

# The oxygen binding site of cytochrome oxidase

## Structural predictions on subunit I from amino acid sequences

Karen G. Welinder and Lars Mikkelsen

*Institute of Biochemical Genetics, University of Copenhagen, Ø. Farimagsgade 2A, DK-1353 Copenhagen K, Denmark*

Received 5 May 1983

Analyses of heme-attached amino acid sequences in known hemoprotein superfamilies provide a basis for prediction of such sequences in hemoproteins of unknown three-dimensional structure. Among 11 histidine residues conserved in subunit I of 3 mammalian and 2 fungal cytochrome oxidases the sequence around His-233 (human) is the most conserved and shows remarkable similarity to the sequence of the oxygen binding site in globins. Furthermore, the gene coding for subunit I in *Saccharomyces cerevisiae* and the gene for leghemoglobin in soybean are both split by introns right after these similar histidine sequences. The predicted distal histidine sequence of subunit I provides for heme  $a_3$  and  $\text{Cu}_{a_3}$  binding and has an extraordinarily high content of aromatic residues. These aromatic groups may serve as a molecular electron capacitor. Transmembrane sequences and electron transfer sequences are proposed.

*Cytochrome c oxidase*      *Cytochrome  $a_3$*       *Globin*      *Exon*      *Heme binding sequence*  
*Membrane buried sequence*

### 1. INTRODUCTION

Interpretation of a variety of chemical, physical and kinetic data on the terminal respiratory complex in several mammalian, fungal and bacterial species is leading to an increasingly consistent structural and functional model [1-4]. Cytochrome *c* oxidase of *Paracoccus denitrificans* provides the simplest complex consisting only of subunits I and II; none the less it apparently performs all essential cytochrome oxidase functions: cytochrome *c* binding (subunit II), electron transport from heme *a* to  $\text{Cu}_a$  (subunit II) and to the heme  $a_3$ - $\text{Cu}_{a_3}$  oxygen binding site (subunit I), and proton pumping [5,6]. Eukaryotic cytochrome oxidases are composed of at least 7 subunits [7,8] and form dimers [3].

Amino acid sequences governing the properties of the cytochrome oxidase metal centers, however, are known only for  $\text{Cu}_a$  [9]. Here, we predict the amino acid sequence at the oxygen binding side of

heme  $a_3$ . Comparison of all conserved histidine sequences in cytochrome oxidase subunits I, II and III with heme-attached sequences in 6 hemoprotein superfamilies revealed one sequence similar to distal histidine of globins, histidine 233 of mammalian subunit I. This finding supports the view that oxygen-binding heme  $a_3$  is part of subunit I and also provides a molecular model for this site. Analyses of distributions of charged and aromatic residues and of hydrophobic and predicted secondary structural patterns in the amino acid sequences of subunit I detects likely transmembrane and electron-transfer sequences. His 376 is pointed at as the most likely proximal histidine residue of heme  $a_3$ .

### 2. METHODS

#### 2.1. Amino acid sequence alignment

The rather closely related sequences of subunit I are easily aligned manually. In more difficult cases

a mutation data scoring matrix MDM78 [10], using a gap penalty of 60, is applied in a computer-aided fashion.

## 2.2. *Hydropathy*

Hydropathy plots are carried out using the hydropathy indices derived by Kyte and Doolittle [11]. Indices of 7 consecutive residues are averaged.

## 2.3. *Prediction of secondary structure*

The results of three predictive methods are presented, Chou and Fasman [12], Lim [13], Garnier, Osguthorpe and Robson [14]. Decision constants in the ROB program are  $DC_H = -75$  and  $DC_S = -87.5$ , and  $DC_S$  is changed to 50 in ROB(A) [14].

## 2.4. *Predicting heme attached sequences*

The apoprotein part of a hemoprotein may be described in terms of structural sites of specific subfunctions: Heme plane covers, heme edge covers, and structures interacting with cooperative subunits or domains of separate function. Some hemoproteins have a structural site for membrane attachment essential to its intracellular organisation.

Detailed three-dimensional structures are known for members of 6 hemoprotein superfamilies: cytochrome *c* types, cytochrome *b<sub>5</sub>* types, 4 helix bundle types (cytochrome *b-562* and *c'*), globins, peroxidases and catalase. Analyses of the 12 heme plane covers revealed that, generally, each cover is composed of 2 nearly linear peptide segments, that these segments follow in sequence: linear-turn-linear, that 17 of 22 segments contain helix, that 9 of 12 iron linked residues are histidine, and that open heme plane covers often have Leu-X-X-His and closed covers Leu-X-X-X-His. Catalase is outstanding in several respects (in preparation).

Classification of an unknown hemoprotein from its amino acid sequence includes:

- (i) Alignment of homologous proteins and identification of the most conserved sequences;
- (ii) Plotting of hydropathic character;
- (iii) Composite prediction of secondary structure;
- (iv) Tracing invariable histidine residues and ranking according to sequence conservatism;

- (v) Comparison of conserved histidine sequences with prototype heme attached histidine sequences of hemoprotein superfamilies;
- (vi) Evaluation of predicted results and comparison with experimental data.

## 3. RESULTS

### 3.1. *Sequence characteristics and patterns*

The amino acid sequences of cytochrome oxidase subunit I from 3 mammalian and 2 fungal species were largely derived from their nucleic acid sequences (fig.1). Of 508 positions occupied in all 5 species, 253 or 50% are conserved. The sequence homology is most extensive between residues 190 and 252 reaching 80%. The 3 mammalian enzymes are 89% identical and cytochrome oxidase subunit I is the most conserved among the 13 mitochondrially coded proteins in man and cow [16]. This structural conservatism reflects the many functional and structural restraints imposed on cytochrome oxidase subunit I.

The amino acid composition of subunit I shows an unusually high proportion of aromatic residues. Of 508 positions, 14% are invariably occupied by aromatics, including 11 invariable histidine residues. Noteworthy is the absence of conserved cysteine residues which excludes the likelihood of any disulfide bridges or cysteine ligands of Cu and Fe.

The distribution of charges is remarkable in two respects:

- (1) Conserved charged residues occur as pairs of opposite sign. There are 13 such pairs. All but 2 Asp + Glu and 3 Arg + Lys are paired.
- (2) Most pairs are spaced by 25 residues or a multiple of 25.

### 3.2. *Transmembrane segments*

The hydropathic character along the 5 amino acid sequences is shown in fig.2. It detects membrane- and protein-buried hydrophobic segments and solvent-exposed hydrophilic segments [11]. The most hydrophilic segments are likely antigenic determinants [20]. Fig.2 discloses a remarkable regularity with many long, conserved hydrophobic segments, very unlike globular proteins. Comparison with the predicted reverse turns in fig.3 suggests that hydrophobic segments larger than 20 residues and terminated by conserved turns

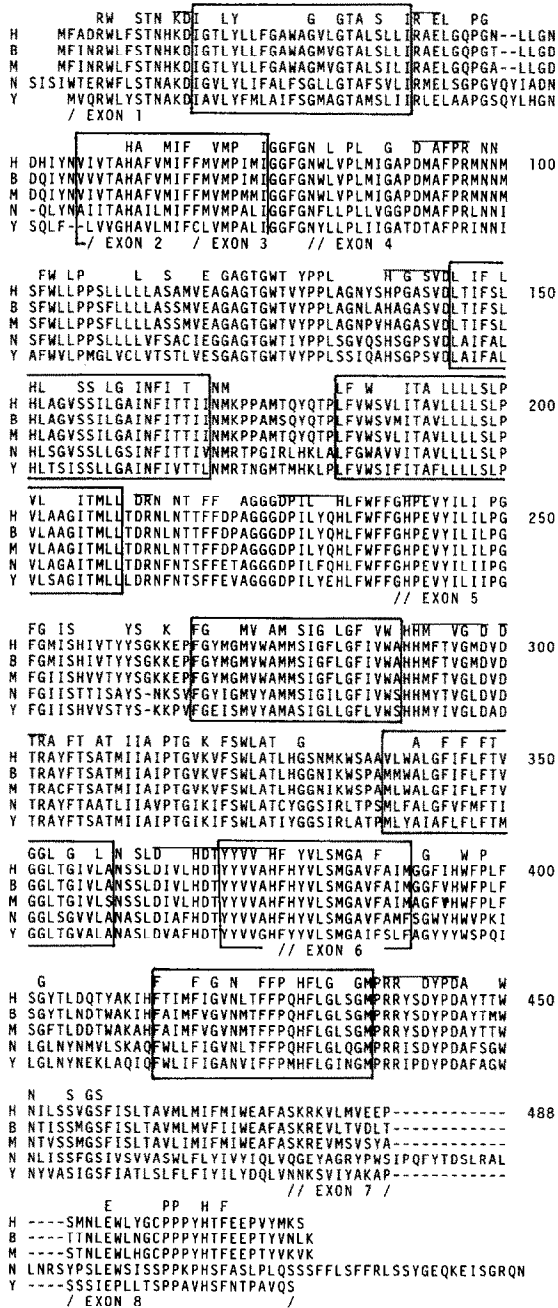


Fig.1. Alignment of the amino acid sequences of cytochrome oxidase subunit I of: H, human [15]; B, bovine [16]; M, mouse [17]; N, *Neurospora crassa* [18]; Y, yeast (*Saccharomyces cerevisiae* strain D273-10B) [19]. Residues are shown in the one-letter notation. Mammalian residue numbers are indicated. The first row indicates residues identical in all and conserved charge pairs are barred. Predicted transmembrane sequences are enclosed.

are probable transmembrane sequences. The predicted transmembrane segments are on the average 22 residues in size, and are frequently flanked by the observed charge pairs (table 1, fig.1).

Prediction of helix and  $\beta$ -strand secondary structure in membrane embedded parts of proteins is presently unreliable, since we know only the crystal structure at high resolution of one such protein, bacteriorhodopsin, which happens to contain transmembrane helices, and since predictive algorithms are derived from globular proteins [21]. The predictions in fig.3 favour  $\beta$ -strand. But in case of bacteriorhodopsin the  $\beta$ -strand potential of Chou and Fasman correlated very well (0.48) with membrane-embedded helices, whereas the helix potential resulted in random correlation (0.01) [21]. Predictive algorithms for globular proteins should still apply to protein structure folded in concert with protein and solvent.

X-ray diffraction of 'membranous cytochrome oxidase' indicated a high proportion of helix [28]. The average size of predicted transmembrane sequences on helical form is  $22 \times 1.5 \text{ \AA}$ ; i.e., similar to the lipid bilayer.

3.3. Analysis of histidine sequences

The ranking after sequence conservatism of 11 invariable histidine residues in subunit I of 5 species is shown in table 2. His 233 and 240 are clearly parts of an important active site. Comparison to heme-attached sequences in hemoprotein superfamilies shows a clear similarity with the heme plane cover at the oxygen binding site of globins (fig.4). His 151 and His 376 have slight, but random, resemblance to closed heme plane covers. The sequence around His 376 shares features with the heme attached sequences of cytochrome *b<sub>5</sub>*.

None of the histidine sequences resembles the Cu binding histidine sequences of presently known Cu proteins (Cu-Zn superoxide dismutase and azurin-related proteins).

3.4. Gene structure

In globins, structural sites of specific subfunctions, are reflected in the globin split-gene structure (review by Blake [22], references herein and [23]). Exons 2 and 3 of the leghemoglobin gene of soybean [24] show a perfect match with the globin

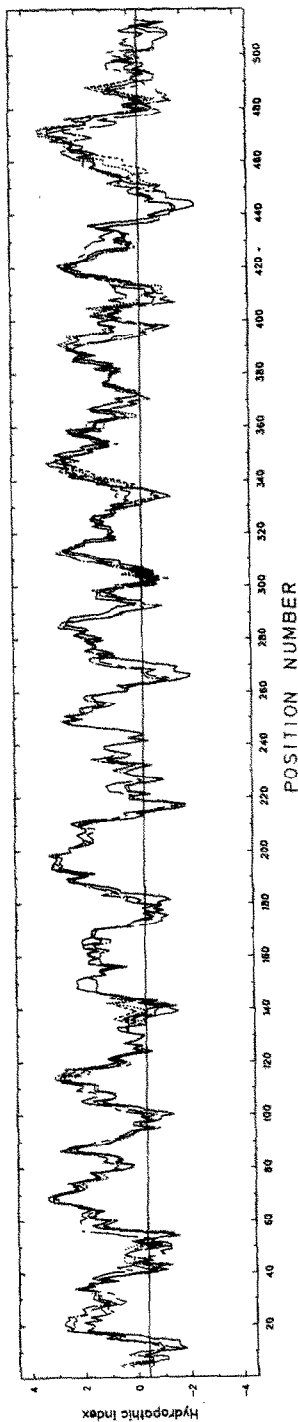


Fig.2. Hydrophobic character of human (—), mouse (---), bovine (· · ·), *N. crassa* (- · - ·) and yeast (· · · ·) cytochrome oxidase subunit I sequences. Hydrophobic is positive, hydrophilic negative. Sequences extra to *N. crassa* are left out. Positions 1-46 correspond with mammalian residue numbers and positions 49-516 with mammalian residues 47-514.

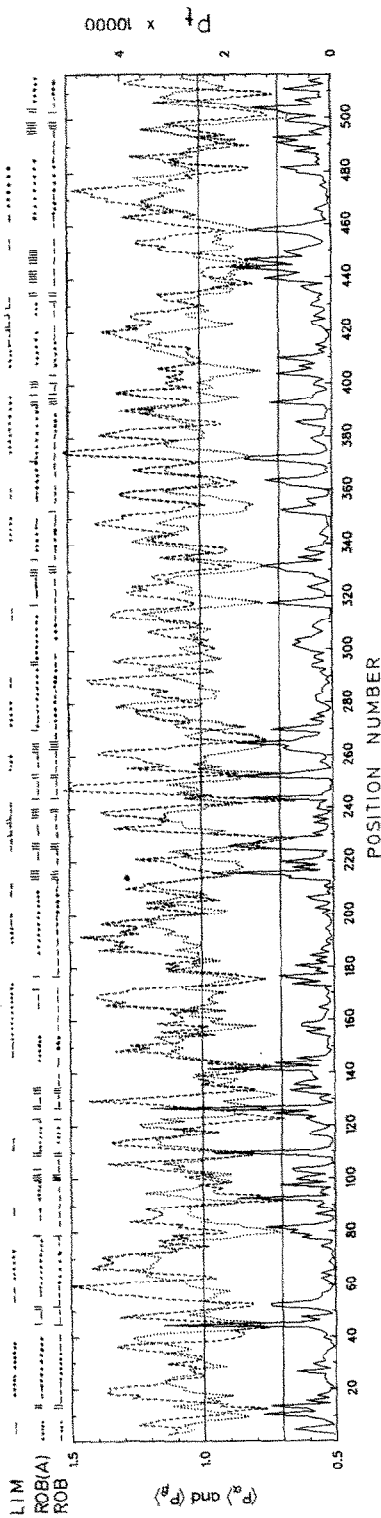


Fig.3. Predicted secondary structure of cytochrome oxidase subunit I. Helix (· · ·),  $\beta$ -strand (- · - ·) and turn (—) or (' ' ' '). Positions are explained in fig.2. Plots show the Chou and Fasman  $\alpha$ - and  $\beta$ -potentials and turn probability ( $P_t$  scale to the right) averaged over the 5 sequences. LIM shows predictions identical in 4 out of 5 sequences. ROB and ROB(A) show automatic, composite predictions on the 5 sequences. ROB(A) reduces  $\beta$ -strand predictions.

Table 1  
Proposed transmembrane sequences in subunit I

Residues	Size	Conserved		Charge
		Total	Aromatic	
15-37	23	39%	2	0
56-75	20	45%	2	1 His
(100-118; turn at				
107-108)	19	32%	2	0
145-169	25	64%	3	1 His
183-210	28	71%	2	0
268-289	22	64%	4	0
(303-327; turn at				
315-316)	25	72%	3	1 Lys
338-359	22	50%	3	0
371-390	20	70%	7	1 His
414-436	23	57%	6	1 His
(459-477	19	0%	2	1 Glu)

heme plane covers. In *Saccharomyces* the cytochrome oxidase subunit I gene is split into 6-10 exons depending on the strain [25]. Amazingly, the cytochrome oxidase subunit I gene is split 6 residues after the proposed distal histidine like the leghemoglobin gene which is split 7 residues after distal histidine (fig.1,4). This we consider a third indication that His 233 of subunit I does indeed function at the oxygen binding site of cytochrome

Table 2

Ranking of invariable histidine residues in cytochrome oxidase subunit I according to sequence conservatism

Residue no. (mammalian)	Conserved (in 30 residues)	
	Totally	Trp + Phe + Tyr
233	90%	8
240	90%	7
376	73%	6
368	73%	5
290	70%	4
291	70%	4
429	70%	6
138	67%	3
151	67%	2
61	47%	4
503	17%	1

	Heme			
HEMOGLOBIN ALPHA	/MF SFPTTKTYFPHF-DLS	-GSAQ	KAHGKKVADAL	
HEMOGLOBIN BETA	/LLVYYPWTQRFF SFGDL	.7	GNPKYKAHGKKVL SF	
MYOGLOBIN	/LFK HPETLEKFDKFKHL	.7	ASEDLKXHG TVL.TAL	
LEGHEMOGLOBIN	/ILEKAPAAKDLFSFL	NGVDPTNPKLTAHAEK FGL/		
CO1	LLTDRN NT FF	AG----	GGDPILYQHLFWFFG/	
GLOBIN AND CO1	LL	tFFd	G	G P L H FG
GLOBIN HELICES	BBBBCCCCC			EEEEEEEEEEEEEE

Fig.4. Alignment of distal histidine sequences of 4 globin families [10] and proposed distal histidine (His 233 in mammals) of cytochrome oxidase subunit I (CO1). Amino acid residues conserved  $\geq 50\%$  are shown in the one-letter notation. Variable residues are blank. Sequence gaps are indicated by (-) or (,) and exons by (/) [22-24]. Small letters indicate common variable residues. Globin helix C and part of helices B and E are shown in the last line.

oxidase. Furthermore, a common evolutionary origin is probable.

### 3.5. Hypothetical model of the oxygen binding site

The similarity of the very most conserved histidine sequence (His 233) of subunit I to globin distal sequences, suggests that His 233 functions as distal histidine in tertiary structural environments similar to globins. Comparable hydrophobic character and predicted turns at this site in subunit I and globins (not shown) support this view.

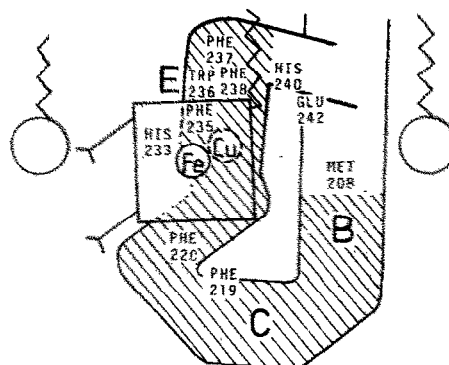


Fig.5. Model of the proposed oxygen binding, heme  $\alpha 3$ -Cu $_{a3}$  site in cytochrome oxidase viewed from proximal histidine and normal to the heme plane and the lipid bilayer. Zig-zag lines indicate part of hydrocarbon chains. The shaded protein fold is homologous to globins and includes the analogues of helices B, C and E as indicated. The turn after helix E and the Cu atom are placed to leave the oxygen entrance open at the propionic edge of heme.

The alignment (fig.4) shows that subunit I lacks the short D helix of globins like  $\alpha$ -hemoglobin and leghemoglobin. The 4 C-terminal residues of the globin helix B analogue in subunit I, terminate a proposed transmembrane segment (fig.1), which like heme  $a_3$  [28], is perpendicular to the membrane. The turns separating helices B and C, and helices C and E are clearly predicted in subunit I (fig.3). The Phe CD1 in globins (the first residue after helix C in fig.4) belongs to the only 2 invariable residues in globins [10] and is in close contact with heme. In subunit I helix E is shortened by a turn that presumably returns the peptide chain to the active site (fig.5). This model provides Met 208, Tyr 231, Gln 232, His 233, His 240, Glu 242 and Tyr 244 as potential ligands of  $Cu_{a_3}$ . However, the distance of the only sulphur atom present (Met 208) makes a possible sulphur bridging of Fe and Cu, proposed from EXAFS measurements [26], unlikely. However, the orientation of heme in cytochrome oxidase subunit I may deviate more from globins than the variations actually observed among globins.

### 3.6. Conserved aromatic residues and electron transfer

The predicted oxygen binding site of cytochrome oxidase subunit I is lined by 9 aromatic residues (219, 220, 231, 233, 235–238, 240), suggesting that  $\pi$ -type interactions may be important in electron transfer and storage. This leads to the more speculative proposal, that experimentally observed proximal histidine of heme  $a_3$  [27] may also be in an aromatic environment. The sequence of His 376 includes 9 conserved aromatic residues (368, 371, 372, 376–379) and is, therefore, in all respects the best candidate for proximal heme attachment (table 2). The proposed transmembrane segment, residues 414–436, does also contain an extraordinarily high proportion of conserved aromatics and might therefore take part in electron transfer from cytochrome subunit II to subunit I.

### ACKNOWLEDGEMENTS

The LIM program was written by Dr Hans Lenstra; the ROB program by Dr Jean Garnier. This work was supported by the Danish Natural Science Research Council (11-2172).

### REFERENCES

- [1] Azzi, A. (1980) *Biochim. Biophys. Acta* 594, 231–252.
- [2] Wikström, M., Krab, K. and Saraste, M. (1982) *Cytochrome Oxidase. A Synthesis*, Academic Press, London, New York.
- [3] Deatherage, J.F., Henderson, R. and Capaldi, R.A. (1982) *J. Mol. Biol.* 158, 487–499.
- [4] Deatherage, J.F., Henderson, R. and Capaldi, R.A. (1982) *J. Mol. Biol.* 158, 501–514.
- [5] Ludwig, B. (1980) *Biochim. Biophys. Acta* 594, 177–189.
- [6] Solioz, M., Carafoli, E. and Ludwig, B. (1982) *J. Biol. Chem.* 257, 1579–1582.
- [7] Verheul, F.E.A.M., Draijer, J.W., Dentener, I.K. and Muijsers, A.O. (1981) *Eur. J. Biochem.* 119, 401–408.
- [8] Kadenbach, B. and Merle, P. (1981) *FEBS Lett.* 135, 1–11.
- [9] Steffens, G.J. and Buse, G. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* 360, 613–619.
- [10] Schwartz, R.M. and Dayhoff, M.O. (1978) in: *Atlas of Protein Sequence and Structure* (Dayhoff, M.O. ed) vol.5, suppl.3, pp.353–358, National Biomedical Research Foundation, Washington DC.
- [11] Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105–132.
- [12] Chou, P.Y. and Fasman, G.D. (1978) *Adv. Enzymol.* 47, 45–148.
- [13] Lim, V.I. (1974) *J. Mol. Biol.* 857–894.
- [14] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97–120.
- [15] Anderson, S., Bankier, A.T., Barrell, B.G. et al. (1981) *Nature* 290, 457–465.
- [16] Anderson, S., De Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982) *J. Mol. Biol.* 156, 683–717.
- [17] Bibb, M.J., Etten, R.A.V., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) *Cell* 26, 167–180.
- [18] Burger, G., Scriven, C., Machleidt, W. and Werner, S. (1982) *EMBO J.* 1, 1385–1391.
- [19] Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E., Tzagoloff, A. and Macino, G. (1980) *J. Biol. Chem.* 255, 11927–11941.
- [20] Hopp, T.P. and Woods, K.R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 3824–3828.
- [21] Argos, P., Rao, J.K.M. and Hargrave, P.A. (1982) *Eur. J. Biochem.* 128, 565–575.
- [22] Blake, C. (1983) *Trends Biochem. Sci.* 8, 11–13.
- [23] Blanchetot, A., Wilson, V., Wood, D. and Jeffreys, A.J. (1983) *Nature* 301, 732–734.
- [24] Jensen, E.Ø., Paludan, K., Hyldig-Nielsen, J.J., Jørgensen, P. and Marcker, K.A. (1981) *Nature* 291, 677–679.

- [25] Hensgens, L.A.M., Arnberg, A.C., Roosendaal, E., Van der Horst, G., Van der Veen, R., Van Ommen, G.-J.B. and Grivell, L.A. (1983) *J. Mol. Biol.* 164, 35-58.
- [26] Powers, L., Chance, B., Ching, Y. and Angiolillo, P. (1981) *Biophys. J.* 34, 465-498.
- [27] Stevens, T.H. and Chan, S.I. (1981) *J. Biol. Chem.* 256, 1069-1071.
- [28] Blasie, J.K., Erecińska, M., Samuels, S. and Leigh, J.S. (1978) *Biochim. Biophys. Acta* 501, 33-52.