Isolation and analysis of the genes for cytochrome *c* oxidase in *Paracoccus denitrificans*

Mirja Raitio, Tuulikki Jalli and Matti Saraste

Department of Medical Chemistry, University of Helsinki, Siltavuorenpenger 10, 00170 Helsinki 17, Finland

Communicated by M. Wikström

Synthetic oligonucleotide probes were used to clone two loci from the chromosomal DNA of *Paracoccus denitrificans* that contain the genes for cytochrome *c* oxidase (cytochrome *aa*₃). One locus seems to contain four or five genes probably forming an operon. Two of these code for the oxidase subunits II and III. Three open reading frames are found between the COII and COIII genes. The other locus codes for the subunit I. A short open reading frame is found upstream of this gene. All three subunits of the *Paracoccus* enzyme show remarkable homology to the corresponding subunits of the mitochondrial cytochrome oxidase. Possible protein products of the open reading frames have not yet been identified.

Key words: cytochrome oxidase/DNA sequencing/*Paracoccus*/subunits I,II and III

Introduction

Oxidative phosphorylation in bacteria is adapted to environmental conditions. The respiratory chains can be 'branched' and simultaneously use different electron acceptors such as nitrate and oxygen. Bacteria are often armed with several different terminal oxidases which are differently expressed depending on the oxygen concentration in the growth medium (Jones, 1982; Poole, 1983). Aerobic bacteria can be divided into two classes; those that can synthesize cytochrome *c* and possess a cytochrome *aa*₃-type oxidase (EC 1.9.3.1) and those that cannot. This oxidase is functionally and structurally homologous to the mitochondrial cytochrome *c* oxidase (Ludwig, 1980, 1987). Only the bacteria that have this enzyme are able to harvest maximal energy from reduction of oxygen in oxidative phosphorylation. As mitochondria, they can employ all three 'coupling sites' in energy transduction (see Jones, 1982; Poole, 1983).

Cytochrome *aa*₃ contains two haem A groups which have different properties when bound to the apoprotein, and two copper ions. One of the latter (Cu_B) forms the oxygen-reducing active centre together with the haem iron of cytochrome *a*₃ (see Wikström *et al.*, 1981, 1985). Dioxygen is bound between this centre and reduced by four electrons coming from cytochrome *c* via the redox centres Cu_A and cytochrome *a*. One proton per electron is consumed in formation of water. This proton is taken from the 'inside' — the matrix side in mitochondria or the cytoplasmic side of the membrane in bacteria — and meets the electron coming from the opposite side ('outside') at the active centre. The catalytic reaction is coupled to translocation of an extra proton across the membrane. Thus 2 H⁺/e⁻ are taken up from the inside (see Wikström and Saraste, 1984).

In mitochondria cytochrome oxidase is made of two classes of proteins. Three genes in the mtDNA code for the major subunits I, II and III (COI, COII and COIII). They form the func-

tional core of the complex; the metal centres are known to be located in COI and COII (Wikström *et al.*, 1985). The rest of the subunits are encoded by genes located in the nucleus. The number of these cytoplasmic subunits varies in eukaryotes; it may be up to 10 in mammals (Kadenbach and Merle, 1981; Saraste, 1983; Capaldi *et al.*, 1986). Our knowledge on the prokaryotic cytochrome *c* oxidase (Ludwig, 1987) indicates that only the mitochondrially coded proteins have homologous counterparts in the bacterial enzyme. The enzyme isolated from *Paracoccus denitrificans* has been characterized as a two-protein complex of COI and COII (Ludwig and Schatz, 1980). We have now found a gene for the third subunit in the genome (Saraste *et al.*, 1986) and argue that it must be part of the enzyme *in situ*.

The cytochrome oxidase genes in *Paracoccus* are divided into two loci. In addition to the genes coding for the oxidase subunits, there seem to be other as yet unidentified genes adjacent to them. The oxidase loci should be controlled in a synchronized manner. The expression of cytochrome *aa*₃ is known to be active

Fig. 1. Synthetic oligonucleotide probes for isolation of the genes for COI and COII. Two different probes (A and B) were designed for COII and one (C) for COI. The alignment of the mitochondrial protein sequences (see Wikström *et al.*, 1985) and the peptide sequences determined for the *Paracoccus* proteins (Steffens *et al.*, 1983) is shown; the numbers refer to the location of the residue in the bovine protein (Anderson *et al.*, 1982). Mixed probes were synthesized (see Materials and methods) complementary to the coding strand. Note that probe A is not made redundant at the first 3' residue and that the COI sequence deduced from DNA is FIVWAH for the *Paracoccus* peptide shown in C (see Figure 4). A is a 19-mer and a mixture of 32 variants, B is a 17-mer with 32 variants and C a 17-mer with 16 variants.

Fig. 2. Restriction maps, subcloning and sequencing strategies of *cox1* (A) and *cox2* (B). (A) The locations of five open frames are shown relative to the restriction sites. The *SalI*, *BamHI* and *SalI/BamHI* fragments purified and subcloned into M13mp18 and mp19 are shown; a truncated version of the 4.0-kb *SalI* fragment, corresponding to an end of the insert in a lambda clone (see Saraste *et al.*, 1986) is marked with * in 'Sequencing strategy' (a). This and the 1.7-kb *SalI* fragment were isolated in replicative form, the inserts released and fragmented by sonication. The sequences obtained in this way are shown in (b). Other sequences originate from the ends of the *ExoIII*-digested clones (c) or were obtained by specific priming (d). The consensus for the two strands is shown on the bottom. (B) The location of ORF4 and the COI gene are shown together with the adjacent restriction sites. *SalI*, *PstI/BamHI*, *SmaI/BamHI* and several *SphI* fragments were subcloned as in (A). *ExoIII* deletion was used to generate a few clones from two *SphI* fragments (b), and the rest of the sequence solved with synthetic primers (c).

only when the oxygen concentration is high in the growth medium (John and Whatley, 1977; van Verseveld *et al.*, 1981; Poole, 1983). The isolated clones will enable us to study this regulation of expression in detail. This cloning also forms the basis for *in vitro* mutagenesis on this enzyme in future.

Results and Discussion

Cloning of the oxidase genes

We used synthetic oligonucleotide probes for isolation of the oxidase genes from a genomic library constructed into a lambda

Fig. 3. Nucleotide sequence of *cox1*. Restriction sites shown in Figure 2 are underlined; the genes are labelled above the probable initiation codons; ribosomal binding sites are double underlined; the assumed promoter site is boxed and the terminator shown by double arrows followed by a wavy line. The GATA sequence in the promoter is flanked by inverted repeats shown by double arrows. GCCGCC and GGCGGC sequences elsewhere in the 5' region are also indicated.

vector. This was possible since *Escherichia coli*, the host, does not possess a cytochrome *aa₃*-type oxidase (see Jones, 1982).

Two mixed probes were designed for the subunit II gene on the basis of the published peptide sequences of the *Paracoccus* enzyme (Steffens *et al.*, 1983) and evolutionary conservation in the mitochondrial proteins (probes A and B in Figure 1). Positive recombinants were picked up using probe A in the first screening and both A and B in parallel in the second one; hybridization conditions are described in Materials and methods. Several recombinants contained two common *SalI* fragments in the inserts. In Southern blots a 1.7-kb fragment hybridized to our probe A and a 4.0-kb *SalI* fragment to the probe B (not shown). These and several other restriction fragments were purified by electrophoresis and subcloned into M13 vectors (Figure 2A). The gene for COII was first located by sequencing the ends of the 1.7-kb *SalI* fragment.

The nucleotide sequence of this fragment and most of the 4.0-kb *SalI* fragment was determined (Figures 2A and 3). The COII and COIII genes could be easily defined since their products have a clear homology to the equivalent mitochondrial proteins (Figure 5). Three open reading frames (ORFs) were found between these genes.

We were unable to detect the gene for the major subunit, COI, in the flanking regions of this first oxidase locus (*cox1*) and thus had to design a third probe (C in Figure 1) and to isolate the COI gene separately. Again the probe-binding 3.5-kb *SalI* fragment and other overlapping restriction fragments were isolated from the positive lambda clones and subcloned into the M13 vectors (Figure 2B). The COI gene was located by sequencing one end of the 3.5-kb *SalI* fragment: a *SalI* site resides very close to the 5' end of the COI gene (Figure 4).

Interpretation of the DNA sequences

The 4887 bases of the DNA sequence at the first locus extending from a *SalI* site to a *BamHI* site (Figure 2A) are shown

in Figure 3. The COII gene is shown to be the first coding frame in this sequence. In its upstream region there are ORFs (not shown) but codon usage in these does not follow the 'genomic strategy' in *Paracoccus* (Table I). All five frames (COII, ORF1, ORF2, ORF3 and COIII) begin at ATG codons preceded by Shine–Dalgarno sequences (underlined in Figure 3; Shine and Dalgarno, 1974). Only ORF1 has a less clear consensus for this ribosomal binding site. The ATG triplet at position 1971 is the first potential initiation codon for ORF1, and the protein sequence is thus predicted to begin there. Note that ORF1 and ORF2 overlap by a single nucleotide at position 2810 and ORF2 and ORF3 by four nucleotides at 3076–3079 (Figure 3).

There are three reasons why ORF1, ORF2 and ORF3 may be assumed to be proper genes. First, they are flanked by the identified genes for the oxidase subunits II and III and appear to form an operon together with these. Second, the frames have an initiation codon preceded by a ribosomal binding site, and third, the codon usage is similar to the identified cytochrome oxidase genes (Table I). ORF2 is, however, a short frame and it is possible that the sequence between the 3' end of ORF1 and 5' end of ORF3 is non-coding.

There is some ambiguity about the exact location of the 5' end in the COII gene. Several in-frame ATG and GTG triplets are possible. However, ATG at the position 992 is one of the first and preceded by a clear Shine–Dalgarno sequence.

The sequence of 3221 nucleotides at the second locus (*cox2*) is shown from a *PstI* to a *SphI* site in Figure 4. Again the COI gene is easily identified by the homology of its protein product to the subunit I in the mitochondrial enzyme (Figure 5). Upstream from this gene there is an ORF which has an ATG triplet preceded by a Shine–Dalgarno sequence (ORF4 in Figure 4), and the frame again follows the proper codon usage.

The promoters have not yet been identified. However, in the front of the COII and COI genes there are similar sequences (GATA boxed in Figures 3 and 4) that might represent promoter

P A R A C O C C U S

Fig. 5. Comparison of the *Paracoccus* COI, COII and COIII with the corresponding subunits of the mitochondrial enzymes from beef and yeast (Anderson *et al.*, 1982; Thalenfeld and Tzagoloff, 1980 respectively). A computer program DIAGON employing a scoring matrix for amino acid substitutions (see Staden, 1982) was used. The scoring level for each comparison was adjusted so that the double matching probability (McLachlan, 1971) is closed to 1×10^{-5} ; the points shown are at this level or above it. Note that the residue scales are different from each comparison matrix. Deletions and insertions are indicated by breaks in the diagonal line.

Apparent termination signals for transcription in *cox1* and *cox2* are typical inverted repeats followed by T/A-rich sequences (Platt, 1981), beginning at nucleotides 4541 and 2738 in Figures 3 and 4, respectively. If the promoter sites are assumed correctly, the first locus is transcribed as a multicistronic mRNA of ~ 3650 bases and the second one as a mRNA of ~ 2080 bases coding

for the COI alone. We are currently trying to identify these mRNAs from aerobically cultivated *Paracoccus* cells.

Codon usage

In Table I the codons of three oxidase genes (A) and of four ORFs (B) are pooled separately. Both tables show how the codon choice

Table I. Codon usage in the cytochrome oxidase genes (A) and the adjacent ORFs (B) (note the strong bias against A/T at the third position of the triplets)

(A)				
F TTT	8. S TCT	0. Y TAT	26. C TGT	1.
F TTC	75. S TCC	8. Y TAC	21. C TGC	8.
L TTA	0. S TCA	2. * TAA	1. * TGA	1.
L TTG	5. S TCG	26. * TAG	0. W TGG	39.
L CTT	7. P CCT	2. H CAT	18. R CGT	1.
L CTC	12. P CCC	15. H CAC	18. R CGC	19.
L CTA	0. P CCA	1. Q CAA	6. R CGA	0.
L CTG	83. P CCG	48. Q CAG	27. R CGG	6.
I ATT	2. T ACT	0. N AAT	3. S AGT	1.
I ATC	79. T ACC	39. N AAC	26. S AGC	10.
I ATA	0. T ACA	1. K AAA	3. R AGA	0.
M ATG	47. T ACG	20. K AAG	24. R AGG	1.
V GTT	7. A GCT	2. D GAT	13. G GGT	10.
V GTC	50. A GCC	63. D GAC	26. G GGC	70.
V GTA	1. A GCA	7. E GAA	12. G GGA	2.
V GTG	43. A GCG	41. E GAG	26. G GGG	14.
(B)				
F TTT	4. S TCT	0. Y TAT	13. C TGT	0.
F TTC	36. S TCC	4. Y TAC	4. C TGC	7.
L TTA	0. S TCA	1. * TAA	0. * TGA	4.
L TTG	3. S TCG	16. * TAG	0. W TGG	16.
L CTT	2. P CCT	2. H CAT	9. R CGT	4.
L CTC	8. P CCC	15. H CAC	3. R CGC	22.
L CTA	0. P CCA	1. Q CAA	3. R CGA	0.
L CTG	46. P CCG	21. Q CAG	17. R CGG	6.
I ATT	1. T ACT	0. N AAT	4. S AGT	0.
I ATC	30. T ACC	26. N AAC	17. S AGC	9.
I ATA	0. T ACA	0. K AAA	3. R AGA	0.
M ATG	23. T ACG	13. K AAG	17. R AGG	1.
V GTT	1. A GCT	2. D GAT	11. G GGT	3.
V GTC	27. A GCC	49. D GAC	20. G GGC	40.
V GTA	2. A GCA	0. E GAA	4. G GGA	2.
V GTG	29. A GCG	44. E GAG	22. G GGG	10.

is strikingly biased. As in other bacterial genes rare codons such as AGA and AGG for arginine and ATA for isoleucine are avoided (see Maruyama *et al.*, 1986). But a general strategy in *Paracoccus* is to avoid the A/T ending in triplets (Table I). This is very clear in most of the families; the only exceptions are tyrosine, histidine, aspartate and glutamate. The codon bias is probably simply a reflection of a general high G/C content of the genome: a bias simply against A/T ending is difficult to reconcile in the context of translation accuracy or efficiency (for discussion, see Holm, 1986).

Unidentified genes

ORF1 appears to code for an integral membrane protein. Its amphipathic profile (Figure 6) indicates that it possibly contains nine long hydrophobic segments (the first segment at the N terminus is, however, shorter than the others), and this protein is probably mostly buried in the cell membrane.

The ORF3 protein contains an N-terminal hydrophobic segment (Figure 6E) that could be a signal sequence. It is preceded by several basic residues which is a characteristic of the prokaryotic signals (Vlasuk *et al.*, 1983; Watson, 1984). Thus it might be transported across the membrane and be a periplasmic protein. The ORF2 protein would have only 89 residues and contain a single long hydrophobic segment in its N-terminal half. The ORF4 protein in *cox2* would have 111 amino acid residues and no continuous hydrophobic segment.

Fig. 6. Amphipathic profiles of the gene products. The Kyte and Doolittle Index (1982) was used; the window length in the calculation was 11. Hydrophobic sequences point upwards; the mean amphipathicity is shown by the horizontal line. Note that the sequence scale is different for each protein. (A–E) COI, COII, COIII, the ORF1 protein and the ORF3 protein respectively. Probable transmembrane segments are marked with *, and S.S. indicates a putative signal sequence.

The function of these proteins has not yet been clarified. No clearly homologous proteins were found in a search through a protein information data base. One possibility is that some of them are involved in haem A biosynthesis. Haem A, the prosthetic group entirely devoted to cytochrome *aa*₃ (Koland *et al.*, 1984) is unique in two aspects: it has a formyl group that replaces a methyl and a long lipophilic isoprenoid (1-hydroxyfarnesylethyl) chain that replaces a vinyl group and renders the structure amphiphilic (Caughey *et al.*, 1975; see Holm *et al.*, 1987). The modification of haem may thus require at least two enzymes that catalyze these reactions.

Cytochrome oxidase subunits

Isolated *Paracoccus* cytochrome oxidase was originally described as a two-subunit complex of COI and COII. The spectral and catalytic properties of this isolated enzyme are very similar to the mitochondrial cytochrome oxidase (Ludwig and Schatz, 1980). Only the net translocation of protons, clearly demonstrated *in vivo* (van Verseveld *et al.*, 1981), has been difficult to repeat *in vitro* in reconstitution experiments. The H⁺/e⁻ stoichiometry in successful experiments has been maximally 0.6 (Solioz *et al.*, 1982). A similar situation has prevailed in the eukaryotic oxidase when its subunit III is removed (Saraste *et al.*, 1981; Penttilä, 1983; Sarti *et al.*, 1985; Finel and Wikström, 1986). We have now identified the gene coding for COIII. There is no reason why this protein would not be synthesized and associated with the enzyme *in situ*. We cannot say, however, if it is controlled differently from the other genes (see Ludwig, 1987), in which case there should be an internal promoter in *cox1*.

Recently it has been possible to modify a purification procedure (Berry and Trumpower, 1985) and to isolate a three-protein *Paracoccus* cytochrome oxidase. This preparation contains a third subunit that has similar electrophoretic migration to the subunit III in the mitochondrial enzyme. This protein is furthermore labelled by radioactive dicyclohexyl carbodiimide (DCCD) — a characteristic of COIII (Prochaska *et al.*, 1981) — and may turn out to be the subunit III (T.Haltia and M.Finel, unpublished). The DCCD-reactive membrane-embedded glutamic acid residue (Prochaska *et al.*, 1981) is present in the *Paracoccus* COIII (Saraste *et al.*, 1986).

The *Paracoccus* COIII is predicted to have seven transmembrane segments (Figure 6C) as the mitochondrial COIII. Compared with the latter it has two insertions which are both predicted to reside on the periplasmic surface in the folded topography (not shown).

A part of the COII sequence has been published earlier (Ludwig, 1987) and agrees well with the sequence reported here. Subunit II has four invariant features. It has a transmembrane hairpin (Figure 6B), a well-conserved aromatic sequence in the C-terminal domain, two strictly invariant carboxylic acid residues and four ligands assumed to bind Cu_A (see accompanying paper). In *Paracoccus* COII appears to have an extended N and C terminus as compared with the mitochondrial proteins (Figure 5). The former extension is hydrophobic (Figure 6B) and has all the characteristics of a signal sequence. The presence of a signal sequence has been confirmed by comparison with the N terminus of the mature protein (Steinrücke *et al.*, 1987). It reflects the fact that the N terminus must be translocated to the periplasmic surface of the membrane (von Heijne, 1986; see Holm *et al.*, 1987).

COI has probably 12 transmembrane segments (Figure 6A). In *Paracoccus* two long insertions (horizontal shifts in the line in Figure 5) are found again in the predicted periplasmic domain. This protein is the most conserved among the oxidase subunits.

Two haems and Cu_B are probably bound to it (Holm *et al.*, 1987).

In the accompanying paper we try to relate these conserved structural features to the functional properties of cytochrome oxidase. We thus formulate hypotheses that may be tested by *in vitro* mutagenesis. This requires, however, that a suitable expression system for cytochrome *aa* is developed, as *E. coli* is unable to synthesize haem A (Koland *et al.*, 1984).

Materials and methods

A genomic library of *P. denitrificans* S1657 was prepared to lambda vector EMBL3 (Frischauf *et al.*, 1983) using partially *Sau*3A-digested DNA inserts. Oligonucleotides were synthesized from β -cyanoethyl phosphoramidites in an Applied Biosystems 381 A synthesizer.

Hybridization

Oligonucleotide probes were labelled by polynucleotide kinase and equimolar amounts of [γ -³²P]ATP (>5000 Ci/mmol, Amersham) and free label separated by column chromatography on Sephadex G-25. Replica filters were prepared and treated as follows (see Wallace *et al.*, 1981). Pre-hybridization was carried out at 60°C for 2–3 h in 0.9 M NaCl, 0.09 M Tris-HCl, 1 mM EDTA, pH 7.5, containing 0.1 mg/ml yeast tRNA (Boehringer), 10 × Denhardt's solution and 0.5% Na-lauroylsarcosine. All hybridization reactions for replica filters were performed in plastic Petri dishes and for Southern blots in sealed plastic bags which were not shaken. They were carried out overnight at the appropriate temperatures and begun by adding ~5 pmol of labelled probe to 3 ml of fresh hybridization solution in each dish for the replica filters or 50 pmol to 10 ml for Southern blots. The *T*_m range for mixed probes was estimated as described by Mason and Williams (1985), and hybridization carried out close to *T*_m (min)–5. For probes A, B and C (Figure 1) the best hybridization temperatures were 56, 44 and 48°C respectively. Washings were made at the same temperatures using 6 × SSC (0.9 M NaCl, 0.09 M Na-citrate, pH 7.0) containing 0.5% Na-lauroylsarcosine. Three 5-min washes normally removed the background.

Sequencing strategy

The dideoxynucleotide method (Sanger *et al.*, 1980; Biggin *et al.*, 1983) was used to determine DNA sequences of the inserts cloned in M13 vectors. A mixture of strategies was used in generating the clones for sequencing. First, the ends of subcloned restriction fragments were sequenced. Secondly, in *cox1* one particular clone (see legend to Figure 2A) was isolated in the replicative form, the insert released, purified by electrophoresis, fragmented by sonication and cloned in the *Sma*I-cut M13mp8 for sequencing (Deininger, 1983). As an alternative way to produce subclones, *Exo*III deletion (Henikoff, 1984) was used for the 1.7- and 4.0-kb *Sall* fragments in *cox1* and two *Sph*I fragments in *cox2* (see Figure 2). Finally specific primers were synthesized and used to fill the remaining gaps in the sequence. They were also used to overlap restriction sites and to solve problematic regions at the final stages of sequencing.

Acknowledgements

We are grateful to J.Knowles and J.E.Walker and their colleagues for help and advice at the initial stages of this work, and to M.Wikström for encouragement and critique. Katri Sundell has given excellent technical assistance and Anneli Sundström and Hilka Vuorenmaa great help during the preparation of this manuscript. We are financially supported by research contracts from the Academy of Finland (S.R.C) and the Sigrid Jusélius Foundation.

References

- 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.
- Berry, E.A. and Trumpower, B.L. (1985) *J. Biol. Chem.*, **260**, 2458–2467.
- Biggin, M.D., Gibson, T.J. and Hong, G.P. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.
- Capaldi, R.A., Gonzalez-Halphen, D. and Takamiya, S. (1986) *FEBS Lett.*, **207**, 11–17.
- Caughey, W.S., Smythe, G.A., O'Keefe, D.H., Maskasky, J.E. and Smith, M.L. (1975) *J. Biol. Chem.*, **250**, 7602–7622.
- Deininger, P.L. (1983) *Anal. Biochem.*, **135**, 247–263.
- Finel, M. and Wikström, M. (1986) *Biochim. Biophys. Acta*, **851**, 99–108.
- Frischauf, A.-M., Lehrach, H., Poutska, A. and Murrey, N.J. (1983) *J. Mol. Biol.*, **170**, 827–842.
- Hawley, D.K. and McClure, W.R. (1983) *Nucleic Acids Res.*, **11**, 2237–2255.
- Henikoff, S. (1984) *Gene*, **28**, 351–359.
- Holm, L. (1986) *Nucleic Acids Res.*, **14**, 3075–3087.
- Holm, L., Saraste, M. and Wikström, M. (1987) *EMBO J.*, **6**, 2819–2823.

- John, P. and Whatley, F.R. (1977) *Biochim. Biophys. Acta*, **463**, 129–153.
- Jones, C.W. (1982) *Bacterial Respiration and Photosynthesis. Aspects on Microbiology*. Vol. 5, Thomas Nelson & Sons, Walton-on-Thames.
- Kadenbach, B. and Merle, P. (1981) *FEBS Lett.*, **135**, 1–11.
- Koland, J.G., Miller, M.J. and Gennis, R.B. (1984) *Biochemistry*, **23**, 1051–1056.
- Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.
- Ludwig, B. (1980) *Biochim. Biophys. Acta*, **594**, 177–189.
- Ludwig, B. (1987) *FEMS Microbiol. Rev.*, **46**, 41–56.
- Ludwig, B. and Schatz, G. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 196–200.
- Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986) *Nucleic Acids Res.*, **14**, r151–r197.
- Mason, P.J. and Williams, J.G. (1985) In Hames, B.D. and Higgins, S.J. (eds), *Nucleic Acid Hybridisation. A Practical Approach*. IRL Press, Oxford, pp. 113–137.
- McLachlan, A.D. (1971) *J. Mol. Biol.*, **61**, 409–424.
- Penttilä, T. (1983) *Eur. J. Biochem.*, **133**, 355–361.
- Platt, T. (1981) *Cell*, **24**, 10–23.
- Poole, R.K. (1983) *Biochim. Biophys. Acta*, **726**, 205–243.
- Prochaska, L.J., Bisson, R., Capaldi, R.A., Steffens, G.C.M. and Buse, G. (1981) *Biochim. Biophys. Acta*, **637**, 360–373.
- Ptashne, M. (1986) *Nature*, **311**, 697–701.
- Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B. (1980) *J. Mol. Biol.*, **143**, 161–178.
- Saraste, M. (1983) *Trends Biochem. Sci.*, **8**, 139–142.
- Saraste, M., Penttilä, T. and Wikström, M. (1981) *Eur. J. Biochem.*, **115**, 261–268.
- Saraste, M., Raitio, M., Jalli, T. and Perämaa, A. (1986) *FEBS Lett.*, **206**, 154–156.
- Sarti, P., Jones, M.G., Antonini, G., Malatesta, F., Colossimo, A., Wilson, M.T. and Brunori, M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4876–4880.
- Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
- Soloz, M., Carafoli, E. and Ludwig, B. (1982) *J. Biol. Chem.*, **257**, 1579–1582.
- Staden, R. (1982) *Nucleic Acids Res.*, **10**, 2951–2961.
- Steffens, G.C.M., Buse, G., Oppliger, W. and Ludwig, B. (1983) *Biochem. Biophys. Res. Commun.*, **116**, 335–340.
- Steinrücke, P., Steffens, G.C.M., Pankus, G., Buse, G. and Ludwig, B. (1987) *Eur. J. Biochem.*, in press.
- Thalenfeld, B.E. and Tzagoloff, A. (1980) *J. Biol. Chem.*, **255**, 6173–6180.
- Van Verseveld, H.W., Krab, K. and Stouthamer, A.H. (1981) *Biochim. Biophys. Acta*, **635**, 525–534.
- Vlasuk, G.P., Inouye, S., Ito, H., Itakura, K. and Inouye, M. (1983) *J. Biol. Chem.*, **258**, 7141–7148.
- Von Heijne, G. (1986) *EMBO J.*, **5**, 3021–3027.
- Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H. and Itakura, K. (1981) *Nucleic Acids Res.*, **9**, 879–894.
- Watson, M.E.E. (1984) *Nucleic Acids Res.*, **12**, 5145–5164.
- Wikström, M. and Saraste, M. (1984) In Ernster, L. (ed.), *Bioenergetics. New Comprehensive Biochemistry*. Elsevier, Amsterdam, Vol. 9, pp. 49–94.
- Wikström, M., Krab, K. and Saraste, M. (1981) *Cytochrome Oxidase. A Synthesis*. Academic Press, London.
- Wikström, M., Saraste, M. and Penttilä, T. (1985) In Martonosi, A.N. (ed.), *The Enzymes of Biological Membranes*. Plenum, New York, Vol. 4, pp. 111–149.

Received on May 19, 1987