*Protein Science* (1997), *6*:808–815. Cambridge University Press. Printed in the USA.


# Architecture of helix bundle membrane proteins: An analysis of cytochrome *c* oxidase from bovine mitochondria

ERIK WALLIN,[1] TOMITAKE TSUKIHARA,[2] SHINYA YOSHIKAWA,[3] GUNNAR VON HEIJNE,[1] AND ARNE ELOFSSON[1]

[1]Department of Biochemistry, Stockholm University, S-106 91 Stockholm, Sweden
[2]Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita 565, Japan
[3]Department of Life Science, Himeji Institute of Technology, Kamigohri Akoh, Hyogo 678-12, Japan

(RECEIVED November 18, 1996; ACCEPTED January 22, 1997)

**Abstract**

We have analyzed the structure of mitochondrial cytochrome *c* oxidase in terms of general characteristics thought to be important for describing the architecture of helix bundle membrane proteins. Many aspects of the structure are similar to what has previously been found for the photosynthetic reaction center and bacteriorhodopsin. Our results lead to a considerably more precise general picture of membrane protein architecture than has hitherto been possible to obtain.

**Keywords:** *c* oxidase; cytochrome; membrane protein

Two basic structural types of integral membrane proteins have been defined so far: helix bundle proteins and $\beta$-barrel proteins (Cowan & Rosenbusch, 1994). Five unrelated high-resolution structures are available for the former group: bacteriorhodopsin (Grigorieff et al., 1996), the bacterial photosynthetic reaction center (Deisenhofer et al., 1985; Allen et al., 1987; Yeates et al., 1987), a plant light-harvesting complex (Kühlbrandt et al., 1994), a bacterial light-harvesting complex (McDermott et al., 1995), and the recently determined bacterial and mitochondrial cytochrome *c* oxidases (Iwata et al., 1995; Tsukihara et al., 1995; Tsukihara et al., 1996).

Up to now, characterization of conserved structural features in the helix bundle membrane proteins has been difficult due to the lack of structural data, and only the photosynthetic reaction center has been analyzed in any depth (Rees et al., 1989a, 1989b). With the availability of the cytochrome *c* oxidase structure, the situation has changed. This is by far the largest complex to date, with, in the case of the mitochondrial oxidase (m-COX), a total of 28 transmembrane helices. We now report a first analysis of this structure in terms of residue distributions and sequence variability. Our results allow a refined description of the general features of helix bundle membrane proteins, in particular regarding the distribution of various classes of amino acids relative to the ends of the transmembrane helices and the lipid bilayer. We also show that the expected correlations between degree of lipid exposure and sequence variability as well as between the degree of lipid exposure and average hydrophobicity hold for m-COX.

Reprint requests to: Arne Elofsson, Department of Biochemistry, Stockholm University, S-106 91 Stockholm, Sweden; e-mail: arne@biokemi.su.se.

## Results

An overall view of the structure of m-COX is shown in Figure 1. The complex consists of 13 subunits, 10 of which span the membrane one or more times, bringing the total number of transmembrane helices to 28. The three core subunits I–III are encoded within the mitochondrial genome, while the remaining 10 are encoded by nuclear genes and imported into the organelle.

### *Distribution of secondary structure and location of lipid molecules*

The dominating structural feature in helix bundle membrane proteins is, of course, the transmembrane helices themselves (Fig. 1). This is further illustrated in Figure 2 (bottom panel), where the distribution of residues in different secondary structures (helix, sheet, irregular) along the *z*-axis perpendicular to the membrane plane is shown. The central $\pm 12$ Å region is almost 100% $\alpha$-helix, whereas a more balanced distribution of secondary structure typical of globular proteins is fully established about 30 Å from the center of the membrane. There is a region with mostly irregular structure around $\pm 20$–30 Å, between the transmembrane helices and the globular peripheral domains, that contains the loops connecting the transmembrane helices to each other (cf., Fig. 1).

The location of the ends of the transmembrane helices is shown in Figure 2 (middle panel). Helix ends are found in a region about 15–25 Å from the center of the membrane.

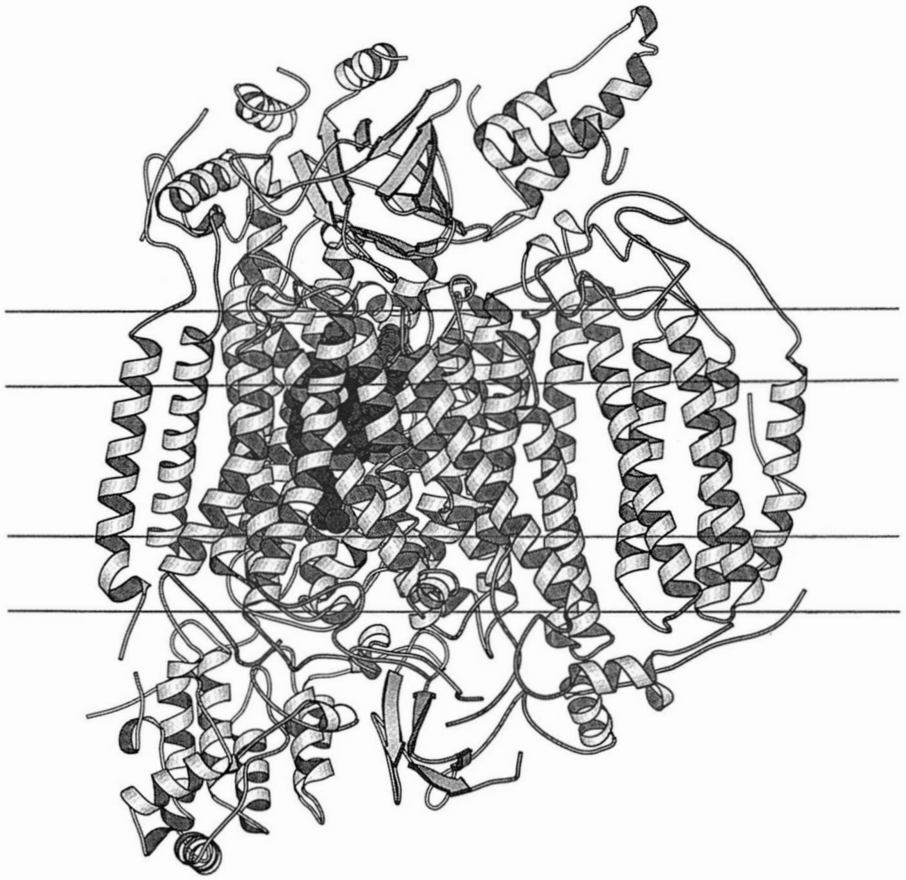

**Fig. 1.** Mitochondrial cytochrome *c* oxidase (Tsukihara et al., 1996). Only one subunit of the dimeric X-ray structure is shown. Heme groups are shown as space-filling models. The ±10 Å and ±20 Å regions are indicated. The inter-membrane space side is up. The picture was made using MOLSCRIPT (Kraulis, 1991).

Eight phospholipid molecules have been identified in the structure, and their locations are shown in Figure 2 (top panel), providing a rough indication of the extension of the lipid bilayer in vivo.

The hydrophobicity profile perpendicular to the membrane (Fig. 3, top panel), is broadly consistent with the location of the lipid bilayer suggested by the location of lipid molecules and the transmembrane helices. The mean hydrophobicity varies little over a region extending ~10 Å on either side of the center of the membrane, and then drops to a value typical for the extramembranous parts over the next 10 Å [a very similar picture is obtained when atom-based (Eisenberg & McLachlan, 1986) instead of residue-based hydrophobicities are used, not shown].

The curves in Figure 3 (bottom panel) show, respectively, the mean hydrophobicity for residues that are buried in the protein interior (<20% exposed surface area), exposed to an intermediate degree (between 20% and 50% exposed surface area), and fully exposed on the surface of the protein (>50% exposed surface area). In the membrane (±10 Å region), the lipid-exposed residues are on average more hydrophobic than those of intermediate exposure, which in turn, are more hydrophobic than the buried ones (cf., Fig. 9). In the parts located outside the bilayer, the buried residues have a much higher mean hydrophobicity than the other two classes. Buried residues within the membrane have a distinctly higher mean hydrophobicity than buried residues in the nonmembranous domains.

*Distribution of individual residue types*

Distribution profiles perpendicular to the membrane for buried and fully exposed hydrophobic residues (Phe, Val, Leu, Met, Ile), charged residues (Asp, Glu, Arg, Lys), polar aromatic residues (Trp, Tyr), amidated residues (Asn, Gln), and Pro are shown in Figure 4. Moving out from the center of the membrane towards the periphery, there is a regular succession of different residue types on both sides of the membrane. The central ±10 Å region is dominated by hydrophobic residues. Around ±12–15 Å there is a concentration of polar aromatic residues (Phe, in contrast, has the same distribution as the aliphatic residues, not shown) and amidated residues (Asn, Gln), closely followed by the appearance of charged residues around ±15–20 Å. Even further out, Pro peaks around ±25–30 Å, i.e., just outside the helix ends. Other residues such as Ser, Thr, Ala, and Gly are more or less evenly distributed across the structure (data not shown). It is noteworthy that fully exposed Trp and Tyr residues are concentrated in the lipid headgroup region (around ±10–14 Å) and that all Pro residues in the central domain (±10 Å) are buried.

The distribution of individual residue types in the central membrane domain (±10 Å) between buried, intermediately exposed, and fully exposed locations is shown in Figure 5. Although the differences are small, some trends are, nevertheless, apparent. Aliphatic residues (Leu and Ala in particular) tend to be more prominent in exposed locations, aromatic residues (Phe, Trp, Tyr) appear

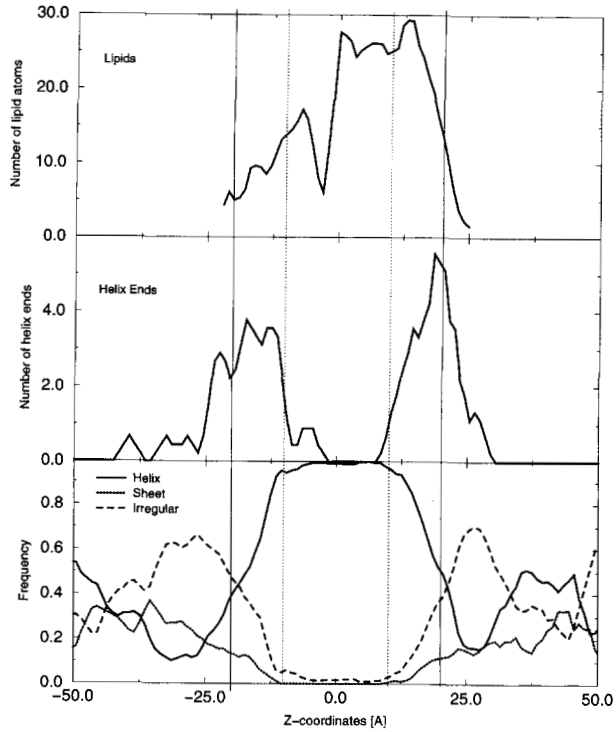


**Fig. 2.** Secondary structure and location of lipid molecules in the m-COX dimer. The location of the middle of the membrane-embedded domain is set to zero, and distances are given in Å. Top panel: lipid mass distribution (number of atoms) in 1 Å slices cut out in the plane of the membrane. Middle panel: location of the ends of the transmembrane helices (number of helix ends) in 1 Å slices. Bottom panel: fraction of secondary structures (helix, sheet, irregular) in 1 Å slices. All curves were smoothed by five-point, hat-shaped running window averaging (see Materials and methods). The vertical lines mark the regions defined by distances of ±10 Å and ±20 Å from the center of the membrane (see text).

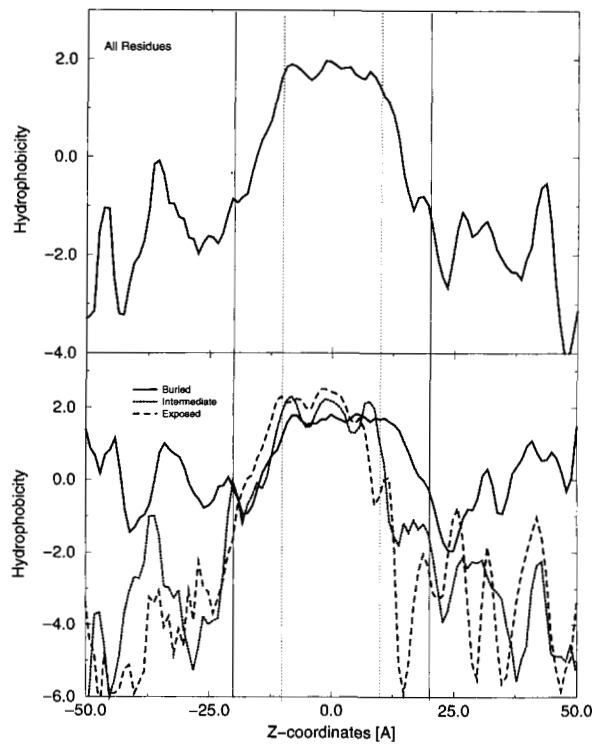


**Fig. 3.** Top panel: average hydrophobicity in 1 Å slices smoothed as in Figure 2. Bottom panel: average hydrophobicity for buried (continuous line), intermediately exposed (dotted line), and fully exposed (broken line) residues.

to prefer intermediately exposed locations, and the rare charged and amidated residues are found exclusively among the buried or intermediately exposed residues. Gly and Pro are also more often buried than exposed. Interestingly, Ser and Thr are quite abundant in the membrane domain and have no clear preferences for buried or exposed locations.

From these results, one can propose a highly symmetrical consensus structure that should be representative for large transmembrane proteins like cytochrome *c* oxidase: a central, helical region of about 25 Å rich in aliphatic residues and Phe, an "aromatic belt" composed of Trp and Tyr (together with Asn and Gln) similar to but less prominent than that seen in the β-barrel porin structures (Weiss et al., 1991), a region near the helix ends with charged residues, and helix breaking Pro residues just outside the helices in a region rich in irregular structure. The lipid-exposed surface is somewhat enriched for aliphatic residues compared to the buried residues in the membrane-embedded domain, whereas charged and amidated residues, together with Gly and Pro, are rarely exposed to lipids.

The only clear asymmetry between the two sides of the cytochrome *c* oxidase structure is related to the "positive inside" rule (von Heijne, 1986): the matrix-facing parts of the mitochondrially encoded subunits have a higher content of Arg and Lys residues than have the parts facing the inter-membrane space (data not shown). This is not true for the nuclearly encoded subunits, however (cf., Gavel & von Heijne, 1992).

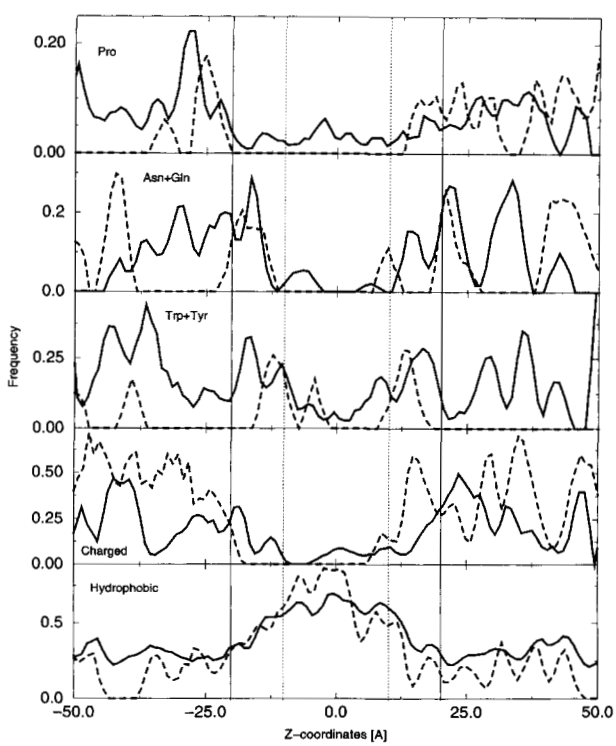


**Fig. 4.** Residue frequencies for buried (full line) and exposed (dashed line) residues in 1 Å slices smoothed as in Figure 2.

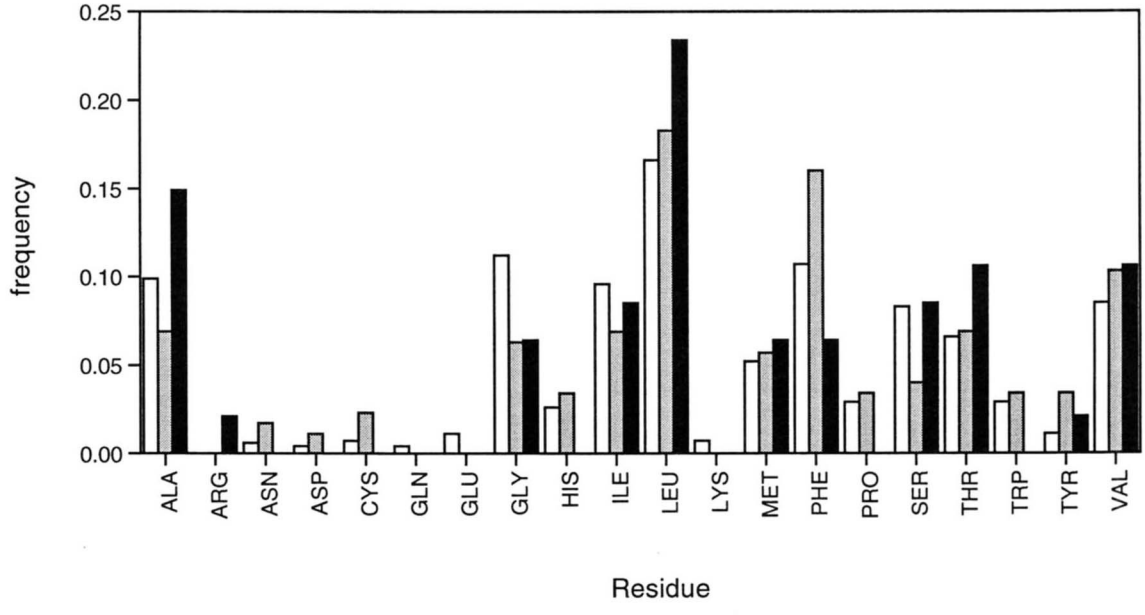


**Fig. 5.** Residue frequencies in the central ±10 Å region for buried (white bars), intermediately exposed (light gray bars), and fully exposed (black bars) locations.

*Residue variability in the membrane domain*

To evaluate the correlation between residue variability and degree of lipid exposure (Rees et al., 1989b), a set of mammalian COX sequences were aligned with the bovine m-COX subunits and the degree of conservation in each position was estimated by calculating the information content (see Materials and methods). As expected, the buried residues in the membrane domain (±10 Å) are on average less variable than the intermediately exposed and exposed ones (Fig. 6). Eighty-two percent of the buried residues, 60% of the intermediately exposed residues, and only 38% of the fully exposed residues are invariant (data not shown).

A relatively large fraction of the surface residues in the membrane domain of the core subunits I, II, and III are buried underneath the peripheral, nuclearly encoded subunits. These buried residues have a lower variability than the core subunit residues that are lipid-exposed (information content equals 3.95 and 3.89, respectively), again consistent with the expected correlation.

*Amino acid preferences around helix ends*

Helices in globular proteins have distinct N- and C-terminal amino acid preferences (Richardson & Richardson, 1988). Because the transmembrane helices in m-COX protrude outside the core of the lipid bilayer (cf., Figs. 2 and 3), similar preferences may be expected also in this case, although other characteristics of the helices such as length and overall amino acid composition are very different. This indeed appears to be the case. Pro is found just outside the N-terminus and in the N1–N2 positions but not among the C-terminal helical residues or in the C-1 position (Fig. 7), as in globular proteins (MacArthur & Thornton, 1991). Asp, Gly, Asn, and Ser, all typical N-cap residues in globular proteins (Richardson & Richardson, 1988; Doig & Baldwin, 1995), are found in 64% of the N-1 positions (data not shown). Gly is also very frequent in the C-2 position (32%), again as found in globular proteins (Richardson & Richardson, 1988; Aurora et al., 1994).

*Other membrane proteins: The photosynthetic reaction center and bacteriorhodopsin*

Of the other helix bundle membrane proteins, only bacteriorhodopsin (BR) and the photosynthetic reaction center (PRC) have a sufficiently large number of transmembrane helices to be useful in this kind of statistical analysis. In order to assess the generality of

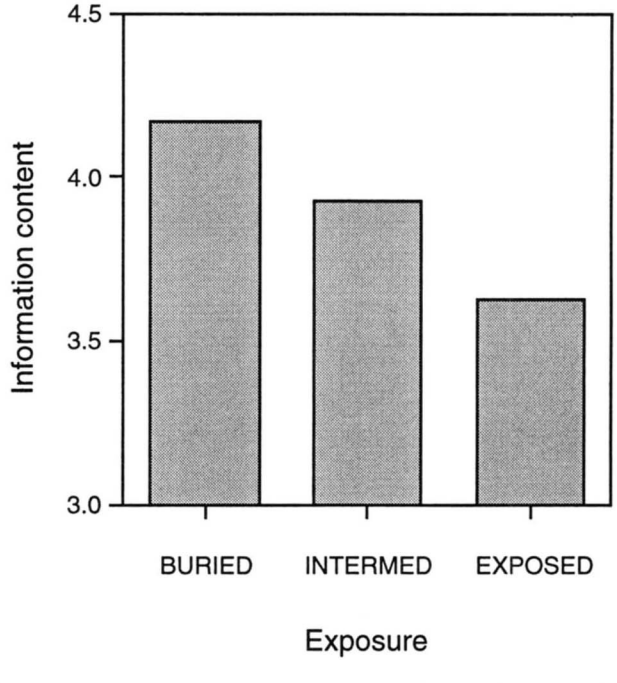


**Fig. 6.** Residue variabilities measured as average information content (see Materials and methods) in the central ±10 Å region for buried, intermediately exposed, and fully exposed locations in the m-COX dimer based on multiple alignments of mammalian m-COX subunits.

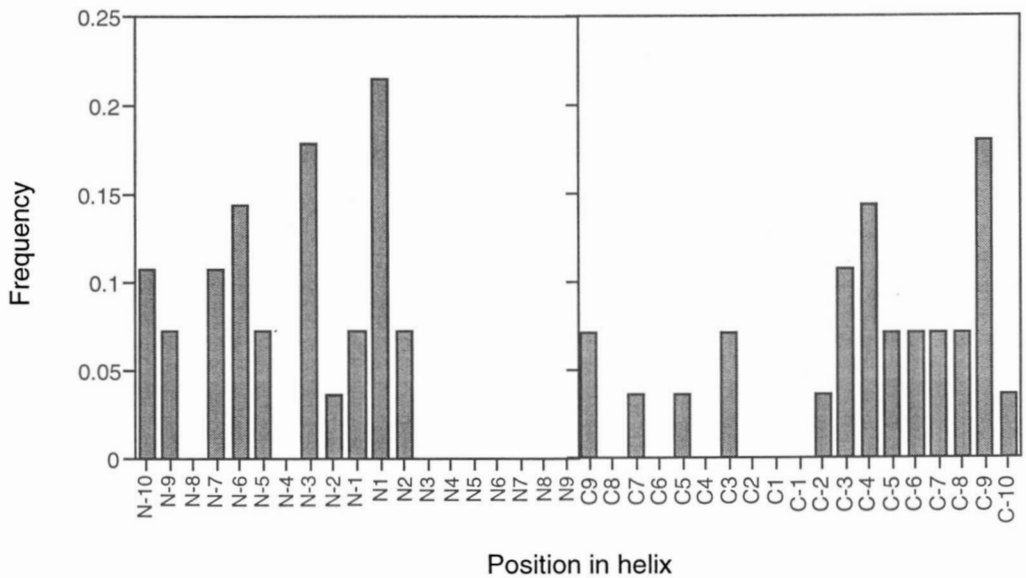


**Fig. 7.** Frequency of proline residues near the ends of transmembrane helices. Residues in helical conformation have positive numbers.

the findings reported above, the same analysis was performed on these two proteins.

The hydrophobicity profiles of BR and PRC are very similar to cytochrome *c* oxidase (Fig. 8). In general, the individual residue distributions are also in accordance with the consensus structure suggested above (data not shown). The overall residue preferences for the central membrane domain (±10 Å) versus the peripheral domains outside the ±20 Å limits for BR, PRC, and m-COX taken together are shown in Table 1. All the aliphatic residues plus Phe and Trp are highly enriched in the membrane domain, whereas the charged and amidated residues plus Pro predominate in the peripheral domains. These preference parameters are similar to widely used hydrophobicity scales (Kyte & Doolittle, 1982; Engelman et al., 1986; Rao & Argos, 1986; von Heijne, 1992; Persson & Argos, 1994) with correlation coefficients of 0.8–0.9 (data not shown).

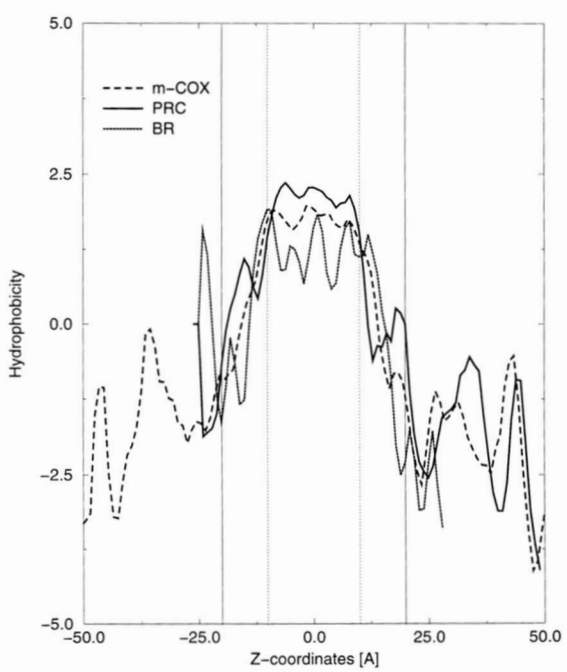


**Fig. 8.** Average hydrophobicities of m-COX (broken line), PRC (continuous line), and BR (dotted line) in 5 Å slices.

**Table 1.** *Overall preference parameters for distribution of residues between the central ±10 Å and the peripheral domains more than 20 Å from the center of the membrane for m-COX (monomeric form), BR, and PRC taken together*[a]

| Residue | Preference |
|---|---|
| Phe | 1.30 |
| Leu | 0.86 |
| Ile | 0.86 |
| Ala | 0.56 |
| Trp | 0.52 |
| Met | 0.44 |
| Val | 0.33 |
| Thr | 0.18 |
| Gly | 0.16 |
| Ser | 0.16 |
| Tyr | −0.39 |
| Cys | −0.41 |
| His | −0.70 |
| Pro | −1.30 |
| Asn | −1.71 |
| Asp | −2.05 |
| Arg | −2.10 |
| Lys | −2.69 |
| Glu | −2.78 |
| Gln | −3.00 |

[a]Preference parameters were calculated as $\ln(p)$, where $p = f_i^{mem}/f_i^{out}$, $f_i^{mem}$ is the frequency of residue type $i$ for $z < \pm 10$ Å, and $f_i^{out}$ is the frequency of residue type $i$ for $z > \pm 20$ Å (c.f., Fig. 1).

The mean hydrophobicity of the buried residues in the central hydrophobic region (±10 Å) is lower than that of the exposed residues in all three proteins (Fig. 9). A similar pattern is seen for the trimeric β-barrel protein OmpF (Cowan et al., 1992), except that the buried residues in the β-sheet tend to be less hydrophobic than the buried residues in the helix bundle proteins. For BR, the low hydrophobicity of the buried residues results mainly from a string of charged residues that are thought to act as a conduit for protons through the membrane (Grigorieff et al., 1996).

We conclude that the three proteins analyzed here have very similar architectures and that the proposed consensus structure should be representative for helix bundle membrane proteins in general.

## Discussion

m-COX is by far the largest integral membrane protein of known structure. Furthermore, although it does contain two heme groups, these are in contact with only a small fraction of the residues and the architecture is thus dominated by protein–protein and protein–lipid interactions. From a structural point of view, m-COX should thus be representative of many if not most helix bundle membrane proteins.

In this communication, we have analyzed the m-COX structure in terms of general characteristics such as secondary structure, residue distributions both between the membrane-embedded and non-membranous domains of the protein and between buried versus lipid-exposed locations, and residue variability. The results agree in broad terms with previous analyses of the photosynthetic reaction center (Rees et al., 1989b), but add considerable precision to the overall picture of membrane protein architecture.

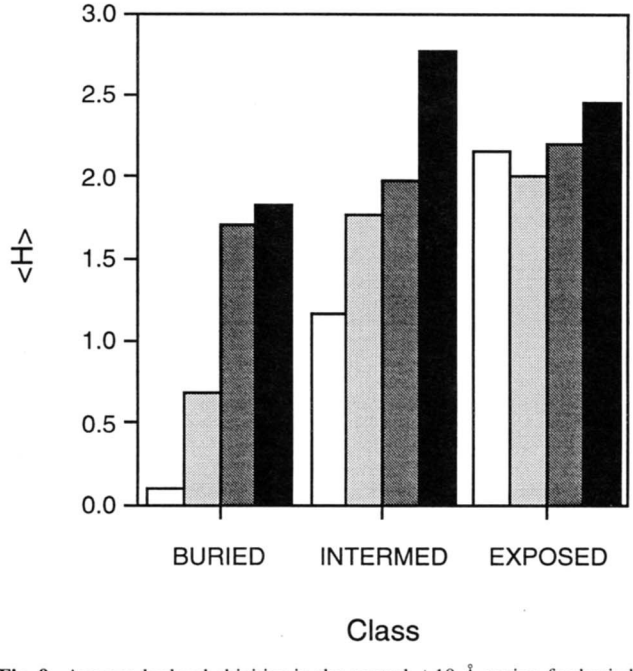


**Fig. 9.** Average hydrophobicities in the central ±10 Å region for buried, intermediately exposed, and fully exposed locations in the trimeric β-barrel outer membrane porin OmpF (white bars), BR (light gray bars), m-COX dimer (gray bars), and PRC (black bars). For OmpF, only residues defined as belonging to the β-sheet and not exposed to the central pore were included (see Materials and methods).

The bundle of 28 transmembrane α-helices in m-COX extends over a region of about ±20 Å from the center of the membrane. Hydrophobic residues are concentrated within a narrower, ~20 Å slice that is symmetrically flanked by regions containing an increased concentration of Tyr, Trp, Asn, and Gln residues (cf., Pawagi et al., 1994; von Heijne, 1994). Charged residues are mainly located outside the central hydrophobic and Tyr-Trp regions, near the ends of the transmembrane helices. Proline residues are relatively infrequent—although often conserved—in the transmembrane helices but become quite abundant just outside the helical region. The amino acid composition around the ends of the transmembrane helices is very similar to what has been observed in helices in globular proteins, and prediction methods intended to identify helix ends in globular proteins may thus be applicable also to membrane proteins.

The mean hydrophobicity of the residues in the central hydrophobic section increases with their degree of lipid exposure. Leu and Val are somewhat more common among the exposed residues, whereas the charged and amidated residues (together with Gly and Pro) are rarely exposed to lipids. This is a somewhat different pattern than found in a recent study where the degree of exposure of residues in a large number of transmembrane segments extracted from the SwissProt database was estimated based on helical amphiphilicity (Samatey et al., 1995), but agrees with previous analyses of proline residues in bacteriorhodopsin and the photosynthetic reaction center (Woolfson & Williams, 1990; von Heijne, 1991).

A good correlation between lipid exposure and sequence variability is apparent in m-COX, just as in other integral membrane proteins (Rees et al., 1989a; Taylor et al., 1994). Variability and hydrophobic moment analysis thus appear to be generally valid first steps in modeling helix–helix interactions in helix bundle membrane proteins.

## Materials and methods

All calculations were performed with the PASSAR program available from the authors at http://www.biokemi.su.se/export.html.

### *PDB files*

The PDB files 2BRD (bacteriorhodopsin—BR), 1PSS (photosynthetic reaction center—PRC), 2OMF (OmpF porin), and a preliminary version of 1OCC (mitochondrial cytochrome *c* oxidase—m-COX) were used.

### *Surface area calculations*

The solvent accessible surface area for each atom (Conolly, 1983) was calculated with the fast method developed by Le Grand and Merz (Le Grand & Merz, 1993). The fraction of the surface area of a residue exposed to the surrounding media was calculated by comparing the exposed surface area in the protein with the exposed surface area for the residue in the tri-peptide Gly-X-Gly (Bowie et al., 1991). Residues were divided into three classes (Rees et al., 1989b): residues with more than 50% of the surface area exposed to the surrounding media were defined as exposed, residues with between 20% and 50% of their surface area exposed were defined as intermediately exposed, and residues with less than 20% of their surface area exposed were defined as buried. All heteroatoms present in the PDB files, except for those in water and lipid molecules,

were included in the surface area calculations. The dimeric m-COX structure was used in all surface area calculations.

For the outer membrane $\beta$-barrel protein OmpF, only residues defined as belonging to the $\beta$-sheet (Cowan et al., 1992) were included in the calculations and residues exposed to the central pore were excluded. All residues belonging to the exposed class in the central $\pm 10$ Å region thus face the lipid, while buried residues are either internal to the monomer or situated in the trimer interface.

*Orientation of the proteins along a common z-axis*

Each transmembrane helix was approximated by a vector running between the average position of the $C_\alpha$ atoms of the three first residues of the helix and the average position of the $C_\alpha$ atoms of the three last residues. The vectors were oriented pointing towards the intermembrane space side in m-COX and the periplasmic side in BR and PRC. A global orientation vector was calculated as the vector sum of the individual transmembrane helix vectors, and the whole molecule was oriented with this vector oriented along the $z$-axis, i.e., perpendicular to the plane of the membrane. The $z = 0$ plane was defined as the midpoint in the $z$-direction of the part of the molecule with an average hydrophobicity $\geq 1.5$ (see below) (cf. Figs. 3 and 9). For m-COX, the orientation was calculated for the dimer structure.

The OmpF $z$-axis was oriented as in the PDB file and the $z = 0$ plane was defined by hand to coincide with the middle of the lipid-exposed face of the $\beta$-barrel.

*Calculation of hydrophobicity*

The hydrophobicity of each residue was taken from the GES hydrophobicity scale (Engelman et al., 1986) and was assigned to the $z$-coordinate of the corresponding $C_\alpha$ atom. The average hydrophobicity profile along the $z$-axis was calculated by summing the hydrophobicities for all residues with $z$-coordinates within 1 Å slabs, followed by smoothing using a 5 Å wide, hat-shaped window (Claverie & Daulmiere, 1991). Thus, for a slab centered at $z$-coordinate $i$ (measured in Å), the smoothed average hydrophobicity was calculated as:

$$\langle H(i)\rangle = \tfrac{1}{9}(3h(i) + 2(h(i-1) + h(i+1)) + h(i-2) + h(i+2))$$

where $h(i)$ is the average hydrophobicity of all residues contained within the 1 Å slab centered at $z$-coordinate $i$. $\langle H(i)\rangle$ values were also calculated separately for exposed, intermediately exposed, and buried residues.

*Secondary structure assignments*

The positions of the ends of the transmembrane helices in m-COX were taken from Figure 2 from Tsukihara et al. (1996). Other elements of secondary structure were calculated by the DSSP program (Kabsch & Sander, 1983).

*Calculation of amino acid distributions*

Each amino acid was assigned to the $z$-coordinate of the $C_\alpha$ atom. The number of buried and exposed residues of each type in each 1 Å slab in the $z$-coordinate were calculated and the corresponding frequencies were obtained by dividing this number by the total number of buried or exposed residues in the slice. A 5 Å wide, hat-shaped window was used to smooth the plots as described above.

*Multiple alignments and calculation of residue variability*

The mammalian COX sequences used to calculate residue variability were obtained by searching the latest cumulative update (as of August 15, 1996) of SwissProt (Bairoch & Boeckmann, 1991) with FASTA version 2.0u4 (Pearson & Lipman, 1988) using the default settings gap-penalty = $-12/-2$, ktup = 2, optcut = 25, cgap = 37 and the BLOSUM50 matrix.

Each of the bovine m-COX subunits were used separately as a query sequence. The results were filtered to include only mammalian sequences. In order to produce reliable alignments, only sequences that were more than 60% identical to the bovine query sequence were selected. The number of sequences used in the variability analysis were: subunit I—11, subunit II—54, subunit III—11, subunit IV—4, subunit VIa—7, subunit VIc—4, subunit VIIa—4, subunit VIIb—2, subunit VIIc—4, subunit VIII—2.

The sequences for each subunit were aligned using Clustal W (Thompson et al., 1994). Only positions in the alignment that correspond to the bovine sequence were analyzed, i.e., insertions in sequences from other organisms were ignored. For each position, the information content (or non-randomness) $I$ was calculated as follows (Schneider et al., 1986):

$$I_n = I_0 + \Sigma(f(a,n) * \log_2 f(a,n))$$

where $f(a,n)$ is the frequency of amino acid $a$ in position $n$, $I_0 = \log_2(20) = 4.32$ bits, and the summation is over all 20 amino acids. A completely conserved position has $I = I_0$ and a position where all 20 amino acids are equally frequent has $I = 0$.

The membrane portion of the structure was analyzed with respect to information content by calculating average values for buried, intermediately exposed, and fully exposed residues (see above). Only residues in the central $\pm 10$ Å of the membrane spanning region were included.

Two more residue classes in the $\pm 10$ Å region were created by selecting residues in the mitochondrially encoded subunits that (i) are buried under nuclear encoded subunits and become fully exposed when the nuclearly encoded subunits are removed, and (ii) are fully exposed to the lipid bilayer irrespective or the presence or absence of the nuclearly encoded subunits.

## Acknowledgments

This work was supported by grants from the Magnus Bergvall Foundation to AE, from the Swedish Technical Sciences Research Council to AE and GvH, from the Swedish Natural Sciences Research Council, the Swedish Cancer Foundation, and the Göran Gustafsson Foundation to GvH, and from the Research for the Future, JSTS, to T.T.

## References

Allen J, Feher G, Yeates T, Komiya H, Rees D. 1987. Structure of the reaction center from *Rhodobacter sphaeroides* R-26: The protein subunits. *Proc Natl Acad Sci USA 84*:6162–6166.

Aurora R, Srinivasan R, Rose GD. 1994. Rules for $\alpha$-helix termination by glycine. *Science 264*:1126–1130.

Bairoch A, Boeckmann B. 1991. The SwissProt protein sequence data bank. *Nucleic Acids Res 19*:2247–2249.

Bowie JU, Luthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known 3-dimensional structure. *Science 253*:164–170.

Claverie J-M, Daulmiere C. 1991. Smoothing profiles with sliding windows: Better to wear a hat! *CABIOS 7*:113–115.

Conolly MJ. 1983. Analytical molecular surface calculation. *J Appl Crystallogr 17*:548–558.

Cowan SW, Rosenbusch JP. 1994. Folding pattern diversity of integral membrane proteins. *Science 264*:914–916.

Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP. 1992. Crystal structures explain functional properties of two *E. coli* porins. *Nature 358*:727–733.

Deisenhofer J, Epp O, Miki K, Huber R, Michel H. 1985. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas* viridis at 3Å resolution. *Nature 318*:618–624.

Doig AJ, Baldwin RL. 1995. N- and C-capping preferences for all 20 amino acids in $\alpha$-helical peptides. *Protein Sci 4*:1325–1336.

Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding. *Nature 319*:199–203.

Engelman DM, Steitz TA, Goldman A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu Rev Biophys Biophys Chem 15*:321–353.

Gavel Y, von Heijne G. 1992. The distribution of charged amino acids in mitochondrial inner membrane proteins suggests different modes of membrane integration for nuclearly and mitochondrially encoded proteins. *Eur J Biochem 205*:1207–1215.

Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R. 1996. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J Mol Biol 259*:393–421.

Iwata S, Ostermeier C, Ludwig B, Michel H. 1995. Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature 376*:660–669.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*:2577–2637.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Kühlbrandt W, Wang DN, Fujiyoshi Y. 1994. Atomic model of plant light-harvesting complex by electron crystallography. *Nature 367*:614–621.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol 157*:105–132.

Le Grand S, Merz KJ. 1993. Rapid approximation to molecular surface area via the use of booleans look-up tables. *J Global Optimiz 3*:49–66.

MacArthur MW, Thornton JM. 1991. Influence of proline residues on protein conformation. *J Mol Biol 218*:397–412.

McDermott G, Price SM, Freer AA, Hawthornthwaite-Lawless AM, Papiz MZ, Cogdell RJ, Isaacs NW. 1995. Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. *Nature 374*:517–521.

Pawagi AB, Wang J, Silverman M, Reithmeier RAF, Deber CM. 1994. Transmembrane aromatic amino acid distribution in P-glycoprotein. A functional role in broad substrate specificity. *J Mol Biol 235*:554–564.

Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA 85*:2444–2448.

Persson B, Argos P. 1994. Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J Mol Biol 237*:182–192.

Rao JK, Argos A. 1986. A conformational preference parameter to predict helices in integral membrane proteins. *Biochim Biophys Acta 869*:197–214.

Rees DC, DeAntonio L, Eisenberg D. 1989a. Hydrophobic organization of membrane proteins. *Science 245*:510–513.

Rees DC, Komiya H, Yeates TO, Allen JP, Feher G. 1989b. The bacterial photosynthetic reaction center as a model for membrane proteins. *Annu Rev Biochem 58*:607–633.

Richardson JS, Richardson DC. 1988. Amino acid preferences for specific locations at the ends of $\alpha$-helices. *Science 240*:1648–1652.

Samatey FA, Xu C, Popot J-C. 1995. On the distribution of amino acid residues in transmembrane $\alpha$-helix bundles. *Proc Natl Acad Sci USA 92*:4577–4581.

Schneider TD, Stormo GD, Gold L, Ehrenfeucht A. 1986. Information content of binding sites on nucleotide sequences. *J Mol Biol 188*:415–431.

Taylor WR, Jones DT, Green NM. 1994. A method for $\alpha$-helical integral membrane protein fold prediction. *Proteins Struct Funct Genet 18*:281–294.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*:4673–4680.

Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S. 1995. Structures of metal sites of oxidized bovine heart cytochrome *c* oxidase at 2.8 Å. *Science 269*:1069–1064.

Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S. 1996. The whole stucture of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. *Science 272*:1136–1144.

von Heijne G. 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J 5*:3021–3027.

von Heijne G. 1991. Proline kinks in transmembrane $\alpha$-helices. *J Mol Biol 218*:499–503.

von Heijne G. 1992. Membrane protein structure prediction—Hydrophobicity analysis and the positive-inside rule. *J Mol Biol 225*:487–494.

von Heijne G. 1994. Membrane proteins: From sequence to structure. *Annu Rev Biophys Biomol Struct 23*:167–192.

Weiss MS, Abele U, Weckesser J, Welte W, Schiltz E, Schulz GE. 1991. Molecular architecture and electrostatic properties of a bacterial porin. *Science 254*:1627–1630.

Woolfson DN, Williams DH. 1990. The influence of proline residues on $\alpha$-helical structure. *FEBS Lett 277*:185–188.

Yeates TO, Komiya H, Rees DC, Allen JP, Feher G. 1987. Structure of the reaction center from *Rhodobacter sphaeroides*: Membrane–protein interactions. *Proc Natl Acad Sci USA 84*:6438–6442.