Eur. J. Biochem. *164*, 111–115 (1987)

The primary structure of cytochrome c_1 from *Neurospora crassa*

Jürgen RÖMISCH[1], Maximilian TROPSCHUG[2], Walter SEBALD[3] and Hanns WEISS[1]

[1] Institut für Biochemie, Universität Düsseldorf
[2] Institut für Physiologische Chemie, Universität München
[3] Gesellschaft für Biotechnologische Forschung, Braunschweig-Stöckheim

(Received September 1/December 5, 1986) – EJB 86 0934

The primary structure of the cytochrome c_1 subunit of ubiquinol-cytochrome-*c* reductase from mitochondria of *Neurospora crassa* was determined by sequencing the cDNA of a bank cloned in *Escherichia coli*. From the coding region the sequence of 332 amino acids, corresponding to the molecular mass of 36496 Da, was derived for the precursor protein. The mature protein, the N terminus of which was previously sequenced [Tsugita et al. (1979) in *Cytochrome oxidase* (King, T. E. et al., eds) pp. 67–77, Elsevier, New York], consists of 262 amino acids and has the molecular mass of 29908 Da including the heme. The sequence contains an N-terminal hydrophilic part of 211 residues, which carries the heme, a hydrophobic stretch of 15 residues, which is assumed to anchor the protein to the membrane, and a C-terminal hydrophilic part of 36 residues. The N-terminal presequence of 70 amino acids contains 9 positive charges but only 1 negative charge and is characterized by a stretch of 20 uncharged residues.

Cytochrome c_1 is a subunit of ubiquinol–cytochrome-*c* reductase (cytochrome reductase), the proton-translocating electron-transfer complex III of the mitochondrial system of oxidative phosphorylation (for a review see [1]). The subunit accepts an electron from the Rieske iron-sulfur protein of the enzyme and passes it to cytochrome *c*. For interaction with cytochrome *c* the domain of cytochrome c_1 which carries the heme, must be exposed to the outer surface of the mitochondrial inner membrane [2–4]. Direct evidence that the larger part of cytochrome c_1 projects into the intermembrane space of mitochondria came from electron microscopic studies of membrane crystals prepared of cytochrome reductase and a subcomplex of the enzyme and from chymotryptic digestion of the isolated cytochrome c_1. The electron microscopic studies showed that the part of cytochrome reductase which contains the cytochrome c_1 protrudes 3–4 nm out of the membrane [5, 6]. Limited proteolysis of isolated cytochrome c_1 gave rise to a water-soluble cytochrome c_1 preparation of the apparent molecular mass of 24 kDa as compared to 31 kDa of the membrane-bound subunit. The water-soluble polypeptide still carries the heme and interacts with cytochrome *c* [7]. The amphiphilic character of the subunit was also predicted from the sequences determined previously for cytochrome c_1 of bovine heart [8], yeast [9] and the photosynthetic bacterium *Rhodopseudomonas sphaeroides* [10].

In this article we report on the sequence of the precursor and the mature form of cytochrome c_1 from *Neurospora crassa*. We discuss homologies between the *Neurospora* cytochrome c_1 and cytochrome c_1 from other organisms with regard of the mechanisms of the import of the precursor protein into mitochondria, the arrangement of the mature protein in the membrane and the interaction of the protein with cytochrome *c*.

Correspondence to J. Römisch, Institut für Biochemie der Universität Düsseldorf, Universitätsstraße 1, D-4000 Düsseldorf 1, Federal Republic of Germany

Enzymes. Ubiquinol–cytochrome-*c* reductase or cytochrome reductase (EC 1.10.2.2).

MATERIALS AND METHODS

Strains, materials and methods for identification of cDNA are described elsewhere [11–13]. A *Neurospora* cDNA bank was created by synthesizing single-stranded cDNA from polyadenylated RNA, double-strand synthesis and preparation of cDNA recombinant plasmids pBR 322, cloned in *Escherichia coli* strain 5K. cDNA was identified by hybridization selection of mRNA, cell-free protein synthesis and immunoprecipitation using antibodies against cytochrome c_1. Further clones were identified by three rounds of colony/filter hybridization. cDNA was sequenced by the Maxam and Gilbert method [14], except for modification at dA and dG, which was performed according to Burton [15]. Modification at dT was performed by the method of Rubin and Schmid [16].

Cytochrome reductase was prepared from *Neurospora* mitochondria as in [17]. Cytochrome c_1 was isolated by preparative SDS gel electrophoresis and antibodies against cytochrome c_1 were raised in rabbits as described elsewhere [18].

RESULTS

Isolation of cloned cDNA

The cDNA of the first clone J1 contains 104 base pairs corresponding to nucleotides 533–636 of the coding sequence (Fig. 2). Three cDNA clones, from a total of 50000 clones, hybridized with cytochrome c_1 cDNA probes. Since no full-length clone was found overlapping cDNA fragments were sequenced (Fig. 1). With the cDNA insert of clone J1, the clones J2, M1 and M2 were identified. J2 extends to the d(ATG) start codon, M1 to the d(TGA) stop codon. No poly(A) tail was found.

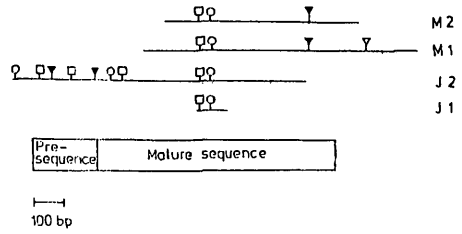

Fig. 1. *Schematic representation of the mRNA coding for cytochrome* c_1 *and of cDNA inserts used for sequencing.* Restriction sites were marked by following symbols: (▼) *Hin*fI, (□) *Alu*I, (▽) *Sfa*NI and (○) *Sau*96

The primary structure of the precursor protein and the mature subunit

The coding region of the cytochrome c_1 cDNA contains 996 nucleotides corresponding to a precursor protein of 332 amino acids with a molecular mass of 36496 Da (Fig. 2). The presumed start codon d(ATG) is preceded by a short sequence d(CACC) (not shown), similar to the start codons of other *Neurospora* mRNAs [11, 13, 19–23]. The presequence consists of 70 amino acids corresponding to a molecular mass of 7220 Da. This value agrees well with the apparent value of 7000 Da deduced from SDS gel electrophoresis of the mature

```
          1                                                  10
          MET LEU ALA ARG THR CYS LEU ARG SER THR ARG THR PHE ALA SER ALA LYS ASN GLY ALA
   1- 60  ATG CTG GCG AGG ACC TGC CTG CGC TCG ACG CGC ACC TTT GCC AGC GCG AAA AAT GGC GCC

                                                             30
          PHE LYS PHE ALA LYS ARG SER ALA SER THR GLN SER SER GLY ALA ALA ALA GLU SER PRO
  61-120  TTC AAA TTT GCC AAG CGT TCG GCT TCC ACC CAG AGC TCC GGC GCC GCT GCC GAG TCT CCC

                                                             50
          LEU ARG LEU ASN ILE ALA ALA ALA ALA ALA THR ALA VAL ALA ALA GLY SER ILE ALA TRP
 121-180  CTC CGC CTG AAC ATT GCT GCC GCC GCT GCC ACC GCC GTC GCC GCC GGC TCG ATC GCC TGG

                                                             70  *
          TYR TYR HIS LEU TYR GLY PHE ALA SER ALA MET THR PRO ALA GLU GLU GLY LEU HIS ALA
 181-240  TAC TAC CAT CTC TAT GGA TTC GCT TCC GCC ATG ACT CCG GCT GAG GAA GGT CTC CAT GCT

                                                             90
          THR LYS TYR PRO TRP VAL HIS GLU GLN TRP LEU LYS THR PHE ASP HIS GLN ALA LEU ARG
 241-300  ACC AAG TAC CCC TGG GTC CAC GAA CAG TGG CTC AAG ACC TTT GAT CAC CAA GCT CTT CGC

                                                            110
          ARG GLY PHE GLN VAL TYR ARG GLU VAL CYS ALA SER CYS HIS SER LEU SER ARG VAL PRO
 301-360  AGA GGT TTC CAG GTG TAC CGT GAG GTC TGC GCC TCG TGC CAC TCT CTC AGC AGA GTC CCC

                                                            130
          TYR ARG ALA LEU VAL GLY THR ILE LEU THR VAL ASP GLU ALA LYS ALA LEU ALA GLU GLU
 361-420  TAC CGC GCT CTT GTC GGC ACC ATC TTG ACC GTC GAT GAG GCC AAG GCT CTT GCC GAA GAG

                                                            150
          ASN GLU TYR ASP THR GLU PRO ASN ASP GLN GLY GLU ILE GLU LYS ARG PRO GLY LYS LEU
 421-480  AAC GAG TAC GAC ACC GAG CCC AAC GAC CAG GGC GAG ATC GAG AAG CGC CCC GGC AAG CTT

                                                            170
          SER ASP TYR LEU PRO ASP PRO TYR LYS ASN ASP GLU ALA ALA ARG PHE ALA ASN ASN GLY
 481-540  TCC GAC TAC CTC CCC GAT CCC TAC AAG AAC GAT GAG GCC GCC CGT TTC GCC AAC AAC GGT

                                                            190
          ALA LEU PRO PRO ASP LEU SER LEU ILE VAL LYS ALA ARG HIS GLY GLY CYS ASP TYR ILE
 541-600  GCC CTT CCT CCC GAT CTC AGC TTG ATC GTC AAG GCC CGC CAC GGT GGC TGC GAC TAC ATC

                                                            210
          PHE SER LEU LEU THR GLY TYR PRO ASP GLU PRO PRO ALA GLY ALA SER VAL GLY ALA GLY
 601-660  TTC TCC CTC CTT ACC GGC TAT CCC GAT GAG CCT CCC GCT GGC GCT TCC GTT GGC GCC GGC

                                                            230
          LEU ASN PHE ASN PRO TYR PHE PRO GLY THR GLY ILE ALA MET ALA ARG VAL LEU TYR ASP
 661-720  CTT AAC TTC AAC CCC TAC TTC CCC GGT ACC GGT ATC GCC ATG GCC CGC GTC CTC TAC GAC

                                                            250
          GLY LEU VAL ASP TYR GLU ASP GLY THR PRO ALA SER THR SER GLN MET ALA LYS ASP VAL
 721-780  GGC CTC GTC GAC TAC GAG GAT GGC ACC CCC GCC TCC ACC TCC CAG ATG GCC AAG GAT GTT

                                                            270
          VAL GLU PHE LEU ASN TRP ALA ALA GLU PRO GLU MET ASP ASP ARG LYS ARG MET GLY MET
 781-840  GTT GAG TTC CTC AAC TGG GCT GCT GAG CCC GAG ATG GAC GAC CGC AAG CGC ATG GGT ATG

                                                            290
          LYS VAL LEU VAL VAL THR SER VAL LEU PHE ALA LEU SER VAL TYR VAL LYS ARG TYR LYS
 841-900  AAG GTT CTG GTT GTC ACC TCT GTG CTC TTC GCC TTG AGC GTC TAT GTT AAG CGT TAC AAG

                                                            310
          TRP ALA TRP LEU LYS SER ARG LYS ILE VAL TYR ASP PRO PRO LYS ARG PRO PRO PRO ALA
 901-960  TGG GCT TGG CTC AAG TCG AGG AAG ATC GTC TAC GAT CCC CCC AAG AGT CCA CCA CCC GCC

                                                            330
          THR ASN LEU ALA LEU PRO GLN GLN ARG ALA LYS SER ***
 961-999  ACT AAT TTA GCA TTG CCA CAA CAA AGG GCA AAG TCA TGA
```

Fig. 2. *Nucleotide sequence of cDNA and deduced amino acid sequence of the precursor protein of cytochrome* c_1. Start of the mature protein is marked by an asterisk

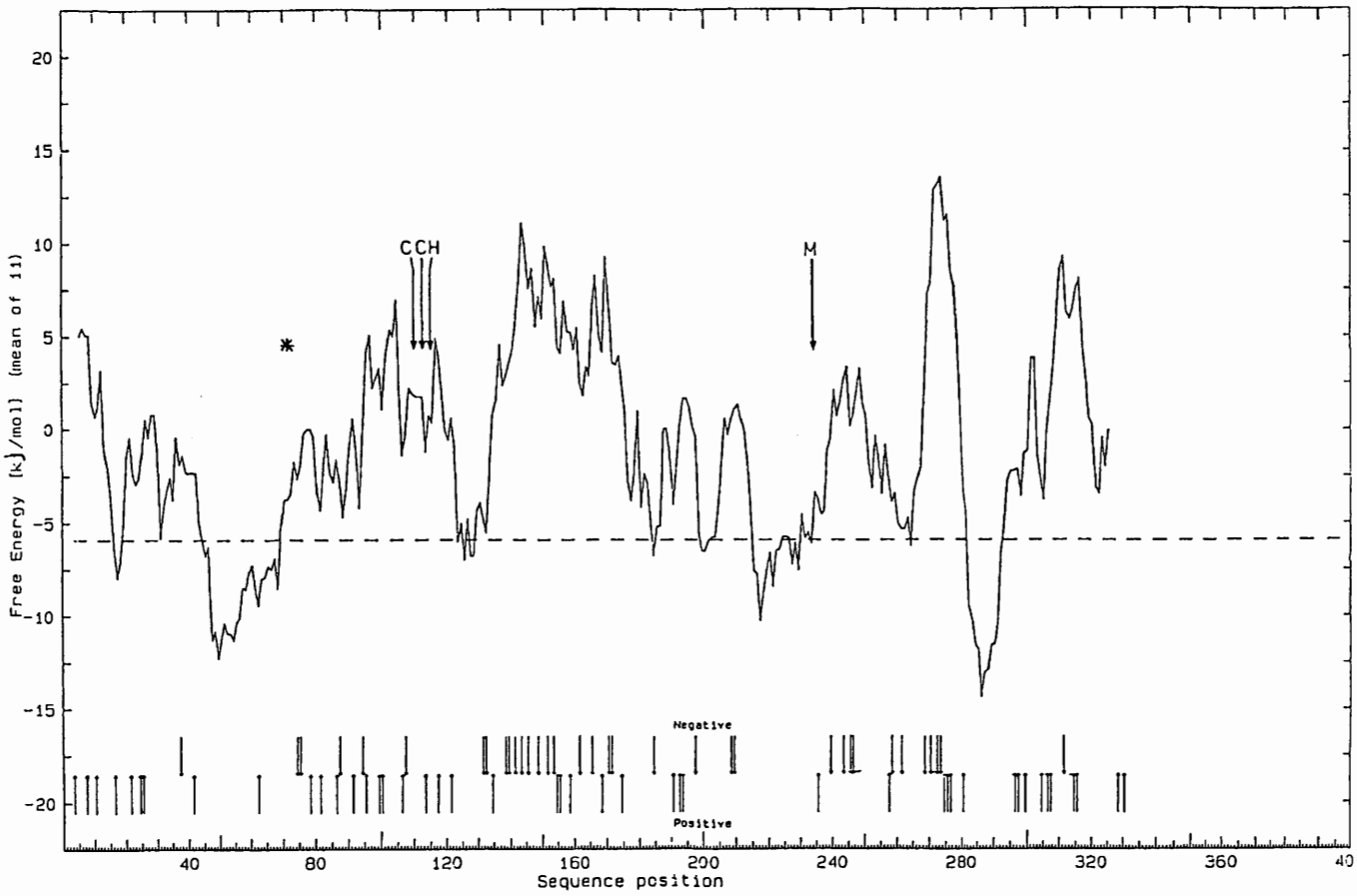

Fig. 3. *Polarity profile of the precursor protein of cytochrome* c_1 *from* N. crassa. The gain of free energy during transition of a 11-residue segment from water into the membrane is calculated for all sequence positions according to von Heijne [34]. Mean hydrophobicity per residue is indicated by the dotted line, negative and positive charges by arrows at the bottom. Start of the mature protein at position 71 is marked by an asterisk and the positions of the 2 Cys (C), 1 His (H) and 1 Met (M) involved in the binding of heme by arrows

cytochrome c_1, which migrates as a 31-kDa protein, and of the precursor migrating as a 38-kDa protein [18]. The presequence is basic containing three Lys, five Arg and one His. Glu-38 is the only negatively charged amino acid of the presequence. From the 70 residues 20 are Ala, some of them being clustered. Remarkably there is a stretch of 20 uncharged residues from Leu-43 to Trp-62, framed by the positively charged Arg-42 and His-63.

Start of the mature protein at Met-71 was deduced from the N-terminal amino acids sequence of isolated cytochrome c_1 reported by Tsugita et al. [24]. This partial sequence, however, differs from our sequence by amino acids in positions 1 and 9 and an additional Gly in 7. This difference we cannot explain.

With regard to the distribution of charged and hydrophobic residues (Fig. 3), the mature protein shows three distinguishable regions: a large hydrophilic N-terminal part of 211 residues (Met-71 to Lys-281), a non-polar stretch of 15 residues (Val-282 to Val-296) and a polar part of 36 amino acids C-terminal (Lys-297 to Ser-332). The molecular mass of the mature subunit including the heme is 29908 Da. By summing the mole fractions of polar residues [25] a polarity of 43% results. As reported elsewhere [7] the non-polar stretch plus the C-terminal polar part can readily be clipped off by chymotrypsin. The cleavage site for chymotrypsin in the sequence was derived from the difference in the amino acid composition of the detergent-bound cytochrome c_1 and the water-soluble cytochrome c_1 preparation. Digestion at Phe-263 or Trp-266 would give rise to a water-soluble preparation of molecular mass 22307 Da or 22720 Da respectively; and, for the clipped C-terminal polypeptide, 7601 Da or 7188 Da.

DISCUSSION

The amino acid sequence of the mature *Neurospora* cytochrome c_1 is 62% homologous to that of yeast [9], 56% to that of bovine heart [8] and 31% to that of *R. sphaeroides* [10] (Fig. 4). The heme-binding region of the protein, located near the N terminus, is more conservative and the homologies between the *Neurospora*, yeast, bovine, *Rhodopseudomonas* and *Paracoccus* proteins [26] amount to 87%, 73%, 67% and 67% respectively. The heme is covalently linked by Cys-110 and Cys-113, probably chelated by His-114 as fifth ligand. Most likely the sixth ligand of the heme is Met-234 [27], which is also conserved in the sequences of the yeast, bovine and *Rhodopseudomonas* protein. The sites for interaction of bovine heart cytochrome c_1 and horse heart cytochrome *c* were positioned in two negatively charged regions of the hydrophilic domain (from Asp-62 to Glu-84 and from Asp-170 to Asp-177), which are found to be conserved in the different organisms [28, 29]. The stretch of 15 uncharged amino acids, located near the C terminus, most probably anchors the protein to the membrane.

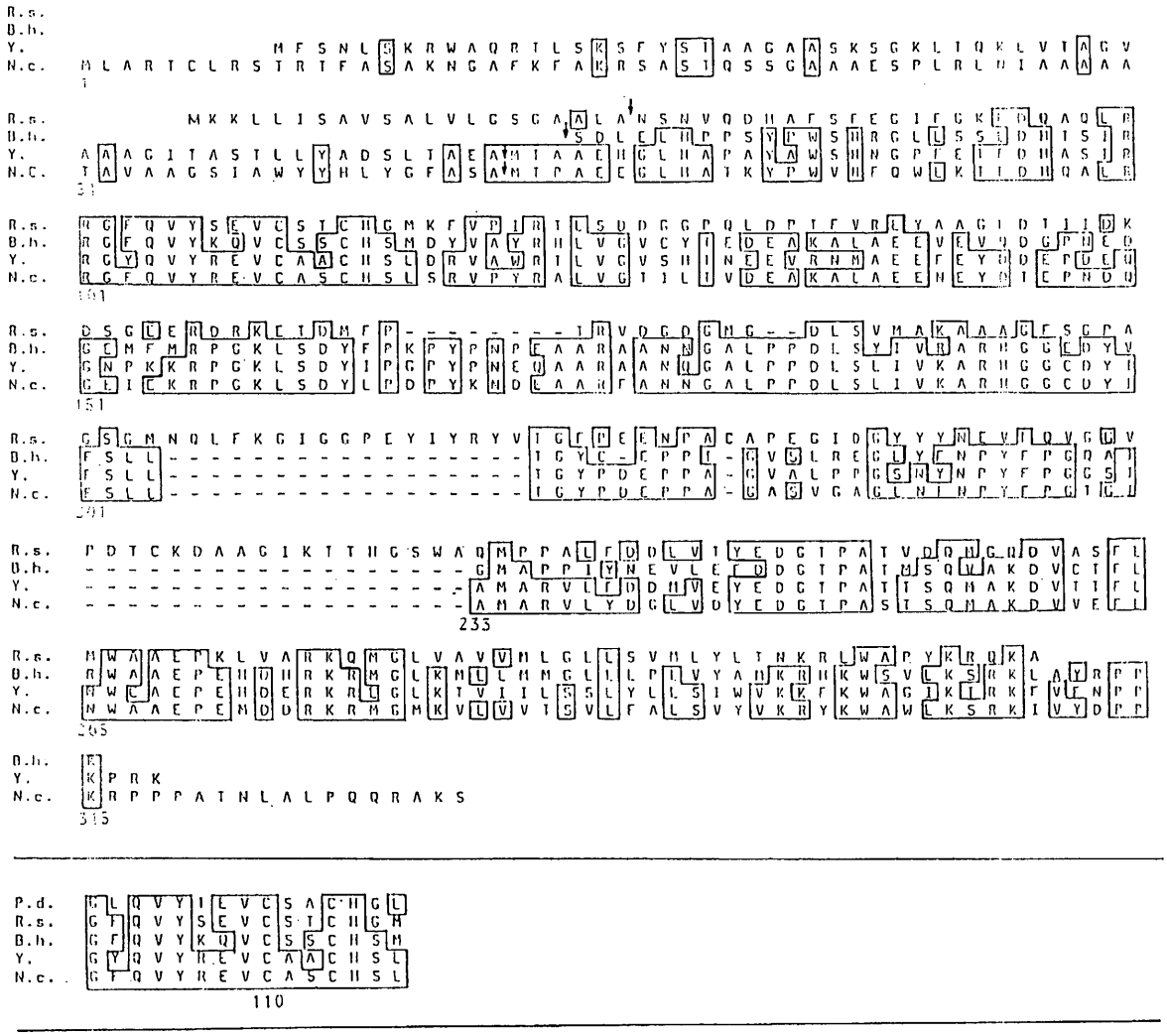

Fig. 4. *Comparison between the amino acid sequences of the precursor cytochrome* c_1 *from* Neurospora *(N. c.) and yeast (Y.), and the mature cytochrome* c_1 *from bovine heart (B. h.) and* R. sphaeroides *(R. s.).* The numbers pertain to the sequence of the *Neurospora* preprotein. Start of the mature proteins is indicated by arrows. The lower panel compares heme-binding regions of *Neurospora*, yeast, bovine heart, *Rhodopseudomonas* and *Paracoccus* (P. d.). Regions of homology with *Neurospora* cytochrome c_1 are boxed

Regarding the import mechanism, the precursor protein probably binds to an 'import receptor', located at the mitochondrial outer membrane [30–32]. Part of the N-terminal precursor penetrates a junction of outer and inner membrane [33]. As discussed elsewhere [31], the strongly positive charge of the presequence, which is found in many precursors imported into mitochondria [35], could channel the preprotein into the bilayer by an electrophoretic effect. A stretch of uncharged amino acids in the presequence is assumed to anchor the precursor in the membrane, while the N terminus becomes exposed to the matrix. Although there is no homology between the presequences of yeast and *Neurospora* cytochrome c_1, the distribution of charged and hydrophobic amino acids is remarkably similar. They contain a stretch of 19 or 20, respectively, uncharged residues framed by 2 positive residues. This stretch could span the membrane during import, as assumed for yeast preprotein [9]. The cleavage by a matrix protease leaves an intermediate import product 3000 Da shorter than the precursor protein. Therefore the first cleavage site is probably located in the region of 11 non-polar amino acids exposed to the mitochondrial matrix, agreeing well with the yeast precursor. After attachment of heme the intermediate can be cleaved by a protease located in the intermembrane space. Anchored in the bilayer the mature cytochrome c_1 can be assembled with the other subunits of cytochrome reductase.

We would like to thank Regina Bienert for excellent technical assistance. This work was supported by the *Deutsche Forschungsgemeinschaft*, the *Fonds der Chemischen Industrie* and the *Genzentrum München*.

REFERENCES

1. Hatefi, Y. (1985) *Annu. Rev. Biochem. 54*, 1015–1069.
2. König, B. W., Osherhoff, N., Wilms, J., Muijers, A. O., Dekker, H. L. & Margoliash, E. (1980) *FEBS Lett. 111*, 395–398.
3. Bosshard, H. R., Zürrer, M., Schägger, H. & von Jagow, G. (1979) *Biochem. Biophys. Res. Commun. 96*. 250–258.
4. Yu, C. A., Yu, L. & King, T. E. (1973) *J. Biol. Chem. 248*, 528–533.
5. Leonard, K., Wingfield, P., Arad, T. & Weiss, H. (1981) *J. Mol. Biol. 149*, 259–274.

6. Karlsson, B., Hovmöller, S., Weiss, H. & Leonard, K. (1983) *J. Mol. Biol. 165*, 287–302.
7. Li, Y., Leonhard, K. & Weiss, H. (1981) *Eur. J. Biochem. 116*, 199–205.
8. Wakabayashi, S., Matsubara, H., Kim, C. H., Kawai, K. & Kink, T. E. (1980) *Biochem. Biophys. Res. Commun. 97*, 1548–1554.
9. Sadler, I., Suda, K., Schatz, G., Kaudewitz, F. & Haid, A. (1984) *EMBO J. 3*, 2137–2143.
10. Gabellini, N. & Sebald, W. (1986) *Eur. J. Biochem. 154*, 569–579.
11. Viebrock, A., Perz, A. & Sebald, W. (1982) *EMBO J. 1*, 565–571.
12. Viebrock, A., Perz, A. & Sebald, W. (1983) *Methods Enzymol. 97*, 254–260.
13. Arends, H. & Sebald, W. (1984) *EMBO J. 3*, 377–382.
14. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
15. Burton, K. (1967) *Methods Enzymol. 12A*, 222–224.
16. Rubin, C. M. & Schmid, C. W. (1980) *Nucleic Acid Res. 8*, 4613–4619.
17. Weiss, H. & Kolb, H. J. (1979) *Eur. J. Biochem. 99*, 139–149.
18. Teintze, M., Slaughter, M., Weiss, H. & Neupert, W. (1982) *J. Biol. Chem. 257*, 10364–10371.
19. Harnisch, U., Weiss, H. & Sebald, W. (1985) *Eur. J. Biochem. 149*, 95–99.
20. Kinnaird, J. H. & Fincham, J. R. S. (1983) *Gene 20*, 253–260.
21. Alton, N. K., Buxton F., Patel, V., Giles, N. H. & Vapnek, D. (1982) *Proc. Natl Acad. Sci. USA 79*, 1955–1959.
22. Schechtman, M. G. & Yanofski, C. (1983) *J. Mol. Appl. Genet. 2*, 83–99.
23. Woudt, L. P., Pastink, A., Kempers-Veenstra, A. E., Jansen, A. E. M., Mager, W. H. & Planta, R. J. (1983) *Nucleic Acid Res. 11*, 5247–5360.
24. Tsugita, A., Gregor, J., Kubota, J. & v. D. Broek, R. (1979) in *Cytochrome oxidase* (King, T. E., Ovii, Y., Chance, B. & Okonuki, K. eds) pp. 67–77, Elsevier/North Holland, New York.
25. Capaldi, R. A. & Vanderkooi, G. (1972) *Proc. Natl Acad. Sci. USA 69*, 930–932.
26. Ludwig, B., Suda, K. & Cerletti, N. (1983) *Eur. J. Biochem. 137*, 597–602.
27. Kaminsky, L. S., Chiang, Y. L. & King, T. E. (1975) *J. Biol. Chem. 250*, 7280–7287.
28. Stonehuerner, J., O'Brien, P., Geren, L., Millett, F., Steidl, J., Yu, L. &Yu, C. (1985) *J. Biol. Chem. 9*, 5392–5398.
29. Broger, C., Salardi, S. & Azzi, A. (1983) *Eur. J. Biochem. 131*, 349–352.
30. Neupert, W. & Schatz, G. (1981) *Trends Biochem. Sci. 6*, 1–4.
31. Hay, R., Böhni, P. & Gasser, S. (1984) *Biochim. Biophys. Acta 779*, 65–87.
32. Zwizinski, C., Schleyer, M. & Neupert, W. (1984) *J. Biol. Chem. 259*, 7850–7856.
33. Schleyer, M. & Neupert, W. (1985) *Cell 43*, 339–350.
34. Heijne, v. G. (1981) *Eur. J. Biochem. 120*, 275–278.
35. Heijne, v. G. (1986) *EMBO J. 5*, 1335–1342.