ELSEVIER

Biochimica et Biophysica Acta 1275 (1996) 145–150

BBA Biochimica et Biophysica Acta

Short sequence-paper

Molecular cloning, DNA sequence and transcriptional analysis of the *Rhodospirillum molischianum* B800/850 light-harvesting genes

Lothar Germeroth, Helmut Reiländer, Hartmut Michel *

Max-Planck-Institut für Biophysik, Abteilung Molekulare Membranbiologie, D-60528 Frankfurt / M., Germany

Received 15 January 1996; accepted 19 February 1996

Abstract

The amino acid sequences of the B800/850 light-harvesting proteins from *Rhodospirillum molischianum* were determined by Edman degradation. On the basis of these amino acid sequences, two degenerated oligonucleotides were synthesized and used for PCR of genomic DNA. The resulting 150 bp DNA fragment was cloned, sequenced and used for subsequent Southern blot analysis of digested genomic DNA. A 2.3 kbp *Eco*RI fragment strongly hybridized to the probe and a size selected genomic library from genomic DNA was constructed. One clone scored positive during screening of the library with the PCR-fragment and subsequent DNA sequence analysis of the clone revealed the presence of three A-genes (A1A2A3) encoding α-polypetides and of two B-genes (B1B2) encoding β-polypeptides of the B800/850 complex. The arrangement of the different genes are B1A1, B2A2 and A3 where only B1 and B2 are preceded by typical Shine-Dalgarno sequences. In addition, typical nucleotide sequences for a rho-independent termination of transcription are located downstream of the genes A1 and A2. The deduced amino acid sequences revealed that the α-genes encoded for identical polypeptides, whereas the deduced β-polypeptides differed in their amino acid sequence at four positions. Transcriptional operon analysis revealed that the genes A1B1 and A2B2 are both dicistronically transcribed, whereas the gene A3 is not.

Keywords: B800/850 complex; Light-harvesting; Nucleotide sequence

The function of light-harvesting complexes in photosynthetic organisms is mainly to absorb visible and near infrared radiation and to mediate energy transfer to the reaction center, where the primary charge separation of photosynthesis takes place [1]. The B800-850 light-harvesting complexes of the purple bacteria typically consist of two polypeptides, the α- and the β-subunits which noncovalently bind bacteriochlorophylls and carotenoids (for review see Ref. [2]). The α,β-heterodimer with its corresponding pigments form higher oligomers [3,4] which can be purified after solubilisation of the intracytoplasmic membrane with detergents.

The α- and β-genes of the B800-850 complexes from several different photosynthetic bacteria have been cloned and sequenced (see e.g., Refs. [5,6]) and the genes coding for the α- and β-subunits appear to be organized in the same way in various organisms. The genes are clustered in the so-called puc-operon and are dicistronically transcribed. In some photosynthetic bacteria the structural genes coding for the α,β-polypeptides are preceded by silent genes thought to be involved in the assembly of the oligomeric light-harvesting complexes [7,8]. In the cases of *Rhodopseudomonas (Rps.) palustris* and *Rps. acidophila*, multi-gene families were found which code for slightly different B800/850 α,β-polypeptides [9,10]. The purification of three different α- and two different β-polypeptides of the B800/850 complex from *Rps. palustris* indicates that different gene products of the multi-gene family are assembled to form the native light-harvesting complexes.

The different light-harvesting complexes from *Rs. molischianum* strain DSM 119 have been characterized recently [11,12]. Only one α- and one β-polypeptide could be purified from the isolated B800/850 complex. The amino acid sequences of the α,β-polypeptides and the bacteriochlorophyll binding sites were analysed by Raman spectroscopy and were found to be more similar to the B870 light-harvesting complexes than to the B800/850 complexes from other purple bacteria. In addition, the isolated B800/850 light-harvesting complex from *Rs. molischianum* is an octamer [13], which was the first deviation from the generally accepted model that

* Corresponding author. Fax: +49 69 96769423.

PII S0005-2728(96)00017-5

B800/850 complexes associate to hexamers [2]. However, the X-ray structure analysis of the B800/850 complex from *Rps. acidophilum* indicated that the crystallized complex is a nonamer of α, β-heterodimers [14]. Moreover, crystals of the B800/850 complex from *Rs. molischianum* were obtained diffracting to 2.4 Å resolution [15]. The space group $P42_12$ was determined indicating a four-fold symmetry of the B800/850 complex from *Rs. molischianum* which is in good agreement with the octameric association.

The α- and β-polypeptides of *Rs. molischianum* were purified as described [11] and the amino acid sequences of the two peptides were determined by Edman degradation. On the basis of these amino acid sequences, two degenerated oligodesoxynucleotides were designed for subsequent PCR-cloning of the corresponding gene region. In a polymerase chain reaction performed with genomic DNA isolated from *Rs. molischianum*, a 150 bp DNA fragment could be amplified. We cloned the fragment into pBluescript and determined the nucleotide sequence. The amino acid sequences which could be deduced from the resulting DNA sequence were identical to the ones determined chemically from the polypeptides. A ^{32}P-labeled RNA probe produced from the 150 bp DNA fragment was used for Southern analysis of *Eco*RI, *Bam*HI and *Eco*RI/*Bam*HI digested genomic DNA from *Rs. molischianum*. The probe hybridized to a 2.3 kbp *Bam*HI fragment and a 3.6 kbp *Eco*RI fragment. A Southern analysis performed with double digested genomic DNA (*Bam*HI and *Eco*RI) also gave a signal at 2.3 kbp, indicating that the light-harvesting genes are encoded on the 2.3 kbp *Bam*HI fragment. A genomic library of *Rs. molischianum* constructed from size fractionated *Bam*HI-fragments (2–2.5 kbp) into pBluescript was subsequently screened by colony hybridization. From 2000 colonies screened one clone scored positive and was identified to contain the 2.3 kbp genomic DNA fragment.

In order to obtain the complete information about the *Rs. molischianum* B800/850 gene organization, the complete nucleotide sequence of this fragment was determined by the dideoxy-chain termination method [16,17] (Fig. 1). The combined sequence information spanned a 2301 bp DNA fragment of which the nucleotide sequence of both strands was determined (Fig. 2). Computer analysis of the sequence revealed that the B800/850 complex genes from *Rs. molischianum* strain DSM 119 is organized as a multigene family coding for three α-polypeptides (A-genes) and two β-polypeptides (B-genes). The arrangement of the different genes is B1A1, B2A2 and A3. Only the two B-genes are preceded by typical Shine-Dalgarno consensus sequences (5′-GGAG-3′) which also have been described for the light-harvesting genes from other photosynthetic bacteria [5]. Two nine base pair repeats (5′-GTCATG-GCG-3′) are located in the upstream region of gene B1 whereas comparable sequences are absent in case of B2. Previously it has been suggested that a repeat of 7 bp

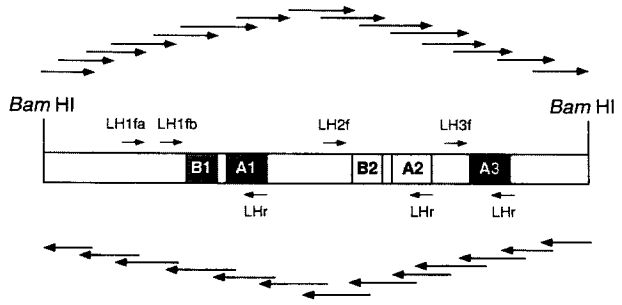

Fig. 1. Sequencing strategy of the 2.3 kbp genomic *Bam*HI DNA fragment of *Rs. molischianum* encoding the B800/850 light-harvesting genes. Arrows show the direction of reading, relative length and positions of sequenced DNA fragments either derived by restriction with appropriate restriction enzymes or using the exonuclease III/mung bean nuclease method. Open reading frames corresponding to the different α,β genes are indicated within the map. Position of primers used for PCR based analysis of the transcripton of the genes are given above (LH1fa, LH1fb, LH2f, LH3f) or below (LHr), respectively.

located about 150 bp upstream from the start site of the puc-operon from *Rb. capsulatus* might be involved in transcription initiation. We assume that the repeats present in the *Rs. molischianum* sequence may serve the same function. 30 bp downstream of the genes A1 and A2 two palindromic sequences followed by a short run of T-residues can be detected. This arrangement is typical for rho-independent termination of transcription. The distances of these sequences from the stop codon of the A genes are comparable to the distance reported for *Rps. palustris* [18].

The A3 gene is not preceded by a corresponding B3 gene. This is a unique situation among all puc-operons from photosynthetic bacteria sequenced so far. The A3 gene exhibits no classical Shine Dalgarno sequence observed in ease of the A1 and A2 genes. Most likely the A3 gene originated from a gene duplication starting directly in-between the AB-gene cluster.

The amino acid sequences deduced for the two α-polypeptides are identical, whereas these for the β-polypeptides differ in four amino acid positions (B_1Ile15-B_2Val15; B_1Glu19-B_2Ala19; B_1Lys22-B_2Gln22; B_1Ile30-B_2Val30; Fig. 2). The amino acid sequence of the α-subunits as deduced from the nucleotide sequence deviates from the chemically determined sequence [11] in position 43. The DNA sequence indicates the presence of a Phe in this position, whereas a Ser was found by chemically sequencing. Since the molecular mass determined by electrospray mass spectroscopy [11] of the α-subunit agrees with the amino acid sequence deduced from the nucleotide sequence within one Dalton, we conclude that the chemically determined sequence is not correct in position 43. The charged amino acids at positions 19 (asp) and 22 (lys) in the β_1-polypeptide are of special interest, because they are in close proximity to histidine-18, which is conserved among most β-polypeptides from other purple bacteria (for comparison see Ref. [2]). Histidine-18 was thought to bind to the Mg-ion of the monomeric bacteriochlorophyll, which

(a)

```
      10                  30                  50
agaatgttggcggtcagcaggctgatttccgccgcggggtaggcggcgcggaggcgtccgccgc
    70                  90                 110
caccaatccgccgatccccggcatgttaacgtcgaccagcacgtaatcaggcggcagaatcgcg
  130                 150                 170                 190
gcggcggtcagggcctcctcgccgctggcttcctcgacgatccgccagcccggtcgcagccgct
                210                 230                 250
cgatcaccgcgtcccggtcatcatccgcgcccacccggctgtcatccaccaatagaacggtttg
            270                 290                 310
tcactcatcgctggcccggtttctccaaggacatgttcatcgcggtacggcgtacgtatcgacg
        330                 350                 370
acctgacggagtctatgcttgagggagggagcgccggcaagggcaaagtccccatccctggccc
    390                 410                 430
tccctgccgccaggggcgaatttcccgttccgaccggttgccggctcccgtagattcgaattga
                    470                 490                 510
cacccgcgagcgacgatcgaagagtaccgcctgttacgggggctggccaaaggggaggcaaggc
                530                 550                 570
ggcagggcagtgcgagggaacggtgccggggtatccagacgttcccgggtgttttcacgtcctg
            590                 610                 630
tgctgatcttcgtcaatggctgctttttcttccgtcgaaatggatgttgtgcgtcggtgcgggc
        650                 670                 690
tgttcaatcccggaccggagcataaatgagagtaatcatggtgtctggcctgggcgcgtgcggc
    710                 730                 750
tggtcgcgccggccaacccggagagggtgactgttttcgtcccgaatggctgcgatctctgcgc
                    790                 810                 830
cgcgggatgggaccgagttggtcatggcgtctaggcgtcatggcgaacccgccgcgcgggag
                850                 870                 890
gtgcgcggcaaccaacaatggagtgtagacacatggctgaaagaagcttgtcgggcctgaccgag
                                  M   A   E   R   S   L   S   G   L   T   E

            910                 930                 950
gaagaggcgatcgcggtccacgaccagttcaagaccaccttctccgctttcatcatcctggccg
 E   E   A   I   A   V   H   D   Q   F   K   T   T   F   S   A   F   I   I   L   A

        970                 990                1010
ccgtcgcgcacgtgctggtttgggtctggaagccctggttctgatttctcgacacacttcttag
A   V   A   H   V   L   V   W   V   W   K   P   W   F   *

   1030                1050                1070
ggggataatcctatgagcaatccgaaggacgactacaagatttggctggtcatcaatccgtcga
             M   S   N   P   K   D   D   Y   K   I   W   L   V   I   N   P   S

                    1110                1130                1150
cctggctgcccgtgatctggatcgttgccaccgtcgtcgcgatcgccgtgcatgccgccgtcct
T   W   L   P   V   I   W   I   V   A   T   V   V   A   I   A   V   H   A   A   V   L

                1170                1190                1210
ggccgctccgggcttcaactggatcgccctcggcgccgcgaagagcgccgcgaagtaagtcact
  A   A   P   G   F   N   W   I   A   L   G   A   A   K   S   A   A   K   *

            1230                1250                1270
ctttttagagtgtctgactgagggGTTGCGGACGGAACCGAGtttcggCTCGGTTCCGTCCGCA
        1290                1310                1330
ACttttctggggtcgtctcccgccccgccggccgcgccctggtgccccggccccgcccccggc
   1350                1370                1390
tctgctccggcttttgccgtgccgccgccatccggcctctgccccgctcttccgcgccatcacg
                    1430                1450                1470
ccccctcgttcgccaggggttcgtctctcactgccttcgccgcgtgaaagttgcgagtcagggg
                1490                1510                1530
gagattcacgttgactccgcccggccccggtgcaacattacaggccgttaatgaggacttgcag
```

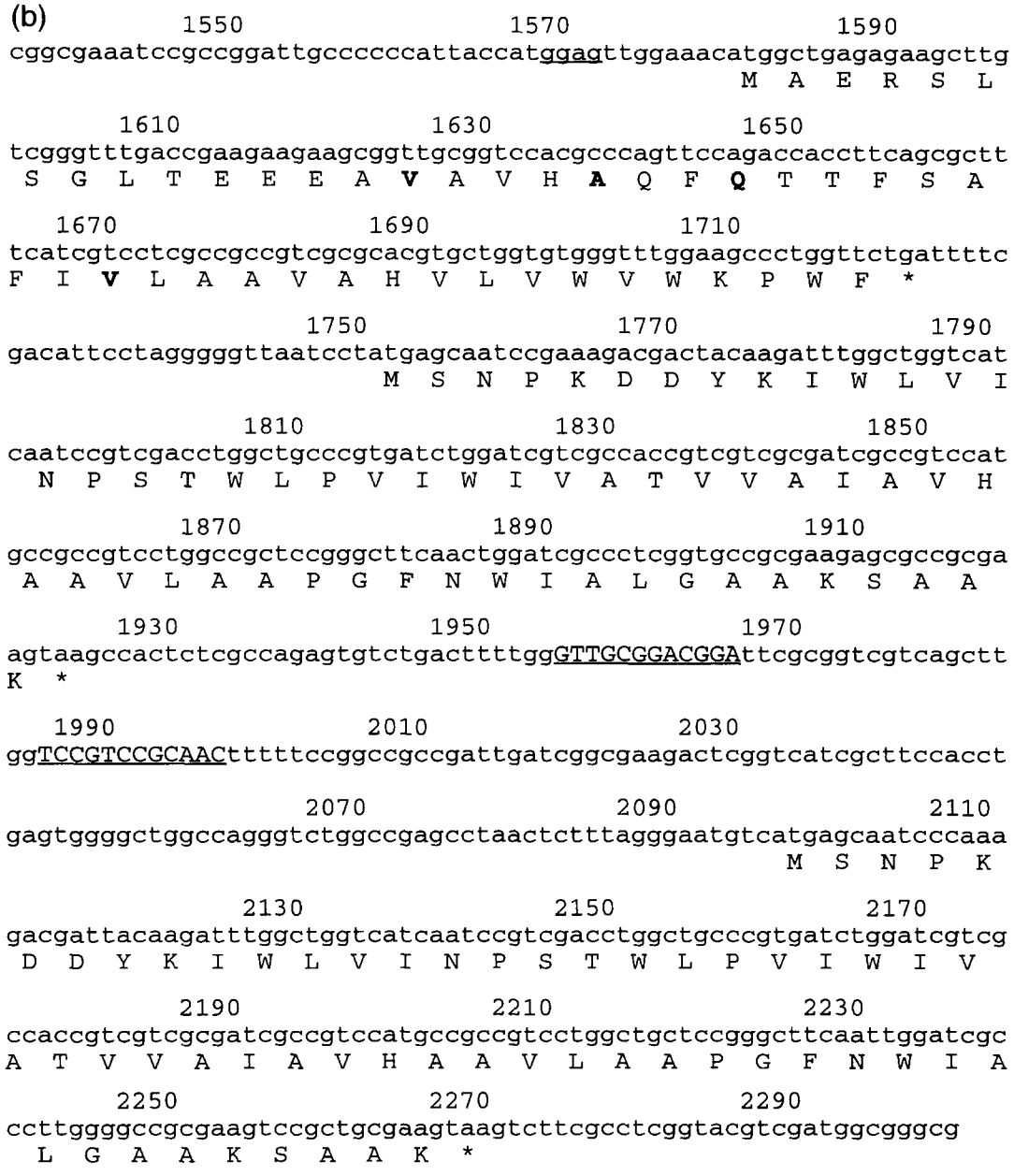

Fig. 2. Nucleotide sequence of the 2.3 kbp genomic fragment and the deduced amino acid sequence of the α-, β-polypeptides. The nucleotide sequence is numbered from the 5′ to the 3′ end. Potential Shine-Dalgarno sequences are underlined. Repeats within the nucleotide sequence are given in bold and are double underlined. Nucleotides in palindromic sequences most probably responsible for *rho*-independent termination of transcription are presented in upper letter code and are underlined. The deduced amino acid sequences of the polypeptides are shown below the nucleotide sequence in one letter code. Asterisks denote stop codons in the respective reading frames. Amino acid residues in the β-gene which are discussed in the text are given in bold.

may cause a characteristic absorption at 800 nm in the absorption spectrum of the B800/850 light-harvesting complexes. However, the X-ray structure analysis showed that this is not the case [14]. Instead of this it seems to stabilize the secondary structure of the protein surrounding the monomeric bacteriochlorophyll by hydrogen bonding. B1Gln19 and B1Lys22 are exchanged into non-charged amino acids (ala-19; gln-22) in the β_2-polypeptide. This indicates that the charged amino acids in position 19 and 22 do not contribute to the characteristic absorption band at 800 nm, but that these amino acids may be involved in electrostatic interaction between the α,β-polypeptides forming the unique octameric B800/850 light-harvesting complex of *Rs. molischianum*. This is in agreement with the fact that β-polypeptides of other purple bacteria, which are more similar to the B800/850 light-harvesting complex from *Rps. acidophilum*, contain non-charged amino acids at positions 19 and 22.

We examined the expression of the genes by Northern blot analysis which revealed that only one band with a size

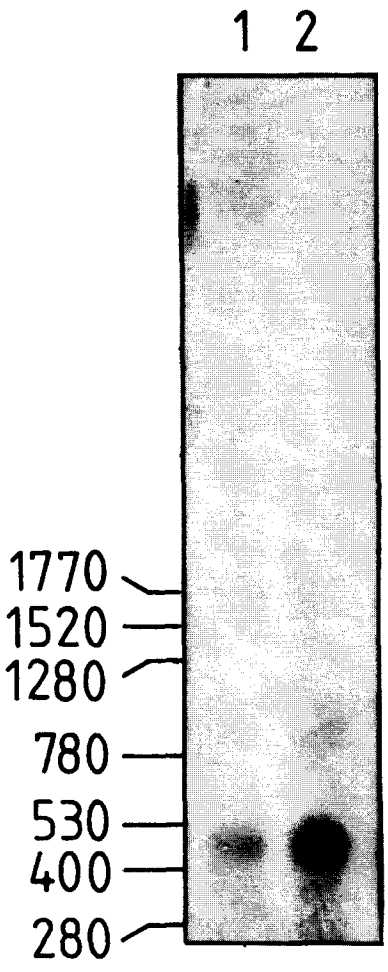

Fig. 3. Northern hybridization of total RNA from *Rs. molischianum*. Lane 1, 20 μg total mRNA; lane 2, 40 μg total mRNA. The RNA was size-fractionated on a formaldehyd agarose gel and then hybridized with a radioactive probe containing part of the A_1B_1 gene sequence.

of 450 bp hybridized to the probe (Fig. 3). The size of the transcript supports the assumption that the two AB gene clusters are bicistronically transcribed. Nevertheless, from this analysis it was not clear whether the A_1B_1 gene cluster and/or the A_2B_2 gene cluster are transcribed. In order to determine which cluster is transcribed, we used a PCR-based approach. Total RNA was prepared and a first strand DNA was synthesized by reverse transcription using primer (LHr) which fits to the three A genes. To discriminate which gene doublet becomes transcribed, PCRs with the single-stranded cDNA and different sets of specific primers were performed (see Fig. 1 for localization of the primers; primer combinations were: LH1fa-LHr, LH1fb-LHr, LH2f-LHr and LH3f-LHr). We were able to obtain PCR-products corresponding to the genes A_1B_1 (primer combination: LH1fb-LHr) and A_2B_2 (primer combination: LH2f-LHr), whereas no PCR-product of the A_3-gene (primer combination: LH3f-LHr) could be detected.

This result suggests that the two dicistronic mRNAs possess a similar if not identical size, which is in contrast to *Rb. sphaeroides* [7] and *Rps. palustris* [18], where two B800/850 mRNAs of significantly different length could be observed. In the case of *Rb. capsulatus* [20], two transcripts of one B800/850 BA gene cluster differing only 15 bases in length were detected. A similar situation could exist in *Rs. molischianum*, because the technique we used for Northern blot analysis was not sufficient to resolve very small size differences of the transcripts.

Although both genes coding for the β-polypeptides were transcribed, only one β-polypeptide could be purified from the isolated B800-850 complex [11]. One possible explanation for this is that we lost the B2-gene product during the purification procedure of the B800/850 complex. But we did not observe a significant loss of B800/850 complex during the purification. More likely is the possibility that both highly hydrophobic β-polypeptides coelute in one peak in RP-HPLC purification. If the β_2-polypeptide is only present in small amounts (< 10%), electrospray mass spectra and amino acid sequencing will fail to detect the β_2-polypeptide. This situation would suggest a different regulation of both dicistronically transcribed BA-clusters. On the transcriptional level it would be possible that the half-lives of the mRNA from B1A1 and B2A2 are different. This could be due to a similar mechanism reported for the puf operon of *Rb. capsulatus* [19] with respect to comparable palindromic sequences of both, the puf BA gene of *Rb. capsulatus* and the puc B_1A_1 genes of *Rs. molischianum*. Moreover different frequencies of initiation of transcription could contribute to different mRNA levels in the cells.

This work was supported by the Max-Planck-Geselschaft and the Fonds der Chemischen Industrie.

References

[1] Van Grondelle, R. (1985) Biochim. Biophys. Acta 811, 147–195.
[2] Zuber, H. and Brunisholz, R. (1991) in Structure and Function of Antenna Polypeptides and Chlorophyll-protein complexes: Principles and Variability in Chlorophylls (Scheer, H., ed.) pp. 627–703, CRC Press, Boca Raton.
[3] Hunter, C.N., Pennoyer, J.D., Sturgis, J.N., Farrely D. and Niederman, R.A. (1988) Biochemistry 27, 3459–3467.
[4] Boonstra, A.F., Visschers, R.W., Calkoen, F., Van Grondelle, R., Van Bruggen E.F.J. and Boekema, E.J. (1993) Biochim. Biophys. Acta 1142, 181–188.
[5] Youvan, D.C. and Ismail, S. (1985A) Proc. Natl. Acad. Sci. USA 82, 58–62.
[6] Kiley, P.J. and Kaplan, S. (1987) J. Bact. 169, 3268–3275.
[7] Lee, J.K., Kiley, P.J. and Kaplan, S. (1989) J. Bact. 171, 3391–3405.
[8] Tichy, H.V., Albien, K.U., Gad'on, N. and Drews, G. (1991) EMBO J. 10, 2949–2955.
[9] Tadros, M.H. and Waterkamp, K. (1989) EMBO J. 8, 1303–1308.
[10] Gardiner, A.T., MacKenzie, R.C., Barett, S.J., Kaiser, K. and Cogdell, R.J. (1992) in Research in Photosynthesis, Vol. 1, pp. 77–80, (N. Murata, ed.), Kluwer, Netherlands.
[11] Germeroth, L., Lottspeich, F., Robert, B. and Michel, H. (1993) Biochemistry 32, 5615–5621.
[12] Boonstra, A.F., Germeroth, L. and Boekema, E.J. (1994) Biochim. Biophys. Acta 1184, 227–234.

[13] Kleinekofort, W., Germeroth, L., Van den Broek, J.A., Schubert D. and Michel H. (1992) Biochim. Biophys. Acta 1140, 102–104.
[14] McDermott, G., Prince, S.M., Freer, A.A., Hawthornthwaite-Lawless, A.M., Papiz, M.B., Cogdell, R.J. and Issacs, N.W. (1995) Nature 374, 517–521.
[15] Michel, H. (1991) in Crystallization of membrane proteins (Michel. H., ed.), pp. 82–83, CRC Press, Boca Raton.
[16] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[17] Henikoff, S. (1984) Gene 28, 352–359.
[18] Tadros, M.H. (1990) in Molecular biology of Membrane-bound Complexes in Phototrophic Bacteria (Drews, G. and Dawes, E.A., eds.), pp. 19–31, Plenum Press, New York.
[19] Chyi-Ying A.C. and Belasco J.G. (1990) J. Bact. 172, 4578–4586.
[20] Zucconi, A.P. and Beatty, J.T. (1988) Posttranscriptional Regulation by Light of the Steady-State Levels Mature B800-850 Light-Harvesting Complexes in *Rhodobacter capsulatus*.