Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure analysis of bovine heart cytochrome *c* oxidase at 2.8 Å resolution

Takashi Tomizaki,[a]† Eiki Yamashita,[a] Hiroshi Yamaguchi,[a]‡ Hiroshi Aoyama,[a]§ Tomitake Tsukihara,[a]* Kyoko Shinzawa-Itoh,[b] Ryosuke Nakashima,[b] Reiko Yaono[b] and Shinya Yoshikawa[b]

[a]Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita 565, Japan, and [b]Department of Life Science, Himeji Institute of Technology, and CREST, Japan Science and Technology Corporation (JST), Kamigohri, Akoh, Hyogo 687-12, Japan

† Present address: ESRF, 41 Avenue des Martyrs, 38027 Grenoble CEDEX 01, France.
‡ Present address: School of Science, Kwansei Gakuin University, 1-1-155, Uegahara, Nishinomiya, Hyogo 662-8501, Japan.
§ Present address: Harima-RIKEN, Kamigohri, Akoh, Hyogo 687-12, Japan.

Correspondence e-mail:
tsuki@protein.osaka-u.ac.jp

Received 18 July 1997
Accepted 1 May 1998

PDB Reference: bovine heart cytochrome c oxidase, 1occ.

The crystal structure of bovine heart cytochrome *c* oxidase has been determined at 2.8 Å resolution by the multiple isomorphous replacement (MIR) method with three heavy-atom derivatives. An asymmetric unit of the crystal has a molecular weight of 422 kDa. Eight heavy atoms as main sites of a CH_3HgCl derivative were clearly located by solving the difference Patterson function. The electron density obtained by the MIR method was refined by density modification, consisting of solvent flattening, histogram matching and non-crystallographic symmetry averaging. The enzyme exhibits a dimeric structure in the crystal. Out of 3606 amino-acid residues in 26 subunits in the dimer, 3560 residues were located in the electron-density map. The structure was refined by *X-PLOR*. The final R factor and the free R factor were 0.199 and 0.252 at 2.8 Å resolution, respectively. One monomer in the dimeric structure with a stronger packing interaction has a lower averaged temperature factor than the other, by 16 Å^2. The region ±12 Å from the centre of the transmembrane part is almost 100% α-helix, despite the glycine residue content being as high as 7.1% in the transmembrane region. The residues around haem a of animals have evolved away from those of bacteria in contrast with the residues of the haem a_3. The hierarchy of the structural organization of the enzyme complex has been proposed on the basis of intersubunit interactions.

1. Introduction

Bovine heart cytochrome *c* oxidase is a large multicomponent membrane protein complex consisting of 13 different polypeptide subunits, two haems a and two copper sites as redox active centres, and one zinc and one magnesium (Malmström, 1990; Fergason-Miller & Babcock, 1996). This membrane protein complex is a typical hybrid protein with three subunits encoded by a mitochondrial DNA and the other ten subunits by a nuclear DNA. The molecular mass of the enzyme calculated for the protein moiety is 204 005 kDa. This enzyme reduces O_2 to H_2O coupled with proton pumping across the mitochondrial inner membrane (Malmström, 1990; Fergason-Miller & Babcock, 1996). Knowledge of the crystal structure of the enzyme is crucial for understanding the reaction mechanism of this enzyme. We have reported the structures of metal centres and protein moieties at 2.8 Å resolution, and have proposed an intramolecular electron-transfer path, proton channels, a water channel and oxygen channels (Tsukihara *et al.*, 1995, 1996). Here we report procedures used in the crystal structure analysis of the enzyme, structural features of the enzyme as a membrane protein, a possible

Table 1
Summary of the intensity data.

The first frame of series A is assigned to the standard data in the scaling of whole frames. The results, except for the first and final shots of each series, are omitted from the table for convenience. Six different series of oscillation photographs were taken from three crystals. Series A, B and C were taken by one crystal, series D and E by another crystal, series F by the other crystal. ΔB is in Å^2 and is a relative temperature factor against the first frame of series A. R factor is evaluated by the equation $R = \sum\sum|\langle I(hkl)\rangle - I_i(hkl)|/\sum\sum I_i(hkl)$.

Series	Number of shots	Shot	Scale	ΔB	Mosaicity	Reflections	$\langle I/\sigma(I)\rangle$	R factor
A	10	First	1.000	0.0	0.073	4664	21.1	0.058
		Final	0.981	3.7	0.068	4601	22.2	0.075
B	20	First	1.011	−1.0	0.062	4880	19.9	0.068
		Final	1.045	7.2	0.094	4945	19.0	0.078
C	23	First	1.218	−2.0	0.073	5132	17.1	0.060
		Final	1.689	7.2	0.123	4943	14.6	0.057
D	22	First	1.358	−1.3	0.070	4675	21.2	0.066
		Final	1.055	6.6	0.071	4843	20.8	0.078
E	17	First	1.830	−3.4	0.077	4965	19.8	0.060
		Final	2.034	9.8	0.119	4949	17.6	0.080
F	15	First	0.351	1.3	0.286	3459	25.3	0.083
		Final	0.201	4.7	0.289	2123	26.6	0.097

mechanism for structure organization of the hybrid protein, molecular evolutional aspects of the haem pockets and molecular packing in the crystal.

2. Experimental and structure determination

2.1. Intensity data collection and processing

The crystal prepared by the method described previously (Tsukihara *et al.*, 1995) belongs to the orthorhombic space group $P2_12_12_1$ with cell dimensions of a = 189.1 (1), b = 210.5 (1) and c = 178.6 (1) Å. Intensity data for the native crystals and heavy-atom derivatives were collected on the BL6A2 station at the Photon Factory, by the oscillation method using a Weissenberg camera for macromolecules (Sakabe, 1983) equipped with an imaging plate (Fuji Film Co. Ltd). The cylindrical cassettes with radii of 540 or 870 mm were filled with helium gas to increase the signal-to-noise ratio. X-rays of wavelength 1.00 Å were used with a 0.1 mm diameter collimator. Each oscillation image was taken with an exposure time of 40–100 s without alignment of the rotation axis. The oscillation range of each image was 1.0° without any overlap between adjacent images. The experimental hatch was kept at 280–283 K during the X-ray experiments. A fresh position was used for every 15–35 images and rotation ranges of 20–100° were obtained from one crystal. Each oscillation frame was processed by *DENZO* (Otwinowski, 1993). Post-refinement of the crystal-setting parameters and mosaicity were carried out by *SCALEPACK* (Otwinowski, 1993), and intensity data sets were scaled by the method of Hamilton *et al.* (1965). Local scaling was applied to provide an absorption correction in order to detect Bijvoet differences owing to Zn,

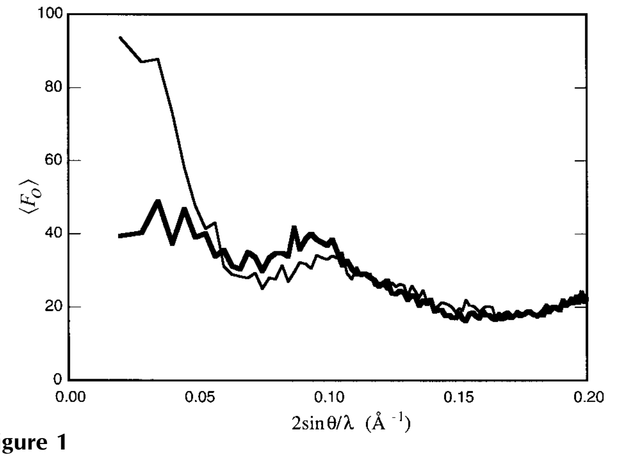


Figure 1
Statistics of observed structure factors of the native crystal. Structure amplitudes for the reflections with $h + l$ even and with $h + l$ odd were averaged within each resolution shell, and were plotted on an arbitrary scale against $2\sin\theta/\lambda$ as a thin and heavy line, respectively. Observed structure amplitudes for $h + l$ even are significantly larger than those for $h + l$ odd in the resolution range <20 Å.

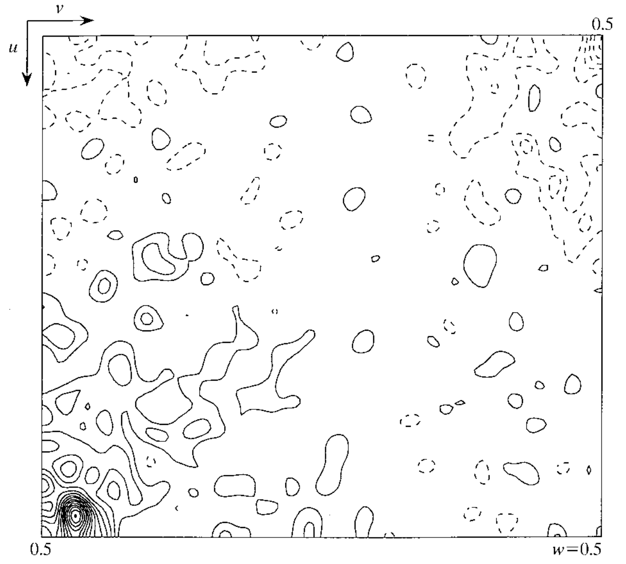


Figure 2
A Patterson function section (w = 0.5) of the native crystal calculated at 5 Å resolution, ranging from 0.0 to 0.5 of the fractional unit cell. Contours are drawn at an equal interval 1σ starting from 1σ. The peak with 14σ at (0.48, 0.03, 0.47) is the highest, except for the origin peak with 187σ, which implies that the crystal is a quasi B-centred lattice.

Table 2
Intensity data collection and phase determination by the multiple isomorphous replacement method.

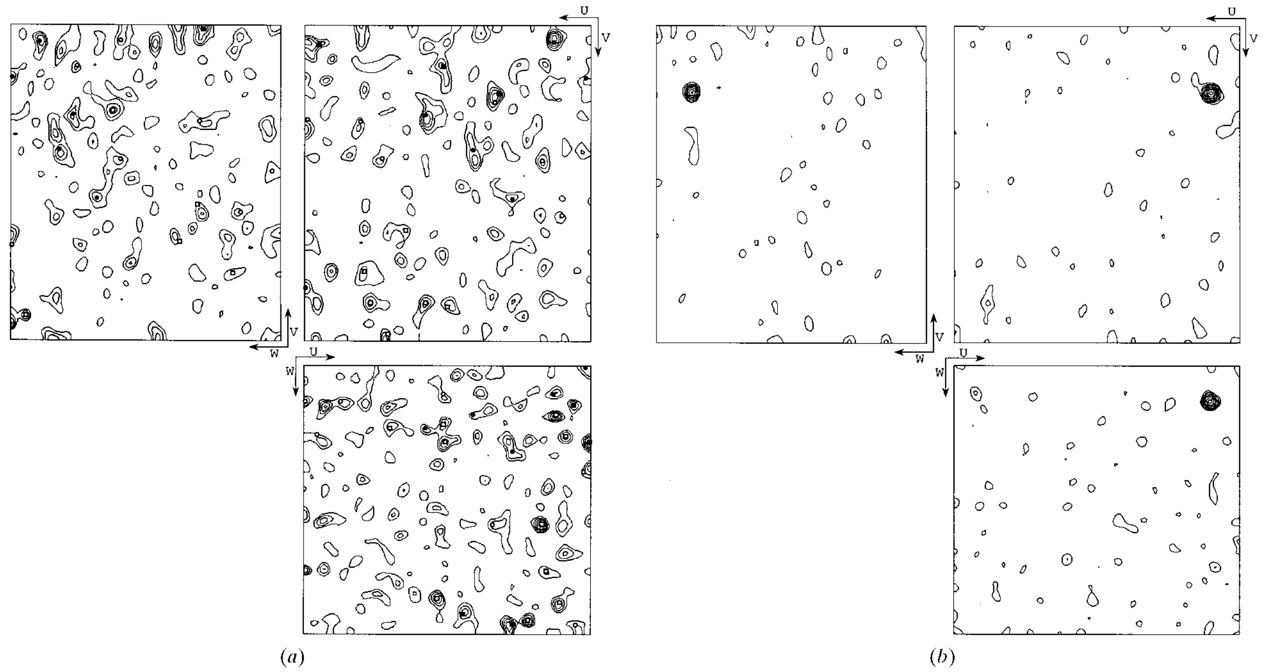
Redundancy is the number of observed reflections for each independent reflection. $R_{\rm merge}$, $\sum_h \sum_i |I(h,i) - \langle I(h)\rangle| / \sum_h \sum_i I(h,i)$, where $I(h, i)$ is the intensity value of the ith measurement of h and $\langle I(h)\rangle$ is the corresponding mean value of $I(h)$ for all i measurements; the summation is over the reflections with $I/\sigma(I) > 1.0$. $R_{\rm iso}$, $\sum |F_{\rm PH} - F_{\rm P}| / \sum F_{\rm PH}$, where $F_{\rm PH}$ and $F_{\rm P}$ are the derivative and the native structure-factor amplitudes, respectively. $R_{\rm Cullis}$, $\sum \big||F_{\rm PH} - F_{\rm P}| - |F_{\rm H}({\rm calc})|\big| / \sum |F_{\rm PH} - F_{\rm P}|$, where $F_{\rm H}$(calc) is the calculated heavy-atom structure factor. The summation is over the centric reflections only. Phasing power is the root-mean-square (r.m.s.) isomorphous difference divided by the r.m.s. residual lack of closure.

	Native	IrCl(I)	IrCl(II)	CH_3HgCl
Experiments				
oscillation angle (°)	1	1	1	1
exposure time (s)	40, 160	80	28	46
overlap (°)	0	0	0	0
Number of crystals	3	1	1	1
Number of frames	107	91	93	91
Resolution range (Å)	100–2.8	100–3.0	100–3.0	100–3.0
Observed reflections	518054	319358	381828	328484
Independent reflections	159742	125036	133245	127548
$I/\sigma(I)$	19.5	10.0	21.3	12.4
Averaged redundancy	3.2	2.6	2.9	2.6
Completeness (%)	90.1	86.3	93.0	88.3
$R_{\rm merge}$ (%)	7.8	9.0	6.2	8.5
$R_{\rm iso}$ (%)		7.3	5.7	9.0
Number of sites		9	1	23
$R_{\rm Cullis}$		0.90	0.86	0.79
Phasing power		0.38	0.54	0.74

Cu and Fe atoms in the native crystal. The results of the processing and scaling are summarized in Table 1. The crystal mosaicity and overall temperature factor increased gradually with exposure time. 92 frames each with a mosaicity less than 0.12° and with relative temperature factors lower than 10 $Å^2$ *versus* the standard frame were merged at first. Although an additional 15 frames had higher mosaicity, these frames gave low R factors and were included in the native data set in order to increase the overall completeness and redundancy. A total of 518 054 observed reflections from 107 frames were merged into 159 742 independent reflections with an $R_{\rm merge}$ of 7.8% at 2.8 Å resolution.

The observed structure factors of reflections with $h + l$ even are significantly smaller than those with $h + l$ odd at resolutions lower than 20 Å, as shown in Fig. 1, implying that molecules in the asymmetric unit are related by a pseudo B-centred symmetry. A Patterson function calculated at 5 Å resolution shows the pseudo translational vector at (0.479, 0.031, 0.500) with a height 0.08 times that of the origin peak as shown in Fig. 2. The crystal contains two molecules in an asymmetric unit and according to Matthews (1968) it contains 70% solvent by volume and its V_m is 4.14 $Å^3$ Da^{-1}.

Examining more than 50 heavy-atom derivatives with a MAC Science imaging-plate detector (DIP2000) on a Rigaku rotating-anode X-ray generator (RU300), we found more than



(*a*) (*b*)

Figure 3
Harker sections of difference Patterson functions of (*a*) the CH_3HgCl derivative and (*b*) the $IrCl_6$(II) derivative, ranging from 0 to 0.5 of the fractional unit cell. Closed circles are Harker peaks. Open squares indicate interatomic vectors on the Harker sections or their neighbouring sections. In (*a*) contours are drawn at an equal interval of 1σ starting from 1σ. Vectors between any two positions among the 32 main sites in the unit cell are depicted in the maps. Most vectors are located at the separated peaks. (*b*) The map contoured at 2σ from 2σ shows prominent Harker peaks representing a single main site of the heavy atom. The difference Patterson map is as clear as that of F1ATPase with a size of 372 kDa at 6.5 Å resolution (Abrahams & Leslie, 1996).

ten heavy-atom derivatives exhibiting significant intensity changes. Any derivative prepared in solutions containing such heavy-atom clusters as Mo_6Cl_{14} and di-μ-iodobis-(ethylenediamine)-di-platinum(II) were non-isomorphous judging from the magnitude of the pseudo *B*-centred translation vector in the Patterson maps. Three isomorphous derivative data sets were obtained using conventional heavy-atom reagents. Two Na_2IrC_6 derivatives were prepared by soaking crystals in heavy-atom solutions of 0.5 and 0.1 m*M*, respectively, and the other was a CH_3HgCl derivative. Each derivative data set was collected from one crystal. Data-collection statistics of the three derivative crystals as well as the native crystals are summarized in Table 2. Other statistics have been given in a previous paper (Tsukihara *et al.*, 1995).

Table 3
Restraining parameters for metal centres.

1 kcal = 4.184 kJ.

Bonds	Distance (Å)	Energy (kcal mol^{-1})	Angles	Distance (°)	Energy (kcal mol^{-1})
Cu—Cu	2.59	500.0	Cu—S^{γ}(Cys)—Cu	69.0	70.0
Cu—S^{γ}(Cys)	2.31	500.0	S^{γ}(Cys)—Cu—S^{γ}(Cys)	111.0	70.0
S^{γ}(Cys)—S^{γ}(Cys)	3.59	500.0			
Cu—S^{δ}(Met)	2.54	500.0			
Cu—N(His)	1.95	500.0			
Fe—N(His)	1.95	270.2			
Zn—S^{γ}(Cys)	2.60	500.0	S^{γ}(Cys)—Zn—S^{γ}(Cys)	109.5	70.0
			Zn—S^{γ}(Cys)—C^{β}(Cys)	100.0	70.0
Mg—O(Asp)	2.60	500.0	O(Asp)—Mg—O(Glu)	109.5	70.0
Mg—O(Glu)	2.60	500.0	N(His)—Mg—O(Glu)	109.5	70.0
Mg—N(His)	2.74	500.0	N(His)—Mg—O(Asp)	109.5	70.0
Mg—O(water)	2.69	500.0	O(water)—Mg—O(Glu)	109.5	70.0
			O(water)—Mg—N(His)	109.5	70.0

Table 4
Real-space correlation coefficients between the ideal and the MIR/*DM* electron densities (upper) and between the ideal and the $(2F_o - F_c)$ electron densities (lower).

Roman numerals with or without a letter represent subunits; pe with a number, phosphatidyl ethanolamine; pg with a number, phosphatidyl glycerol; ch with a number, cholate.

	MIR/*DM* map				$(2F_o - F_c)$ map			
	Monomer *A*		Monomer *B*		Monomer *A*		Monomer *B*	
Molecules	Main chain	Side chain	Main chain	Side chain	Main chain	Side chain	Main chain	Side chain
I	0.91	0.88	0.90	0.87	0.92	0.89	0.91	0.88
II	0.89	0.85	0.87	0.83	0.90	0.87	0.88	0.83
III	0.91	0.87	0.90	0.86	0.90	0.87	0.89	0.85
IV	0.88	0.82	0.83	0.77	0.91	0.86	0.85	0.79
Va	0.89	0.85	0.87	0.81	0.90	0.85	0.88	0.82
Vb	0.88	0.82	0.87	0.80	0.89	0.84	0.88	0.81
VIa	0.87	0.82	0.84	0.81	0.89	0.84	0.86	0.82
VIb	0.85	0.83	0.82	0.78	0.88	0.85	0.84	0.81
VIc	0.87	0.78	0.86	0.76	0.89	0.81	0.86	0.77
VIIa	0.87	0.79	0.82	0.74	0.88	0.81	0.82	0.75
VIIb	0.85	0.79	0.79	0.74	0.89	0.83	0.86	0.77
VIIc	0.88	0.80	0.84	0.77	0.91	0.85	0.87	0.78
VIII	0.89	0.79	0.84	0.74	0.91	0.84	0.87	0.75
Whole subunits	0.90	0.85	0.89	0.84	0.91	0.87	0.89	0.84
pe1†	0.86		0.84		0.85		0.84	
pe2†	0.87		0.84		0.86		0.84	
pe3†	0.77		0.74		0.78		0.78	
pe4†	0.80		0.79		0.81		0.80	
pe5†	0.79		0.75		0.80		0.78	
pe6†	0.84		0.80		0.85		0.80	
pg1†	0.88		0.88		0.89		0.79	
pg2†	0.79		0.79		0.80		0.79	
pg3†	0.78		0.78		0.84		0.82	
ch1	0.88		0.89		0.92		0.90	
ch2	0.92		0.92		0.92		0.94	
ch3	0.81		0.74		0.80		0.75	
ch4	0.79		0.78		0.75		0.74	

† The full lengths of the hydrocarbon tails of pe3, pe4 and pe6 are not included in the calculation, since the ends of the tails are not visible in the electron-density maps. The numbers of C atoms in the hydrocarbon tails excepting the carboxyl groups for each phospholipid are as follows: pe1, 17 and 19; pe2, 17 and 19; pe3, 11 and 5; pe4, 9 and 19; pe5, 17 and 19; pe6, 9 and 19; pg1, 15 and 17; pg2, 15 and 17; pg3, 15 and 17.

2.2. Phase determination

Determination of heavy-atom sites of each derivative was initiated by solving each difference Patterson function at 5 Å resolution with the program *PROTEIN* (Steigemann, 1992). Eight and three heavy-atom sites in an asymmetric unit were assigned independently for the CH_3MgCl derivative and the $IrCl_6$(I) derivative prepared in 1 m*M* K_2IrCl_6 solution, respectively, as interpreted from the difference Patterson maps shown in Fig. 3(*a*). The third derivative of $IrCl_6$(II) prepared in 0.1 m*M* K_2IrCl_6 solution contains only one main site of Ir clearly shown in the difference Patterson map (Fig. 3*b*). One of two possible absolute configurations was chosen for convenience by the difference Fourier method phased by the CH_3MgCl derivative.

Heavy-atom parameters of the three derivatives were refined and phases were calculated using data to 3.0 Å resolution with *MLPHARE* (Otwinowski, 1991) of the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). In spite of the low figures of merit, 0.56 at 5.0 Å resolution and 0.33 at 3.0 Å resolution, an electron-density map at 5 Å resolution clearly showed a 28-helix bundle, four haems and a dimeric structure in the asymmetric unit. Phases determined by the multiple

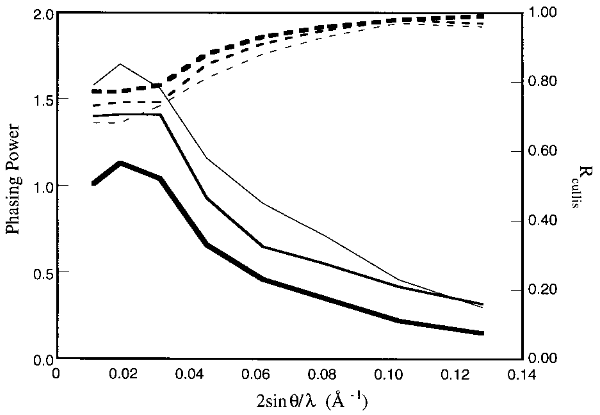


Figure 4
$R_{\rm Cullis}$ (solid lines) and phasing powers (broken lines) of the three derivatives are plotted as functions of $\sin\theta/\lambda$. Heavy, medium-weight and thin lines represent the CH_3HgCl, $IrCl_6$(I) and $IrCl_6$(II) derivatives, respectively. $R_{\rm Cullis}$ and phasing power are defined in Table 2.

Table 5
The averaged temperature factors of 13 subunits in each monomer.

Temperature factor $B = \exp[-B(\sin\theta/\lambda)^2]$.

	B^* of monomer A			B of monomer B		
Subunits	Main chain	Side chain	All	Main chain	Side chain	All
I	14.5	14.8	14.6	26.8	26.6	26.7
II	22.3	24.3	23.3	41.6	43.5	42.6
III	19.8	21.2	20.5	29.0	31.3	30.2
IV	25.5	29.0	27.3	60.6	63.2	61.9
Va	29.0	31.5	30.2	39.7	42.2	41.0
Vb	32.1	37.2	34.5	44.4	48.3	46.3
VIa	46.0	45.5	45.8	52.2	52.8	52.5
VIb	37.9	38.6	38.3	54.3	55.1	54.7
VIc	32.7	39.8	36.3	51.3	57.5	54.5
VIIa	31.6	40.4	35.9	50.2	56.3	53.2
VIIb	34.2	35.6	34.9	57.3	62.7	60.0
VIIc	27.6	31.9	29.8	51.8	58.1	55.0
VIII	28.3	32.5	30.3	56.0	63.2	59.5
Monomer	24.0	26.2	25.10	39.8	42.2	41.0
Dimer	31.9	34.20	33.05			

isomorphous replacement (MIR) method (Green *et al.*, 1954) were refined by solvent flattening (Wang, 1985) and histogram matching (Zhang & Main, 1990) with the program *DM* of *CCP*4. The phases were extended from 5.0 to 2.8 Å in 200 small steps, using a protein mask determined by the method of Wang (1985) with a sphere of 10 Å and 70% solvent content. MIR phases were combined with the refined phases during the phase extension. A free *R* factor (Brünger, 1992) for 5% of the reflections in each resolution shell was reduced from 0.53 to 0.29. A difference Fourier map of each derivative was calculated using the refined phases in order to find remaining heavy-atom sites. The heavy-atom parameters were refined again by *MLPHARE* (Otwinowski, 1991). The CH_3HgCl derivative has 23 heavy-atom sites and the highest phasing power of 0.74 and the lowest $R_{\rm Cullis}$ of 0.79 among the three derivatives. The $IrCl_6$(I) derivative contains one main site and eight minor sites. The $IrCl_6$(II) derivative has only one heavy-atom site at the same position as the main site of the $IrCl_6$(I) derivative.

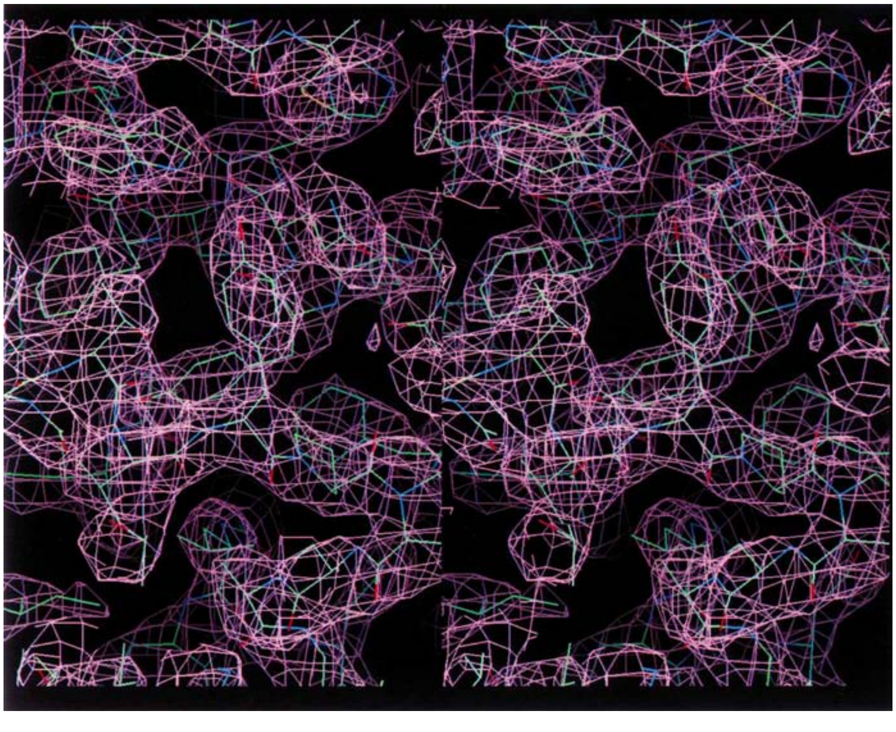


Figure 5
Stereoscopic drawing of MIR/*DM* electron density together with a wire model of the enzyme complex. Cages are drawn at the 2.35σ level. C, N and O atoms are coloured in green, blue and red, respectively. The model fits the cages well. The phenol groups of the tyrosine residues are clearly distinguished from the phenyl groups of the phenylalanine residues in the map.

The phasing powers and $R_{\rm Cullis}$ values at the final stage of the refinement are plotted *versus* resolution in Fig. 4. The MIR phases based on the new heavy-atom parameters were refined by the same procedure as before, but a sphere of 8 Å was applied instead of 10 Å. The free *R* was reduced to 0.28 at 2.8 Å resolution. A molecular envelope for non-crystallographic symmetry (NCS) averaging was generated using *MAP VIEW* of *PHASES* (Furey, 1980) and the overlap between envelopes related by crystallographic symmetry operations was inspected manually. The relation between the two molecules in the asymmetric unit was determined using *RAVE* (Kleywegt & Jones, 1994) and the correlation coefficient $\{\sum[(\rho_1 - \langle\rho_1\rangle)(\rho_2 - \langle\rho_2\rangle)]/\sum[(\rho_1 - \langle\rho_1\rangle)^2(\rho_2 - \langle\rho_2\rangle)^2]\}^{1/2}$ was 0.78. The phase-refinement procedure including NCS averaging (Bricogne, 1974, 1976) reduced the free *R* factor to 0.26 and raised the correlation coefficient to 0.91. The phase angles obtained by *DM* deviate by 45 and 71° from the initial MIR phases in the resolution ranges of 15–5 and 5–3 Å, respectively.

Table 6
Heavy-atom parameters and distances between heavy atoms and protein atoms.

Q is occupancy in arbitrary units. Heavy-atom locations are represented by M, C, and T matrix side, intermembrane side and transmembrane space, respectively.

Derivative	x	y	z	Q	B	Monomer and subunits	Residues	Atoms	Distances (Å)	Location
$IrCl_6$(I)	0.275	0.447	0.783	2.05	48	*A*-VIIc	Glu16	OE1	2.81	C
								OE2	2.05	
	0.204	0.394	0.780	0.73	167	*A*-I	Met1	SD	1.69	C
						A-VIIc	His2	O	3.91	
	0.242	0.074	0.743	0.86	233	*B*-I	Met1	SD	2.25	C
						B-VIIc	His2	O	3.34	
	0.018	0.449	0.430	0.67	185	*A*-III	Cys115	SG	3.67	C
	0.021	0.020	0.071	0.58	177	*B*-III	Cys115	SG	3.27	C
	0.001	0.426	0.714	0.44	99	*A*-III	Glu236	OE1	1.69	M
						A-III	His148	NE2	2.35	
								CE1	1.86	
	0.032	0.040	0.785	0.51	129	*B*-III	Glu236	OE1	1.71	M
						B-III	His148	NE2	2.87	
								CE1	2.62	
	0.470	0.068	0.281	0.51	158	*A*-III	Met152	SD	1.98	M
	−0.001	0.036	0.779	0.51	168	*B*-III	Met152	SD	2.24	M
$IrCl_6$(II)	0.275	0.447	0.783	2.08	44	*A*-VIIc	Glu16	OE1	2.81	C
								OE2	2.05	
CH_3HgCl	0.393	0.429	0.308	2.16	84	*A*-Vb	Cys60	SG	2.41	M
						A-Vb	Cys85	SG	3.26	
						A-I	Lys514	NZ	3.00	
								CE	2.95	
						A-Vb	Tyr89	OH	3.76	
						A-Vb	Thr87	OG1	3.23	
						A-Vb		Zn	3.02	
	0.353	0.099	0.205	1.86	75	*B*-Vb	Cys60	SG	2.13	M
						B-Vb	Cys85	SG	3.66	
						B-I	Lys514	NZ	2.71	
								CE	2.98	
						B-Vb	Tyr89	OH	3.10	
						B-Vb	Thr87	OG1	3.29	
						B-Vb		Zn	3.21	
	0.283	0.488	0.277	1.76	34	*A*-I	Cys498	SG	2.30	M
	0.245	0.041	0.250	1.52	51	*B*-I	Cys498	SG	3.10	M
	0.181	0.363	0.671	1.72	69	*A*-VIIa	Met36	SD	3.98	T
								O	3.79	
						A-III	Leu22	CD2	2.65	
						A-VIIa	Leu40	CD2	2.92	
	0.206	0.107	0.853	1.50	69	*B*-VIIa	Met36	SD	4.15	T
								O	3.66	
						B-III	Leu22	CD2	2.78	
						B-VIIa	Leu40	CD2	2.83	
	0.225	0.350	0.538	1.11	48	*A*-VIIc	Phe55	O	3.53	C
						A-VIIc	Trp52	O	3.14	
	0.237	0.118	0.990	1.22	79	*B*-VIIc	Phe55	O	3.15	C
						B-VIIc	Trp52	O	3.01	
	0.166	0.439	0.100	0.80	75	*A*-VIIb	Cys25	SG	2.10	T
	0.148	0.092	0.440	0.70	91	*B*-VIIb	Cys25	SG	1.99	T
	−0.001	0.427	0.718	0.78	76	*A*-III	His232	NE2	3.20	M
								CE1	2.32	
						A-III	Glu236	OE1	2.05	
						A-III	His148	NE2	2.86	
								CE1	2.46	
	0.031	0.042	0.785	0.75	90	*B*-III	His232	NE2	3.33	M
								CE1	2.39	
						B-III	Glu236	OE1	1.88	
						B-III	His148	NE2	2.71	
								CE1	2.31	
	0.377	0.406	0.276	0.69	138	*A*-Vb	Asn66	OD1	3.39	M
	0.194	0.359	0.577	0.59	111	*A*-III	Met33	SD	2.72	C
						A-VIIa	Cys49	SG	3.53	
								CB	2.47	
	0.461	0.064	0.315	0.52	62	*A*-III	His149	NE2	2.76	M
								CE1	2.09	
	0.490	0.467	0.187	0.56	86	*B*-III	His149	NE2	2.98	M
								CE2	2.20	
	0.432	0.105	0.429	0.47	78	*A*-VIa	Cys31	SG	1.99	C
	0.450	0.425	0.077	0.53	88	*B*-VIa	Cys31	SG	1.49	C

Table 6 (continued)

Derivative	x	y	z	Q	B	Monomer and subunits	Residues	Atoms	Distances (Å)	Location
	0.243	0.489	0.149	0.47	72	*A*-I	His413	NE2	2.38	T
								CE1	2.42	
						A-I	Met468	SD	3.88	
						A-I	Met417	SD	3.81	
						A-I	Met390	SD	3.59	
								CE	2.39	
	0.219	0.043	0.379	0.43	89	*B*-I	His413	NE2	2.38	T
								CE1	2.42	
						B-I	Met468	SD	3.88	
						B-I	Met417	SD	3.81	
						B-I	Met390	SD	3.59	
								CE	2.39	
	0.367	0.427	0.301	0.47	35	*A*-Vb	Cys82	SG	1.10	M
						A-Vb	Ser84	OG	2.41	
						A-Vb	Cys62	SG	2.79	
						A-Vb		Zn	2.07	
	0.328	0.103	0.215	0.47	49	*B*-Vb	Cys82	SG	0.82	M
						B-Vb	Ser84	OG	2.74	
						B-Vb	Cys62	SG	3.07	
						B-Vb		Zn	2.03	
	0.176	0.444	0.418	0.28	67	*A*-II	Tyr105	OH	3.91	C
						A-II	His102	CD2	1.85	

2.3. Model building and refinement

The electron-density map (MIR/*DM* map) refined by the density-modification procedures will be referred to as the MIR/*DM* electron-density map in this paper, and it was used for model building with *TurboFrodo* (Jones, 1978) and *X-Fit* in *QUANTA* (CTC Laboratory Systems Co., Ltd). The metal sites of Cu$_A$, Cu$_B$, haem a, haem a$_3$ and Zn were clearly distinguished from those of phosphate groups in the native anomalous difference Fourier map calculated at 4.5 Å resolution with the coefficients of $m[F_o(hkl) - F_o(hkl)]\exp[-i(\alpha p - \pi/2)]$ as shown in a previous paper (Tsukihara *et al.*, 1995). Most of the carbonyl groups were identified by slightly protruding shapes in the electron density, and most of the phenol groups of tyrosine residues were distinguished from phenyl groups of phenylalanine residues in the electron-density map as shown in Fig. 5. Out of 1803 amino-acid residues of 13 subunits in the monomer, 1780 residues were located in the electron-density map. Five phosphatidyl ethanolamines, three phosphatidyl glyceroles, two cholic acids and

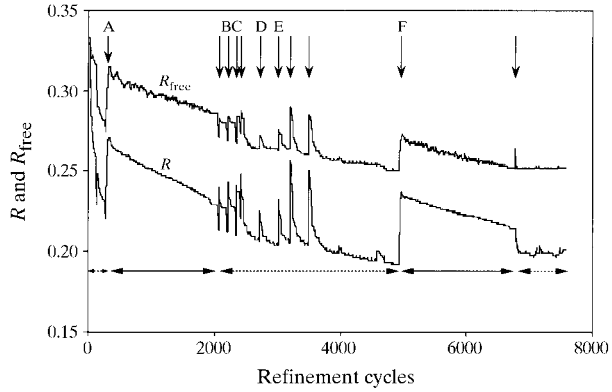


Figure 6
A summary of the *X-PLOR* refinement at 2.8 Å resolution is shown by plotting conventional R factors and free R factors *versus* refinement cycle, where $R = \sum \big||F_o(hkl)| - |F_c(hkl)|\big| / \sum |F_o(hkl)|$, summation for the conventional R is taken for all the reflections included in the refinement and the summation for the free R is done for the reflections excluded from the refinement. The structural models of the protein, haem a, haem a$_3$, Cu$_A$, Cu$_B$, Zn, Mg, phospholipids and cholates were refined. In the intervals shown by dotted lines with arrows at both ends, conventional refinement for positional and temperature factors was carried out. The structure was refined by simulated annealing during the cycles shown by solid lines with arrows at both ends. The model was revised manually at each stage denoted with a vertical arrow. At the stage A before simulated annealing, five amino-acid residues were added to the model. At the stage B, two phospholipids were added. In addition, four amino-acid residues were built at stage C. Two phospholipids and two cholic acids found in an $(F_o - F_c)$ map were refined at stage D. Another phospholipid, two cholates and a decylmartoside were added to the structure at stage E. The parameters for the protein were changed from those of Brooks *et al.* (1983) to those of Engh & Huber (1991) at stage F.

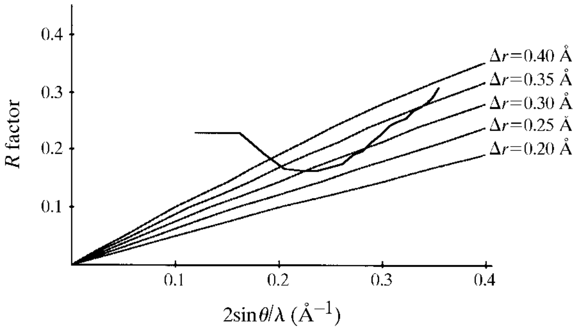


Figure 7
Luzzati plots (Luzzati, 1952) show the distribution of R factors as a function of 2sin(θ)/λ. The heavy line is the R-factor distribution of the observed structure factors. The thin lines are those theoretically estimated with mean absolute displacements (Δr) of 0.20, 0.25, 0.30, 0.35 and 0.40 Å in atomic position. The mean absolute error in atomic position of the enzyme is around 0.30 Å. The R factors in the high-resolution range are higher than those expected from 0.30 Å. This is because of the low accuracy of the observed structure factors in the high-resolution range.

one decylmaltoside in the monomer unit were also assigned in the MIR/*DM* map. Species of phospholipids were identified from the shape of electron-density contours corresponding to the polar head groups. All the bonds in the hydrocarbon tails of phospholipids were treated as single bonds because it was difficult to distinguish a double bond from a single bond at the resolution used. A total of 15 033 atomic positions in one monomer, defined as monomer *A*, were transformed by the NCS operation to generate the monomer *B* in the asymmetric unit.

The crystal structure was refined with *X-PLOR* (Brünger *et al.*, 1987) and by manual rebuilding in the MIR/*DM* electron-density map. A summary of the structure refinement is shown in Fig. 6. Ten cycles of positional and temperature-factor refinements were repeated with *X-PLOR* under structural and non-crystallographic symmetry restraints, respectively. Restraining parameters for the metal centres are given in Table 3. The polypeptide stereochemical restraints used were those of Brooks *et al.* (1983) during the early stages of refinement but were later replaced by the tighter parameters of Engh & Huber (1991). At the final stage of refinement the atomic positions of the metal centres and ligand atoms were inspected in the difference Fourier map calculated with coefficients of $(F_o - F_c)\exp(i\alpha_c)$, where F_o and F_c are the observed and calculated structure amplitudes, respectively, and α_c is the phase angle of the calculated structure factor.

The conventional *R* factor and the free *R* factor calculated for 5% of the data were reduced to final values of 0.199 and 0.252, respectively. The root-mean-square deviations from standard values of bond lengths and angles of the refined structure are 0.012 Å and 1.73°, respectively. The mean absolute error in atomic position estimated by Luzzati plots (Luzzati, 1952) is around 0.3 Å as shown in Fig. 7. Ramachandran plots (Ramakrishnan & Ramachandran, 1965) for all the residues of the dimer are shown in Fig. 8. The main chains of the enzyme have quite reasonable structures with 99.0% of non-glycine residues in the most favoured or the additional allowed φ/ψ regions. Out of 3304 non-glycine residues determined, eight residues are in disallowed regions in the φ,ψ plot.

Real-space correlation coefficients (Brändén & Jones, 1990) were evaluated by the program *OVERLAP MAP* of *CCP*4 between an ideal electron density calculated with the coefficients of $(F_c)\exp(i\alpha_c)$ and the MIR/*DM* map, and between the ideal density and the $(2F_o - F_c)$ one composed of the coefficients $(2F_o - F_c)\exp(i\alpha_c)$. The real-space correlation coefficients for individual side chains of the dimer are shown in Fig. 9, their averaged values for 26 subunits as well as those of phospholipids and cholates are listed in Table 4. The averaged correlation coefficients between the ideal and the MIR/*DM* electron densities are 0.79–0.91 for main chains and 0.74–0.88 for side chains. Those between the ideal and the $(2F_o - F_c)$ electron densities are 0.82–0.92 for the main chain and 0.75–0.89 for the side chains. Six side chains, Ser1 of subunit Va,

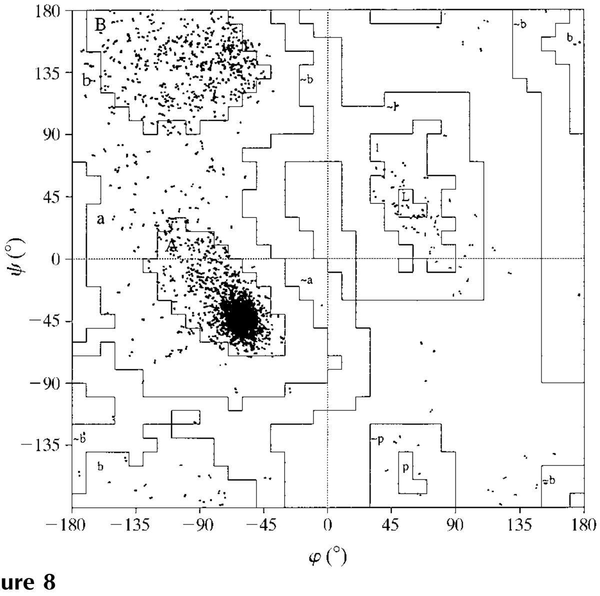


Figure 8
Ramachandran plots (Ramakrishnan & Ramachandran, 1965) for 3560 amino-acid residues The following eight residues are in disallowed regions: Gln59 (subunit II), Asp158 (subunit II), Thr2 (subunit III) and Ser2 (subunit Vb) of both molecules *A* and *B*. Out of 3058 non-glycine residues, 89.2% are in the most favoured regions, 9.8% in additional allowed regions, 0.7% in generously allowed regions and 0.3% in disallowed regions.

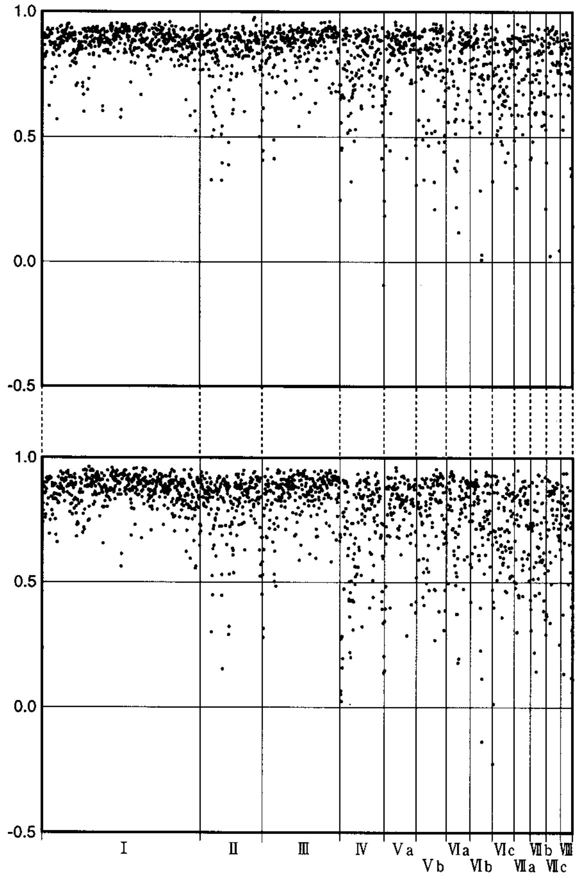


Figure 9
Averaged real-space correlation coefficients between the ideal and the MIR/*DM* electron densities for side chains of non-glycine residues. Roman numerals denote subunit names. The coefficients for the monomer *A* are plotted in the upper part, and those for the monomer *B* in the lower part.

Cys39 and Glu40 of subunit VIb in monomer *A*, and Ser1 of subunit Va, Glu40 of subunit VIb and Ser1 of subunit VIc in monomer *B*, exhibit negative correlation coefficients between the ideal and the MIR/*DM* electron densities, where the numbering system of subunits has been given in Tsukihara *et al.* (1996). Their correlation coefficients between the ideal and the $(2F_o - F_c)$ electron densities are positive but lower than 0.5. These residues may be disordered in the crystal. The correlation coefficients between the ideal and the MIR/*DM* electron densities for phospholipids and cholates are in the range 0.74–0.92. These structures are determined well except for flexible hydrocarbon chains of phospholipids which remain unknown.

The average temperature factor for each subunit and that of each monomer are given in Table 5. Subunits I and III have lower temperature factors than other subunits on the molecular surface in each monomer unit. The temperature factor of the molecule *A* is lower than that of the molecule *B* by 16 Å^2.

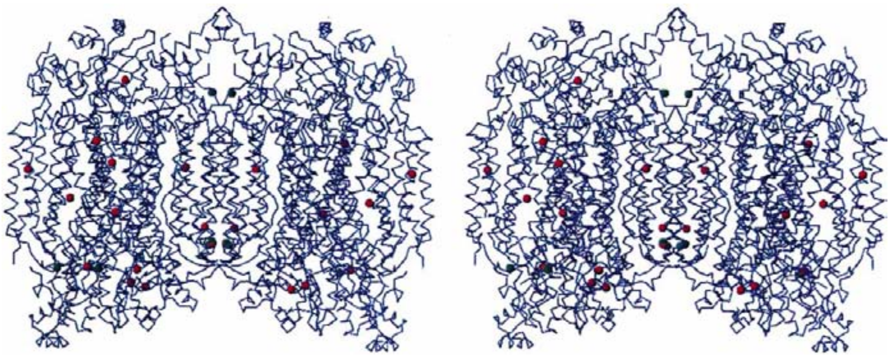


Figure 10
A stereoscopic drawing of the heavy-atom sites together with the cytochrome *c* oxidase dimer. Green and red balls represent Hg and It sites, respectively. The dimeric structure of the enzyme is 130 Å in height and is depicted by wire models of Cα atoms. Several heavy atoms are located in the central transmembrane parts abundant in α-helices.

3. Results and discussion

3.1. Solving the difference Patterson function and locations of heavy atoms

In general, the larger the number of heavy-atom sites in the unit cell of a derivative crystal, the more difficult it is to solve the difference Patterson function because of overlapping interatomic vectors. Although the CH_3HgCl derivative contains as many as eight main sites per asymmetric unit, most of the interatomic vectors between the main sites are clearly located as separate peaks in the difference Patterson map at 5 Å resolution as shown in Fig. 3(*a*). This is partly

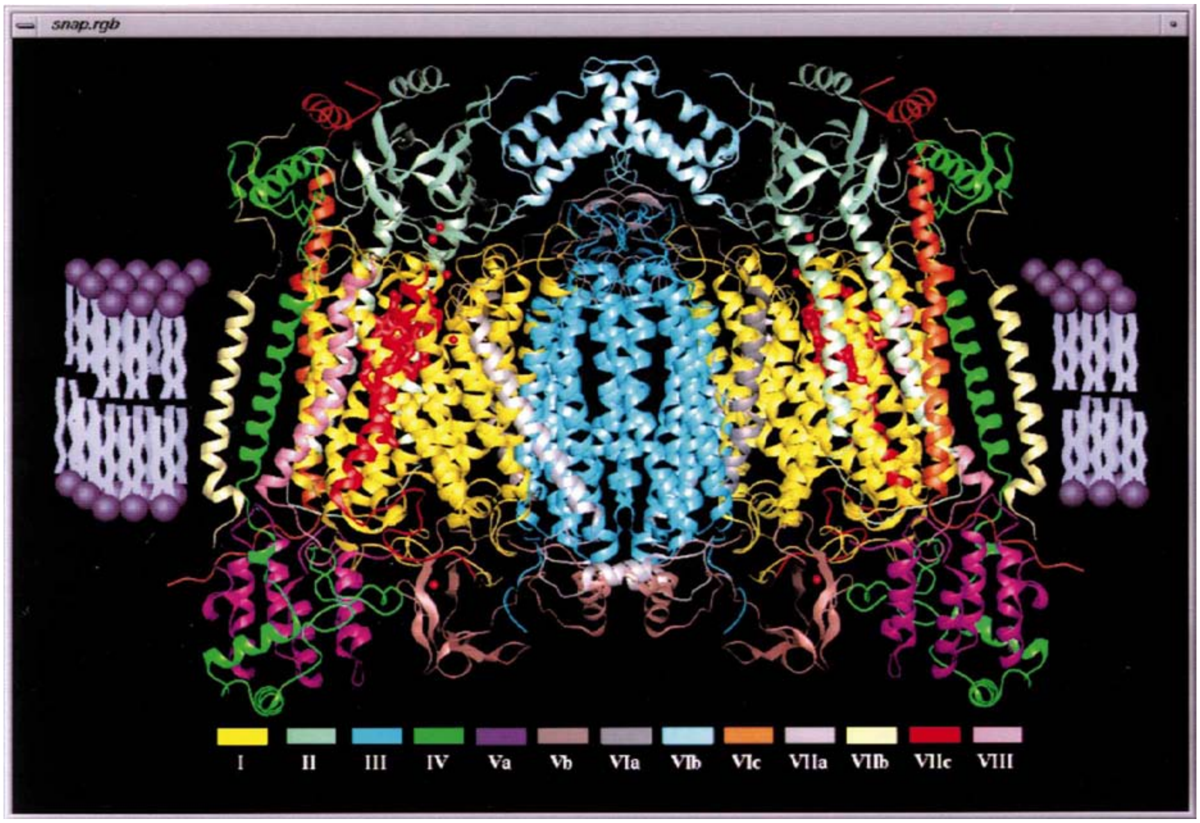


Figure 11
The dimeric bovine cytochrome *c* oxidase drawn as ribbons in the lipid bilayer with about 40 Å thickness. The upper extra-membrane side is the intermembrane space; the lower, the matrix space. The colour of each subunit is shown at the bottom of the figure. The transmembrane region is abundant in α-helices.

because of the large cell dimensions of the crystal compared with those of other protein crystals. Thus, it is not as difficult to solve the difference Patterson function for a large unit cell containing multiple heavy-atom sites as it is for a small unit cell.

Table 6 shows the final parameters of heavy atoms and the distances to nearby amino acids and metal ions. Out of 23 heavy-atom sites in the derivative crystal of CH_3HgCl, ten pairs of equivalent sites are located in both monomers *A* and *B* and three sites in monomer *A* alone. The $IrCl_6$(I) derivative contains a main site near Glu16 (subunit VIIc) of the monomer *A* and eight minor sites near four equivalent residues of both monomers *A* and *B*. The heavy atoms are located not only in the extra-membrane parts of the molecules but also in the transmembrane part. The heavy atoms are also not only on the molecular surface, but several are found inside the complex. The location of the heavy atoms shown in Fig. 10 implies that water-soluble reagents can penetrate the complex and the space between monomers.

3.2. Efficiency of each step of the density-modification procedure

Solvent flattening, histogram matching and NCS averaging were included in the present density modification. Four different procedures of density modification were carried out to evaluate the effectiveness of each step of the density modification. The averaged real-space correlation coefficient between the ideal and the MIR electron densities was 0.39 for the side-chain atoms. The value increased to 0.57 after the solvent flattening. The histogram matching increased the correlation coefficient from 0.57 to 0.66. A correlation coefficient of 0.83 was obtained by the solvent flattening followed by the NCS averaging. The electron density was slightly improved to give correlation coefficients of 0.85 for the side-chain atoms and 0.89 for the main-chain atoms by the additional procedure of histogram matching. The solvent flattening and NCS averaging were very effective in improving the electron density. The high solvent content of the crystal definitely contributes to the effectiveness of the solvent-flattening procedure.

3.3. Overall structure

Two monomers related by a pseudo twofold axis produced by combination of a crystallographic 2_1 axis and a pseudo *B* symmetry form a dimeric structure. The side view of the dimer is depicted by ribbon drawings of Cα atoms in Fig. 11. Wallin *et*

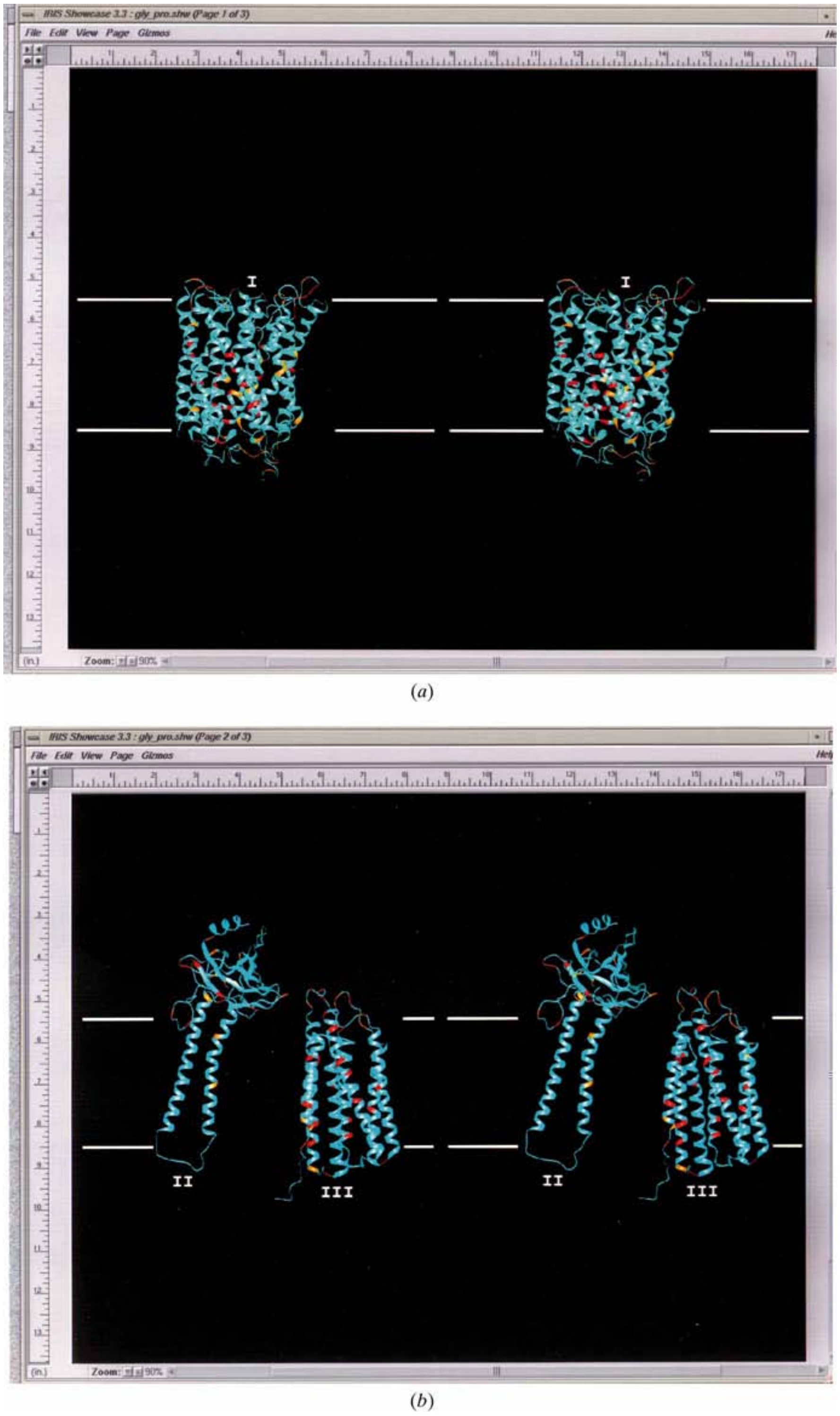
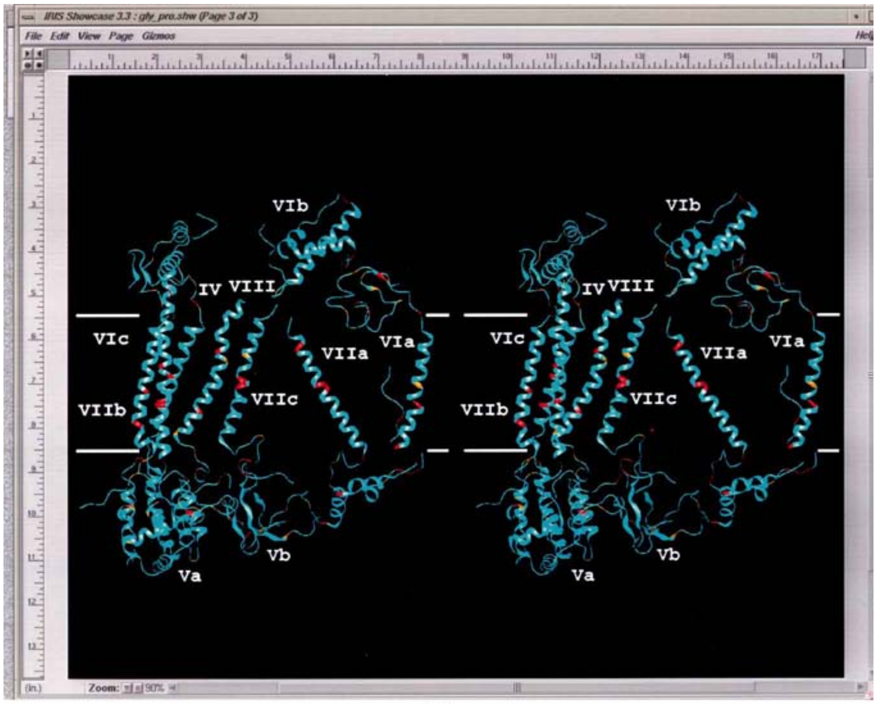


Figure 12
Stereoscopic ribbon drawings of the Cα trace of subunit I (*a*), subunits II and III (*b*), and all the nuclear subunits (*c*). Glycine, proline and the other residues are drawn in red, yellow and blue, respectively. Transmembrane α-helices are kinked near proline residues.



(*c*)

Figure 12 (continued)

al. (1997) analysed the structure of the enzyme in terms of secondary structure, hydrophobicity, amino-acid species and phospholipids along the molecular axis perpendicular to the membrane plane. The centre of the transmembrane part was defined as the midpoint of the central part of the molecule with an average hydrophobicity higher than 1.5, where the hydrophobicity of each residue was taken from the GES hydrophobicity scale (Engelman *et al.*, 1986). The central region ±12 Å from the centre of the transmembrane part is almost 100% α-helix and the extremely hydrophobic region extends 10 Å on either side of the centre of the membrane. Irregular structures are prominent in the region ±20–±30 Å. Of 14 transmembrane proteins whose structures have been determined at resolutions higher than 3.5 Å, nine proteins, two photosynthetic reaction centres (Deisenhofer *et al.*, 1984; Allen *et al.*, 1987), two cytochrome *c* oxidases (Iwata *et al.*, 1995; Tsukihara *et al.*, 1996), two light-harvesting antenna complexes (McDermott *et al.*, 1995; Koepke *et al.*, 1996), a cytochrome bc_1 complex (Xia *et al.*, 1997), a plant light-harvesting complex (Kuhlbrandt *et al.*, 1994) and a bacteriorhodopsin (Grigorieff *et al.*, 1996; Pebay-Peyroula *et al.*, 1997; Kimura *et al.*, 1997) fold in the α-helical conformation in the transmembrane region. The other five proteins, three porins (Weiss *et al.*, 1991; Cowan *et al.*, 1992), a maltoporin (Schirmer *et al.*, 1995) and an α-haemoplysin (Song *et al.*, 1996), have an antiparallel β-barrel crossing the membrane. Although a glycine residue is a typical α-helix-breaking amino acid in the case of soluble proteins, glycine residues amounting to 7.1% in the transmembrane α-helices of the cytochrome *c* oxidase, in contrast with a low content of 0.9% in the extra-membrane α-helices of the enzyme, do not break transmembrane α-helices as depicted in the ribbon drawings of Cα atoms in Figs. 12(*a*), 12(*b*) and 12(*c*). Since an amide group of the transmembrane α-helix is included in intrahelical hydrogen bonds and excluded from interaction with the hydrophobic surrounding, the α-helical structure consisting of hydrophobic residues including glycine and proline residues is more stable than the irregular structures in the membrane. Proline residues in the transmembrane region do not break helical structures but induce kinked structures of the helices as shown in Figs. 12(*a*), 12(*b*) and 12(*c*). These structures indicate a high propensity to α-helix formation in the membrane.

Distributions of charged residues at both sides of the membrane are shown schematically in Fig. 13. The mitochondrially encoded core subunits are more negatively charged at the intermembrane side than at the matrix side. This asymmetric charge distribution across the membrane is energetically favourable for a gradient of proton concentration across the membrane, the concentration being

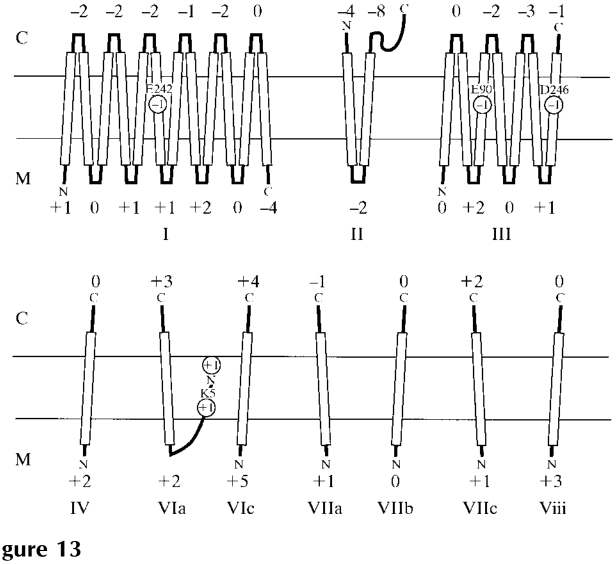


Figure 13

Schematical drawings of protein foldings of transmembrane helices and charge distributions of bovine cytochrome *c* oxidase. Each transmembrane helix is depicted by an open column. The region between two horizontal lines is the central, extremely hydrophobic, region with 20 Å thickness. Amino and carboxyl termini of each subunit are shown by the small letters N and C, respectively. Giving +1 to arginine and lysine residues and the amino group of the amino terminus, and −1 to aspartate and glutamate residues and the carboxyl group of the carboxyl terminus, we evaluated net charges for the intermembrane (C) and the matrix (M) sides. Amino termini of subunits I, II and III are acetylated. Three acidic groups, Glu242 of subunit I and Glu90 and Asp246 of subunit III, and two basic groups, Lys5 and the amino terminus of subunit VIa, are buried inside the central region of the membrane. All the transmembrane subunits encoded by the nuclear DNA are neutral or positive in overall charge.

Table 7
Intersubunit interactions given by numbers of residue pairs with an atomic contact shorter than 4.0 Å.

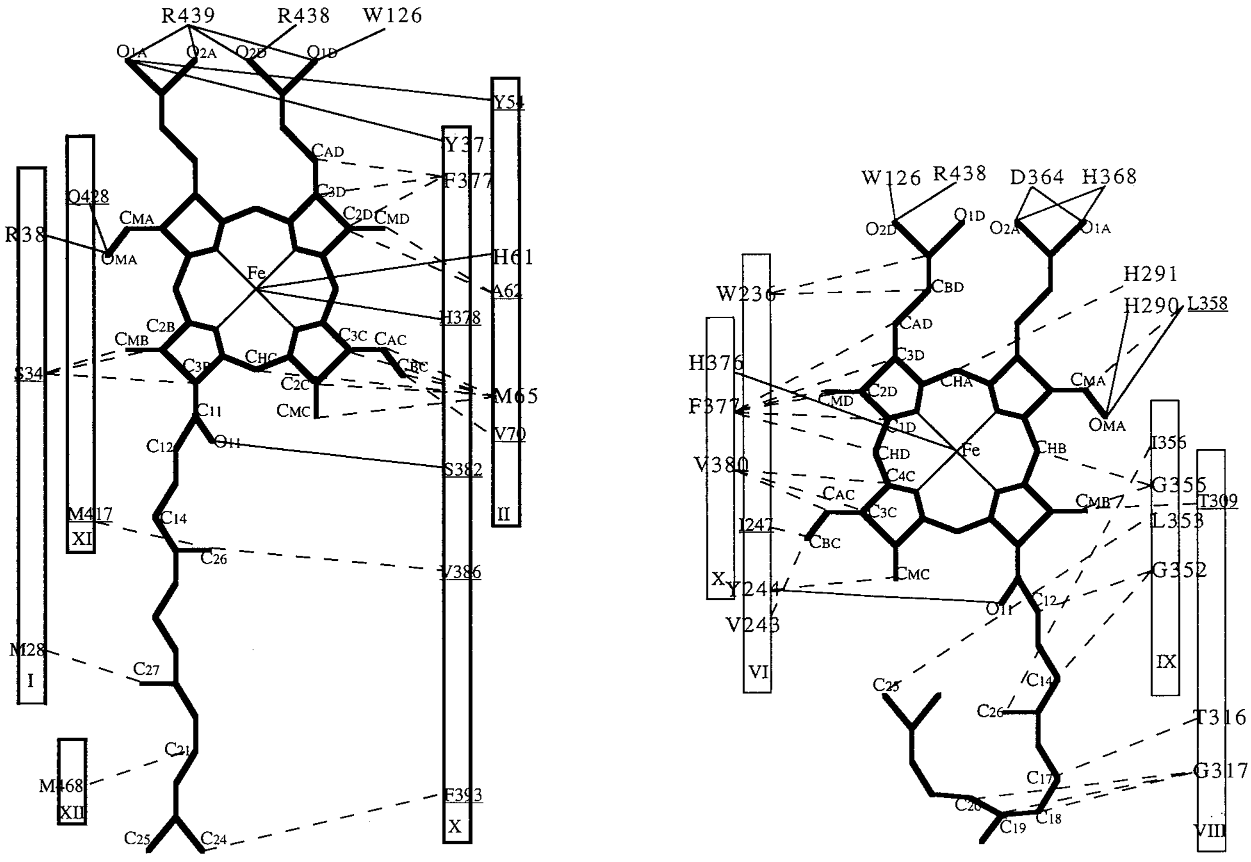
Figures in the upper right represent interactions within monomer *A*; those of the diagonal and the lower left (in bold) represent interactions between the two monomers *A* and *B*. Since the intermonomer interactions *A* to *B* and *B* to *A* are estimated independently, two values are given.

	I	II	III	IV	Va	Vb	VIa	VIb	VIc	VIIa	VIIb	VIIc	VIII
I	*	326	249	112		162	38	16	5	14	21	189	116
II		*		104	33			159	200		11		
III			*			116	194	23		128		1	
IV				*	203	18			40		167		15
Va					*	25			77				
Vb						**13**	13			23			
VIa	**89, 73**					**3, 1**	*	28					
VIb								**50**					
VIc									*		2		
VIIa										*		12	
VIIb											*		
VIIc												*	64
VIII													*

higher at the intermembrane side. A similar asymmetric charge distribution has been observed in the reaction centre of *Rhodopseudomonas viridis* (Deisenhofer & Michel, 1989) and has been named the 'positive inside' rule by von Heijne (1986). The nuclear-coded subunits do not follow the rule as has been stated by Gavelo & von Heijne (1992). This suggests that structural organization of the nuclear-coded subunits is determined by different factors from those of the mitochondrially encoded subunits.

3.4. Protein–haems a and a$_3$ interactions

The amino-acid residues interacting with haems a and a$_3$ were identified by a short contact less than 4.0 Å. 20 and 21 amino-acid residues interacting with haems a and a$_3$, respectively, are shown schematically in Figs. 14(*a*) and 14(*b*). Propionate groups interact with the protein moiety though ionic bonds and hydrogen bonds. Other polar groups of carbonyl and hydroxyl groups are hydrogen bonded to amino-acid residues. A lot of non-polar



(*a*) (*b*)

Figure 14
Schematic drawings of amino-acid residues interacting with haem a (*a*) and a$_3$ (*b*) in subunit I of bovine cytochrome *c* oxidase. Haems are illustrated by heavy lines. Helices are depicted by open columns with a helix number defined in Tsukihara *et al.* (1996). Each amino-acid name is given by a single letter with residue number. The amino-acid residues interact with haem groups through coordination bonds (thin lines) and non-polar interactions (broken lines). The residues represented by larger letters are evolutionally identical among procaryotes and eucaryotes; those represented by small letters underlined are conserved in animals; and those represented by small letters are evolutionally variable residues. The residues around haem a are more variable during molecular evolution than those around haem a$_3$.

contacts surrounding both haems stabilize their unique orientations in the protein. Comparing 11 amino-acid sequences of subunit I aligned by Capaldi (1990), we inspected the evolutionary stability of each amino-acid residue along the sequence. Out of 20 residues surrounding the haem a, 18 and eight residues are identical among six animal sequences and all 11 sequences, respectively. The residues around the haem a_3 are more conservative than those of haem a: 20 and 16 of the 21 residues are identical among six animal sequences and all 11 sequences, respectively. That is to say, the residues around the haem a of animals have evolved away from those of bacteria in marked contrast with those around the haem a_3.

3.5. Intersubunit interactions and hierarchy of structural organization of the enzyme complex

Bovine cytochrome *c* oxidase is a typical hybrid protein complex. Three of the 13 different subunits are encoded by the mitochondrial DNA and synthesized in the matrix space of the mitochondrion, and the other ten subunits encoded by the nuclear DNA are synthesized in the cytosolic space of the cell. Intersubunit interactions are likely to reflect the process of subunit assembly of the protein complex in the mitochondrion. Numbers of close contacts are listed in Table 7. Each of the eight nuclear-encoded subunits, except for subunits Va and VIIb, contacts mainly with the mitochondrial core subunits. Subunit Va interacts prominently with subunits IV and VIc, and subunit VIIb with subunit IV. Subunits I, Vb, VIa and VIb take part in intersubunit interactions between two monomers related by pseudo-twofold symmetry and each consisting of 13 subunits. A hierarchy of structural organization of the bovine cytochrome *c* oxidase is proposed on the basis of these intersubunit interactions (Fig. 15). The mitochondrially encoded subunits synthesized in the matrix space are inserted into the inner membrane and the nuclear-encoded subunits are synthesized in the cytosolic space. The mitochondrial subunits assemble themselves into a core structure and each nuclear subunit is transported into the correct space in the mitochondrion. The nuclear subunits IV, Vb, VIa, VIb, VIc, VIIa, VIIc and VIII find their own location on the surface of the mitochondrial core structure. The subunits Va and VIIb assemble into a complex consisting of 11 subunits. A dimeric structure of the enzyme complex is formed by intermonomer interactions among the subunits I, Vb, VIa and VIb. A symbiotic evolution, a hypothesis for an evolutionary pathway for the origin of mitochondria, is reflected in the hierarchy of structural organization of the enzyme complex.

3.6. Crystal structure and temperature factors of the two monomers

The crystal structure of the enzyme complex is projected along the *b* axis in Fig. 16. Monomers *A* and *B* related by a pseudo twofold axis form a dimeric structure with 294 pairs of atomic contacts shorter than 4.0 Å in the crystal. Extra-membrane subunits of a monomer interact with extra-membrane subunits of other monomers. There is no non-polar interaction between two dimers in the crystal as is the case in the photoreaction centre of *R. viridis* (Deisenhofer & Michel, 1989). Monomers *A* form a molecular column where each monomer has 128 close atomic contacts shorter than 4.0 Å with an adjacent monomer on either side. Another column consisting of *B* monomers is formed by 24 close contacts with

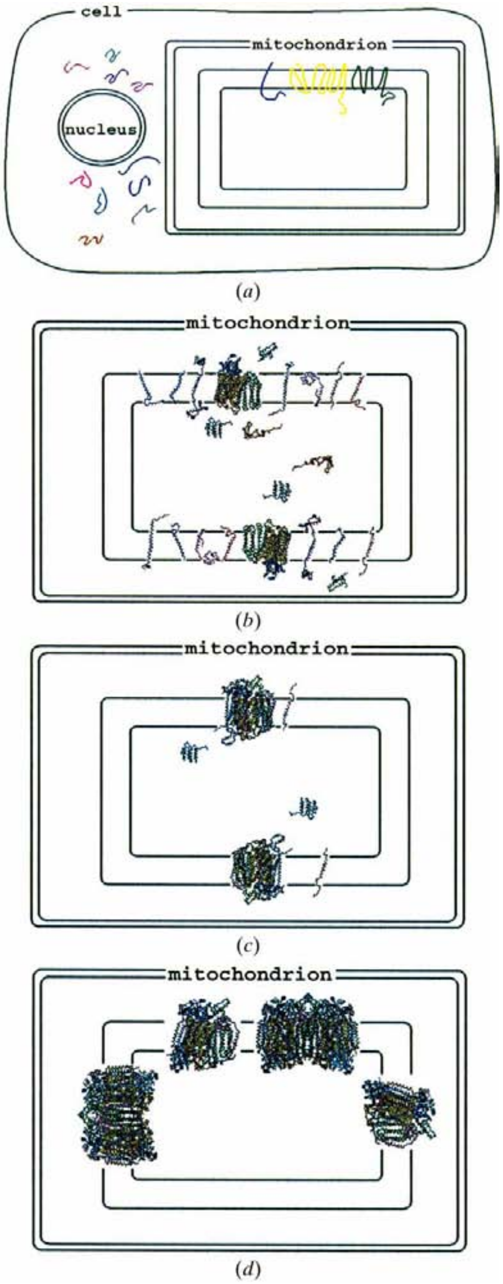


Figure 15
A hierarchy of structural organization of bovine cytochrome *c* oxidase proposed on the basis of intersubunit interactions in the dimeric structure. (*a*) The mitochondrial subunits synthesized in the matrix space are inserted into the inner membrane and the nuclear subunits are synthesized in the cytosolic space. (*b*) The mitochondrial subunits form a core structure and each nuclear subunit is transported in the proper space in the mitochondrion. (*c*) The nuclear subunits IV, Vb, VIa, VIb, VIc, VIIa, VIIc and VIII are located at their own positions on the surface of the mitochondrial core structure. (*d*) Assembly of the subunits Va and VIIb is followed by dimerization of the monomer consisting of 13 subunits.

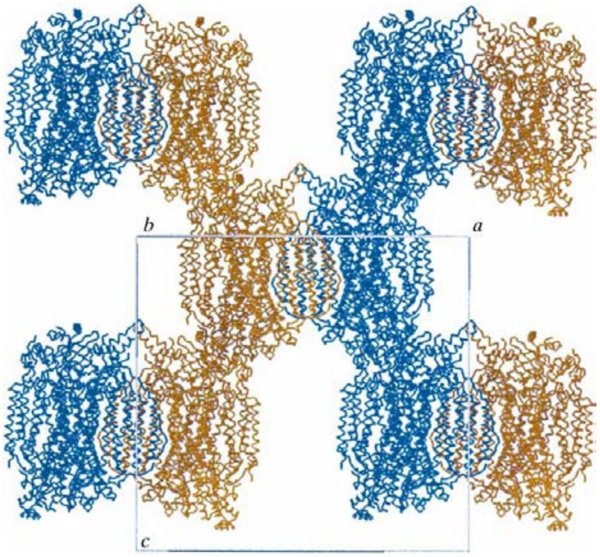


Figure 16
The crystal structure of the enzyme projected on the *ac* plane. Monomers *A* drawn in yellowish green are related to one another by a 2_1 axis along the *c* axis at (1/4, 1/2, *z*), and monomers *B* in light blue, by a 2_1 axis at (3/4, 1/2, *z*). Monomers *A* and *B* form a dimeric structure. There is no non-polar contact between two dimers in the crystal.

two adjacent monomers in the column of *B*. Thus, the monomers *A* in the column at (1/3, 1/2, *z*) are more tightly linked to each other than the monomers *B* in the other column at (3/4, 1/2, *z*). The close contact between the monomers *A* in the column seems to reflect the lower temperature factor and the higher real-space correlation coefficient (Table 4) of the monomer *A*. Atomic parameters of bovine heart cytochrome *c* oxidase are available from the Protein Data Bank.

The present work was supported in part by Grants-in-Aid for Scientific Research (Nos. 06558102 and 08408026 to TT), Grants-in-Aid for Scientific Research on Priority Areas (Bioinorganic Chemistry and Cell Energetics to SY and Nos. 05244102, 06276102 and 08288231 to TT) from the Ministry of Education and Culture of Japan, Grant-in-Aid for 'Research for Future' Program from the Japan Society for the Promotion of Science (JSPS-RFTF96L00503 to TT), Grant-in-Aid for the Proposal-Based Advanced Industrial Technology Development Organization of Japan (B-021 to SY), and the TARA project of Tsukuba University (SY and TT). The work was carried out with the approval of the Photon Factory Advisory Committee, the National Laboratory for High Energy Physics, Japan (proposals 91-050 and 94G-0410). We are grateful to Professor N. Sakabe, Dr A. Nakagawa, Dr N. Watanabe and Dr S. Ikemizu for their expert help with data collection at the Photon Factory.

References

Abrahams, J. P. & Leslie, A. G. W. (1960). *Acta Cryst.* D**52**, 30–42.

Allen, J. P., Feher, G., Yeates, T., Komiya, H. & Rees, D. C. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 6162–6166.

Brändén, C. & Jones, A. (1990). *Nature (London)*, **343**, 687–689.

Bricogne, G. (1974). *Acta Cryst.* A**30**, 395–405.

Bricogne, G. (1976). *Acta Cryst.* A**32**, 832–847.

Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S. & Karplus, M. (1983). *J. Comp. Chem.* **4**, 187–217.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Brünger, A., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Capaldi, R. A. (1990). *Annu. Rev. Biochem.* **59**, 569–596.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R. A., Jansonius, J. N. & Rosenbusch, J. P. (1992). *Nature (London)*, **358**, 727–733.

Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1984). *J. Mol. Biol.* **180**, 385–398.

Deisenhofer, J. & Michel, H. (1989). *EMBO J.* **8**, 2 149–2170.

Engelman, D. M., Steitz, T. A. & Golgman, A. (1986). *Annu. Rev. Biophys. Biophys. Chem.* **15**, 321–353.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Fergason-Miller, S. & Babcock, G. T. (1996). *Chem. Rev.* **96**, 2889–2907.

Furey, W. Jr (1980). *PHASES. Computer programs for X-ray crystal structure determination.* VA Medical Center and University of Pittsburgh, Pittsburgh, USA.

Gavelo, Y. & von Heijne, G. (1992). *Eur. J. Biochem.* **205**, 1207–1215.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287–307.

Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. (1996). *J. Mol. Biol.* **259**, 393–421.

Hamilton, W. C., Rollett, J. S. & Sparks, R. A. (1965). *Acta Cryst.* **18**, 129–130.

Heijne, G. von (1986). *EMBO J.* **5**, 3021–3027.

Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995). *Nature (London)*, **376**, 660–669.

Jones, A. T. (1978). *J. Appl. Cryst.* **11**, 268–272.

Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T. & Fujiyoshi, Y. (1997). *Nature (London)*, **389**, 206–211.

Kleywegt, G. J. & Jones, T. A. (1994). *From First Map to Final Model*, pp. 59–66. Warrington: Daresbury Laboratory.

Koepke, J., Hu, X., Muenke, C., Schulten, K. & Michel, H. (1996). *Structure*, **4**, 581–597.

Kuhlbrandt, W., Wang, D. N. & Fujiyoshi, Y. (1994). *Nature (London)*, **367**, 614–621.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

McDermott, G., Prince, S. M., Freer, A. A., Hawthornthwaite-Lawless, A. M., Papiz, M. Z., Cogdell, R. J. & Isaacs, N. W. (1995). *Nature (London)*, **374**, 517–521.

Malmström, B. G. (1990). *Chem. Rev.* **90**, 1247–1260.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Otwinowski, Z. (1991). *Proceedings of Daresbury Study Weekend*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.

Otwinowski, Z. (1993). *DENZO* and *SCALEPACK. Data processing and scaling programs*. Yale University, New Haven, Connecticut, USA.

Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. (1997). *Science*, **277**, 1676–1681.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Sakabe, N. (1983). *J. Appl. Cryst.* **16**, 542–547.

Schirmer, T., Keller, T. A., Wang, Y. & Rosenbusch, J. P. (1995). *Science*, **267**, 512–514.

Song, L., Hobaugh, M. R., Shustak, C., Cheley, S., Bayley, H. & Gouaux, J. E. (1996). *Science*, **274**, 1859–1866.

Steigemann, W. (1992). *PROTEIN* version 3.1. *A program system for the crystal structure analysis of proteins*. Max-Planck Institute für Biochemie, Martinsried, Germany.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. (1995). *Science*, **269**, 1069–1074.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. (1996). *Science*, **272**, 1136–1144.

Wallin, E., Tsukihara, T., Yoshikawa, S., Von Heijne, G. & Elofsson, A. (1997). *Protein Sci*. **6**, 808–815.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–113.

Weiss, M. S., Kreusch, A., Schiltz, E., Nestel, U., Welte, W., Weckesser, J. & Schulz, G. E. (1991). *FEBS Lett.* **280**, 379–382.

Xia, D., Yu, C., Kim, H., Xia, J., Kachurin, A. M., Zhang, L., Yu, L. & Deisenhofer, J. (1997). *Science*, **277**, 60–66.

Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* A**46**, 377–381.