The Primary Structure of Cytochrome *c*-554 from the Green Photosynthetic Bacterium *Chloroflexus aurantiacus*[†,‡]

Stella Dracheva,[§] JoAnn C. Williams,[§] Gonzalez Van Driessche,[∥] Jozef J. Van Beeumen,[∥] and Robert E. Blankenship*[,§]

Department of Chemistry and Biochemistry, Center for the Study of Early Events in Photosynthesis, Arizona State University, Tempe, Arizona 85287-1604, and Laboratory of Microbiology, State University of Ghent, Ghent 9000, Belgium

Received July 8, 1991; Revised Manuscript Received September 11, 1991

ABSTRACT: The complete nucleotide sequence of the cytochrome *c*-554 gene from the green photosynthetic bacterium *Chloroflexus aurantiacus* has been determined. The derived amino acid sequence showed that the cytochrome precursor protein consists of 414 residues and contains 4 -Cys-X-X-Cys-His- heme binding motifs. The only regions of the cytochrome *c*-554 sequence that were found to be significantly similar to the sequences of cytochromes from other organisms were the heme binding sites. The highest similarity was found with the heme binding segments in the four-heme reaction center cytochrome subunit from the purple photosynthetic bacterium *Rhodopseudomonas viridis*. The importance of this similarity for the evolutionary relationship between *Chloroflexus* and the purple bacteria is discussed.

Chloroflexus aurantiacus is a thermophilic green photosynthetic bacterium found in hot springs around the world (Pierson & Castenholz, 1974; Ward et al., 1989). According to a phylogenetic tree based on 16S rRNA sequence comparisons, it appears to have diverged from other photosynthetic organisms at a very early point in time and thus has a unique evolutionary history (Woese, 1987). However, a phylogenetic tree based on sequences of pheophytin–quinone-type reaction centers places *C. aurantiacus* much closer to the purple bacteria than is indicated by 16S rRNA analysis (Beanland, 1990; Blankenship, 1992). The meaning of this discrepancy is unclear, but may indicate that lateral gene transfer of photosynthesis genes has taken place. A variety of spectroscopic and kinetic data shows that *C. aurantiacus* contains a peripheral chlorosome antenna system similar to that of the green sulfur bacteria, but a reaction center and an integral membrane antenna complex similar to that of the purple photosynthetic bacteria (Blankenship & Fuller, 1986; Zuber, 1987; Amesz, 1991).

The *C. aurantiacus* reaction center consists of two subunits, L and M (Pierson et al., 1983; Shiozawa et al., 1987; Blankenship et al., 1988). The complete DNA sequence of the genes encoding these two proteins revealed considerable homology to the purple bacterial reaction center L and M subunits (Ovchinnikov et al., 1988a,b; Shiozawa et al., 1989). No evidence exists for an H subunit in reaction centers from *C. aurantiacus* (Pierson et al., 1983; Shiozawa et al., 1987; Blankenship et al., 1988).

In purple bacteria, two different types of photosynthetic reaction center complexes are known. In one type, the reaction center complexes have L, M, and H subunits. The L and M subunits are integral membrane proteins and form the core of the reaction center (Deisenhofer et al., 1985; Michel et al., 1986; Allen et al., 1987). The H subunit is attached to the cytoplasmic side of the reaction center and is not required for the initial electron-transfer steps (Debus et al., 1985). In the other type, the reaction center complexes possess a cytochrome subunit with four *c*-type hemes in addition to the L, M, and H subunits. The cytochrome is attached to the periplasmic side of the reaction center and is either tightly bound to the reaction center as in *Rhodopseudomonas viridis* (Deisenhofer et al., 1985) or easily removed by detergent as in *Rhodocyclus gelatinosus* (Fukushima et al., 1988). The tetraheme cytochrome donates an electron to the bacteriochlorophyll dimer ($BChl_2$) that is oxidized in the first step of the light-induced charge separation. Both types of reaction center are widely distributed in the α group of purple bacteria. The reaction center–cytochrome complex is present in all photosynthetic species belonging to the β group (e.g., *Rhodocyclus gelatinosus*) and the γ group (e.g., *Thiocapsa pfennigii*, *Chromatium tepidum*) of purple bacteria for which this determination has been made (Fukushima et al., 1988; Seftor & Thornber, 1984; Nozawa et al., 1987). It has been suggested that the ancestral purple bacterium had a reaction center with a four-heme cytochrome subunit that has been lost in several independent lines (Matsuura & Shimada, 1990).

Recently, tetraheme cytochromes were discovered in the green sulfur photosynthetic bacterium *Chlorobium limicola* and in the Gram-positive bacterium *Heliobacterium chlorum* (Feiler et al., 1989; Nitschke et al., 1990). This result is surprising, because these organisms have a very different type of reaction center complex (Amesz, 1991). The ubiquitous presence of the multiheme cytochrome in all the major groups of anoxygenic photosynthetic prokaryotes indicates that it may be a characteristic of the very earliest reaction centers.

C. aurantiacus also possesses a membrane-bound four-heme cytochrome *c* which does not copurify with the reaction centers but serves as the immediate electron donor to the oxidized $BChl_2$ in whole cells and membrane particles (Bruce et al., 1982; Zannoni & Venturoli, 1988; Freeman & Blankenship, 1990). This cytochrome appears to be similar to the cytochrome subunit of the reaction center from *R. viridis*, for

[†] This research was supported in part by Grant 91-37306-6314 from the Cooperative State Research Service of the U.S. Department of Agriculture, by Grant BRSG 2 S04 RR07112 from the Division of Research Resources and Grant GM45902 from the National Institutes of Health, and by the Belgian National Incentive Program on Fundamental Research in Life Sciences initiated by the Science Policy Programming Department (Contract B10-22). This is Publication No. 84 from the Arizona State University Center for the Study of Early Events in Photosynthesis. The Center is funded by U.S. Department of Energy Grant DE-FG-88-ER13969 as a part of the USDA/DOE/NSF Plant Science Centers Program.

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number M77813.

* To whom correspondence should be addressed.

[§] Arizona State University.

[∥] State University of Ghent.

0006-2960/91/0430-11451$02.50/0

which the primary and three-dimensional structure is known and some thermodynamic and kinetic properties have been investigated (Deisenhofer et al., 1985; Weyer et al., 1987a; Shopes et al., 1987; Alegria & Dutton, 1987; Dracheva et al., 1988; Nitschke & Rutherford, 1989; Knaff et al., 1991). All four hemes of the tetraheme cytochrome from *C. aurantiacus* can be distinguished by their midpoint redox potentials and orientations with respect to the plane of the membrane, but have the same α-band absorbance maximum at 554 nm. (Freeman & Blankenship, 1990; Van Vliet et al., 1991). Accordingly, this cytochrome has been designated cytochrome *c*-554. The existence of a cytochrome that is functionally associated with the reaction center strengthens the similarity of the *C. aurantiacus* reaction center to that of purple bacteria. To yield some additional information on the similarities between the mechanisms of photosynthesis in purple bacteria and *C. aurantiacus*, we have determined the primary structure of cytochrome *c*-554.

Materials and Methods

Enzymes, Chemicals, Cell Growth, and Protein Purification. Restriction endonucleases and other enzymes used in molecular cloning were obtained from Stratagene (La Jolla, CA), New England Biolabs (Beverly, MA), or Boehringer Mannheim (Indianapolis, IN).

[α-^{32}P]dATP and [^{35}S]dATPαS were purchased from DuPont–NEN Products (Boston, MA). All other chemicals were reagent grade.

C. aurantiacus strain J10-fl was grown under high-light conditions in modified medium D (Pierson & Castenholz, 1974) at 55 °C. *Escherichia coli* strain DH5α (Hanahan, 1983) was obtained from Gibco BRL.

Cytochrome *c*-544 from *C. aurantiacus* was isolated according to Freeman and Blankenship (1990).

Isolation of Protein Fragments and Amino Acid Sequencing. The covalently bound hemes were removed from the native protein according to the method of Ambler and Wynn (1973) using mercury chloride in acidic urea. After separation of the apoprotein from hemes and salts in 5% formic acid, an aliquot of 40 nmol of apoprotein was dried and digested with *Staphylococcus aureus* protease at pH 8.0 using an enzyme:substrate ratio of 1:40. A second aliquot of 3 nmol was subjected to partial acid hydrolysis in 2% formic acid for 90 min at 107 °C.

The peptides were separated by reversed-phase HPLC on a Vydac 214TP54 column using a gradient of 0.1% trifluoroacetic acid in water (solvent A) and 0.1% trifluoroacetic acid in acetonitrile (solvent B). Peptides were sequenced using a 477A pulsed liquid sequenator with on-line detection of phenylthiohydantoin (PTH) amino acids (Applied Biosystems).

Isolation of DNA. C. aurantiacus chromosomal DNA was isolated using a modification of the method described by Williams et al. (1983). After lysozyme and protease treatment, the cell mass was extracted with phenol. The aqueous phase was collected and extracted once with phenol/chloroform (1:1, by volume) and once with chloroform. After 2 volumes of ethanol were added, the DNA was spooled out, rinsed with 80% ethanol, dried, and resuspended in 10 mM Tris-HCl/1 mM EDTA, pH 8.0. The DNA was treated with RNase, extracted once with phenol, once with phenol/chloroform, and once with chloroform, and precipitated in ethanol.

Plasmids were prepared using the alkaline lysis method (Sambrook et al., 1989).

Preparation of PCR Probe. The amino acid sequences of several fragments of *C. aurantiacus* cytochrome *c*-554 have been chemically determined (Freeman, 1989; Van Beeumen et al., unpublished results). To obtain a DNA probe which could be used to identify the cytochrome *c*-554 gene, chromosomal DNA encoding a cytochrome *c*-554 fragment of 90 amino acids was amplified by the polymerase chain reaction (PCR) method. This fragment was obtained starting from the amino acid sequences of 2 peptides which were located at the N- and the C-terminal ends of the 90-residue fragment. The peptides, called S42 and H50, were obtained from cleavage of the apoprotein with *S. aureus* protease and with 2% formic acid, respectively. Details of their purification will be published elsewhere (Van Beeumen et al., unpublished results). The amino acid sequence of peptide S42 was Leu-Asn-Pro-Ala-Gly-Asp-Asn-Val-Leu-Asn-Pro-Leu-Tyr-Ala-Tyr-Asn-Lys-Leu-Lys-Ala-Gln-Arg-Met-Leu. The sequence of H50 was Gly-Gln-Ile-Tyr-Asn-Val-Pro-Gly-(?)-Tyr-Thr-(?)-His-Gln-Gly-Asn-Asn-Ile-Pro-Leu-Ala-Ser-Ile-Asn-Gln-Ala-Asn-Ile-Pro-Ser-Gly. Because of the presence of His in H50 and the absence of any PTH residue at the previous Edman step and at the one four steps ahead of histidine, it was assumed that these positions were taken by cysteine, in accordance with the heme binding sequence Cys-X-X-Cys-His of nearly all known cytochromes *c* (Ambler, 1982). The underlined stretches of sequence were used to design the PCR primers as they yielded the least degenerate base sequence upon back-translation. To further reduce degeneracy, the codon usages of the genes encoding the *C. aurantiacus* reaction center L and M subunits (Ovchinnikov et al., 1988a,b; Shiozawa et al., 1989) and 5.7-kDa chlorosome protein (Theroux et al., 1990) were taken under consideration. The primers were synthesized as 2 mixtures, containing 64 and 16 species of oligonucleotides.

The PCR was performed in a reaction containing 0.4 μg of chromosomal DNA in the presence of 100 pmol of each primer and 200 μM dNTPs. Twenty-five cycles of denaturation at 94 °C for 1 min, primer annealing at 50 °C for 2 min, and extension by *Taq* polymerase (Perkin-Elmer Cetus) at 72 °C for 3 min were used. The DNA products from the PCR were fractionated in a 3% NuSieve agarose gel. The appropriate size DNA band (231 base pairs) was cut from the gel and used as a probe for the cytochrome *c*-544 gene.

Labeling of the probe was performed using Klenow enzyme and random oligonucleotide primers obtained in kit form from Boehringer Mannheim. Testing for label incorporation on Whatman DE-81 paper was done as described by Sambrook et al. (1989).

Cloning of the Cytochrome c-544 Gene from Genomic DNA. The location of the cytochrome *c*-544 gene in restriction endonuclease digested *C. aurantiacus* chromosomal DNA was determined by Southern blotting. After agarose gel electrophoresis, the denatured DNA was transferred to Gene Screen Plus (DuPont NEN Products) and hybridized according to the manufacturer's protocol. The DNA from the digest regions which hybridized with the PCR probe was eluted from the agarose gels, ligated into the polylinker region of the plasmid pUC19, and transformed into *E. coli* DH5α cells to create a clone library enriched for the target fragments. *E. coli* cells were screened for ampicillin resistance and absence of β-galactosidase activity. Positive transformants were then further screened by colony hybridization with the same PCR probe.

Restriction fragments of the cloned DNA that hybridized with the probe were incorporated into the phage vectors M13 mp18 and mp19 in order to provide templates for sequence analysis.

DNA Sequence Analysis. Sequencing of single-stranded DNA templates was performed by the dideoxy chain-termi-

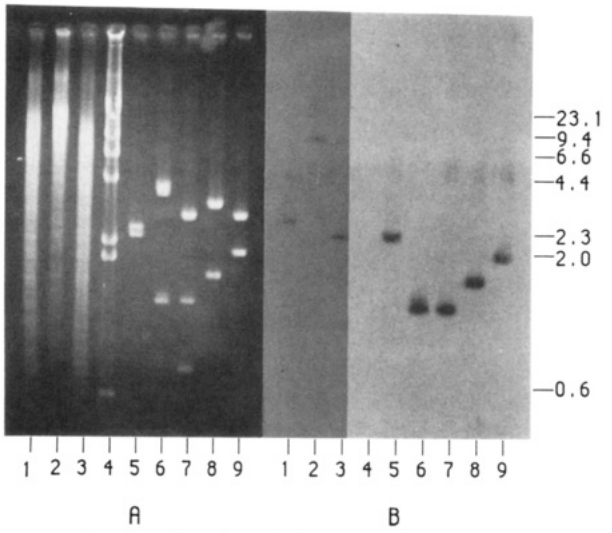

FIGURE 1: Southern blot of restriction enzyme digested *C. aurantiacus* chromosomal DNA and isolated pCEN12 clone plasmid DNA. (A) Ethidium bromide stained agarose gel; (B) autoradiograph of the blotted DNA after hybridization with the PCR probe. Lane 1, *C. aurantiacus* DNA digested with *Nar*I; lane 2, *C. aurantiacus* DNA digested with *Eco*RI; lane 3, *C. aurantiacus* DNA digested with *Eco*RI + *Nar*I; lane 4, λ DNA digested with *Hind*III; lane 5, pCEN12 plasmid DNA digested with *Eco*RI + *Hind*III; lane 6, pCEN12 plasmid DNA digested with *Nru*I; lane 7, pCEN12 plasmid DNA digested with *Nru*I + *Hind*III; lane 8, pCEN12 plasmid DNA digested with *Sal*I + *Hind*III; lane 9, pCEN12 plasmid DNA digested with *Eco*RI + *Bam*HI. The part of the blot containing digested chromosomal DNA (lanes 1–3) was exposed longer than that containing pCEN12 plasmid DNA (lanes 4–9).

nation procedure of Sanger et al. (1977) using modified T7 polymerase (Sequenase). The manufacturer's (USB) instructions were followed. Some regions were sequenced by the use of synthetic oligonucleotide primers. To minimize secondary structure in the sequencing products, gels containing 40% formamide as well as the substitution of a nucleotide analogue for dGTP (dITP) were used.

Computer Analysis of Sequences. DNA and derived protein sequence data were analyzed using the GCG sequence analysis software package version 7.0 (Devereux et al., 1984) and the GENEBEE package of computer programs (Brodsky et al., 1991). The structural analysis of the reaction center cytochrome subunit from *R. viridis* was done using file 1PRC from the Brookhaven databank and a Silicon Graphics Iris workstation with the Quanta software package.

RESULTS

The *C. aurantiacus* chromosomal DNA was resistant to digestion by restriction enzymes as has been previously reported (Robinson et al., 1987; Shiozawa et al., 1989). Only four enzymes, *Eco*RI, *Kpn*I, *Cla*I, and *Nar*I, digested the DNA. The PCR probe (see Materials and Methods) hybridized to different bands obtained by electrophoresis of *C. aurantiacus* chromosomal DNA that had been digested with these four enzymes. The DNA from the 9.0-kb *Eco*RI, 2.8-kb *Nar*I, and 2.4-kb *Eco*RI + *Nar*I digest regions that hybridized with the probe (see Figure 1, lanes 1–3) was ligated into pUC19 which had been restricted with *Eco*RI, *Acc*I, and *Eco*RI + *Acc*I, respectively, and then transformed into *E. coli* cells to create enriched clone libraries. *E. coli* cells which were not sensitive to ampicillin and had no β-galactosidase activity were screened by colony hybridization with the same PCR probe.

Several libraries (900 transformants total) from *Eco*RI and *Nar*I digests were examined, but no clones which hybridized with the probe were found. Only when analyzing the library of *C. aurantiacus* DNA digested with *Eco*RI and *Nar*I simultaneously was the positive clone found. The clone (pCEN12) grew very slowly relative to cells transformed with other plasmids. This indicates that the foreign fragment in pCEN12 plasmid DNA might be expressed and the product is probably toxic to *E. coli* cells.

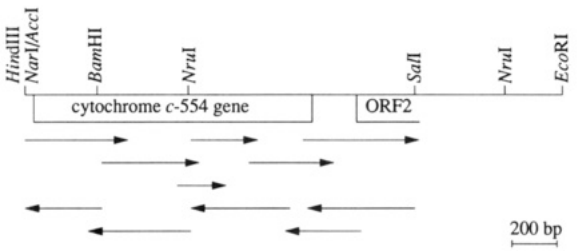

FIGURE 2: Restriction map, open reading frames, and sequencing strategy of the cloned *Nar*I–*Eco*RI fragment, containing the cytochrome *c*-554 gene. Boxed regions indicate the location of the cytochrome *c*-554 gene and ORF2. Arrows show the parts of subclones that were sequenced. The map is drawn so that the transcription of the cytochrome *c*-554 gene and ORF2 would proceed from left to right.

Figure 1 shows a Southern blot of *C. aurantiacus* chromosomal DNA digested with *Nar*I, *Eco*RI, and *Eco*RI + *Nar*I (lanes 1–3) and plasmid DNA from pCEN12 digested with several different restriction enzymes (lanes 5–9). The hybridization pattern indicates that the *Eco*RI–*Nar*I fragment cloned in pCEN12 (lane 5) is the same size as the chromosomal *Eco*RI–*Nar*I fragment that hybridizes with the PCR probe (lane 3). Since the *Nar*I site of the fragment was lost after ligation with the *Acc*I site of pUC19, the adjacent *Hind*III site of the polylinker was used. Strong hybridization of the pCEN12 plasmid DNA insert to the PCR probe implied that the *C. aurantiacus* cytochrome *c*-554 gene (or part of that gene) was situated inside the pCEN12 plasmid.

The restriction map of the pCEN12 plasmid DNA insert and the sequencing strategy are shown in Figure 2. It was found by Southern blotting of restricted plasmid DNA from pCEN12 (see Figure 1, lanes 6–9) that the PCR probe hybridized to the *Sal*I–*Nru*I fragment of the insert. This fragment and later the *Hind*III–*Bam*HI and *Bam*HI–*Nru*I fragments were subcloned into M13 mp18 and M13 mp19 phage vectors. The complete sequence of both strands of the *Hind*III–*Sal*I fragment was obtained. The bulk of the sequencing was performed with a universal M13 primer and the two primers which had been used for preparation of the PCR probe (see Materials and Methods). Two more sequencing primers were synthesized on the basis of the newly obtained nucleotide sequence. The nucleotide sequence of the 1.74-kb *Hind*III–*Sal*I fragment from the pCEN12 plasmid DNA insert and the derived amino acid sequence are shown in Figure 3.

Using a codon frequency table produced by the three published *C. aurantiacus* genes (Ovchinnikov et al., 1988a,b; Shiozawa et al., 1989; Theroux et al., 1990), a codon preference plot was generated by a statistical program written by Gribskov et al. (1984) for identification of potential genes. Two open reading frames (ORFs) were discovered with codon usages similar to that of these *C. aurantiacus* genes (Figure 4). The derived gene product of the ORF1 consisted of 414 amino acid residues and carried 4 heme *c* binding motifs (-Cys-X-X-Cys-His-), as well as the sequence of peptides S42 and H50 (see Materials and Methods). Therefore, the ORF1 encodes the cytochrome *c*-554. A stop codon for the ORF2 inside the *Nar*I(*Hind*III)–*Sal*I fragment was not found, so this putative ORF probably continues after the *Sal*I site. A computer search through the SwissProt (release 17) protein

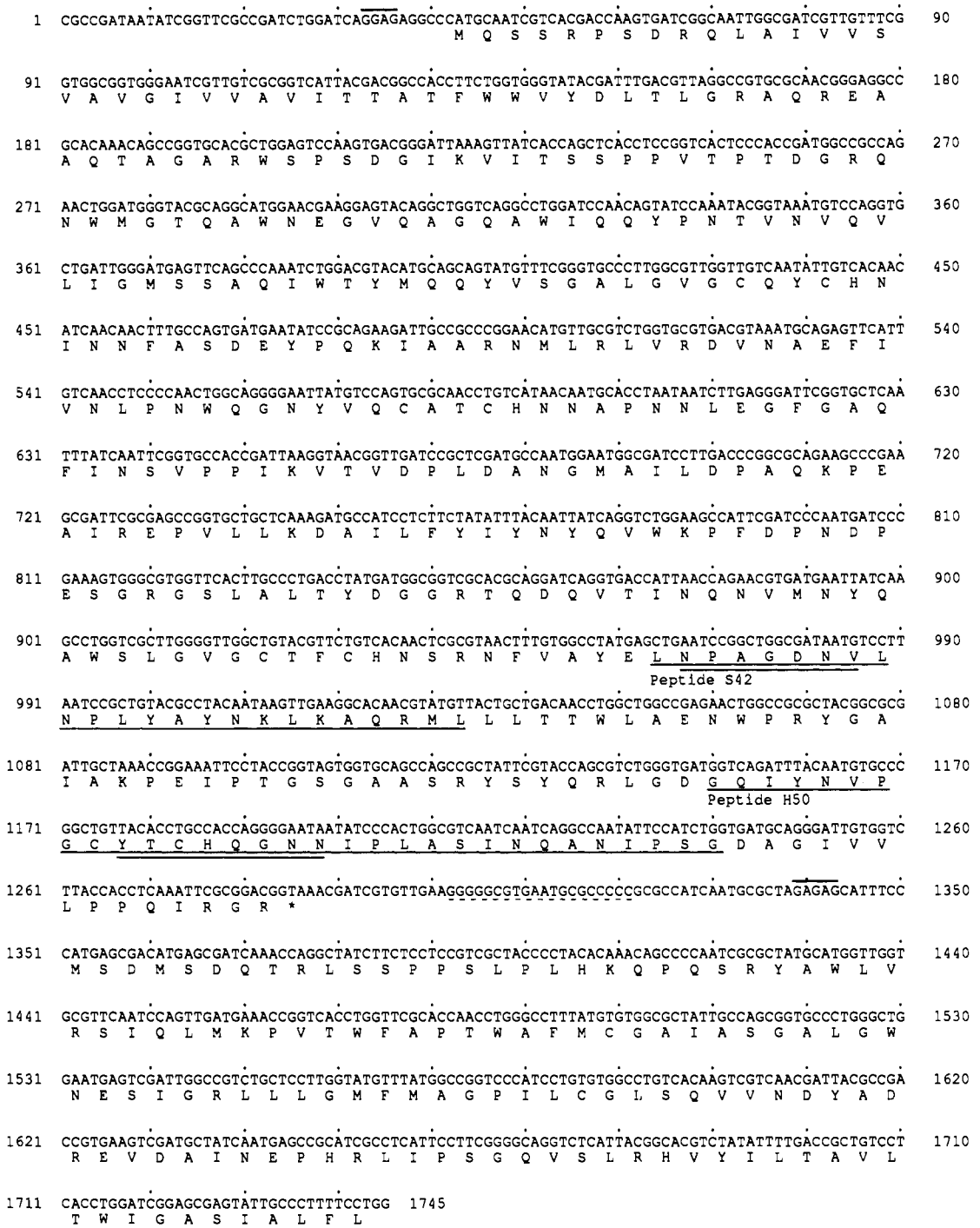

FIGURE 3: Nucleotide sequence and derived amino acid sequence of the cloned *NarI–SalI* fragment, containing the cytochrome *c*-554 gene. The underlined amino acid sequences (peptides S42 and H50) correspond to N- and C-terminal ends of the 90-residue fragment amplified by the PCR method. The double-underlined stretches of S42 and H50 were used to design the PCR primers. Lines above the nucleotide sequence indicate potential Shine–Dalgarno sites. The dashed underlined region contains an inverted repeat which could form a hairpin structure.

databank did not reveal any protein with significant sequence similarities to the ORF2.

Comparison of the derived cytochrome *c*-554 gene product and the chemically determined amino acid sequences of isolated peptides obtained from purified cytochrome *c*-554 protein (Freeman, 1989; Van Beeumen et al., unpublished results) showed that approximately 70% of the amino acid sequence had already been obtained by Edman degradation. The amino acid sequence mostly coincided with the derived gene product, with the exception of the putative N-terminal peptide which was not found in the derived DNA sequence. However, the presence upstream of a Shine–Dalgarno sequence and an in-frame TAA stop codon provides evidence that the ATG start codon at position +43 of the sequence is the beginning of the cytochrome gene. It can be assumed that the preparations of cytochrome *c*-554 protein which were used for chemical sequencing contained some impurities which had been analyzed. The real N-terminal amino acid of the cytochrome must be blocked.

The molecular mass of the cytochrome *c*-554 gene product calculated from the amino acid composition was 45 593. If the four heme groups are included, it is 48 307. Since the mass of the cytochrome *c*-554 signal peptide is unknown, but is probably about 2500 (see Discussion), this value is in rather good agreement with the apparent molecular mass of 43 000 determined by SDS–PAGE (Freeman & Blankenship, 1990).

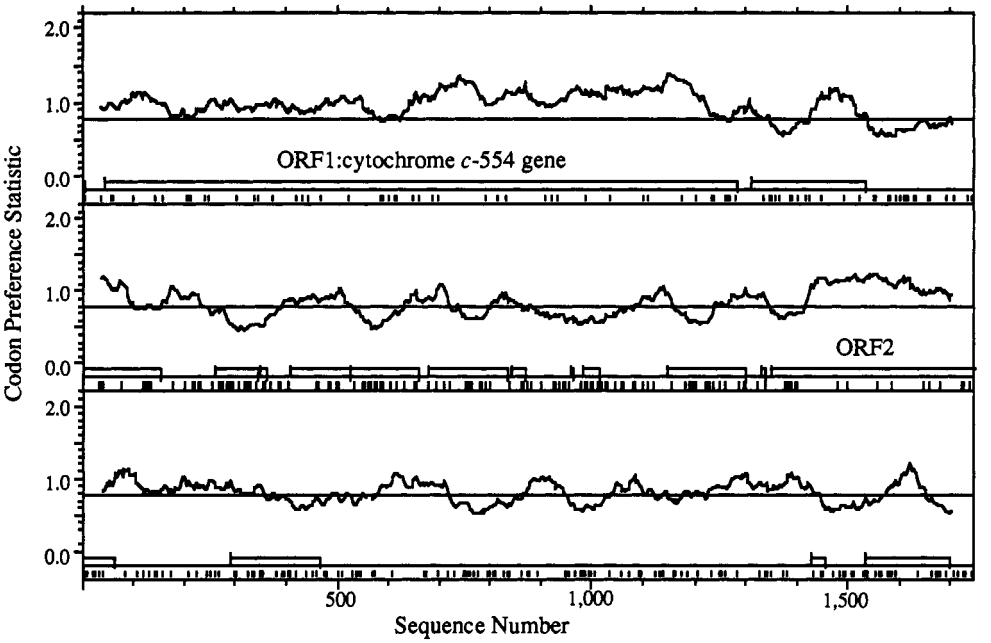

FIGURE 4: Codon preference plots of three different reading frames of the cloned *Nar*I–*Eco*RI fragment. The plots were generated using a program written by Gribskov et al. (1984). The codon frequency table used to produce the plots was obtained from nucleotide sequences of genes coding for reaction center L and M subunits and the 5.7-kDa chlorosome protein from *C. aurantiacus* (Ovchinnikov et al., 1988a,b; Shiozawa et al., 1989; Theroux et al., 1990). The window size and rare codon threshold were set at 25 and 0.1, respectively. Open reading frames (ORFs) are shown as open boxes and rare codons as vertical dashes below the ORFs. Efficiently translated genes appear as peaks above ORFs which contain a small number of rare codons. ORF1 is identified as the cytochrome *c*-554 gene, and ORF2 is an unidentified putative gene. The third reading frame, shown at the bottom of the figure is not predicted to contain any genes. The three reading frames of the complementary strand generated plots similar to that shown in the bottom frame.

The heme iron of *c*-type cytochromes is coordinated by four nitrogen atoms of the porphyrin in the heme plane; the fifth out of plane ligand is a histidine. This histidine in one of the two axial positions occurs in a constant structural motif (-Cys-X-X-Cys-His-). The second axial ligand is almost always histidine or methionine (Moore & Pettigrew, 1990). The derived amino acid sequence of the cytochrome *c*-554 gene contained eight methionine and four histidine residues. Since all four histidines in the cytochrome *c*-554 were found in the heme binding sites, the sixth iron ligands of all four hemes are evidently methionines.

Discussion

The open reading frame of the cytochrome *c*-554 gene was preceded by the Shine–Dalgarno sequence GGAG (Shine & Dalgarno, 1974). Examination of the DNA sequence downstream from the cytochrome *c*-554 gene stop codon showed that it contains a palindromic region which could form a hairpin structure (Figure 3). The cloned *Nar*I–*Eco*RI fragment started 42 bp upstream from the cytochrome *c*-554 gene start codon; therefore, no information about the promoter region could be obtained. Approximately 350 bp 3′ of the gene encoding the M subunit of the *C. aurantiacus* reaction center was sequenced, and four potential ORFs were found downstream of the M gene stop codon (Shiozawa et al., 1990). None of these match the cytochrome *c*-554 gene. Therefore, the cytochrome *c*-554 gene is not adjacent to the genes encoding the L and M subunits of the reaction center as it is in *R. viridis* (Weyer et al., 1987a). Apparently, the cytochrome and reaction center genes are situated in different transcriptional units.

In view of the periplasmic location of cytochrome *c*-554 in vivo, it is likely that this protein is synthesized as a precursor with an amino-terminal signal sequence which is removed during secretion. The length of Gram-negative bacterial signal peptides usually varies between 20 and 30 residues (von Heijne & Abrahmsen, 1989). Analysis of the first 35 amino acids encoded by the cytochrome *c*-554 gene, using the method described by von Heijne (1986), revealed several predicted cleavage sites: after Ala-28, Phe-31, Ala-24, Ile-21, and Gly-20 in order from highest to lowest score. Comparison of the derived N-terminal amino acid sequence of cytochrome *c*-554 with the signal peptide sequence of the *R. viridis* reaction center cytochrome (Figure 5) suggests that the most likely cleavage site is situated after Ala-24. In this case, the signal peptide of the cytochrome *c*-554 would have a "canonical" signal structure with a positively charged amino-terminal region (n-region) followed by a central hydrophobic region (h-region) and a more polar carboxy-terminal region that contains the cleavage site (c-region) (von Heijne & Abrahmsen, 1989). Also, a fairly well-conserved sequence, Val-X-Ala, would be present in positions −3 and −1 of this predicted signal peptide (Franzen et al., 1990). However, only the N-terminal sequence of the mature cytochrome *c*-554 protein, which is not yet available, can definitely establish the cleavage site.

In the *R. viridis* reaction center cytochrome *c*, the first amino acid of the mature protein is cysteine. The cysteine is covalently bound to fatty acids, indicating that the cytochrome is a lipoprotein (Weyer et al., 1987b). The fatty acids seem to anchor the cytochrome in the photosynthetic membrane. The cytochrome amino acid sequence at positions +1 to −3 is known to be the recognition site for the signal peptidase II (Inouye et al., 1983; Polit et al., 1986). This enzyme cleaves the peptide bond of bacterial prolipoproteins between the cysteine which is linked to the diglyceride and its preceding amino acid which is usually glycine (Inouye et al., 1983). This recognition site is clearly absent in the respective part of the derived amino acid sequence of the *C. aurantiacus* cytochrome, because the only cysteines that could be found in the sequence are located in heme binding sites. These data suggest that there is no fatty acid membrane anchor in the *C. aurantiacus* cytochrome. This could at least partially explain why the

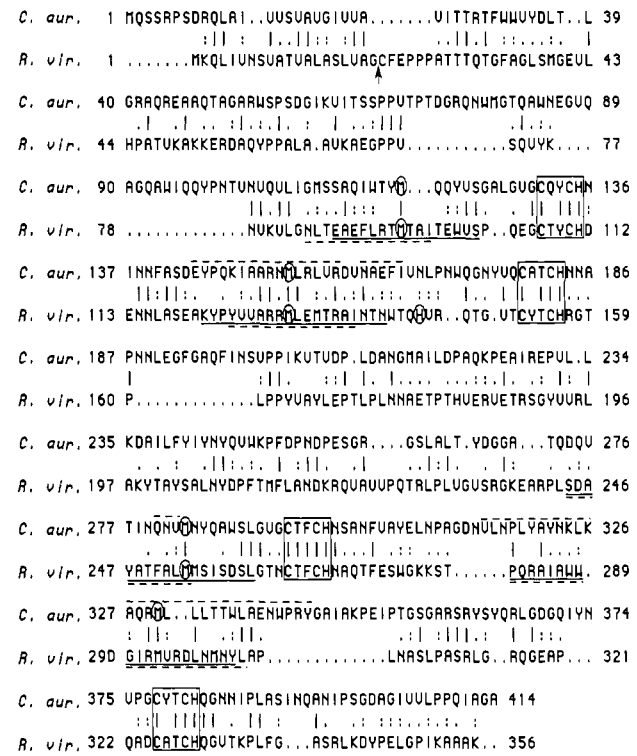

FIGURE 5: Comparison of the primary structures of the precursors of cytochrome *c*-554 from *C. aurantiacus* and the reaction center cytochrome subunit from *R. viridis* (Weyer et al., 1987a). The alignment was generated by the GAP/GCG program (Devereux et al., 1984) with a gap penalty of 2, and slightly modified by hand to maximize the number of invariant amino acid residues. Full vertical lines indicate identical and dashed vertical lines indicate similar amino acid residues. The end of the N-terminal signal peptide of the *R. viridis* cytochrome is indicated by an arrow. The heme binding motifs are boxed. The reported sixth axial ligands of the *R. viridis* cytochrome hemes and predicted ones of the *C. aurantiacus* cytochrome *c*-554 are circled. The α-helical regions adjacent to the heme binding sites, derived from X-ray structural analysis of *R. viridis* reaction center crystals (Deisenhofer et al., 1985; Michel et al., 1986), are indicated by solid lines. The α-helices in these regions that were predicted using the program PROTEIN2/GENEBEE (Brodsky et al., 1991) based on the Garnier method (Biou et al., 1988) for both cytochromes are indicated by dashed lines. The parameters of the program which revealed the best coincidence with the real secondary structure of the *R. viridis* cytochrome were used for the cytochrome *c*-554 structure prediction.

tetraheme cytochrome is easily lost from the reaction center in *C. aurantiacus* but not in *R. viridis*.

Upon comparison of the amino acid sequence of the *C. aurantiacus* cytochrome *c*-554 with the database of all cytochrome sequences, the heme binding sites containing the consensus (-Cys-X-X-Cys-His-) were found as the only significantly similar sequences in a number of cytochromes. Not surprisingly, the highest similarity was revealed with the heme binding segments of the four-heme reaction center cytochrome of *R. viridis*. Alignment of the complete sequences of the two four-heme cytochromes showed 30% identical amino acid residues, and 50% identical plus similar residues. Greater sequence similarity was observed in several segments of these two proteins, including the four heme binding sites and the region of the predicted signal peptide (Figure 5). Generation of the alignment required a number of gaps to be introduced due to the fact that the *C. aurantiacus* cytochrome is 58 amino acid residues longer than the *R. viridis* one. No significant similarity was revealed between the *C. aurantiacus* *c*-554 and one-heme cytochromes, except for the heme binding consensus.

The presence of the four heme binding sites in the *C. aurantiacus* and *R. viridis* cytochromes suggests the possibility of duplications in the genes coding for these proteins, or in their common ancestor. Self-comparison of each of these sequences, using the program ALIGN2/GENBEE, did not reveal pronounced internal periodicity. Specifically, much less similarity was observed between "heme binding domains" (segments surrounding different heme binding sites) in the *C. aurantiacus* cytochrome *c*-554 than between the respective homologous segments in the proteins of *C. aurantiacus* and *R. viridis*. These observations are not in disagreement with the proposed origin of the four-heme cytochromes by two consecutive duplications (Weyer et al., 1987a) but indicate that these duplications might have occurred in the remote past so that the common ancestor of the *C. aurantiacus* and *R. viridis* cytochromes would have already contained four heme binding sites.

The *R. viridis* cytochrome contains four structurally important α helices flanking each of the heme binding sites. Each helix runs parallel to the ring plane of each heme group. The first three of these helices (numbered according to their sequence position) include the Met axial ligands for the adjacent hemes (Weyer et al., 1987a; Figure 5). Secondary structure prediction using the Garnier method (Biou et al., 1988) showed that similarly positioned α helices may exist in the *C. aurantiacus* cytochrome upstream of the last three heme binding sites. Counterparts to the axial ligand Met residues could be identified for the first three cytochrome *c*-554 hemes (Figure 5). The fourth, C-terminal helix, which exists in the *R. viridis* cytochrome and was predicted in the *C. aurantiacus* cytochrome, contains a conserved Met residue which is not an axial ligand in *R. viridis*. Analysis of the *R. viridis* cytochrome X-ray structure revealed that this Met-273 (numbered according to the sequence position in the mature protein) is situated in van der Waals contact with both heme 4 and the actual axial ligand of this heme, His124. Therefore, if the structure of the *C. aurantiacus* cytochrome *c*-554 is similar to that of the *R. viridis* cytochrome, then the axial ligand for heme 4 could be Met330 in the *C. aurantiacus* sequence, since there is no His analogous to His124. The inferred positions of the axial Met ligands for all four hemes in cytochrome *c*-554 are shown in Figure 5.

In general, heme groups possessing a methionine as the sixth ligand tend to be high-potential, and hemes with a histidine as a sixth ligand are low-potential (Moore & Pettigrew, 1990). However, other environmental effects are important, so there are exceptions to this rule. For example, in the *R. viridis* reaction center cytochrome, histidine is found as the sixth ligand only for heme 4 (E_m = +20 mV), whereas methionine is the sixth ligand for hemes 1, 2, and 3, for which the corresponding midpoint redox potentials are −70, +320, and +380 mV (Weyer et al., 1987a; Dracheva et al., 1988). In *C. aurantiacus*, the midpoint potentials of the cytochrome *c*-554 hemes are 0, +120, +220, and +300 mV (Freeman & Blankenship, 1990). In at least one of the hemes from both cytochromes, other environmental influences must lower the redox potential.

A central problem with all multiheme proteins is the identification of particular spectroscopic and redox properties with specific hemes in the structure. All four hemes in cytochrome *c*-554 are distinguishable according to their redox potentials and orientations with respect to the membrane plane, but the hemes are not readily distinguished by the position of their α-band absorbance maxima (Freeman & Blankenship, 1990; Van Vliet et al., 1991). The *C. aurantiacus* and *R. viridis* cytochromes differ from each other in primary structure, size, and spectral and redox properties of hemes and also by heme orientations. Such differences do not allow an unambiguous

assignment of the arrangement of hemes in cytochrome c-554 just by comparison with the *R. viridis* heme positions. A crystal structure of cytochrome *c*-554 is necessary to determine the details of the heme organization.

Alignment of the complete sequences from *R. viridis* and *C. aurantiacus* showed 45% identical amino acid residues for the L subunits of the reaction centers, 40% for the M subunits (Michel et al., 1986; Shiozawa et al., 1989), and 30% for the tetraheme cytochromes (Weyer et al., 1987a). This variability in relatedness probably indicates that the tetraheme cytochromes evolved faster than the reaction center proteins. This difference might be explained by the fact that reaction centers from all photosynthetic bacteria have only one strictly conserved function—to produce a charge separation inside photosynthetic membranes—whereas the reaction center associated tetraheme cytochromes may have more than one role. The presence of low-potential hemes implies that besides donating an electron to the photooxidized special pair, tetraheme cytochromes can also oxidize low-potential substrates. If different organisms utilize different substrates, tetraheme cytochromes may need to be more adaptable than reaction centers.

Despite the functional similarities of the proteins, significant differences in the gene organization of the integral membrane antenna and reaction center complexes exist between *C. aurantiacus* and purple bacteria. In purple bacteria, the genes coding for the α- and β-polypeptides of the B870 (or B1015) antenna complex and the L and M subunits of the reaction center are situated in one operon (Youvan et al., 1984; Williams et al., 1984; Kiley et al., 1987; Berard et al., 1986; Wiessner et al., 1990; Liebetanz et al., 1991). In *R. viridis*, the cytochrome gene is located downstream of the gene for the M subunit (*puf*M). In another purple bacterium, *Erythrobacter* sp. OSH114, which possesses a reaction center tetraheme cytochrome (Matsuura & Shimada, 1990), an ORF was discovered directly downstream of *puf*M (Liebetanz et al., 1991). Partial sequence indicates that this ORF is highly homologous to the cytochrome subunit from the *R. viridis* reaction center. Although the gene organization is only known for two cases, it is tempting to speculate that in those purple bacteria that contain tetraheme cytochromes, the cytochrome gene is adjacent to the genes encoding the antenna and reaction center proteins. In contrast, in *C. aurantiacus*, the genes coding for the reaction center L and M subunits are situated in one operon, while the genes for the cytochrome *c*-554 and B808–866 antenna proteins are apparently in two other transcriptional units (Shiozawa et al., 1990).

According to 16S rRNA sequence analysis (Woese, 1987), *Chloroflexus* diverged from the eubacteria before branching of the purple bacteria from the cyanobacteria. However, comparison of reaction center sequence data places it much closer to the purple bacteria than to photosystem II (Beanland, 1990; Blankenship, 1992). In view of this, it might be assumed that *Chloroflexus* received its photosynthetic apparatus from purple bacteria by lateral gene transfer (or vice versa). The close grouping of photosynthetic genes in purple bacteria in comparison with the more dispersed organization in *Chloroflexus* suggests that the lateral gene transfer may have occurred from purple bacteria to *Chloroflexus* rather than the other way around, although more data are necessary to distinguish between possible evolutionary pathways.

Acknowledgments

We acknowledge Dr. J. Freeman for purification of cytochrome *c*-554, Drs. J. Allen and E. Koonin for helpful suggestions and discussions, Dr. S. Bingham for synthesizing the oligonucleotides, and Mr. A. Drachev for his help with the computer analyses.

References

Alegria, G., & Dutton, P. L. (1987) in *Cytochrome Systems* (Papa, S., Chance, B., & Ernster, L., Eds.) pp 601–608, Plenum Press, New York and London.

Allen, J. P., Feher, G., Yeates, T. O., Komiya, H., & Rees, D. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6162–6166.

Ambler, R. P. (1982) in *From Cyclotrons to Cytochromes* (Robinson, A. B., & Kaplan, N. O., Eds.) pp 263–280, Academic Press, New York.

Ambler, R. P., & Wynn, M. (1973) *Biochem. J. 131*, 485–498.

Amesz, J. (1991) in *Variations in Autotrophic Life* (Shively, J. M. & Barton, L. L., Eds.) pp 99–119, Academic Press, London.

Beanland, T. J. (1990) *J. Theor. Biol. 145*, 535–545.

Bérard, J., Bélanger, G., Corriveau, P., & Gingras, G. (1986) *J. Biol. Chem. 261*, 82–87.

Biou, V., Gibrat, F., Levin, G. M., Robson, B. & Garnier, J. (1988) *Protein Eng. 2*, 185–189.

Blankenship, R. E. (1992) in *Photosynthesis and the Origin of Life* (Hartman, H., Ed.) Neil Patterson Publ., Burlingame, NC (in press).

Blankenship, R. E., & Fuller, R. C. (1986) in *Photosynthesis III* (Staehelin, L. A. & Arntzen, C. J., Eds.) pp 390–399, Springer-Verlag, Berlin.

Blankenship, R. E., Trost, J. T., & Mancino, L. J. (1988) in *The Photosynthetic Bacterial Reaction Center* (Breton, J., & Vermeglio, A., Eds.) pp 119–128, Plenum Press, New York.

Brodsky, L. I., Drachev, A. L., Tatuzov, R. L., & Chumakov, K. M. (1991) *Biopolim. Kletka 7*, 10–14.

Bruce, B. D., Fuller, R. C., & Blankenship, R. E. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6532–6536.

Debus, R. J., Feher, G., & Okamura, M. Y. (1985) *Biochemistry 24*, 2488–2500.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., & Michel, H. (1985) *Nature 318*, 618–624.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.

Dracheva, S., Drachev, L., Konstantinov, A., Semenov, A., Skulachev, V., Arutjunjan, A., Shuvalov, V., & Zaberezhnaya, S. (1988) *Eur. J. Biochem. 171*, 253–264.

Feiler, U., Nitschke, W., Michel, H., & Rutherford, A. W. (1989) Symposium on Molecular Biology of Membrane Bound Complexes in Phototrophic Bacteria, Freiburg, FRG, Abstract P-6, 19.

Franzen, L., Rochaix, J., & von Heijne, G. (1990) *FEBS Lett. 260*, 165–168.

Freeman, J. C. (1989) Ph.D. Thesis, Arizona State University.

Freeman, J. C., & Blankenship, R. E. (1990) *Photosynth. Res. 23*, 29–38.

Fukushima, A., Matsuura, K., Shimada, K., & Satoh, T. (1988) *Biochim. Biophys. Acta* 933, 399–405.

Gribskov, M., Devereux, J., & Burgess, R. (1984) *Nucleic Acids Res. 12*, 539–548.

Hanahan, D. (1983) *J. Mol. Biol. 166*, 557–580.

Inouye, S., Franceschini, T., Sato, M., Itakura, K., & Inouye, M. (1983) *EMBO J. 2*, 87–91.

Kiley, P. J., Donohue, T. J., Havelka, W. A., & Kaplan, S. (1987) *J. Bacteriol. 169*, 742–750.

Knaff, D. B., Willie, A., Long, J. E., Kriauciunas, A., Durham, B., & Millett, F. (1991) *Biochemistry 30*, 1303–1310.

Liebetanz, R., Hornberger, U., & Drews, G. (1991) *Mol. Microbiol. 5*, 1459–1468.

Matsuura, K., & Shimada, K. (1990) in *Current Research in Photosynthesis* (Baltscheffsky, M., Ed.) Vol. 1, pp 193–196, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Michel, H., Weyer, K. A., Gruenberg, H., Dunger, I., Oesterhelt, D., & Lottspeich, F. (1986) *EMBO J. 5*, 1149–1158.

Moore, G. R., & Pettigrew, G. W. (1990) *Cytochrome c. Evolutionary, Structural and Physicochemical Aspects*, Springer-Verlag, Berlin.

Nitschke, W., & Rutherford, A. W. (1989) *Biochemistry 28*, 3161–3168.

Nitschke, W., Agalidis, L., Jubault-Bregler, M., Feiler, U., Liebl, U., & Rutherford, A. W. (1990) C5 Symposium on Charge Transfer in Cytochrome and Chlorophyll Complexes, Montreal, Canada.

Nozawa, T., Trost, J. T., Fukada, T., Hatano, M., McManus, J. D., & Blankenship, R. E. (1987) *Biochim. Biophys. Acta 894*, 468–476.

Ovchinnikov, Yu. A., Abdulaev, N. G., Shmuckler, B. E., Zargarov, A. A., Kutuzov, M. A., Telezhinskaya, I. N., Levina, N. B., & Zolotarev, A. S. (1988a) *FEBS Lett. 232*, 364–368.

Ovchinnikov, Yu. A., Abdulaev, N. G., Zolotarev, A. S., Shmuckler, B. E., Zargarov, A. A., Kutuzov, M. A., Telezhinskaya, I. N., & Levina, N. B. (1988b) *FEBS Lett. 231*, 237–242.

Pierson, B. K., & Castenholz, R. W. (1974) *Arch. Microbiol. 100*, 5–24.

Pierson, B. K., Thornber, J. P., & Seftor, R. E. B. (1983) *Biochim. Biophys. Acta 723*, 322–326.

Pollitt, S., Inouye, S., & Inouye, M. (1986) *J. Biol. Chem. 261*, 1835–1837.

Robinson, S. J., & Redlinger, T. E. (1987) in *Progress in Photosynthesis Research* (Biggins, J., Ed.) Vol. 4, pp 741–744, M. Nijhoff Publishers, The Netherlands.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Seftor, R. E. B., & Thornber, J. P. (1984) *Biochim. Biophys. Acta 764*, 148–159.

Shine, J., & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1342–1346.

Shiozawa, J. A., Lottspeich, F., & Feick, R. (1987) *Eur. J. Biochem. 167*, 595–600.

Shiozawa, J. A., Lottspeich, F., Oesterhelt, D., & Feick, R. (1989) *Eur. J. Biochem. 180*, 75–84.

Shiozawa, J. A., Csiszar, K., & Feick, R. (1990) in *Molecular Biology of Membrane-Bound Complexes in Phototrophic Bacteria* (Drews, G., & Dawes, E., Eds.) pp 11–18, Plenum Press, New York.

Shopes, R. J., Levine, L. M. A., Holten, D., & Wraight, C. A. (1987) *Photosynth. Res. 12*, 165–180.

Theroux, S. J., Redlinger, T. E., Fuller, R. C., & Robinson, S. J. (1990) *J. Bacteriol. 172*, 4497–4504.

Van Vliet, P., Zannoni, D., Nitschke, W., & Rutherford, A. W. (1991) *Eur. J. Biochem. 199*, 317–323.

Von Heijne, G. (1986) *Nucleic Acids Res. 14*, 4683–4690.

Von Heijne, G., & Abrahmsen, L. (1989) *FEBS Lett. 244*, 239–246.

Ward, D. M., Weller, R., Shiea, J., Castenholz, R. W., & Cohen, Y. (1989) in *Microbial Mats: Physiological Ecology of Benthic Microbial Communities* (Cohen, Y., & Rosenberg, E., Eds.) pp 3–15, American Society for Microbiology, Washington, D.C.

Weyer, K. A., Lottspeich, F., Gruenberg, H., Lang, F., Oesterhelt, D., & Michel, H. (1987) *EMBO J. 6*, 2197–2202.

Weyer, K. A., Schäfer, W., Lottspeich, F., & Michel, H. (1987) *Biochemistry 26*, 2909–2914.

Wiessner, C., Dunger, I., & Michel, H. (1990) *J. Bacteriol. 172*, 2877–2887.

Williams, J. C., Steiner, L., Ogden, R. C., Simon, M. I., & Feher, G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6505–6509.

Williams, J. C., Steiner, L., Feher, G., & Simon, M. I. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7303–7307.

Woese, C. R. (1987) *Microbiol. Rev. 51*, 221–271.

Youvan, D. C., Bylina, E. J., Alberti, M., Begusch, H., & Hearst, J. E. (1984) *Cell 37*, 949–957.

Zannoni, D., & Venturoli, G. (1988) in *Green Photosynthetic Bacteria* (Olson, J. M., Ormerod, J. G., Amesz, J., Stackebrandt, E., & Trüper, H., Eds.) pp 135–143, Plenum Press, New York.

Zuber, H., Brunisholz, R., & Sidler, W. (1987) in *Photosynthesis* (Amesz, J., Ed.) pp 233–271, Elsevier Science Publishers B.V. (Biomedical Division), Amsterdam.