 1).

Diverse Statistically Significant Residue Clusters in Protein Structures. In previously published work (*14*, *15*), we introduced methods for detecting particular residues in 3D protein structures that are found in clusters more frequently than one would expect. 3D residue clusters of various types in protein structures may help to pinpoint functionally

important parts of protein structures as well as unusual tertiary or quaternary structural arrangements. It is useful to review concisely our methods for identifying different kinds of 3D residue clusters. Generally, a 3D residue cluster incorporates a collection of special residues from several distinct regions of the primary sequence, which congregate in three dimensions. Possible clusters of interest include a 3D local concentration of charged residues (acidic, basic, or mixed), of histidine, of cysteine-histidine, of cysteine-histidine-methionine, of cysteine-histidine-tyrosine, of aromatic preponderance, and others.

The complete protein structure of N residues is represented by N linear sequences, one starting from each residue generating a distinct sequence of the structure arranged to reflect appropriate 3D "distance" relationships. From each residue a_k in a 3D structure, a sequence $S_k = \{s_{k,1}, s_{k,2}, ..., s_{k,j-1}, ...\}$ is generated with respect to a distance measure as follows.

Method 1 (M1). In the sequence $s_{k,1}$ is a_k, and the jth residue $s_{k,j}(j>1)$ is the next closest residue to any of the residues in the series $s_{k,1}, s_{k,2}, ..., s_{k,j-1}$. Thus, M1 accumulates residues in s_k, reflecting the shape and density of the structure about the starting residue $s_{k,1}$.

Method 2 (M2). Again, $s_{k,1}$ is a_k and the jth 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 the series $s_{k,1}, s_{k,2}, ..., s_{k,j-1}$. In M2, S_k proceeds along a pathway which often reveals channels relevant for transporting suitable molecules or substrates.

Method 3 (M3). Let $s_{k,1} = a_k$; $s_{k,j}(j>1)$ is the next closest in the 3D structure to $s_{k,1}$, distinct from the series $s_{k,1}, s_{k,2}, ..., s_{k,j-1}$. The M3 construction is equivalent to incorporating residues in order of increasing distances from a_k.

Consider a sequence S_k of length N, composed from an alphabet of letters $A = \{a_1, a_2, ..., a_r\}$, generated randomly with corresponding probabilities $\{p_1, p_2, ..., p_r\}$. The letters carry appropriate scores $\{s_1, s_2, ..., s_r\}$. Let M_N be the maximal aggregate segmental score in the sequence. It has been shown that $\text{prob}(M_N \geq s) \approx 1 - \exp(-K^*N e^{-\lambda^* s})$ where the constants K^* and λ^* can be calculated depending on the parameters $\{p_1, ..., p_r\}$ and $\{s_1, ..., s_r\}$ (*14*). The significance level s^* is determined so that $P^* = 1 - \exp(-K^*N e^{-\lambda^* s^*}) = 0.01$. Then the initial segment $\{s_{k,1}, ..., s_{k,n}\}$ of the sequence S_k is deemed a P^* statistically significantly high-scoring segment if the cumulative score of the segment exceeds s^*. For each significant cluster, the value of n is optimized. Examples of scoring regimes for negative charge clusters have (one-letter code) E, D = +2, K, R = −2, and other amino acids = −1; mixed charge cluster have K, R, E, D = +1 and others = −1; and {CHY} cysteine-histidine-tyrosine cluster has C, H, Y = +2 and others = −1 [see Karlin (*16*) for a discussion on specification of scores].

We illustrate the procedure in the case of a histidine-acidic {HED} cluster by Method M1. (1) For a given protein 3D structure, the probabilities $\{p_1, ..., p_{20}\}$ are the frequencies of the amino acids in the sequence. (2) Generate the matrix of sequence arrays $S = \{S_k\}$ for $k = 1, 2, ...$, as described previously. (3) For S_k, using scores H, D, E = +2, K, R = −2, and others = −1, use the scoring theory to determine whether a significant high-scoring segment at 1% level occurs starting with $s_{k,1}$. (4) Apply the test to every S_k sequence. Two significant clusters are considered synonymous if in the smaller cluster at least 50% of its residues are contained in the larger cluster. The cluster with the higher significance evaluation is retained.

RESULTS AND DISCUSSION

Metal Sites and the Second Shell (Figure 2). From the structure determinations, the Cu_A site is consistent with two close copper atoms linked to a magnesium ion. The X-ray crystallographic analyses locate heme *a* and heme a_3 and Cu_B on an approximately straight line parallel to the membrane surface where the two heme planes are orthogonal to the membrane plane. The heme a_3-Cu_B complex forms a molecular pocket to where O_2 molecules diffuse and bind to the heme iron atom and possibly to the copper atom as well (*7*).

Cu_A Environment. Figure 3 for the Cu_A center of 1occ demarcates the metal core (inner region), the ligand array between the two heavy curves, and the second-shell residues beyond the outer curve connecting to ligands.

The binuclear Cu_A center is generally described by itself as the important electron transfer site receiving electrons from cytochrome *c* and transferring them to heme *a* and/or heme a_3-Cu_B (*9*). However, our analysis emphasizes the importance of including the Mg with the Cu_A center along with their ligands and second-shell residues. This notion is further supported by the finding that several statistically significant residue clusters envelop the Cu_A and Mg ions (see below).

The coordination structure of the dicopper Cu_A center {Cu228, Cu229} is reminiscent of a mononuclear copper type I motif which exhibits the ligand array 2H, 1C, and 1M where both histidines bond copper ions in the tautomeric conformation $N_{\delta 1}$ (*1*). At the dicopper Cu_A core, there are two bridging cysteine residues, an individual histidine bonding each copper ion, a more distant methionine ligand for Cu228, but a glutamate E-198 II which functions as a bridging residue coordinating Cu229 and the Mg ion via its carbonyl oxygen and its $O_{\epsilon 1}$ oxygen side chain, respectively. The second oxygen $O_{\epsilon 2}$ of E-198 II appears to interact with the second-shell residue D-173 II. These multiple interactions of E-198 II with Cu229, Mg, and other residues putatively reflect a highly stable configuration and posits E-198 II in this context as a critical residue.

The second shell of the metal group Cu_A-Mg is predominantly hydrophilic. Actually, the hydrophobic components Φ vs the hydrophilic components Π (see Figure 3) consist of Π = {V-159 II, A-174 II, L-160 II, I-199 II} and Π = {W*-104 II, W*-106 II, Y*-440 I, D-158 II, H-102 II, D-173 II, Q-103 II, K-171 II, S-197 II, D-51 I, D-364 I, R-438 I, R-439 I, T-370 I, T-294 I}, respectively. The asterisk in W* and Y* signifies that these residues can be considered hydrophilic as they establish side chain H-bonds with appropriate ligand residues. Hydrogen bond networks between residues of the second shell and ligands in the metal environments are indicated in the legend of Figure 3. The ligands of the dicopper Cu_A are all contained in subunit II, whereas the ligands of Mg-518 I, apart from the bridging bidentate E-198 II, are contained in subunit I. The second-shell residues Y-440 I, R-438 I, D-51 I, D-173 II, and K-171 II are at the interface of subunits I and II within 3.5 Å of a residue of the other subunit.

The bulk of the second shell, 13 residues of 21, are found in coil elements, and five of these residues are surface-

Table 1: Residue Clusters Identified in the 3D Structure of Cytochrome *c* Oxidase[a]

cluster type	cluster content	hydrogen bonds
{CHY} = {Cys-His-Tyr}, M1-D_m, $P^* = 0.00292$	[Y-379 I], [A-375 I], [Y-371 I], T-370 I, [Y-372 I], **heme *a***, [H-61$_\Delta$ I], [H-378$_\Delta$ I], [S-382 I], R-438 I, R-439 I, W-126 I, [V-374 I], *H-368 I*, [D-364 I], **heme a_3**, [H-376$_\Delta$ I], N-360 I, **Mg-518 I**, *E-198 II*, **Cu-229 (Cu_A)**, *H-204 II*, *C-200 II*, *C-196 II*, **Cu-228 (Cu_A)**, {*H-161 II*} 2 C, 6 H, 3Y, 2 R, 1 D, 1 E, 1 N, 1 S, 1 T, 1 W, 1 A, and 1 V 10 residues in α-helices, 1 in β-strands, and 10 in coil elements 10 α, 1 β, and 10 coil; all 21 residue side chains are buried	T-370 I ($O_{\gamma 1}$) and Y-372 I (N), 2.81 Å V-374 I (O) and H-378 I ($N_{\delta 1}$), 2.91 Å C-196 II (S_γ) and E-198 II (O), 3.15 Å R-438 I (O) and H-204 II ($N_{\epsilon 2}$), 3.36 Å
{HED} = {His-Glu-Asp}, M1-d_m, $P^* = 0.00179$	[*H*-429 I], [T-354 I], [H-376$_\Delta$ I], **heme a_3**, [D-364 I], [Y-244 I], [*H-240 I*], **Cu-517 (Cu_B)**, *H-291 I*, *H-290 I*, [T-309 I], W-126 I, [F-377 I], *H-368 I*, **Mg-518 I**, *E-198 II*, D-369 I, T-294 I, {D-173 II}, {S-162 II}, R-438 I, **heme *a***, [H-61$_\Delta$ I], [H-378$_\Delta$ I] 8 H, 3 D, 1 E, 3 T, 1 Y, 1 W, 1 F, 1 S, and 1 R 10 α, 2 β, and 8 coil; 1 exposed and 19 buried	H-240 I ($N_{\epsilon 2}$) and Y-244 I (OH), 3.12 Å H-290 I ($N_{\delta 1}$) and T-309 I ($O_{\gamma 1}$), 2.60 Å T-294 I($O_{\gamma 1}$) and H-368 I ($N_{\epsilon 2}$), 3.28 Å T-294 I ($O_{\gamma 1}$) and D-173 II ($O_{\delta 1}$), 2.71 Å S-162 II (O_γ) and D-173 II ($O_{\delta 1}$), 2.87 Å
{HY} = {His-Tyr}, M1-d_m, $P^* = 0.00292$	Y-129 I, [W-236 I], *H-291 I*, **Cu-517 (Cu_B)**, [*H-240 I*] [Y-244 I], **heme a_3**, [H-376 I], [T-354 I], [Y-372 I], T-370 I, [G-432 I], Y-371 I, **heme a**, [H-61$_\Delta$ I] 4 H, 4 Y, 2 T, 1 W, and 1 G 10 α, 1 β, and 10 coil; all buried	Y-129 I (OH) and W-236 I ($N_{\epsilon 1}$), 2.79 Å H-240 I ($N_{\epsilon 2}$) and Y-244 I (OH), 3.12 Å T-354 I ($O_{\gamma 1}$) and Y-372 I (OH), 3.49 Å T-370 I ($O_{\gamma 1}$) and Y-372 I (N), 2.81 Å
{CH} = {Cys-His}, M1-d_m, $P^* = 0.00262$	*H*-5 V, *G*-3 V, *H*-2 V, *E*-63 VI, *N*-66 VI, *C-62 VI*, **Zn-99 VI**, [*C-60 VI*], {*C-82 VI*}, *C-85 VI* 4 C, 2 H, 1 E, 1 G, and 1 N 0 α, 2 β, and 7 coil; 2 buried and 7 exposed	E-63 VI (N) and N-66 VI ($O_{\delta 1}$), 3.29 Å
{CHY} = {Cys-His-Tyr}, M2-d_m, $P^* = 0.00808$	[*C*-39], *C*-53, [*Y*-56], [*F*-56], [H-37], [*R*-76], [*D*-73], [Y-33], [W-72], [C-29], C-64, [V-69], [Y-60] all in chain VIII 4 C, 1 H, 3 Y, 1 F, 1 W, 1 V, 1 R, and 1 D 11 α, 0 β, and 2 coil; 6 exposed and 7 buried	C-29 (S_γ), and C-64 (S_γ), 2.01 Å C-39 (S_γ) and C-53 (S_γ), 2.39 Å Y-33 (OH) and D-73 ($O_{\delta 1}$), 2.63 Å Y-33 (OH) and R-76 (N_ϵ), 3.26 Å D-73 ($O_{\delta 1}$) and R-76 ($N_{\eta 2}$), 2.76 Å
{KRED} = {Lys-Arg-Glu-Asp} (mixed charge cluster), M2-d_m, $P^* = 0.00011$	*D*-141 IV, *D*-139 IV, *K*-142 IV, {*E*-144 IV}, {W-138 IV}, P-50 XI, [*K*-121 IV], [*D*-125 IV], *W*-53 XI, *E*-52 XI, *K*-51 XI, [*E*-114 IV], [*K*-118 IV], *P*-130 II, *K*-129 II, *E*-132 II, E-137 II, [R-112 IV] 5 K, 1 R, 5 E, 3 D, 2 W, and 2 P 5 α, 2 β, and 11 coil; 4 buried and 14 exposed	K-129 II (N) and E-132 II ($O_{\epsilon 1}$), 2.94 Å K-121 IV (Nζ) and D-125 IV ($O_{\delta 1}$), 2.93 Å K-121 IV (Nζ) and E-52 XI ($O_{\epsilon 1}$), 3.42 Å R-122 IV ($N_{\eta 1}$) and E-137 II ($O_{\epsilon 2}$), 2.61 Å D-141 IV (N) and D-139 IV ($O_{\delta 1}$), 3.08 Å D-139 IV ($O_{\delta 2}$) and K-142 IV (N_ξ), 3.14 Å D-125 IV ($O_{\delta 2}$) and W-53 XI ($N_{\epsilon 1}$), 2.64 Å

[a] See the text for how the clusters are identified. Residues are enclosed in brackets if the residue is in an α-helix and enclosed in braces if the residue is in a β-strand; otherwise, the residue is in a coil element. Italic letters with regular numbers indicate the residue side chain atoms are exposed (side chain accessibility >10%). Residue ligands to Cu_A, Cu_B, or Zn are italic and underlined, while second-shell residues are underlined. Subscript triangles indicate axial ligands to heme *a* or heme a_3. Only non-main chain···main chain hydrogen bonds among residues of the clusters are listed.

exposed, including W-104 II, D-158 II, and H-102 II, all mutually close to each other (W-104−H-102, 2.4 Å; and H-102−D-158, 3.4 Å). These residues proffer opportunities for cytochrome *c* docking for transfer of electrons. In fact, mutagenesis studies of the *P. dentificans* 1ar1 structure attest to the function that W-121 II [corresponds to W-104 II in the bovine mitochondrial (1occ) structure] is an electron entry site to cytochrome *c* oxidase (*17*). The residues proximal to W-104 II, namely, D-158 II and H-102 II, putatively also behave as electron entry positions. We further hypothesize that a mixed charge cluster situated on the surface about 10 Å from these residues contributes in cytochrome *c* docking (see below). Other specific residues proposed (*18*) to be important for electron transfer in 1ar1, either as part of a direct pathway or for docking, are D-193 II (corresponds to D-173 II of 1occ in the second shell of Cu_A), E-218 II of 1ar1 (corresponds to E-198 II, the bridging residue of Cu229 and Mg), and E-126 II of 1ar1 (corresponds to E-109 of 1occ, not part of the Cu_A second shell).

All ligands of Cu_A and 17 of 21 residues of the second shell of Cu_A 1occ correspond structurally and in perfect sequence identity to the Cu_A environment of 1ar1, again attesting to their functional importance (cf. Figure 1). There are two conservative aliphatic replacements involving L with I, which likely contribute equivalently. The small alanine residue A-176 II of (1ar1) belongs to the second shell of Cu_A, whereas the small aligned serine residue S-156 II of (1occ) is more distant (3.7 Å) from the ligand. We do not know the significance, if any, of this difference in residue type and distance. However, D-51 I of the second shell in bovine 1occ that H-bonds to the ligand H-204 possesses *no* counterpart in the second shell about Cu_A of *P. denitrificans* 1ar1. Residue D-51 I is distinguished in several ways. In a reduced state (*19*), D-51 I, at both its side chain carboxylate and backbone amide and carbonyl, putatively, has capacity for at least alternative or simultaneous H-bonding connections to subunit II with residues S-202 II, H-204 II, and S-205 II and H-bonding along the face of a helix connecting D-51 I with N-55 I (see Table 2). It appears that D-51 I from the reduced to the oxidized states can move several angstroms in H-bonding contacts (*19*).

Cu_B Environment (in 1occ). The Cu_B (Cu517 I) I environment is depicted schematically in Figure 4. The exclusive ligands of Cu517 I comprise two consecutive histidine

Table 2: Hydrogen Bonds among Interface Residues of Subunits I and II

hydrogen bond type[a]	residue 1		distance (Å)		residue 2
B	D-50 I[b]	$O_{\delta 2}$[c]	3.31[d]	O_{γ}[c]	S-202 II[b]
	D-51 I	O	2.96	N	N-55 I
	D-51 I	O	2.57	$N_{\delta 2}$	N-55 I
B	D-51 I	N	2.91	O	S-202 II
B	D-51 I	$O_{\delta 2}$	2.84	N	H-204 II
B	D-51 I	$O_{\delta 2}$	2.78	N	S-205 II
B	D-51 I	$O_{\delta 1}$	2.69	O_{γ}	S-205 II
B	D-51 I	$O_{\delta 2}$	2.59	O_{γ}	S-205 II
	N-55 I	$N_{\delta 2}$	3.42	O	R-439 I
	T-294 I	$O_{\gamma 1}$	3.28	$N_{\epsilon 2}$	H-368 I
B	T-294 I	$O_{\gamma 1}$	2.71	$O_{\delta 1}$	D-173 II
B	T-294 I	O	2.98	N	D-173 II
B	V-295 I	O	3.27	$N_{\delta 2}$	N-180 II
B	G-296 I	O	3.01	$N_{\eta 2}$	R-178 II
B	D-300 I	$O_{\delta 1}$	2.72	OH	Y-85 II
Mw	L-324 I	O	3.15	N	H-328 I
M, B	G-329 I	O	2.81	N	H-52 II
M, B	N-360 I	$N_{\delta 2}$	2.85	O_{γ}	S-80 II
Mw	S-361 I	O	2.99	N	I-365 I
M, B	S-362 I	O_{γ}	3.43	O	S-80 II
M, B	S-362 I	O_{γ}	3.47	O_{γ}	S-80 II
	S-362 I	O	3.10	N	I-365 I
B	I-365 I	O	2.71	N_{ζ}	K-171 II
B	H-368 I	O	3.45	N_{ζ}	K-171 II
	D-369 I	O	3.16	N	R-438 I
	F-430 I	O	2.76	N	L-433 I
B	R-438 I	O	3.36	$N_{\epsilon 2}$	H-204 II
B	Y-440 I	OH	2.76	O	C-196 II
B	D-442 I	$O_{\delta 1}$	2.97	$N_{\eta 2}$	R-134 II
B	D-442 I	$O_{\delta 2}$	2.54	$N_{\eta 2}$	R-134 II
B	Y-443 I	O	3.27	$N_{\eta 1}$	R-134 II
	P-444 I	O	3.30	N	Y-447 I
B	D-445 I	$O_{\delta 1}$	2.59	$N_{\eta 1}$	R-134 II
B	W-450 I	$N_{\epsilon 1}$	2.81	O	Q-6 II
Mw	S-80 II	O	2.91	N	L-84 II
	Y-85 II	O	2.91	N	D-88 II
	D-173 II	O	2.96	$N_{\delta 2}$	N-180 II
	I-175 II	O	2.88	N	R-178 II
	R-178 II	N_{ϵ}	2.69	$O_{\delta 1}$	N-180 II
	Q-195 II	O	3.30	O_{γ}	S-197 II
	C-196 II	S_{γ}	3.15	O	E-198 II
	H-204 II	N	3.10	O	G-201 II

[a] For hydrogen bond type, B and M indicate hydrogen bonds formed for specified residues of subunit I and subunit II in a nontransmembrane region and in a transmembrane region, respectively. Mw indicates hydrogen bonds formed by residues from a common subunit of the transmembrane region. Unmarked places refer to hydrogen bonds formed by residues from a common subunit in nontransmembrane regions. [b] Residue designations. [c] Atoms making hydrogen bonds between residue 1 and 2. [d] Hydrogen bond distance.

residues, H-290 I and H-291 I, both ligating via $N_{\epsilon 2}$ and H-240 I ligating via $N_{\delta 1}$. All residues of the second shell of Cu_B for 1occ and 1ar1 are conserved with one exception. Explicitly, F-293 I (1occ) in the second shell of Cu_B is at a distance of 3.3 Å, whereas the corresponding aligned Y-328 I (1ar1) is 3.6 Å from the Cu_B ligands. However, aromatics F and Y often substitute for each other. Moreover, ligand H-403 I (1ar1) of Mg is in the second shell of Cu_B, but the corresponding ligand H-368 I (1occ) of Mg is distant (3.62 Å) from the ligands of Cu_b. A greater disparity is related to the ligand H-326 I (1ar1) of Cu_B lying in the second shell of Cu_A ($\leq$3.4 Å), whereas the aligned ligand H-291 I (1occ) of Cu_B is significantly displaced 4.6 Å with respect to the ligands of Cu_A.

The second shell of 1occ to Cu_B consists of seven buried residues and one exposed residue (W-288 I; see Figure 4).

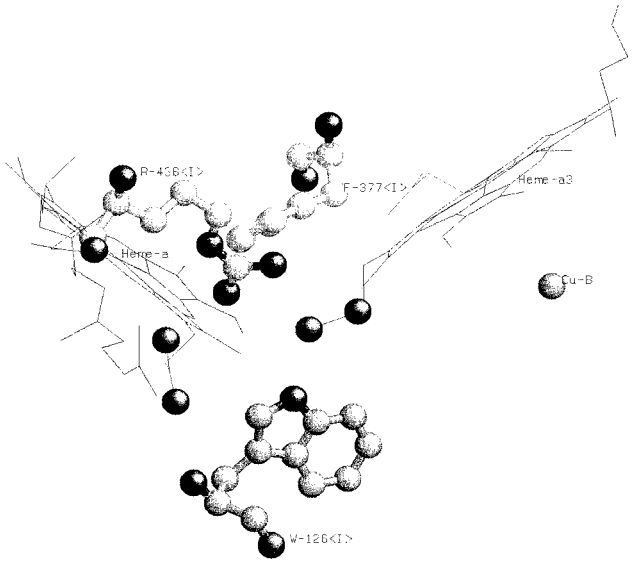

FIGURE 5: Heme *a* and heme a_3 of cytochrome *c* oxidase (1occ). W-126 I and R-438 I form hydrogen bonds (propionate contacts) with both hemes. See the text concerning interpretations.

The single polar residue T-309 I makes a side chain hydrogen bond with H-290 I connecting the side chain atoms $O_{\delta 1}$ and $N_{\delta 1}$. According to recent X-ray crystallographic refinements (*5*, *19*), Y-244 I (Y-280 in 1ar1) of the second shell of Cu517 is actually covalently linked to H-240 I, apparently due to a post-translational modification.

Heme a and Heme a_3 Connections. Figure 5 shows the residues in contact with both heme *a* and heme a_3. Tsukihara et al. (*9*) noted that the two hemes are bridged by three consecutive residues: H-376 I (axial ligand to heme a_3), F-377 I, and H-378 I (axial ligand to heme *a*). F-377 I makes hydrophobic contacts with both hemes. However, we also note that W-126 I and R-438 I in the second shell of Cu_A make two hydrogen bonds each with oxygens (propionate groups) from both heme units (Figure 5) (R-439 I hydrogen bonds to only one heme). These residues, especially W-126 I conserved in structure 1ar1, may be important for holding the hemes in place and may also provide an alternative inter-heme electron transfer pathway. For a broad discussion of the role of the intervening medium in long-range biological electron transfer, see the series of articles in the *Journal of Biological Inorganic Chemistry* (*JBIC*), volume 2, pages 372−404 (1997). The side chain imino nitrogen $N_{\epsilon 1}$ of W-126 I and $O_{2\delta}$ (propionate) of heme a_3 connect at a distance of 3.0 Å, whereas the *backbone nitrogen* of W-126 I H-bonds to $O_{1\delta}$ (a propionate of heme *a*) at 2.9 Å (see Figure 5). Also, R-438 I H-bonds via its side chain $N_{\eta 2}$ to $O_{\delta 1}$ (propionates) at a bond length of 2.7 Å.

Zn Environment (Figure 6). The isolated zinc of unit VI is coordinated by four cysteines {C-60 VI, C-62 VI, C-82 VI, C-85 VI} at bond lengths in the range of 2.4−2.5 Å. The second shell consists of Φ = {W-27 V}, and Π = {*Y-89 VI, S-84 VI, T-87 VI}, {G-86 VI}. The ligands occur in two close pairs CXC and CX_2C in the primary sequence, somewhat similar to that of aspartate carbamoyltransferase (8atc) and β3 alcohol dehydrogenase (1htb) (*2*). CX_2C peptides in cytochrome *c* oxidase and 8atc have similar β−turn−β structure (antiparallel β-sheet). The zinc structure is described as being similar to a zinc finger by Tsukihara et al. (*9*), but the spacing and the secondary structure

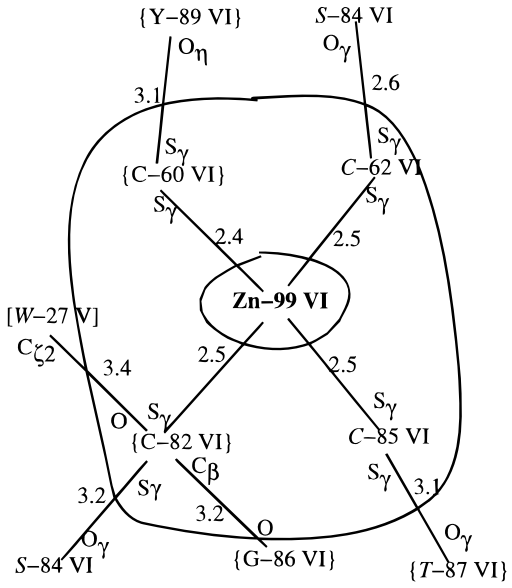

FIGURE 6: Environment of the Zn metal center in 1occ. See Figure 3 regarding the meaning of symbols. The second shell has one residue in an α-helix, three in *β*-strands, and one in a coil element. Three residues are solvent-exposed, and two are buried.

determinations are not consistent with this assignment. The role of the zinc center is unknown, although it likely augments the structural stability of the 1occ complex. Apropos, the zinc ion is not present in the bacterial (*P. dentifricans*) cytochrome *c* oxidase homologue (*10*).

Ca Environment (Figure 7). The calcium environment (Ca-7) in 1ar1 is displayed in Figure 7a (in the legend H-bonding interactions are listed). All four ligand residues are proximal in the sequence ...ElqHpGvQ..., where letters in uppercase indicate ligand residues. There is a similar ligation pattern found in the sequence ...DcfVqGcD... coordinating Ca-320 in the peanut peroxidase structure (1sch). In both cases, the first ligand binds calcium as a bidentate using main chain and side chain atoms. The second and third ligands bind via carbonyl atoms. The fourth ligand binds through side chain atoms. In both structures, the sequence preceding this calcium binding group up to the first ligand forms an α-helix which interacts with a heme. The calcium [or possibly sodium; see Yoshikawa et al. (*19*)] environment Ca-7 of 1occ is modeled from the alignment of Figure 1 and from structure superpositions, revealing only three corresponding ligands. The second shell of Ca-7 of 1ar1 is contained in subunit I only, whereas for Ca-7 of 1occ, the second shell involves residues of subunits I and IV.

Distinctive Residue Clusters of 1occ. The 1occ and 1ar1 structures contain many statistically significant residue clusters (compare with refs *1*, *11*, and *14*) described explicitly in Table 1. These include a {CHY} cluster surrounding the Cu_A−Mg complex core, a {HED} (Table 1) cluster surrounding metals Mg-518 I and Cu_B517 I and the two heme units, a {HY} cluster about the two heme units, a {CH} cluster about the zinc site, and a mixed charge cluster {KRED} away from the metals, situated mostly on the surface of the protein structure at the interface of subunits II, IV, and XI. This latter cluster contains seven H-bonds and includes at least three salt bridges connecting D-139 IV with K-142 IV, K-121 IV with D-125 IV, and E-137 II with R-122 IV. The mixed charge cluster at hand may mediate and stabilize quaternary protein formation. Examples of this function include glutathione *S*-transferase (2gst), catalase (8act), and southern bean mosaic virus coat protein (4sbv) (*15*). A mechanism for this stability proposes relatively long-range electrostatic attractions that help orient the surfaces (*20*). Hydrophobic interactions and hydrogen bonding subsequently act to strengthen the interface contacts. In this vein, it was observed that the most frequent interunit specific residue contacts involve charged residues of opposite sign and secondarily nonspecific hydrophobic interactions (*21*). This result is consistent with many observations on antibody−antigen pairs, hormone−receptor contacts, and large protein complexes where multichain stabilization is nucleated by electrostatic interactions. Electrostatic interactions also facilitate important processes such as protein sorting, translocation, localization, orientation, oligomerization, and binding to protein and other molecules. The {KRED} mixed charge cluster is approximately 10 Å from the proposed electron entry site, W-104 II (see above), and mostly on the same face of subunit I also putatively participates in cytochrome *c* docking.

Cytochrome *c* oxidase effects the reduction of O_2 to H_2O coupled with proton pumping, processes that require channels for proton and O_2 movements and also electron transfer pathways through the membrane. The precise pathways are unknown, although different possibilities have been discussed (*3*, *5*, *8*, *19*, *22*, *23*). It is believed that transfer of protons occurs in a hydrophilic environment (e.g., along the surface of a hydrophilic helix), most effectively through hydrogen bonds (*3*). Notably, the interface of subunits I and II constitutes a hydrophilic pathway possibly relevant for proton conduction. Explicitly, the interface between subunits I and II of 1occ (each such residue within 3.5 Å of some side chain or backbone atom of the other subunit) exhibits in subunit I 18 of 33 hydrophilic residues, including 10 charged (six D, three R, and one K) and seven aromatic residues (three Y, three F, and one W). At this interface, subunit II provides 19 of 34 hydrophilic residues with many opportunities for hydrogen bonds within and between subunits I and II. The identity of the interface residues and hydrogen bond connections [ascertained by program HBOND (*24*)] is displayed in Table 2. H-bonds within and between the two subunits and along and separated from the helical transmembrane span are given. The statistics of these interface residues with respect to aligned positions of the two structures are as follows. In subunit I, 28 of 36 are conserved, including ligands H-368 and D-369 of Mg and second-shell residues T-294, R-438, R-439, and Y-440 of Cu_A−Mg. A key interface residue of I not conserved is D-51. In subunit II, 21 of 35 are conserved, including ligands C-196, E−198, and H-104 and second-shell residues K-171, D-173, S-197, I-199, and G-201 of Cu_A−Mg. The preponderance of hydrophilic residues and hydrogen bond potential at the interface offers an attractive route for the possibility of water conduction and proton transfer and/or appropriate for H-bonding to hold subunit I and subunit II.

The {CHY} cluster (see Table 1) that encompasses the Cu_A−Mg metal core includes both heme units and presents H-bond connections which can mediate electron transfer from Cu_A to the heme sites. The histidine-acidic {HED} cluster of Table 1, replete with H-bond contacts, surrounds the Cu_B metal, the Mg518 site, and the two heme units, further

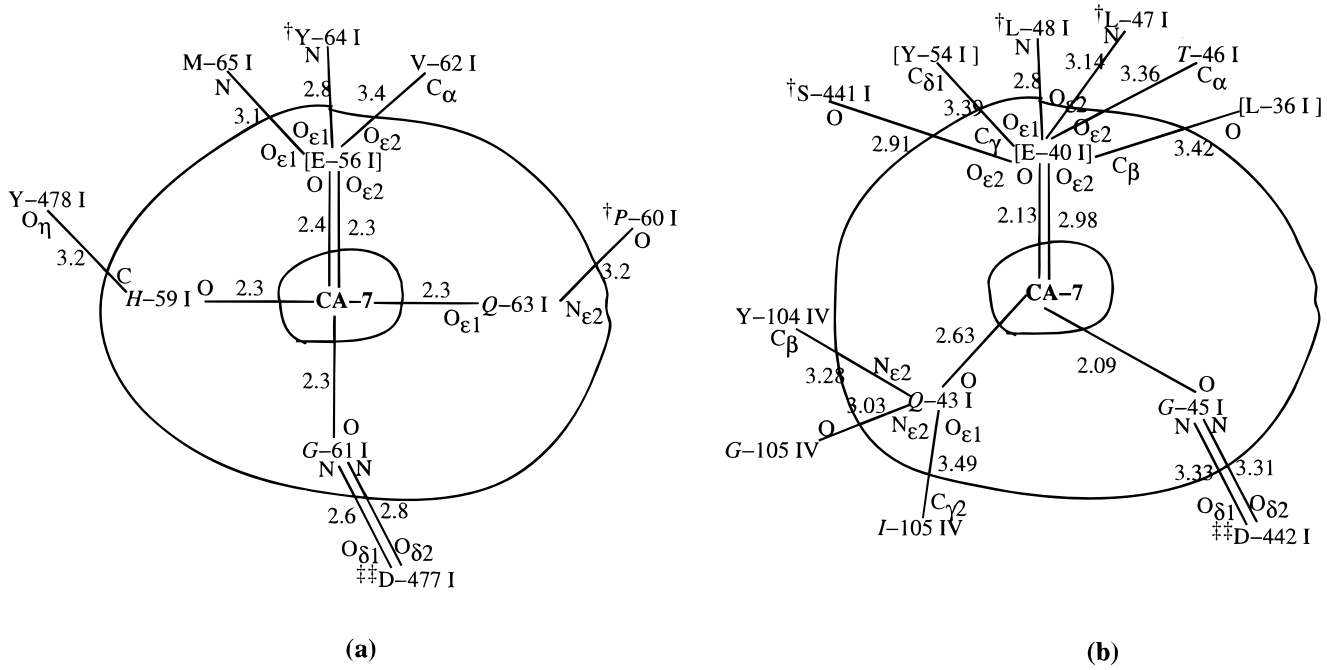

FIGURE 7: (a) Environment of the calcium (Ca-7) metal center (1ar1). See Figure 3 regarding the meaning of symbols. All residues of the second shell are in coil elements. Two residues are solvent-exposed, and five are buried. Non-main chain···main chain hydrogen bonds between ligand residues and second-shell residues include E-56 I···Y-64 I, G-61 I···D-477 I, and Q-63 I···P-60 I. (b) The environment of the modeled calcium (Ca-7) center (1occ). This calcium was modeled on the basis of the structural homology between 1occ and 1ar1. We superimpose the structures of peptide residues 56−62 of unit I in structure 1ar1 on the corresponding peptide residues 40−46 of structure 1occ. Only main chain atoms were used in the superposition. The modeled calcium ion was placed at the resulting position of the calcium ion Ca-7 of 1ar1.

emphasizing the importance of Mg in association with Cu_A and Cu_B. Previously, we have suggested means by which W-126 I and R-438 I in this cluster H-bond to oxygen atoms (propionate groups) of both heme molecules (see also Figures 4 and 5). The {HED} cluster involves no small aliphatic residues but contains many histidine residues and three key aromatic residues. We suggest these may facilitate the electron transfer through Cu_A, Mg, heme, and Cu_B. The three heme axial histidines and other histidine residues, in aggregate eight, are included in this cluster. The significant numbers of acidic nonmetal ligand residues in the {HED} cluster and of tyrosine nonmetal ligand residues in the {CHY} cluster about the metal centers attest to their role in stabilizing the metal environment and in channeling electrons and protons to and from the metal centers. There is also a statistically significant histidine-tyrosine {HY} cluster about the two heme molecules (Table 1).

A second {CHY} cluster comprised mostly of exposed residues occurs in subunit VIII centered on two pairs of cysteine disulfide bridges. This may be a source of stability for the structure. The zinc site is encompassed by a {CH} = {Cys-His} cluster near the surface of the 1occ structure.

The *P. denitrificans* structure (1ar1) also shows an {HED} cluster encompassing the Mg, heme a_3, heme *a*, and Cu_B sites, with 17 total residues, of which 14 have identical counterparts in the {HED} cluster of 1occ (data not shown). Also, 1ar1 contains a {CHY} cluster about metals Cu_A and Mg, and heme *a*, including 11 of 13 residues coincident in alignment with residues of the {CHY} cluster of 1occ. 1ar1 also contains a mixed charge cluster traversing subunits II and III. For the *P. denitrificans* 1ar1 structure, Witt et al. (*25*), on the basis of mutagenesis studies, highlight five acidic residues (D-257 I, E-126 II, D-135 II, D-159 II, and D-178 II) as potential docking sites for cytochrome *c*. However, all but three pairs of these residues are ≥10 Å apart (exceptions are D-135 II−D-159 II, 3.8 Å; D-135 II−E-126 II, 7.5 Å; and D-257 I−D-178 II, 6.7 Å). No acidic cluster at the site of cytochrome *c* docking to cytochrome oxidase is detected by our methods (*15*).

SUMMARY

The residue clusters overlapping the metal environment and other residue clusters distant from the metal environments may contribute to substrate channeling, enhancing stability, redox reactions, protein−protein interactions, and quaternary structure formation and suggest possible pathways for electron transfer, O_2 diffusion, H_2O movement, and proton pumping. The foregoing analysis also suggests attractive candidate residues for mutagenesis. These could include generally residues of the second shell about the different metal centers, especially in the second shell of Cu_A, such as W-126 I (W-164 I in 1ar1) which putatively connects and stabilizes heme *a* with heme a_3. Similar interpretations apply to R-438 I (R-473 I of 1ar1). Residue W-104 (W-121 II in 1ar1) has been described as an electron entry site to the cytochrome *c* oxidase (*17*). The proximal exposed residues D-158 II (within 5.3 Å of W-104) and H-102 II (within 2.4 Å of W-104) of the second shell of CuA might also be tested for this capability. The residues at the interface of subunits I and II (see the text and Table 2) and especially the flexible D-51 I may be important in a proton pumping pathway.

ACKNOWLEDGMENT

We are happy to acknowledge valuable discussions with Drs. David Randall of Stanford University and R. Gennis of the University of Illinois.

REFERENCES

1. Karlin, S., Zhu, Z.-Y., and Karlin, K. (1997) The extended environment of mononuclear metal centers in protein structures, *Proc. Natl. Acad. Sci. U.S.A. 94*, 14225−14230.
2. Karlin, S., and Zhu, Z.-Y. (1997) Classification of mononuclear zinc metal sites in protein structures, *Proc. Natl. Acad. Sci. U.S.A. 94*, 14231−14236.
3. Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S. (1996) The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 A, *Science 272*, 1136−1144.
4. Ostermeier, C., Iwata, S., and Michel, H. (1996) Cytochrome *c* oxidase, *Curr. Opin. Struct. Biol. 6*, 460−466.
5. Ostermeier, C., Harrenga, A., Ermler, U., and Michel, H. (1997) Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome *c* oxidase complexed with an antibody Fv fragment, *Proc. Natl. Acad. Sci. U.S.A. 94*, 10547−10553.
6. Saraste, M. (1990) Structural features of Cytochrome Oxidase, *Q. Rev. Biophys. 23*, 331−366.
7. Ferguson-Miller, S., Babcock, G. T. (1996) Heme/copper terminal oxidases, *Chem. Rev. 96*, 2889−2909.
8. Malmström, B. G. (1998) Cytochrome oxidase: pathways for electron tunneling and proton transfer, *J. Biol. Inorg. Chem. 3*, 339−343.
9. Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S. (1995) Structure of metal sites of oxidized bovine heart of cytochrome *c* oxidase at 2.8 Å, *Science 269*, 1069−1064.
10. Iwata, S., Ostermeier, C., Ludwig, B., and Michel, H. (1995) Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*, *Nature 376*, 660−669.
11. Karlin, K. D., Zhu, Z.-Y., and Karlin, S. (1998) The environment of multinuclear copper metal linkages in protein structures, *J. Biol. Inorg. Chem. 3*, 172−187.
12. Yamashita, M. M., Wesson, L., Eisenman, G., and Eisenberg, D. (1990) Where metal ions bind in proteins, *Proc. Natl. Acad. Sci. U.S.A. 87*, 5648−5652.
13. Christianson, D. W., and Fierke, C. A. (1996) Carbonic anhydrase: Evolution of the zinc binding site by nature and by design, *Acc. Chem. Res. 29*, 331−339.
14. Karlin, S., and Zhu, Z.-Y. (1996) Characterizations of distinctive residue clusters in protein 3D structures, *Proc. Natl. Acad. Sci. U.S.A. 93*, 8344−8349.
15. Zhu, Z., and Karlin, S. (1996) Clusters of charges residues in protein three-dimensional structures, *Proc. Natl. Acad. Sci. U.S.A. 93*, 8350−8355.
16. Karlin, S. (1994) Statistical studies of biomolecular sequences: Score based methods, *Philos. Trans. R. Soc. London, Ser. B 344*, 391−401.
17. Witt, H., Malatesta, F., Nicolett, F., Brunori, M., and Ludwig, B. (1998) Tryptophan 121 of subunit II is the electron entry site to cytochrome *c* oxidase in *Paracoccus denitrificans*, *J. Biol. Chem. 273*, 5132−5136.
18. Lappalainen, P., Watmough, N. J., Greenwood, C., and Saraste, M. (1995) Electron transfer between Cytochrome *c* and the isolated CuA domain: Identification of substrate binding residues in Cytochrome *c* Oxidase, *Biochemistry 34*, 5824−5830.
19. Yoshikawa, S., Shinzawa-Itoh, K., Nakashima, R., et al. (1998) Redox coupled crystal structural changes in bovine heart Cytochrome *c*-Oxidase, *Science 280*, 1723−1729.
20. Gray, H. B., and Ellis, W. R., Jr. (1994) in *Bioinorganic Chemistry*, (Bertini, I., Gray, H. B., Lippard, S. J., and Valentine, J. S., Eds.) pp 315−364, University Science Books, Mill Valley, CA.
21. Brocchieri, L., and Karlin, S. (1996) How are close residues of protein structures distributed in primary sequence? *Proc. Natl. Acad. Sci. U.S.A. 92*, 12136−12140.
22. Yoshikawa, S. (1997) Beef heart cytochrome *c* oxidase, *Curr. Opin. Struct. Biol. 7*, 574−579.
23. Gennis, R. B. (1998) Multiple proton conducting pathways in cytochrome oxidase and a proposed role for the active-site tyrosine, *Biochim. Biophys. Acta 1365*, 241−248.
24. Overington, J. P., Johnson, M. S., Sali, A., and Blundell, T. L. (1990) Tertiary structural constraints on protein evolutionary diversity: Template, key residues and structure prediction, *Proc. R. Soc. London, Ser. B 241*, 132−145.
25. Witt, H., Malatesta, F., Nicoletti, F., Brunori, M., and Ludwig, B. (1998) Cytochrome c binding site on cytochrome oxidase in *Paracoccus denitrificans*, *Eur. J. Biochem. 25*, 367−373.

BI981390T