Nucleotide sequence of the gene for apocytochrome *b*-559 on the spinach plastid chromosome: implications for the structure of the membrane protein

Reinhold G. Herrmann, Juliane Alt, Barbara Schiller, W.R. Widger[+] and William A. Cramer[+]

Botanisches Institut der Universität, Universitätsstr. 1, 4 Düsseldorf 1, FRG and [+]*Department of Biological Sciences and Biochemistry, Purdue University, West Lafayette, IN 47907, USA*

Received 13 August 1984

Analysis of a 0.6 kb fragment of the spinach plastic chromosome adjacent to the 3′ end of the apocytochrome *f* gene has disclosed two uninterrupted reading frames of 83 and 39 triplets, the product of the first one being apocytochrome *b*-559. The first 27 predicted amino acid residues had been verified by protein sequence analysis and the molecular mass of 9390 Da derived from the amino acid sequence deduced here is close to that of the authentic protein. The two genes are transcribed by a bicistronic RNA in a direction opposite to that of cytochrome *f*, and their translation stop/potential ribosome binding sites overlap. Features of the two genes resembling those of bacterial genes include putative tetra- or pentanucleotide Shine–Dalgarno sequences, Pribnow boxes, '–35' promotor consensus sequences and possibly a transcription termination region. Both gene structure and products of DNA- or RNA-programmed cell-free translation preclude that apocytochrome *b*-559 is made as a precursor. The amino acid sequence includes only one histidine residue located in a predicted secondary structure of strong hydrophobicity which indicates the intriguing possibility that more than one protein chain must cooperate in heme binding of this cytochrome.

Plastid DNA *Cytochrome* b-*559* *Amino acid sequence* *Hydropathy* *Thylakoid membrane* *Spinach*

1. INTRODUCTION

The lack of understanding of the physiological function of the chloroplast cytochrome *b*-559 has led to the accumulation of a great deal of phenomenological data concerning its properties [1]. The high potential redox state of this cytochrome in situ is known to be unusually labile, and the different functions of this cytochrome that have been proposed without resolution of the different possibilities, are summarized in [2]. They include, (i) redox function in water splitting, (ii) participation as an electron and proton carrier in the main chain, (iii) a redox carrier in a cycle around photosystem II, (iv) a redox carrier on a side path or branch from the plastoquinone pool, (v) reduced *b*-559 as a proton acceptor for the S_3, and possibly S_2, state of the water-splitting complex, and (vi) overlap or interaction of cytochrome *b*-559 with the DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea) binding site. An understanding of the unique properties of cytochrome *b*-559 in situ, and further insight into its function, would presumably be facilitated by structural information. In the case of the other chloroplast *b* cytochrome, cytochrome b_6, such information became available [3] when its sequence was completed [4], compared with the highly homologous cytochrome *b* sequences of mitochondrial complex III [5–9], and a unique prediction made of the position in the membrane of the two heme-binding sites [3,10]. Therefore, it was of interest to determine the nucleotide sequence of the gene for the chloroplast cytochrome *b*-559. This became possible when the gene was localized on the spinach plastid chromosome [11] and facilitated by determination of an N-terminal 27 residue sequence [12].

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

2. MATERIALS AND METHODS

2.1. *Recombinant DNA technique*

The parental plasmid pWHsp404 containing the 8.0 kilobase pair (kbp) primary fragment BamHI-4 of spinach plastid DNA as an insert, or the 5 EcoRI subclones of the 4.5 kbp overlap between fragments BamHI-4 and SalI-3 [11], designated E1 to E5 (inserts 2.25, 0.85, 0.50 or 0.27 kbp), were used. The construction and selection of these plasmids has been outlined previously [11]. Plasmid DNA was isolated on a large or small scale by previously described methods [13,14]. Restriction enzymes SalI, EcoRI, BamHI, TaqI, RsaI, HinfI, DdeI, Sau3A and Ava II were purchased from Boehringer (Mannheim) or Biolabs (Bad Schwalbach, Taunus). The restriction fragments were analysed on 0.8–1.8% agarose or 4–6% polyacrylamide slab gels and stained with ethidium bromide for visualization under UV light. Excised inserts or their partial digestion products were recovered by electroelution and purified by phenol extraction and ethanol precipitation after concentration with isobutanol [15].

2.2. *DNA sequence analysis*

Excised DNA fragments prepared as above were digested with appropriate restriction endonucleases to obtain small secondary fragments for sequence analysis. These fragments were labelled at their 3′ termini by fill-in synthesis using appropriate α-^{32}P-labelled deoxyribonucleoside triphosphates (Amersham, Braunschweig) and the large fragment of DNA polymerase I (Boehringer, Mannheim) as previously described [16]. The radioactive ends were recovered by single-strand separation [17] or after digestion with a second restriction endonuclease in appropriate polyacrylamide slab gels. The DNA fragments were eluted from excised and macerated gel pieces by diffusion in 0.25 M NaCl and purified by chromatography on Elutip columns (Schleicher and Schüll, Dassel) without any ethanol precipitation step. Base-specific chemical modification (C, T + C, A > C, and G) was performed according to [17] and partial degradation products were separated by electrophoresis on 0.2 mm polyacrylamide slab gels (4, 5 or 20%; 40 or 80 × 20 cm). Further resolution of the autoradiograms was achieved by drying the gels prior to exposure.

3. RESULTS AND DISCUSSION

The previous study has shown that in vitro synthesis of apocytochrome *b*-559 is directed by the 8 kbp fragment BamHI-4, or by a 1400 base mRNA selected with this fragment from purified chloroplast RNA, of spinach [11]. The derivative EcoRI clones of BamHI-4 did not contain the entire protein-coding sequence. Hybrid selection using these subclones allowed assignment of the functional transcript to the overlap of 0.65 (E3) and 0.50 (E4) kbp EcoRI fragments within the 4.5 kbp segment shared by fragments BamHI-4 and SalI-3 (20.3 kbp; fig.1).

From the outlined results it seemed likely that the 1.0 kbp segment bounded by the 0.65 and 0.50 kbp EcoRI fragments together would encompass the cytochrome gene and flanking sequences. This DNA segment was therefore sequenced [17] according to the scheme diagrammed in fig.1. Initial information was obtained from analysing all EcoRI restriction sites within this region. These data allowed the location of new restriction endonuclease cleavage sites by computer analysis. Also, a DNA sequence corresponding to the determined N-terminal amino acid sequence [12] was found on the 650 bp EcoRI fragment. To find fragment overlaps and the carboxyl terminus of the gene, digests of the 4.5 kbp SalI/BamHI fragment obtained with appropriate endonucleases were subjected to Southern blot analysis [18] and probed with the 0.65 and 0.50 kbp EcoRI sequences nick-translated [19] to high specific activity. Selected fragments were subsequently used to confirm and extend the existing DNA sequence. In this way both strands and the entire protein-coding region with its flanking regions were sequenced. The approach confirmed the location of the *b*-559 apocytochrome gene and the arrangement of the EcoRI secondary fragments of 0.85–0.65–0.50–0.27 kbp in the order given (fig.1).

Computer analysis of the 600 nucleotide interval revealed two uninterrupted reading frames of 83 and 39 triplets. The non-coding strand of this segment, transcription polarity, and derived amino acid sequence of these genes are depicted in fig.2. The region coding for apocytochrome *b*-559 and the correct reading phase of this protein were identified within this sequence by comparison with the known amino-terminal sequence. This frame starting with-

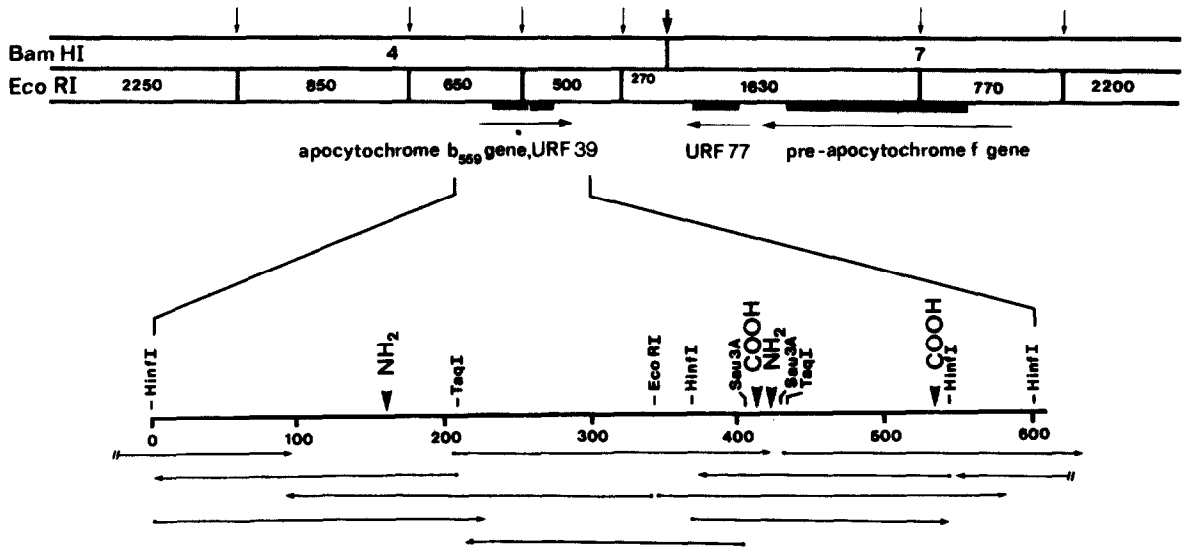

Fig.1. Restriction map and organization of the genes for the cytochromes *b*-559, *f* and two unidentified reading frames (URF 39, URF 77) of the spinach plastid chromosome. The approximate locations of restriction sites for the enzymes BamHI (fragments designated in decreasing size) and EcoRI (distances given in base pairs) are shown. Some fragment sizes have been refined (cf., [11]). The position of genes is indicated by closed boxes and their direction of transcription by arrows below the map. The lower expanded part of the diagram illustrates the sequence assay strategy. Numbering is in base pairs. Only relevant restriction cleavage sites are indicated. Arrows represent the direction and extent of individual DNA sequencing assays. The sites of 3′ end labelling are indicated by points; slash marks mean that the sites of labelling are not shown. The locations of the amino-terminal and carboxy-terminal coding regions for *b*-559 and URF 39 are marked.

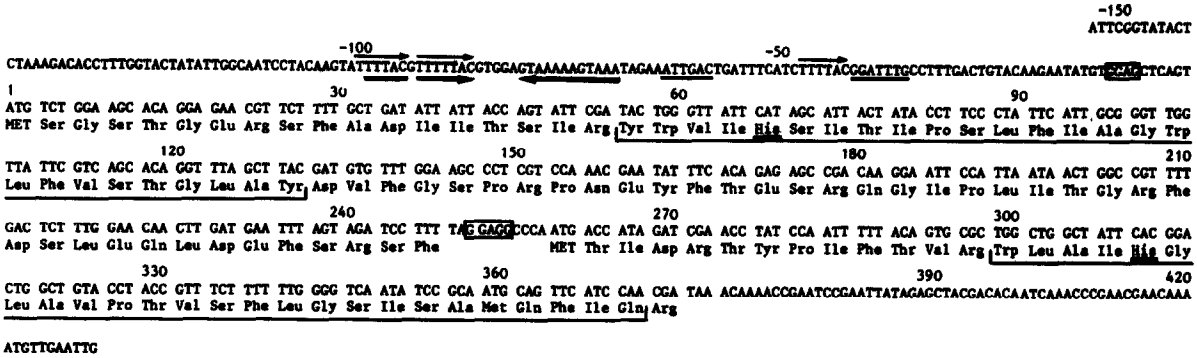

Fig.2. The nucleotide sequence of the cytochrome *b*-559 gene, URF 39 and flanking regions from spinach. The sequence of the non-transcribed strand of the genes are arranged in codons and the corresponding amino acids indicated. The sequence presents approximately 600 nucleotides of the map shown in fig.1. Nucleotides are numbered in the 5′–3′ direction with the initiative methionine in the *b*-559 gene designated +1. Possible ribosome binding sites are boxed, inverted, tandem and direct repeats are indicated by arrows, and histidine residues are underlined. Hydrophobic regions which might span the membrane are indicated by horizontal brackets.

in the 0.65 kbp EcoRI fragment at an ATG codon 185 bases upstream of the RI site separating the 0.65 and 0.50 kbp fragments, and terminating at the TAG amber codon 64 bp within the latter, defines a coding region of a polypeptide of 83 amino acid residues. Only one of the 6 possible reading frames across the 0.65/0.50 kbp fragment junction was free of stop codons over approximately 250 bp.

In accord with hybrid selection data [11], fragments E3 and E4 harbor 185 base pair N- and 64 bp C-terminal sequences of the gene. The calculated molecular mass of this polypeptide, 9390 Da, is in good agreement with molecular mass determinations (9–10 kDa) for the apocytochrome [11,12,20, 21]. We conclude, therefore, that this DNA is the apocytochrome *b*-559 gene. This result also established that the direction of transcription of the gene is from the 0.65 kbp EcoRI fragment towards fragment BamHI-7, that is, divergent from that of the adjacent genes for cytochrome *f* (fig.1; [22,23]) and for ribulose bisphosphate carboxylase/oxygenase large subunit [24].

Of the deduced residues, 26 are confirmed by protein sequence analysis of spinach apocytochrome *b*-559 [12]. Residue 23, which was not resolved by protein sequencing, is histidine. The TCT triplet that encodes the N-terminal serine of the protein is preceded by methionine. The sequence is numbered from the ATG codon. Just upstream of this methionine is a potential ribosome-binding site GGAG [25] located 10 nucleotides prior to the ATG triplet (fig.2). This tetranucleotide is complementary to a sequence at the 3′ end of 16 S plastid rRNA of maize [26] and tobacco [27] and of 16 S rRNA of *Escherichia coli* [25]. We consider, therefore, that translation of the apocytochrome of spinach starts at this residue rather than at one of the 11 amino acids proximal to the next in-frame stop codon (which do not include methionine). Furthermore, within the transcribed but untranslated segment, at positions −40 to −34 (GGATTTG) and −62 to −57 (ATTGAC), typical prokaryotic-like motifs resembling Pribnow box and '−35' consensus sequences of *E. coli* promotors [28] are found. This organization, in our opinion, precludes the possibility that apocytochrome *b*-559 is decoded as a larger precursor that has been claimed for spinach in [29]. Such a precursor would also not be in accord with data obtained by cell-free translation in DNA- and RNA-programmed systems [11]. To date, only two of 18 chloroplast proteins synthesized on organelle ribosomes and coded for by plastome genes are made as precursors, one being apocytochrome *f* with an N-terminal extension of 35 amino acids [22,23,30]. The other is the '32 kDa' herbicide-binding protein associated with the photosystem II reaction center [31] which probably carries a C-terminal excess [32].

The second open reading frame in this sequence, beginning at an ATG codon at position 261 and ending at a TAA ochre codon (position 378) had the same direction as the *b*-559 gene but was read in different phase. The size of this (hydrophobic) 39-residue protein, 4.4 kDa, is smaller than that of the *b* cytochrome. Its nature is unknown; no significant homology with the sequence from any other protein monitored was found. DNA sequences specific to this reading frame, e.g., the 55 bp Sau3A-RsaI fragment, nucleoties 270–325, hybridize to or select the same 1.4 kb large RNA species, as do sequences specific for cytochrome *b*-559 (TaqI-EcoRI, positions 52–184 or EcoRI-Sau3A, positions 185–242). No indication of smaller or larger RNA species hybridizing to protein-coding sequences of any of these proteins was seen [11]. Apparently, the two genes are co-transcribed into a single RNA species and, though we have not explicitly shown that the second protein is decoded, we anticipate that the transcript is used as a bicistronic RNA in vivo. The third nucleotide (G) of the UAG stop codon of the cytochrome *b*-559 coding sequence together with the following 4 bases may comprise a pentanucleotide Shine–Dalgarno sequence (GGAGG) for the second protein. No open reading frame is apparent in the segment overlapping into fragment BamHI-7 (a reading frame of 77 codons specifying a basic protein begins 280 bp distal to the gene for cytochrome *f*; fig.1,2). The narrow gene spacing and transcriptional organization is reminiscent of that of the clustered genes for the ATP synthase β and ϵ subunits [33], and for the '32 kDa-like' protein and the 44 kDa chlorophyll *a* apoprotein of the photosystem II reaction center [34]. The former genes overlap with 4 bases of their protein-coding sequence in spinach [35] and maize [36], the latter with 50 bp in spinach [34] and contain cryptic Shine–Dalgarno sequences within the beta and '32-kDa'-like protein structural genes, respectively.

Neither the gene for apocytochrome *b*-559 nor the smaller unidentified reading phase show prokaryotic-like transcription termination secondary structures [28] immediately distal to their translation stop codons. However, sequences capable of forming relatively stable stem-loop structures follow further downstream (position 394–439, $\Delta G = -9.4$ kcal; position 826–859, $\Delta G = -16.4$ kcal;

calculated according to [37] and within URF 77 ($\Delta G = -21.6$ kcal). This suggests either that the genes use signals that are different or that the transcript terminates further downstream, halfway between or closer to the cytochrome *f* gene at one of the indicated stem and loop structures (not shown).

Codon utilization for apocytochrome *b*-559 and the unidentified protein, like other plastic genes, occurs non-randomly but does not deviate from the general code (fig.3). Although approximately 50 of the possible 61 codons are used, specific choices are preferred for some amino acids, generally reflecting a preference for A and T in the third codon position as would be expected from the relatively low GC content of plastid DNA (37 mol%).

The 83 amino acid coding sequence was localized using antibody to an HPLC-purified 9–10-kDa polypeptide from spinach cytochrome *b*-559 [12]. The critical residues in this sequence for heme coordination are His, Met, and Lys. The nucleotide sequence shown in fig.2 confirms the conclusion from the amino acid composition analysis that the polypeptide contains one histidine residue, but shows no lysines and no internal methionines, thus removing some ambiguity in the low levels of these residues present in the original composition analysis. The coordination of the heme is *bis*-histidine, as determined by EPR and Raman spectroscopic analysis of the cytochrome in situ, the purified cytochrome, and a *bis*-histidine model compound (G. Babcock et al., in preparation). The *bis*-histidine coordination implies that the protein must contain at least two polypeptides, since the 83-residue polypeptide contains only one histidine residue. This histidine, at position 23, is located toward the N-terminal side of a 26-residue span of hydrophobic amino acids determined according to [38]. The simplest structure of the cytochrome *b*-559 protein would be a homodimer of the 83-residue component. The presence of the contiguous 39-residue sequence also containing one histidine residue in a hydrophobic region suggests the alternate possibility of a heterodimeric structure made of the 83- and 39-residue polypeptides.

1st	2nd T	2nd C	2nd A	2nd G	3rd
T	Phe 5/2	Ser 3/1	Tyr 1/1	Cys 0/0	T
	Phe 3/1	Ser 2/1	Tyr 2/0	Cys 0/0	C
	Leu 3/0	Ser 0/1	Stop 0/1	Stop 0/0	A
	Leu 1/1	Ser 0/0	Stop 1/0	Trp 2/1	G
C	Leu 1/0	Pro 2/1	His 1/0	Arg 3/0	T
	Leu 0/0	Pro 0/0	His 0/1	Arg 0/1	C
	Leu 1/0	Pro 2/1	Gln 2/1	Arg 2/2	A
	Leu 0/2	Pro 0/0	Gln 0/1	Arg 0/0	G
A	Ile 7/2	Thr 2/0	Asn 0/0	Ser 2/0	T
	Ile 0/1	Thr 1/3	Asn 1/0	Ser 5/0	C
	Ile 2/2	Thr 3/1	Lys 0/0	Arg 1/0	A
	Met 1/2	Thr 0/0	Lys 0/0	Arg 0/0	G
G	Val 1/1	Ala 2/2	Asp 3/1	Gly 2/0	T
	Val 1/0	Ala 0/0	Asp 1/0	Gly 1/0	C
	Val 0/1	Ala 0/1	Glu 4/0	Gly 4/1	A
	Val 1/1	Ala 1/0	Glu 1/0	Gly 0/1	G

Fig.3. Codon usage for cytochrome *b*-559/URF 39 from the spinach plastid chromosome (cf. fig.1,2).

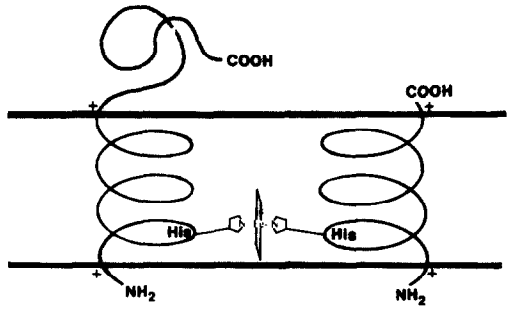

Fig.4. Predicted two-chain model (heterodimer version) for cytochrome *b*-559 in the thylakoid membrane. The hydrophobic amino acid residues traversing the membrane are arranged in α-helices. The folding of the external amino acids and the number of polypeptide chains are purely schematic.

For either case a well-defined model for the structure of cytochrome *b*-559 emerges from these sequence studies (fig.4). The heme-containing cytochrome would be at least a dimer that spans the hydrophobic membrane core one time, with the *bis*-histidine heme ligands 5 residues from the aqueous phase on one side of the membrane defined by the presence of an arginine residue. This would be the first example of a dimeric or oligomeric membrane-bound cytochrome with the heme cross-linking the polypeptide chains, and might explain the usual lability of this cytochrome and of its redox properties.

ACKNOWLEDGEMENTS

We are grateful to Dr P. Westhoff for help in the preparation of this manuscript, and to Ms Gabriele Schewe for skilful technical assistance. This work was supported with financial aid from

the Stiftung Volkswagenwerk, by Forschungsmittel des Landes Nordrhein/Westfalen and by the Deutsche Forschungsgemeinschaft (grant He 693).

REFERENCES

[1] Cramer, W.A. and Whitmarsh, J. (1977) Ann. Rev. Plant Physiol. 28, 133–172.
[2] Cramer, W.A., Whitmarsh, J. and Widger, W.R. (1981) in: Photosynthesis II. Electron Transport and Phosphorylation (Akoyunoglou, G. ed.) pp. 509–522, Internat. Science Serv., Philadelphia, PA.
[3] Widger, W.R., Cramer, W.A., Herrmann, R.G. and Trebst, A. (1984) Proc. Natl. Acad. Sci. USA 81, 674–678.
[4] Heinemeyer, W., Alt, J. and Herrmann, R.G. (1984) Curr. Genet. 8, 543–549.
[5] Anderson, S., Bankeir, A.T., Barrell, B.C., De Bruijn, M.H.L., Coulson, A.R., Dronin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schrier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature 290, 457–465.
[6] Anderson, S.L., De Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1981) J. Mol. Biol. 156, 683–717.
[7] Bibb, M.J., Van Etten, R.A., Wraight, C.T., Walberg, M.W. and Clayton, D.A. (1981) Cell 26, 167–180.
[8] Nobrega, F.G. and Tzagaloff, A. (1980) J. Biol. Chem. 255, 9828–9837.
[9] Waring, R.B., Davies, R.W., Lee, S., Grisi, E., McPhail Berks, M. and Scazzocchio, C. (1981) Cell 27, 4–11.
[10] Saraste, M. (1984) FEBS Lett. 166, 367–372.
[11] Westhoff, P., Alt, J., Widger, W.R., Cramer, W.A. and Herrmann, R.G. (1984) Plant Mol. Biol., in press.
[12] Widger, W.R., Cramer, W.A., Hermodson, M., Meyer, D. and Gullifor, M. (1984) J. Biol. Chem. 259, 3870–3876.
[13] Alt, J., Winter, P., Sebald, W., Moser, J.G., Schedel, R., Westhoff, P. and Herrmann, R.G. (1983) Curr. Genet. 7, 129–138.
[14] Westhoff, P., Alt, J., Nelson, N., Bottomley, W., Bünemann, H. and Herrmann, R.G. (1983) Plant Mol. Biol. 2, 95–107.
[15] Stafford, D.W. and Beiber, D. (1975) Biochim. Biophys. Acta 378, 18–21.
[16] Morris, J. and Herrmann, R.G. (1984) Nucl. Acids. Res. 12, 2837–2850.
[17] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[18] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[19] Rigby, P.W.J., Dieckmann, J., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.
[20] Koenig, F. and Lindberg Møller, B. (19892) Carlsberg Res. Commun. 47, 245–262.
[21] Metz, J.G., Ulmer, G., Bricker, T.M and Miles, D. (1983) Biochim. Biophys. Acta 725, 203–209.
[22] Alt, J., Westhoff, P., Sears, B.B., Nelson, N., Hurt, E., Hauska, G. and Herrmann, R.G. (1983) EMBO J. 2, 979–986.
[23] Alt, J. and Herrmann, R.G. (1984) Curr. Genet. 8, 551–557.
[24] Zurawski, G., Perrot, B., Bottomley, W. and Whitfeld, P.R. (1981) Nucl. Acids Res. 9, 3251–3270.
[25] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.
[26] Schwarz, Z. and Kössel, H. (1980) Nature 283, 739–742.
[27] Tohdoh, N. and Sigiura, M. (1982) Gene 17, 213–218.
[28] Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319–353.
[29] Zielinski, R.E. and Price, C.A. (1980) J. Cell Biol. 85, 435–445.
[30] Willey, D.L., Auffret, A.D. and Gray, J.G. (1984) Cell 36, 555–562.
[31] Grebanier, A.E., Coen, D.M., Rich, A. and Bogorad, L. (1978) J. Cell Biol. 78, 734–746.
[32] Marder, J.B., Goloubinoff, P. and Edelman, M. (1983) J. Biol. Chem. 259. 3900–3908.
[33] Westhoff, P., Nelson, N., Bünemann, H. and Herrmann, R.G. (1981) Curr. Genet. 4, 109–120.
[34] Alt, J., Morris, J., Westhoff, P. and Herrmann, R.G. (1984) Curr. Genet., submitted.
[35] Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1982) Proc. Natl. Acad. Sci. USA 79, 6260–6264.
[36] Krebbers, E.T., Larrinua, I.M., McIntosh, L. and Bogorad, L. (1982) Nucl. Acids Res. 10, 4985–5002.
[37] Tinoco, J., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1983) Nat. New Biol. 246, 40–41.
[38] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.