Nucleotide sequence of the clustered genes for apocytochrome b6 and subunit 4 of the cytochrome b/f complex in the spinach plastid chromosome

Wolfgang Heinemeyer, Juliane Alt, and Reinhold G. Herrmann

Botanisches Institut der Universitäts, Universitätsstr. 1, D-4000 Düsseldorf 1, Federal Republic of Germany

Summary. A 2.4 kilobase-pair segment of the spinach plastid chromosome carrying the genes for apocytochrome b6 and subunit 4 of the thylakoid membrane cytochrome b/f complex has been analysed by DNA sequencing and Northern blot analysis. The nucleotide sequence reveals two uninterrupted open reading frames of 211 and 139 triplets coding for two hydrophobic proteins of 23.7 kd (cytochrome b6) and 15.2 kd (subunit 4). The genes are located on the same strand and are separated from each other by 1018 untranslated base pairs. They map adjacent to the gene for the P₆₈₀ chlorophyll a apoprotein of the photosystem II reaction center. The three genes appear to be under common transcriptional control and the transcripts post-transcriptionally modified. The deduced amino acid sequences of cytochrome b6 and subunit 4 both exhibit significant homology with published sequences from mitochondrial b cytochromes (42 kd) suggesting that these functionally equivalent polypeptides in photosynthetic and respiratory electron transport chains arose monophyletically.

Key words: Plastid DNA – Cytochrome *b6* gene – Amino acid sequence – Hydropathy – Thylakoid membrane – Transcript modification – Evolution – Spinach

Introduction

The cytochrome bf complex of chloroplast thylakoid membranes and the cytochrome bc_1 complex on the inner mitochondrial membrane are functionally and structurally related multisubunit assemblies (Hauska et al. 1983). They act as proton-translocating plastoquinolplastocyanin and ubiquinol-cytochrome c oxidoreductases in photosynthetic and oxidative respiratory electron transport chains, respectively. Both contain c- and b-type cytochromes and high potential iron-sulfur centers. The mitochondrial complex in addition has several low molecular weight core polypeptides (Capaldi 1982).

Current analysis of the molecular organization and complex biogenesis of energy transducing membranes proceeds mainly along two paths: (i) DNA sequence analysis directed towards the derivation of the primary and secondary structures of the membrane proteins, as well as identification of segments involved in the biogenesis of nuclear and organelle-coded components, and (ii) spectral and functional identification of individual components. Among the four proteins constituting the spinach bf complex only one, the Rieske FeS protein (19 kd) is encoded by nuclear DNA. The other three, cytochromes f (34 kd) and b6 (23 kd), and a 17 kd protein (subunit 4) of unknown function originate in the plastome (Alt et al. 1983). By contrast, in mitochondria only a single component of the bc_1 complex, cytochrome b (ca. 42 kd), is derived from an organelle gene (e.g., Nobrega and Tzagoloff 1980).

The genes for the cytochromes f, b6 and subunit 4 have been localized in the large single-copy segment on the plastid chromosomes of spinach, *Oenothera* and tobacco (Alt et al. 1983; Herrmann et al. 1983a). Although they are encoded by the same strand, they are not all part of one operon. Only the genes for the latter two proteins map close to one another and appear to be under common transcriptional control. We have sequenced the entire DNA region containing these genes including flanking and intercistronic regions, allowing us to compare these sequences to those of their mitochondrial equivalents. The organization of the three genes displays features that have not previously been demonstrated to occur in plastid chromosomes. In this paper we report

Offprint requests to: R. G. Herrmann



Fig. 1. Restriction map of the spinach plastidDNA segment encoding the genes for cytochrome b6, subunit 4, and P₆₈₀ chlorophyll a apoprotein, and summary of the extent and direction of individual sequence determinations (*small arrows*). The map was drawn to scale with use of data described in Alt et al. (1983); Westhoff et al. (1983); Morris and Herrmann (1984) and in this paper. Restriction sites for the various enzymes (left) are designated as previously described (Herrmann et al. 1983). The fragments are designated in decreasing size (SalI, XhoI, BamHI) or given in bp (EcoRI). The small, closely spaced BamHI-fragments have been renumbered (cf. Alt et al. 1983; Westhoff et al. 1983). The direction of transcription is marked by an *arrow*

results for cytochrome b6 and subunit 4; the evaluation of the structural consequence of the protein sequence has been the subject of a separate paper (Widger et al. 1984). Information on the gene for apocytochrome fwhich is located some 12 kbp away and transcribed into separate RNA species will be presented in the accompanying communication. Preliminary accounts of this work have been published (Heinemeyer 1983; Herrmann et al. 1983b, c).

Experimental procedures

The recombinant plasmid pWHsp208 carrying the 5.2 kbp primary fragment SalI-8 of spinach plastid DNA was used in these experiments. The fragment was cloned in the vector pBR322 and subcloned in pBR325 as previously described (Alt et al. 1983; Herrmann et al. 1980, 1983c). Restriction endonucleases were purchased from Boehringer (Mannheim; BamHI, BgII, EcoRI, HpaII, PvuI, SalI, TaqI) or New England Biolabs (Bad Schwalbach; AvaII, DdeI, HaeIII, HincII, HinfI, KpnI, RsaI, XbaI, XhoI). DNA digestions with restriction endonucleases and gel electrophoresis were performed as previously described (Herrmann and Whitfeld 1982). The 3' termini of the restriction fragments were labelled with $(\alpha^{32}P)$ -dXTP and Klenow fragment (Boehringer) filling-in the ends, after fragment purification by chromatography on Elutip columns (Schleicher and Schüll, Dassel). DNA was sequenced by the method developed by Maxam and Gilbert (1980) as outlined in Morris and Herrmann (1984).

Results and discussion

The location and transcription polarity of the genes for cytochrome b6 (23 kd), subunit 4 (17 kd) and the P₆₈₀

chlorophyll a apoprotein of the photosystem II reaction center ("51 kd" protein) on a 7 kbp segment of the spinach chromosome, together with the relevant Sall, BamHI, EcoRI and XhoI sites is depicted in Fig. 1. The genes have been mapped by cell-free translation of appropriate recombinant DNAs carrying specific fragments of spinach plastid DNA or of chloroplast RNA fractions hybrid-selected with these fragments (Alt et al. 1983; Westhoff et al. 1983). Derivative clones, spanning the entire 7 kbp DNA segment, selected mRNA that directed the synthesis of all three polypeptides in vitro suggesting that these loci are active and form an operon-like transcriptional unit (Alt et al. 1983; Morris and Herrmann 1984). The genes for cytochrome b6 and subunit 4 are completely contained within the 5.2 kbp fragment SalI-8 of spinach plastid DNA. Plasmid pWHsp208 containing this fragment directs the synthesis of both proteins in an E. coli S30 in vitro transcription-translation system (Alt et al. 1983). A more detailed restriction map for SalI-8 was constructed by the method of Smith and Birnstiel (1976), as well as by double and triple digestions, and was used to guide DNA sequence analysis as indicated in Fig. 1.

The sequence of the DNA segment encoding cytochrome b6 and subunit 4 was determined using the procedure of Maxam and Gilbert (1980). Translation of this sequence disclosed the presence of only two reading frames large enough to code for both proteins. These reading frames of 211 and 139 codons are located on the same strand and are separated from each other by an untranslatable sequence of 1018 bp. No open reading

HUTAHATATA	TACCOTT	TTAAT	TAGT	ATT	GAGT	GTGT	CTGC	TTGA	GCTG	TACG	AGAT	GAAA	GTCT	CATA	TACA	GTTC	TTA	GGGGG	GTTT	AATA	GTT1	ГАССТ	ATCI	TCAAT	AAA	tc1
1 <u>ATG A</u> TT GGT MET Ile Gly	TCG AAG Ser Lys	AAC Asn	GTC Val	TCG Ser	AGA Arg	30 TTC / Phe /	AGG Arg	CGA Arg	TTG Leu	CGG Arg	ATG Met	ATA Ile	ATA Ile	ACT Thr	AGT Ser	6C AAA Lys	TAT Tyr	GTT Va∙l	CCT Pro	CCT Pro	CAT His	GTC Val	AAC Asn	ATA Ile	TTT Phe	90 TAT Tyr
TGT CTA GGC Cys Leu Gly	GGA ATT Gly Ile	ACG Thr	CTT Leu	ACT Thr	TGT Cys	120 TTT Phe I	0 TTA Leu	GTA Val	CAA Gln	GTA Val	GCT Ala	ACA Thr	GGG Gly	TTT Phe	GCG Ala	15 ATG Met	O ACT Thr	TTT Phe	TAC Tyr	TAT Tyr	CGT Arg	CCA Pro	ACC Thr	GTT Val	ACT Thr	180 GAT Asp
GCT TTT GCT Ala Phe Ala	TCT GTT Ser Val	CAA Gln	TAT Tyr	ATA Ile	ATG Met	21 ACT Thr	0 GAA Glu	GTC Val	AAC Asn	TTT Phe	GGT Gly	TGG Trp	TTA Leu	ATA Ile	CGA Arg	24 TCC Ser	O GTT Val	CAT His	CGG Arg	TGG Trp	TCG Ser	GCA Ala	AGT Ser	ATG Met	ATG Met	270 GTT Val
CTA ATG ATG Leu Met Met	ATC CTC Ile Lec	CAT His	GTA Val	TTT Phe	CGT Arg	30 GTG Val	0 TAT Tyr	CTT Leu	ACC Thr	GGT Gly	GGG Gly	TTT Phe	AAA Lys	AAA Lys	CCT Pro	33 CGA Arg	GAA G1u	TTG Leu	ACT Thr	TGG Trp	GTT Val	ACA Thr	GGC Gly	GTG Val	GTT Val	360 CTG Leu
GGT GTA TTG Gly Val Leu	ACC GCC Thr Ala	TCT Ser	TTT Phe	GGA Gly	GTA Val	39 ACT Thr	0 GGT Gly	TAT Tyr	TCC Ser	TTA Leu	CCT Pro	TGG Trp	GAC Asp	CAA Gln	ATT Ile	42 GGC Gly	20 TAT Tyr	TGG Trp	GCA Ala	GTA Val	AAA Lys	ATT Ile	GTA Val	ACA Thr	GGC Gly	450 GTA Val
CCG GAT GCT Pro Asp Ala	ATT CC Ile Pro	GTA Val	ATA Ile	GGA Gly	TCG Ser	48 CCT Pro	0 TTG Leu	GTA Val	GAG Glu	TTA Leu	TTA Leu	CGT Arg	GGA Gly	AGT Ser	GCT Ala	51 AGT Ser	IO GTG Val	GGA Gly	CAA G1n	TCT Ser	ACT Thr	TTG Leu	ACT Thr	CGT Arg	TTT Phe	540 TAT Tyr
AGT TTA CAC Ser Leu <u>His</u>	ACT TT Thr Phe	GTA Val	TTG Leu	CCT Pro	CTT Leu	57 CTT Leu	0 ACT Thr	GCC Ala	GTA Val	TTT Phe	ATG Met	TTA Leu	ATG Met	CAC His	TTT Phe	60 CTA Leu	DO ATG Met	ATA Ile	CGT Arg	AAA Lys	CAA Gln	GGT Gly	ATT Ile	TCT Ser	GGT Gly	6 <u>30</u> CCC Pro
TTA TAG AATT		AGAA	AAGG	TATA			TTT	TAAT	CAAT	CATI	TATO	CACT	, FGAG(GGAA	GAACA	ATA	GTAT	TTCAT	TGC	TĂČAŢ	GTA	TGGA	TAT	TGAA	AATA	ATAA
750 TCCATGTATTTC	GACATT	гтсст	TCAA	стсси	780 ATAA1	TTTG	TAT	TAG	TAT	ITAAG	CATA	8 ACATO	IO CACT <i>I</i>	AGTTO	GAAG	STAA	гтстл	ACGA	ATG/	840 \AAA1	GGA	TAT	GGGA	GTGT	GTGA	CTTG
870	AGGCCGTO	CAGG	ΤΑΤΑΊ	TATCO	900 TTTT) ICTGC	CACA	TTGA	AATT	CATA	СТСА		930 GTTC		CATI	төтс	CAA	CGCC	GTAG	960 AAGT	AAG1	гсста	TAGA	GAG	GATA	GTT
990				+		-	~						1000							40	00					
COTTOCTTOC	CACAATO			ATC		120	тсла	стел	TCC	TCA	ceee	ото	105U) NONTO	CACT	COM	тет	TCAT	ттта		180 1711	ΓΛΛΤΟ	TAC	\TTT	гатт	TATT
GGTTCGTTTGA/	GAGAATO		CTATO	GATC#		020 CTGAA 1140	TCAT	GTGA	TGCA	TGAA	CGGG	стс	1050 GTAA 117	, АGATC /О	CAGT	CGAA	TGTO	GTGAT	TTT		80 ATA1 200	TAATO	TAG/	3 3	IGTT	TATT
GGTTCGTTTGAA 1110 ACTAATTTGAAT 1230			GCATI		ACCC 	120 CTGAA 1140 CTTGC 1260	TCAT ATTO	GTGA	TGCA	TGAA	ACGGG ATAC		117 GTAA GGAC	GATO	CAGT	CGA#	TGTO VICT/	GTGAT	TTTO ACAA		ATA1 200 AAC1 1320	TAATO		3 3 1111	GTAA	
GGTTCGTTTGAA 1110 ACTAATTTGAAT 1230 TAAACCCTTTAT			GCATI GCATI			120 TGAA 140 TTGC 1260 TTTA	TCAT ATTO TAGA	GTGA ACCT	TGCA TGATC	TGAA TATO	ACGGG ATAC CCAA		117 GGAC 117 GGAC 12	GATO	CAGT A <u>CTA</u> TAAA	CGAA GGGA CACC	ATGTO ATCT/	AAAGA CATAA	ACAA GGTA		ATA1 200 AAC1 1320 ACG4			ATTTA	GTT GTAAI GTTG	TATT CAAG ATCA
GGTTCGTTTGAA 1110 ACTAATTTGAAT 1230 TAAACCCTTTAT 1350 TGATCAACTTTA		GAAAT GAAAT TATAT	CTATO SCATI ATAA1 GGGTO			120 140 1260 1260 1111 138 1114	TCAT ATTO TAGA 0 2 TTGT	GTGA ACCT AAGT AAGA	IGATO IGATO ICAAA	TGAA TATG ATC1 GAAT1	ACGGG GATAC TCCAA			AGATO YO STGAA 290 GGAGA 410 ATAA1	CAGT A <u>ACTA</u> TAAA	CGAA AGGGA ACACC	ATGTO ATCT/ CAATO	AAAGA CATAA AAAGI	ACAA AGGTA		ATA1 200 AAC1 1320 ACG4 144 ATA4	FAATO FAGGO ACCCO 10 AATAO	CTAGA CGATA 4 GAAAA GGTTO	ATTTA ATTAC AGCAC	TGTT GTAAI GTTG/ GTTG/	TATT CAAG ATCA GAAC
GGTTCGTTTGAA 1110 ACTAATTTGAAT 1230 TAAACCCTTTAT 1350 TGATCAACTTTA 1470 CTTTATAGATG	AGAGAATO AGTATTO ATTAAGCC D GATAAGA		GCATI GCATI ATAAI GGGTC	CATC TEATI TTTC 2 GTTGA Met	ACCC TCCT TCCT TCCT TCCT TCCT TCCT TCCT	120 TGAA 1140 TTGC 1260 TTTA 138 TTTAT CCA <i>Pro</i>	TCAT ATTO TAGA 0 2 TTGT TTGT Phe	GTGA ACCT AAGT AAGA GGT <i>GTy</i>	TGATO TGATO TCAAA AGCAO TCT Ser	ATGAA TATCI ATCI GAATI 1! TTT Phe	ACGGG ATAC CCAA CCTTI 510 GAT Asp	AGTATO AGTATO TTCCO TCT Ser		AGATO 290 GAGA 410 ATAA1 TTG <i>Leu</i>	CAGT A <u>ACTA</u> TTAAA TTACA AGC <i>Ser</i>	CGAA AGGGA ACACC AACTT CGG <i>Arg</i>	ATGTO A <u>TCT</u> / CAATO ITGG/ Met	AAAGA CATAA AAAGI AAAGI <i>Met</i>	ACAA AGGTA AGGTA AGGA AAA <i>Lys</i>	ATTTA ATTTA ATTTA ATTTA AAT Asn	200 AAC1 1320 ACG4 144 AATA4 540 TAT <i>Tyr</i>	TAGGO TAGGO ACCCO HO AATAO CAT His	GATAGA GATA GAAAA GAAAA GATC Val	ATTTA ATTAC AGCAC CTTAC CGG <i>Arg</i>	TGTT GTAAI GTTG/ GTTG/ CAGG TTC <i>Phe</i>	TATT CAAG ATCA GAAC TTT Phe
GGTTCGTTTGAA ACTAATTTGAAT 1230 TGAACCCTTTAT 1350 TGATCAACCTTA CTTTATAGATG CGG GGG ATG Arg Gly Met	AGAGAATO AGTATTO AGTATTO ATAAGCC GATAAGA 1570 AAT TG AST TG AST TG		CTATE SCATI ATAAI GGGTC ATAT TAT	ATCATI TTTC TTTC TTTC TATG Met AAG Lys	ACCO TCCT TCCT TCCT TCCT TCCT TCCT TCCT	TGAA 140 1260 1260 17TTA 138 1TTAT CCA <i>Pro</i> TCA <i>Ser</i>	TCAT ATTO TAGA 0 2 TTGT TTT Phe CCT Pro	GTGA ACCT AAGT AAGT GGT GIY ATC Ile	TGATO TGATO TCAAA AGCAO TCT Ser CTA Leu	ATGAA TATCI ATCI ATCI ATTI Phe 10 ATA 12e	ACGGG ATAC CCAA CCTTI 510 GAT Asp 500 ACA Thr	AGTA TCT Ser AAAA Lys	TGC Cys AAA Lys	AGATO AGATO ATGAA 290 GAGA 410 ATAA1 TTG Leu CCT Pro	ACTA ACTA ACTA ATAAA TTACA AGC Ser GAC Asp	CGAA AGGGA ACACC AACTT CGG Arg TTG Leu	ATGTO ATCTA CAATO TTGGA ATG Met AAT ABN	AAAGA AAAGA AAAAGI AAAAGI ATG <i>Met</i> Asp	AGGTA AGGTA AGGTA AGGA AAA Lys CCT Pro	ATGAA ATGAA ATGAA ATTTA ATTTA AAT ASN 16 GTA Val	ATA1 200 1320 1320 1320 144 144 144 144 144 144 144 144 144 14	raato raggo accco to aatao cat <i>His</i> Agg <i>Arg</i>	GATA GATA GATA GAAAA GATTO GTC Val GCT Ala	ATTTA AGCAC CTTAC CGG Arg AAA Lys	TGTT GTAAC GTTG GTTG CAGG TTC Phe TTA Leu	TATT CAAG ATCA GAAC TTT Phe GCG Ala
GGTTCGTTTGA/ ACTAATTTGAAI 1230 TGAACCCTTTAT 1350 TGATCAACTTT/ CTTTATAGATG CGG GGG ATG Arg Cly Met Lys Gly MET	AGTAATTC AGTATTC AGTAATTC ATAAGCC D GATAAGA ABAT TG AAT TG ABAT TG ABAT TG AGT AGA GG CA GJ Hij	TATTI AAATI ATAT ACTTI FATAT A ATC JLe F AAT S ASn	CTATC GCATI ATAA1 GGGTC ATAT <i>TAT</i> <i>Tyr</i>	GATCA ICATII ICATII ITTIC 2 STTGA Met AAG Lys TAT Tyr	ACCC TTCCT TTCCT TTCTT AGCAT ASP AAT ASP GGG Gly	GAG GAG GIU	TCAT ATTO TAGA TTGT TTT Phe CCT Pro	GTGA AACCT AAAGT GGT Gly ATC Ile GCG Ala	ATGCA IGATC IGATC ICAAA AGCAA AGCAA AGCAA AGCAA CTA Leu TGG Trp	TGAA CTATC ATC1 CAATC1 11 TTT Phe 11 ATA Ile 11 CCT Pro	ACGGG GATAC TCCAA TCTTT 510 GAT ASP 500 ACA Thr 590 AAT ASn	CTATC CTATC AGTA TTCCC TCT Ser AAA Lys GAT Asp	CGTAA 117 CGGAG 12 TTTTC 12 TTTC 12 TTTC 13 TGC Cys AAA Lys CTT Leu	AGATC AGATC GAGATC GAGAGA 290 GGAGA 410 ATAAT TTG <i>Leu</i> CCT <i>Pro</i> TTA Leu	ACTA AACTA TTAAAA TTACA AGC Ser GAC Asp TAT Tyr	AGGGAA AGGGGA ACACCO CGGG Arg TTG Leu ATT Ile	ATGTC ATCTA CAATC ATGGA Met AAT Asn TTT Phe	GTGAT AAAAGA CATAA AAAAGT ATG <i>Met</i> Asp CCT Pro	AACAA AGGTA AAAA Lys CCT Pro GTA Val	IC GCATG I I ITAGA ATTIA ASD AAT ASD Val GTA Val	ATA1 200 AAC1 1320 AAC1 1320 AC1 1320 AC1 44 AATAA 540 TAT Tyr 530 TTA TA7 TYr 530 TTA Leu 20 ATT 11e	TAATC TAGGC ACCCC ACCCC THIS AGG Arg TTA Leu	CTAGA CGATA GAAAAA GGTTC Val GCT Ala GGT GIY	ATTTI <u>3</u> ATTAC AGCAC CGG Arg AAA Lys ACT Thr	TGTT GTAAA GTTG GTTG CAGG Phe TTC Phe Leu ATT Ile	ATCA GAAC TTT Phe GCG Ala GCG Ala
GGTTCGTTTGA/ ACTAATTGAAT 1230 TAAACCCTTTAT 1350 TGATCAACTTT/ CTTTATAGATG/ CGG GGG ATG Arg GLy Met IAAA GGI ATG Lys GLy MET TGT AAC GTA Cys Asn Val	AGTATTO AGTATTO AGTATTO ATAAGCC D GATAAGA AATTG AATTG AATTG GGG CA GIY His 1750 GGC TI GIY Lei	AAATI AAATI AAATI AATAT AATC AATC AATC A	CTATC GCATT ATAAT GGGGTC TAT TAT TAT Tyr GTT Val	GATCA ICATI ICATI ICATI 2 GTTGA Met AAG Lys TAT Tyr TTA Leu	AACCCC AACCCC TTCCT TTCCT CTATT AGGCAT ASP AAT ASP GGG Gly GAA Glu	120 CTGAA 1140 TTGC 1260 TTTA 138 TTTA Pro TCA Ser GAG Glu CCA Pro	TCAT ATTO TAGA 0 2 TTGT TTGT Phe CCT Pro CCC Pro TCA Ser	GTG/ AACCI AAAGI GAAGI GCI AAAG/ ATC AIa ATG Met	TGATC TGATC TCAAA AGCAG TCT Ser CTA Leu TGG Trp ATT Ile	TGAA TATCI ATCI GAATTI 11 TTT Phe 11 ATA 11 CCT Pro 12 GGC Gly	ACGGG GATAC TCCAA TCCTTT 510 GAT Asp 500 ACA <i>Thr</i> 590 AAT Asn 780 GAG GAG GAG	GCTCC CTATC 4 AGTA TTCCC TCT Ser AAA Lys GAT Asp CCCG Pro	CGTAA 117 CGGAG 12 TTTC 13 TGC Cys AAA Lys CTT Leu GCG Ala	GGATC GGATGAA 290 GGAGA 410 TTG Leu CCT Pro TTA Leu GAT Asp	AACTA AACTA TAAAA TTACA AGC Ser GAC Asp TAT Tyr CCA Pro	AGGGA AGGGGA ACACC AACACC CGGG <i>Arg</i> TTG <i>Leu</i> ATT Ile TTT Phe	ATGTO ATCT/ CAATO ITGG/ Met AATG Asn TTT Phe GCA Ala	GTGAT AAAAGA CATAA AAAAGT AAAAGT AAG CCT Pro ACT Thr	AACAA AGGTA AAAA Lys CCT Pro GTA Val CCT Pro	11 11 11 11 11 11 12 13 14 15 15 16 17 17 17 17 17 17 17 17 17 17	ата ата 200 ааста 13200 ааста 1220 ааста 1220 ааста 1220 ааста 1220 ааста 1220 ааста 1220 ааста 1220 ааста 1220 ааста 1220 ааста 1220 ааста 1220 ааста а ааста а ааста а а а ааста ааста ааста ааста ааста ааста ааста ааста ааста ааста ааста аас	TAATO TAGGO ACCCCO TO ACCCCO TO ACCCCO TO ACCCCO TAATA AGG Arg TTA Leu ATA Ile	GATA GAAAA GAAAA GGTTC GTC Val GCT Ala GGT GIY TTA Leu	ATTTI 3 ATTAC AGCAC CTTAC CGG AAA Lys ACT Thr CCT Pro	IGTT GTAAA GTTGJ CAGG TTC Phe TTA Leu ATT Ile GAA Glu	ATCA GAAC ATCA GAAC GCG Ala TGG Trp
GGTTCGTTTGAA ACTAATTGAA1 1230 TAAACCCTTTAT 1350 TGATCAACTTTA CTTTATAGATG CGG GGG ATG <i>Lys Gly Met</i> TGT AAC GTA Cys Asn Val TAT TIT TIT TTT	AGTAATTC AGTATTC AGTATTC ATAAGCC D GATAAGA AAT TG AGG CA GGG CA GGG CA GGG CA GGG CT GGC TT G1y Lei 1840 CCT GT. Pro Va	A ATC A	CTATE GCATT ATAAT GGGTC ATAT <i>Tyr</i> TAT Tyr TAT Tyr CAA GIn	GATCATI ICATII I	AGCA1 AGCA1 AGCA1 AGCA1 ASP AAT ASP GAA GIU CTT Leu	1140 TTGCA 11260 TTTGC 11260 TTTA 138 TTTA 138 TTTA 138 TTTA 138 TTTA 7 TCA Ser GAG Glu CCA Pro CGT Arg	TCAT ATTO TAG/ 0 2 TTG TTG TTG TTT Phe CCT Pro CCC Pro TCA Ser ACA Thr	GTGA AACCI AAAGI TAAGA GGT <i>Gly</i> ATC <i>Ile</i> GCG AIa ATG Met GTA Val	ATGCA IGATC TCAAA AGCAA TCT Ser CTA Leu TCG Trp TGG Trp TIle CCA Pro	TGAAA CTATC AATC1 11 CATA 11 Phe 10 ATA 11 CCT Pro 11 GGC Gly 11 AAT ASn	ACGGG AATAC TCCAA TCCCAA TCCCAA TCCCAA 500 ACA Thr 5500 AAT ASn 630 GAG GAG GAG GAG GAG GAG GAG GAG GAG CAG C	GCTCC TATCC AGTA TTCCCC TCT Ser AAA Lys GAT ASp CCG Pro TTA Leu	CGTAA 117 CGGAQ 12 CGGAQ 12 TTTTC 15 TGCA Cys AAA Lys CTT Leu GCG Ala TTA Leu	GGATC CO TGAAA 290 GGAGA 410 ATAA1 TTG <i>Leu</i> CCT <i>Pro</i> TTA Leu GAT Asp GGC Gly	AACTA AACTA TAAA TTACA AGC Ser GAC Asp TAT Tyr CCA Pro GTT Val	AGGGAA AGGGGA ACACCO AACTT CGGG Arg TTG Leu ATT Ile TTT Phe CTT Leu	ATGTO ATCT/ CAATO TTGG/ ATGG Met AAT Asn TTT Phe GCA Ala TTA Leu	GTGAT AAAGGA CATAA AAAAGI AAAAGI ATG Met CCT Pro ACT Thr ATG Met	AGGTA AGGTA AAAA Lys CCT Pro GTA Val CCT Pro GCT Ala	ATTTA ATTTA ATTTA ATTTA AAT 18 AAT 18 GTA Val 11 GTA Val 11 STCA Ser	атат 2000 (ААСТ 1320 (АССА 144 1320 (АССА 144 АТТ 1320 ТТА 144 ТАТ 720 АТТ 11е 310 ССА ССА ССА ССА ССА ССА ССА ССА ССА СС	TAATO TAGGO ACCCCC ACCCCC AATAO CATT His AGG Arg TTA Leu ATA Ile CCT Pro	GATA GATA GAAAA GGTT(GGT Ala GGT Ala GGT Ala	ATTTI AGCAC CTTAC CGG Arg AAA Lys ACT Thr CCT Pro GGA GIY	IGTT GTAAI GTAAI GTTG CAGG TTC Phe TTA Leu ATT Ile GAA Glu TTA Leu	ATCA GAAC TTT Phe GCG Ala GCG Ala TGG Trp TTA Leu
GGTTCGTTTGAA ACTAATTTGAAI 1230 TGAACCCTTTAT 1350 TGATCAACTTTA CTTTATAGATG CGG GGG ATG Arg GLy Met CGG GGG ATG Arg GLy Met TGT AAC GTA Cys Asn Val TAT TIT TIT Tyr Phe Phe ACA GTT CCT Thr Val Pro	AGTATTC AGTATTC AGTATTC ATAAGCC D GATAAGA ATTG AST - 20 1660 GGC CA GIY Hi: 1750 GGC TT GIY Lei 1840 CCT GT, Pro Va 1930 TTT TT. Phe Lei	AATCT AATAT AATAT AATCT AATCT AATCT AATCT AATCT AATCT AATCT AATCT AATCT AATTT Phe AGAA GIU	CTATE GCATI ATAA1 GGGTO ATAT TAT TAT Tyr GTT Val CAA GIn AAT	AAG AAG AAG Lys TAT Tyr TTA AIleu ATA Ile	AGCA1 AGCA1 AGCA1 AGCA1 ASP AAT ASP GGG Gly GAA Glu CTT Leu AAT ASn	120 CTGAA 1140 TTTGC 1260 TTTA 138 TTTA 138 TTAT CCA Pro CCA Ser GAG Glu CCA Pro CGT Arg AAA Lys	TCAT ATTC ATTC O 2 TTGT TTT Phe CCT Pro CCC CCC Pro CCC CCC Pro ACA Thr TTC Phe	GTGA ACCI AAGI GGT <i>Gly</i> ATC <i>Gly</i> ATC <i>Gly</i> ATC <i>Gly</i> CA A GCG Ala CA A GIN	ATGCA IGATC IGATC IGAGCAG AGCAG TCT Ser CTA Leu TGG Trp ATT Ile CCA Pro AAT Asn	TGAA TTATC TATCI GAATTI TTT Phe 10 ATA 12 CCT Pro 12 GGC Gly 13 AAT ASn 19 CCA Pro	ACGGG GATAC TCCAA TCCTTT 510 GAT Asp 500 ACA Thr 590 AAT ASn 780 GAG Glu 370 AAG Glu 2370 AAG Clys 5960 TTT Phe	CCG CTATC CTATC AGTA TTCCC TCT Ser AAA Lys GAT ASP CCG Pro TTA Leu CGT Arg	CGTAA 1177 CGGAAG 1177 CGGAAG 12 CGGAG 12 CGGA 12 CGGA 12 CGTA 12 CGTA 12 CGTA 12 CGTA 12 CGTA 12 CGTA 1177 CGGAAG 1177 CG CS CS CS CS CS CS CS CS CS CS CS CS CS	GATC GAGATC GGAGA 290 GGAGA 1410 TTG Leu CCT Pro TTA Leu GAT Asp GGC Gly CCA Pro	AACTA AACTA TAAA TAAA AGC Ser GAC Asp TAT Tyr CCA Pro GTT Val GTA Val	AGGGP AGCACCO AACTTT CGGG Arg TTG Leu ATTT Phe CTTT Leu GCT Ala	ATGTO ATCT/ CAATO TTGG/ ATG Met AAT AAT AAT AAT AL ACA Thr	GTGAI AAA <u>AGF</u> CATAAA AAAAGI AAAAGI <i>ATG Met</i> ACT Thr ATG Met Thr	AACAA AGGTA AGGGTA AAA Lys CCT Pro GTA Val CCT Pro GCT Ala GCT Val	ATTTA ATTTA AAT AAT AAT AAT AAT AAT AAT	ATA1 2000 IAACI 1320 IAACI 1320 ACGA 144 ATAA 540 TAT TYP 530 TTAA Leu 720 ATT GAA GAA GAA GAA 900 GTG 290 TTG	TAATO TAGGO ACCCCC 10 AATAO CAT His AGG Arg TTA Leu ATA Ile CCT Pro GTT Val	GGTT GGTT GGTT GGTT GGT GGT GGT GGT GGT	ATTTI AGCAC CGG Arg AAAA Lys ACT Thr CCT GGA GIY ACC Thr	IGTT GTTAAA GTTGJ GTTGJ CAGGG TTC Phe TTA Leu ATT Ile GAA Glu TTA Leu GTA Val	ATTA AAAG ATCA GAAC TTT Phe GCG Ala GCG Ala TGG Trp TTA Leu GTG Val
GGTTCGTTTGA/ ACTAATTGAA1 1230 TAAACCCTTTAT 1350 TGATCAACTTT/ CTTTATAGATG CGG GGG ATG Arg Gly Met CGG GGG ATG Arg Gly Met TAT AC GTA CGTA Cys Asn Val TAT TTT TTT Tyr Phe Phe ACA GTT CCT Thr Val Pro GCC CTT TGG Ala Leu Trp	AGTATTC AGTATTC AGTATTC ATAAGCC D GATAAGA 1570 AAT TG AAT TG AAT TG AAT TG AAT TG AAT TG GG CA G1y Hi: 1660 GG CA G1y Hi: 1840 CCT GT Pro Va 1930 TTT TT, Phe Lei 2020 TTA GG Leu G1;	AAAATI AATAT. AATAT. AATAT. AATAT. AATC AATC A	GCATT GCATT ATAAT GGGTU ATAAT TAT TAT TAT TAT Tyr GTT Val CAA Gln AAT ASn GGA	ATCA TITIC CATTI TITIC C CATTI C C TTGA Met AAG Lys TAT Tyr TTA Leu ATA Ile GTT Val GCA Ala	AGCAT AGCAT AGCAT AGCAT AGCAT AGCAT AGCAT AGCAT AGCAT AGCAT AGCAT ACA Thr	IZCO TIGAA TIGAA TIGAA 1260 TITA 138 TITA 138 TITA TCA Ser GAG Glu CCA Pro CCA Arg AAAA Lys TTA Leu	TCAT ATTO TAGE O 2 TTGT TTT Phe CCT Pro CCC Pro TCA Ser ACA Thr TTC Phe CCT Phe	GTGA AAGT GAAGT GTA GGT GTA GCG Ala ATC GCG Ala GTA GTA CAA GIN Ile	ATGCA GATC TCAAAA AGCAAA TCT Ser CTA Leu TGG Trp ATT Ile CCA ATT ASn GAT ASp	TGAA TATCI CAATTI CAATTI CAATTI Phe 10 ATA 11 Phe 10 CCT CAAT Asn 10 CCA CT CCA AAT CCA CCT CCA CCT CCA CCT CCA CCT CCA CCT CCA CCT CCA CCT CCA CCT CCA CCT CCA CCT CCA CCA	ACGGG ATAC CCCAA TCTTT 510 GAT 500 ACA 590 ACA 780 GAG GAG GAG GAG GAG GAG GAG GAG GAG GA	CTATC AGTA AGTA TTCCCC TCT Ser AAA Lys GAT Asp CCG Pro TTA Leu CGT Arg CCTA CCTA CCTA CCCC CCCCCCCCCCCCCC	CGTAA CGTAA 117 CGGAG 12 CTTTTC Cys AAA Lys CTT Leu GCG Ala TTA Leu CTT Leu CGT Ala TTA Leu CGT ATA	AGATC AGATC TGAA 290 GAGAGA 1410 TTG Leu CCT Pro TTA Leu GAT ASP GGC G1y CCA Pro TTA Leu	AACTA AACTA AACTA AACTA AACTA AAC Ser TACA Ser TACA ASP TAT Tyr CCA Pro GTT Val GTA Val GCA Gly	AGGGAA AGGGGA AACACC AACTI CGG Arg TTG Leu ATT TTT Phe CTT Leu GCT Ala CTT Leu	ATGTO ATGTO CAATO ITGG/ ATG Met AAT Asn TTT Phe GCA Ala TTA Leu ACA Thr TTT Phe	GTGAI AAAAGA CATAAA AAAAGI AAAAGI AAG Met ACT Thr ATG Met ACT Thr TAA -C-	AGGTA AGGTA AGGTA AGGA AAA Lys CCT Pro GTA Val CCT Pro GCT Ala GTC Val ATTO	ATTTA ATTTA ATTTA AAT AAT AAT AAT AAT A	180 AATA1 200 AAAC1 1320 AAAC1 1320 AAAC1 1320 1320 144 AAAA 540 TAT TYP 530 TTA 200 ATT 310 GAA GAU 900 GTA 900 GTA 900 GTA 900 GTA 900 GTA 900 TTG 200 TAAT	TAATO TAGGO ACCCCO TAATAO CAT His AGG Arg TTA Leu ATA Ile CCT Pro GTT Val	GATA GAAAAA GGTTC GGTC Val GCT Ala GGT Gly TTA Leu GCG Ala GGT Gly AAAAT,	ATTTI AGCAC CTTAC CGG Arg AAA ACT Thr CCT Fro GGA Gly ACC Thr AACA	IGTT GTTAA GTTG CAGG TTC Phe TTA Leu ATT Ile GAA Glu TTA Leu GTA Val CGGC	TATT CAAG ATCA GAAC TTT Phe GCG Ala TGG Trp TTA Leu GTG Val GTGT

Fig. 2. Nucleotide sequence of the genes for cytochrome b6 and subunit 4 as well as of their flanking and intercistronic regions. The sequence of the non-coding (mRNA-like) strand and the inferred amino acid sequences, commencing at the ATG codon at which translation is probably initiated, are shown. The positions of the sequence coding for proteins are presented in triplets, with the corresponding amino acid below. The discussed amino terminal extension of subunit 4 including an in-frame stop codon (nucleotide 1571) is marked by italic letter code. Possible ribosome binding and transcriptional termination sites are boxed and underlined, respectively. Conserved histidine residues in spans II and V (cf. Fig. 6) are boxed. Corresponding segments capable of forming secondary structure are marked by corresponding arrows; alternative possibilities are numbered 1-4

frame of comparable size was found in the segment of approximately 1400 nucleotides spanning the SalI-7/SalI-8 junction into the gene for the P_{680} chlorophyll *a* apoprotein (Morris and Herrmann 1984). These findings precisely located the gene for cytochrome *b6* (cf. Alt et

al. 1983), established that there is a single gene per chromosome for this protein, and confirmed the position of the gene for subunit 4. In accord with hybrid selection data, all three genes are located on the same DNA strand. The interval of 2424 nucleotides starting

	T	T C		A		G] [Т		C		A		G			
т	Phe Phe Leu Leu	13 1 8 6	Ser Ser Ser Ser	4 2 0 4	Tyr Tyr Stop Stop	8 1 0 1	Cys Cys Stop Trp	2 0 0 5	T C A G		T	Phe Phe Leu Leu	9 1 14 2	Ser Ser Ser Ser	0 1 2 0	Tyr Tyr Stop Stop	4 0 1 0	Cys Cys Stop Trp	1 0 0 3	T C A G
с	Leu Leu Leu Leu	4 0 3 2	Pro Pro Pro Pro	7 1 1 1	His His Gln Gln	3 2 5 0	Arg Arg Arg Arg	5 0 3 2	T C A G		с	Leu Leu Leu Leu	5 0 1 0	Pro Pro Pro Pro	8 1 5 1	His His Gln Gln	1 0 2 0	Arg Arg Arg Arg	3 0 0 0	T C A G
A	Ile Ile Ile Met	6 1 7 11	Thr Thr Thr Thr	11 3 3 1	Asn Asn Lys Lys	0 3 5 1	Ser Ser Arg Arg	5 0 1 1	T C A G		A	Ile Ile Ile Met	6 0 2 3	Thr Thr Thr Thr	4 1 4 0	Asn Asn Lys Lys	6 1 2 1	Ser Ser Arg Arg	0 0 0	T C A G
G	Val Val Val Val	7 3 12 3	Ala Ala Ala Ala	5 1 2 2	Asp Asp Glu Glu	2 1 2 1	Gly Gly Gly Gly	7 4 5 2	T C A G		G	Val Val Val Val	5 1 8 1	Ala Ala Ala Ala	2 1 3 4	Asp Asp Glu Glu	3 0 4 2	Gly Gly Gly Gly	3 4 2 2	T C A G
a									1	י ב ו	b								· · · · · ·	L

Fig. 3. Codon usage for the genes for cytochrome $b\delta(a)$ and subunit 4 (b)

164 kp prior to an DdeI site and ending at an AvaII site is shown by Fig. 2; the predicted coding regions are written as triplets and translated.

The amino acid sequence of the 211 residue protein, inferred from the DNA sequence and beginning at the first in-frame methionine, gives a polypeptide chain of 23.7 kd, which matches the molecular weight of 23,000 of cytochrome b6 determined from its electrophoretic mobility in SDS-polyacrylamide gels (Hurt and Hauska 1981; Alt et al. 1983). The sequence shares a 31-38%homology with residues 50-209 of mitochondrial bcytochromes from five different sources (Widger et al. 1984). The derived protein includes four invariant histidine residues (Fig. 2) that probably act as binding sites for two protohemes (Widger et al. 1984; Saraste 1984; see below). This establishes unequivocally that this DNA region codes for cytochrome b6.

The second open reading frame in the sequence (nucleotides 1655 to 2074) is identified as the gene for subunit 4 on the basis of cell-free translation (Alt et al. 1983) and amino acid sequence. The gene has the same polarity as cytochrome b6 but is in a different reading frame. The position of the N-terminal ATG codon (position 1655) for this gene was initially ambiguous. The determined sequence preceding the indicated initiation point contains a potential protein-coding segment for 54 amino acids with four methionine residues but includes an in-frame TGA stop codon at position 1571 (Fig. 2). In order to eliminate the possibility of sequencing and cloning artifacts this segment was analysed from two independently selected pWHsp208 clones, as well as from authentic spinach plastid DNA, using five individual sequence reactions starting from different restriction sites in different orientations. As the TGA triplet was found in each instance, we conclude that the coding region for subunit 4 begins at the indicated methionine and that the molecular weight of the primary translation

product is 15.2 kd. This assignment is further substantiated by hydropathy analysis (see below). If translation were to commence at the first methionine indicated (position 1490), the predicted molecular weight would be much greater than the 17 kd reported as that for the authentic protein and in vitro translation products of spinach chloroplast RNA and plastid DNA (Hurt and Hauska 1981; Alt et al. 1983).

The codon usage for both proteins obeys the general code and is tabulated separately for the two proteins in Fig. 3. Only 5 of the 61 possible codons are not used. There is a 70 and 80% preference for A and T at the third (and first) position for cytochrome b6 and subunit 4, respectively. Of the 18 triplets not used in subunit 4, 14 include C or G in the third codon position. Biased codon usage has been noted for several plastome-coded genes (e.g., Zurawski et al. 1981; Howe et al. 1982) and is probably a consequence of the relatively low CG contents of plastid DNAs (37 mole-%).

The motifs AAAG (b6) and AAAGG (subunit 4) preceding the respective initiation codons (Fig. 2) may operate as ribosomal binding sites (Shine and Dalgarno 1974) as they correspond to conserved sequences of 16S plastid rRNA in maize (Schwarz and Kössel 1980), tobacco (Tohdoh and Sugiura 1982) and prokaryotes (Shine and Dalgarno 1974). The protein-coding regions terminate at TAG amber (b6) and TAA ochre (subunit 4) codons immediately following leucine and phenylalanine codon numbers 211 and 139, respectively. It is unlikely that the small open reading frame (1490–1570) is expressed because it is not preceded by a typical ribosomal binding site.

The transcriptional organization of this DNA segment is complex. More than half a dozen RNA species have been identified by a Northern blot survey (Fig. 4; Morris and Herrmann 1984). These include a large colinear transcript of approximately 6 kb spanning the three



Fig. 4. Autoradiogram of blot hybridization analysis of total cellular RNA from spinach with segments from the spinach plastid chromosome. Denatured (McMaster and Carmichael 1977) RNA (5 μ g per blot) was separated by electrophoresis on a 1.3% agarose slab gel and transferred to nitrocellulose sheets. The hybridization probes were (1) a 427 bp HpaII-HaeIII fragment (residues 453-880), (2) a 207 bp HaeIII-XbaI fragment (residues 881-1088), and (3) a 305 bp XbaI-EcoRI fragment (residues 1098-1394). The probes were labelled by nick-translation (Rigby et al. 1977). RNA sizes are given in kb (left)



Fig. 5. Possible stem and loop structures at the 3' ends of the spinach cytochrome $b\delta$ (a) and subunit 4 (b) genes. Numbering of nucleotides corresponds to that in Fig. 2. The free energies of formation, ΔG , were calculated according to Tinoco et al. (1973)

protein-coding regions and components of 2.5 kb for cytochrome *b6 plus* subunit 4 (Alt et al. 1983) and of 2.1 kb for the P_{680} chlorophyll *a* apoprotein alone (Morris and Herrmann 1984). Remarkably, RNA species

in the range of 2.4–4 kb are also detected hybridising with protein-coding segments from all three genes (Morris and Herrmann 1984) as well as discrete species of 1.4 and 0.9 kb appearing with a 207 bp HaeIII-XbaI segment from within the 1018 bp nontranslated intercistronic spacer (Fig. 4, track 2). The patterns obtained with RNA from etiolated or illuminated material are almost indistinguishable. On the other hand, Western blot analysis (Towbin et al. 1979) shows that only cytochrome bband subunit 4 are stably made in the dark; no trace of the P_{680} chlorophyll *a* apoprotein is detectable under the chosen growth conditions (Westhoff, Alt and Herrmann, unpublished). Apparently the genes for proteins belonging to different membrane complexes can be under common transcriptional control but are not necessarily co-produced. These results indicate post-transcriptional processing, the existence of various transcription initiation and termination points, or both, and suggest strongly that post-transcriptional, translational, or posttranslational control constitutes an essential element in the biogenesis of the thylakoid membrane (Morris and Herrmann 1984). Bearing in mind that the three genes are more than 5 kbp apart, the existence of the last of the above mentioned RNA species might indicate that RNA processing in plastids includes intercistronic splicing (Morris and Herrmann 1984).

The stem-loop structures following the translational termination codons at positions 637–684 (b6; Fig. 5a) and 2105-2150 (subunit 4; Fig. 5b) and extensive secondary structure within the intercistronic spacer (Fig. 2) may, therefore, function in transcriptional termination, RNA processing or attenuation analogous to those involved in corresponding processes in prokaryotic genes (Rosenberg and Kramer 1977; Rosenberg and Court 1979; Yanofsky 1981; Altmann et al. 1982). At first glance, the spacer sequence elements which can form single and multiple stem-loop structure do not exhibit the typical consensus motifs of group I and II introns described for fungal mitochondrial RNA (Michel and Dujon 1983; Langford et al. 1984). The remarkably low free enthalpy (Tinoco et al. 1973) of the stem-loop stuctures distal to the cytochrome b6 gene ($\Delta G = -4.4$ kcal) or within the untranslated spacer when compared to the structure following the gene for subunits 4 ($\Delta G = -26$ kcal; Fig. 5a, b) may reflect a basic difference in the roles these elements may play in the flux of genetic information. Only the latter may function in transcriptional termination analogous to those involved in termination of prokaryotic genes (Rosenberg and Court 1979). This organization may also explain why typical -10 and -35 consensus regions resembling prokaryotic transcription start sites that are generally located upstream of other plastid genes are not found in the vicinity of the two genes. The transcriptional organization outlined here deserves further study.



Fig. 6. Calculated (Kyte and Doolittle 1982) hydropathy profile of cytochrome b6 (b) subunit 4 (c) from spinach and cytochrome b from maize mitochondria (a; by courtesy of Prof. Leaver). The curves represent the averaged hydrophobicity values of a nonopeptide composed of amino acid residue i - 4 to i + 4 plotted against i, where i represents the amino acid number. They are drawn to place corresponding membrane spanning segments at equivalent positions. Roman numbers indicate the proposed membrane-spanning segments (cf. Saraste 1984; Widger et al. 1984). Designation of amino acids is as in Fig. 2. The 5' terminal segment in front of the subunit 4 gene including the translational opal stop codon (position 1571) is marked by an *arrow* (c)

The amino acid sequences of cytochrome b6 and subunit 4 were analysed by computer for secondary structure (Kyte and Doolittle 1982). As would be expected in a membrane protein, there is a high proportion of hydrophobic amino acid residues. When the free energy of transfer of each amino acid residue from a helix in water to a helix in a non-polar solvent lacking hydrogenbonding capacity is integrated, patterns of interspersed hydrophilic and hydrophobic regions emerge. The hydropathy profile, shown in Fig. 6, exhibits five and three hydrophobic regions of about 20 amino acid residues each for b6 and subunit 4, respectively, that probably traverse the membrane. These segments are generally flanked by single or pairs of basic amino acids that could fix the spans by interaction with the negative surface charges of the lipid bilayer. Since the number of spanning segments is not even, the N- and C-terminal ends of the proteins must be on opposite sides of the membrane.

The comparison of the deduced amino acid sequences for both chloroplast proteins with those published for mitochondrial b cytochromes discloses the striking fact that both cytochrome b6 and subunit 4 exhibit significant homology (ca. 40 and 30%, respectively) with N-

and C-terminal parts of the highly homolgous b-proteins from such diverse organisms as Saccharomyces, Neurospora, human, bovine and mouse (Widger et al. 1984). Homology or cross-correlation is even greater in the hydropathy profiles of cytochrome b6 (70–90%, Fig. 6). Conservation includes the four conserved histidine residues at positions 82 and 96 (span II) and 183 and 197 (span V) that could bind two hemes per cytochrome molecule (Fig. 2; Widger et al. 1984), and an acidic amino acid residue in span IV, aspartic acid, in b6 at positon 152, that might be involved in proton pumping across the membrane (Hauska et al. 1983). The greater conservation of cytochrome b6 may reflect its functional role in the catalytic reaction. Subunit 4 may only exert structural and regulatory functions, and constraints on its sequence may be less demanding. Alignment of the sequences in such a way as to maximize homology indicates that the last span in the mitochondrial b cytochromes that is lacking in subunit 4 (Fig. 6a, c), and a deletion of 5-6 amino acid residues between b6 and subunit 4, could account for most of the molecular weight difference observed between the mitochondrial (42 kd) and the sum of the weights of the plastid proteins (39 kd; Widger et al. 1984). The sequence of 54 amino acid residues preceding the indicated initiative methionine codon in the gene for subunit 4 (see above) is higly hydrophilic and must be untranslated as no structural conservation is observed in this segment. These data support strongly that b cytochromes in photosynthetic and respiratory electron transport chains have arisen monophyletically. The gross divergence has either resulted by fusion of genes for two ancestor proteins in mitochondria, or by insertion of an untranslatable segment into a single ancestor gene in plastid chromosomes.

Acknowledgements. The authors are most grateful to Prof. B. Hallick and Prof. W. Cramer for their help in the preparation of this manuscript, and to Prof. Ch. Leaver for providing the sequence for corn mitochondrial cytochrome b before publication. They acknowledge the expert technical assistance of Ms. Barbara Schiller and Ms. Gabriele Schewe. The work was supported by a grant from the Stiftung Volkswagenwerk and by Forschungsmittel des Landes Nordrhein/Westfalen.

References

- Alt J, Westhoff P, Sears BB, Nelson N, Hurt E, Hauska G, Herrmann RG (1983) EMBO J 2:979–986
- Altmann S, Guervier-Takada C, Frankfort HM, Robertson HD (1982) In: Linn SM, Roberts JR (eds) Nucleases. Cold Spring Harbor, New York, pp 243-274
- Capaldi RA (1982) Biochim Biophys Acta 694:291-306
- Hauska G, Hurt E, Gabellini N, Lockau W (1983) Biochim Biophys Acta 726:97–133
- Heinemeyer W (1983) Diplomarbeit, Univ Düsseldorf

- Herrmann RG, Whitfeld PR (1982) In: Edelman M, Hallick R, Chua N-H (eds) Methods in chloroplast molecular biology. Elsevier, North-Holland Amsterdam, pp 451-468
- Herrmann RG, Seyer P, Schedel R, Gordon K, Bisanz C, Winter P, Hildebrandt JW, Wlaschek M, Alt J, Driesel AJ, Sears BB (1980) In: Bücher Th, Sebald W, Weiss H (eds) Biological chemistry of organelle formation. Springer, Berlin Heidelberg New York, pp 97-112
- Herrmann RG, Westhoff P, Alt J, Bisanz C, Heinemeyer W, Olesch B, Nelson N (1983a) In: Swaminathan MS, Chopra VL, Joshi BC, Sharma RP, Bansal HC (eds) Progr 15th Internatl. Congr Genetics, pp 171–172
- Herrmann RG, Westhoff P, Alt J, Bisanz C, Heinemeyer W, Olesch B, Nelson N (1983b) In: Progr 6th Internatl Congr on Photosynthesis, Brussels, vol 2, p 51
- Herrmann RG, Westhoff P, Alt J, Winter P, Tittgen J, Bisanz C, Sears BB, Nelson N, Hurt E, Hauska G, Viebrock A, Sebald W (1983c) In: Cifferi O, Dure L III (eds) Structure and Function of Plant Genomes. Plenum Publ Corp, pp 143-154
- Howe CJ, Auffret AD, Doherty A, Bowman CM, Dyer TA, Gray JC (1982) Proc Natl Acad Sci USA 79:6903-6907
- Hurt E, Hauska G (1981) Eur J Biochem 117:591-599
- Kyte J, Doolittle RF (1982) J Mol Biol 157:105-132
- Langford CJ, Klinz FJ, Donath C, Gallwitz D (1984) Cell 36: 645-653
- Maxam AM, Gilbert W (1980) In: Großman L, Moldave K (eds) Methods in enzymology, vol 65. Academic Press, New York, pp 499-560
- McMaster GK, Carmichael GG (1977) Proc Natl Acad Sci USA 74:4835-4838

- Michel F, Dujon B (1983) EMBO J 2:33-38
- Morris J, Herrmann RG (1984) Nucleic Acids Res 12:2837-2850
- Nobrega FG, Tzagoloff A (1980) J Biol Chem 255:9228-9837
- Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977) J Mol Biol 113:237-251
- Rosenberg M, Court D (1979) Ann Rev Genet 13:319-353
- Rosenberg M, Kramer RA (1977) Proc Natl Acad Sci USA 74: 984–988
- Saraste M (1984) FEBS Lett 166:367-372
- Schwarz Z, Kössel H (1980) Nature 283:739-742
- Shine J, Dalgarno L (1974) Proc Natl Acad Sci USA 71:1342-1346
- Smith HO, Birnstiel ML (1976) Nucleic Acids Res 3:2387-2398
- Tinoco J, Borer PN, Dengler B, Levine MD, Uhlenbeck OC, Crothers DM, Gralla J (1973) Nature (New Biology) 246:40– 41
- Tohdoh N, Sugiura M (1982) Gene 17:213-218
- Towbin H, Staehelin T, Gordon J (1979) Proc Natl Acad Sci USA 76:4350-4354
- Westhoff P, Alt J, Herrmann RG (1983) EMBO J 2:2229-2237
- Widger WR, Cramer WA, Herrmann RG, Trebst A (1984) Proc Natl Acad Sci USA 81:674-678
- Yanofsky C (1981) Nature 289:751-758
- Zurawski G, Perrot B, Bottomley W, Whitfeld PR (1981) Nucleic Acids Res 9:3251-3270

Communicated by C. P. Hollenberg

Received June 7, 1984