JOURNAL OF BACTERIOLOGY, June 1996, p. 3369–3373 Vol. 178, No. 11
0021-9193/96/$04.00+0

Sequence of the *bchG* Gene from *Chloroflexus aurantiacus*: Relationship between Chlorophyll Synthase and Other Polyprenyltransferases†

JOHN C. LOPEZ, SHANNON RYAN, 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

Received 27 December 1995/Accepted 21 March 1996

The sequence of the *Chloroflexus aurantiacus* open reading frame thought to be the *C. aurantiacus* homolog of the *Rhodobacter capsulatus bchG* gene is reported. The BchG gene product catalyzes esterification of bacteriochlorophyllide *a* by geranylgeraniol-PP$_i$ during bacteriochlorophyll *a* biosynthesis. Homologs from *Arabidopsis thaliana*, *Synechocystis* sp. strain PCC6803, and *C. aurantiacus* were identified in database searches. Profile analysis identified three related polyprenyltransferase enzymes which attach an aliphatic alcohol PP$_i$ to an aromatic substrate. This suggests a broader relationship between chlorophyll synthases and other polyprenyltransferases.

BchG in *Rhodobacter capsulatus* is involved in the biosynthesis of bacteriochlorophyll *a*, catalyzing the esterification of bacteriochlorophyllide *a* with geranylgeraniol (10). The alcohol moiety is then reduced in steps to phytol, the chain found in the mature bacteriochlorophyll *a* of most bacteria. Plants use a similar pathway for the biosynthesis of chlorophyll *a* in etiolated membranes (26, 28) with the enzyme chlorophyll synthase. Green plants appear to be esterified directly with phytol (25). Other chlorophylls and bacteriochlorophylls use other substrates in the esterification reaction (27, 29).

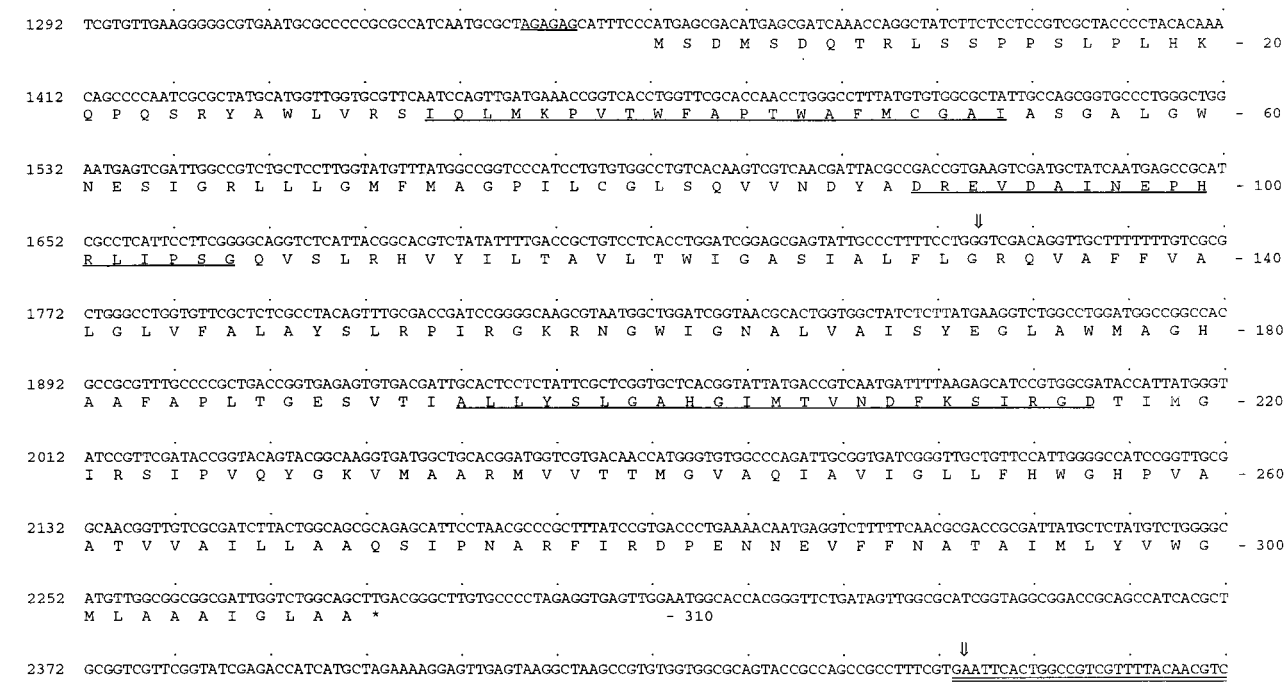

FIG. 1. Nucleotide sequence and derived amino acid sequence of the *C. aurantiacus bchG* gene and surrounding regions. The sequence shown starts slightly after the stop codon from the upstream cytochrome *c*-554 gene. A putative Shine-Delgarno sequence (underlined) lies directly upstream of the predicted start codon. The first arrow above the nucleotide sequence (nucleotide 1745) indicates the end of the previous sequencing work (13). The second arrow above the nucleotide sequence (nucleotide 2463) marks the *Eco*RI site at the end of the cloned *Nar*I-*Eco*RI fragment and the beginning of the pUC19 sequence (double-underlined sequence). The underlined amino acid sequences show the highest levels of identity between the *C. aurantiacus* and *R. capsulatus* BchG proteins in a GAP (17) alignment. The stop codon is indicated by an asterisk.

* Corresponding author. Phone: (602) 965-1439. Fax: (602) 965-2747.

† Publication no. 280 from the Arizona State University Center for the Study of Early Events in Photosynthesis.

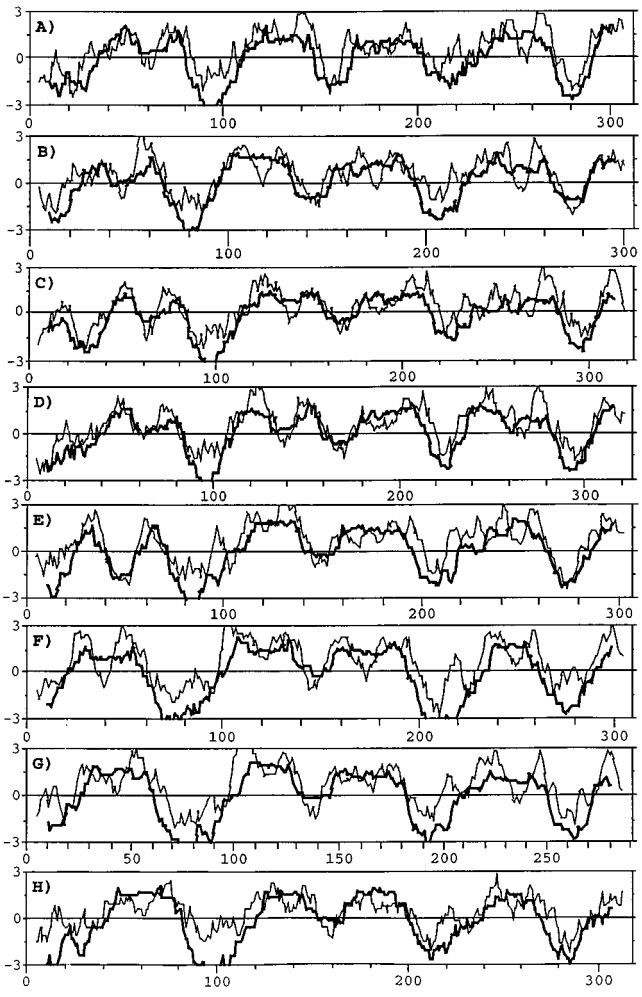

FIG. 2. Hydrophobicity plots for the identified polyprenyltransferases. (A) *C. aurantiacus* BchG (311 amino acids [aa]); (B) *R. capsulatus* BchG (304 aa); (C) chlorophyll synthase G4 from *A. thaliana*, aa 67 to 387 (321 aa); (D) chlorophyll synthase from *Synechocystis* sp. strain PCC6803 (324 aa); (E) the second bacteriochlorophyll synthase, BchG2, from *C. aurantiacus* (305 aa); (F) UbiA from *E. coli* (290 aa); (G) Coq2 from *S. cerevisiae*, aa 57 to 372 (317 aa); (H) MenA from *E. coli* (308 aa). The plots were generated by using the PEPPLOT (17) program and the complete amino acid sequences from all proteins, except the chlorophyll synthase from *A. thaliana* and Coq2 from *S. cerevisiae*. The predictions for these two proteins do not include the proposed transit peptide for each protein or the serine-rich domain in Coq2. The plots show the Kyte and Doolittle (20) (thin lines) and Engelman et al. (14) (thick lines) hydrophobicity predictions. Positive values are hydrophobic, and negative values are hydrophilic.

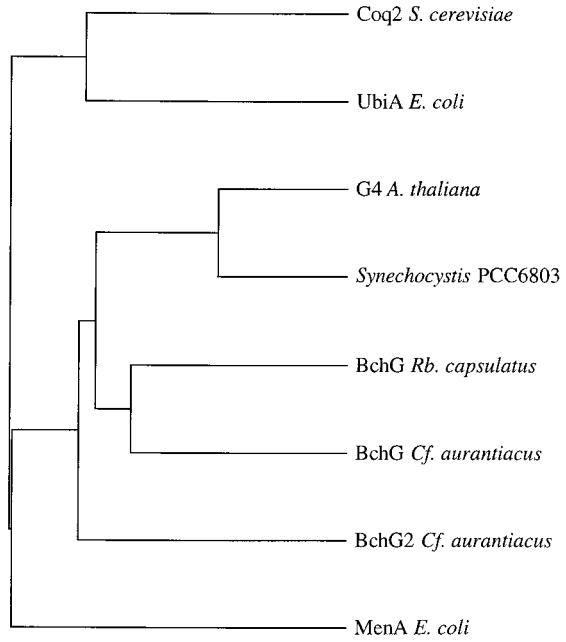

FIG. 3. PILEUP (17) dendrogram of the polyprenyltransferases identified in the legend to Fig. 2.

Chloroflexus aurantiacus is esterified with phytol on bacteriochlorophyll *a* and predominantly with stearol on bacteriochlorophyll *c*. However, there is a low substrate specificity for the esterifying alcohol attached to bacteriochlorophyll *c* in *C. aurantiacus* (15, 21).

C. aurantiacus occupies a unique evolutionary position among photosynthetic bacteria. 16S rRNA sequence comparisons suggest that this organism diverged early from other photosynthetic organisms (34). However, a comparison of pheophytin-quinone-type reaction centers places *C. aurantiacus* much closer to purple bacteria (6, 8). One suggested explanation for this discrepancy involves the lateral transfer of genes involved in photosynthesis. Even within the photosynthetic apparatus, the evolutionary history of *C. aurantiacus* is unclear. While the peripheral chlorosome antenna system is similar to that of green sulfur bacteria, the reaction center and integral membrane light-harvesting antenna complex (LHI) resemble those of the purple photosynthetic bacteria (3, 9, 35). Therefore, an analysis of photosynthetic genes from this organism offers a unique view into the evolution of photosynthesis. We have cloned and sequenced a gene from *C. aurantiacus* that is highly homologous to *bchG* in *R. capsulatus.*

Sequencing of the *C. aurantiacus bchG* gene. The *bchG* gene was isolated as a second open reading frame (*orf1*) discovered during the cloning and sequencing of the cytochrome *c*-554 gene in *C. aurantiacus* (13). *orf1* was partially sequenced but showed no homology to any sequence in the database at that time.

The *C. aurantiacus bchG* gene was sequenced by using a Sequenase version 2.0 sequencing kit (Amersham) according to the manufacturer's protocol for double-stranded sequencing. *Escherichia coli* DH5α containing the previously described pCEN12 plasmid (13) was grown overnight in Luria broth plus ampicillin in 10-ml cultures. Plasmid DNA was isolated by an alkaline lysis procedure and then purified through differential precipitation in a solution of 0.5% cetyltrimethylammonium bromide (Sigma Chemical Co., St. Louis, Mo.) (12). Portions of the final sequence were confirmed with a Perkin-Elmer/Applied Biosystems (Foster City, Calif.) 377 DNA sequencer by using a PRISM DyeDeoxy terminator cycle sequencing kit (Perkin-Elmer/Applied Biosystems). For the plasmids isolated, a QIAGEN-tip 20 (QIAGEN Inc., Chatsworth, Calif.) minicolumn was used according to the manufacturer's instructions.

Figure 1 shows the nucleotide sequence and the predicted amino acid sequence of *bchG* of *C. aurantiacus*. The sequenced region contains an open reading frame of 933 nucleotides, generating a protein with a predicted molecular mass of 33.6 kDa. The predicted amino acid sequence has a high degree of identity with *R. capsulatus* BchG. An alignment was generated by using the GAP program from the Genetics Computer Group analysis package (17) with the default settings. This alignment has a quality score of 243.7 (average random score, 101.8 ± 4.5), a percent similarity of 69.5%, and a percent

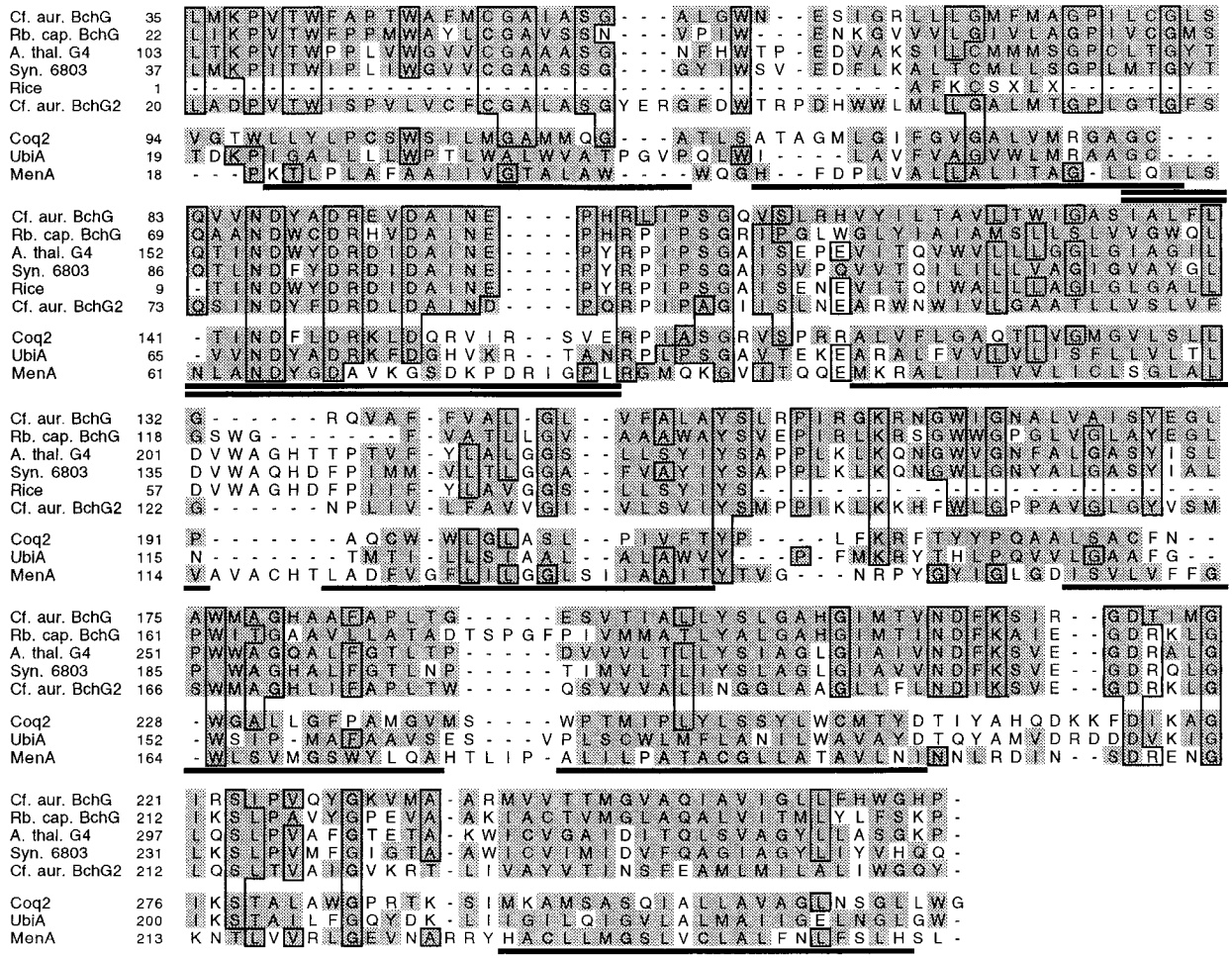

FIG. 4. Amino acid sequence alignment of BchG proteins from *C. aurantiacus* (Cf. aur.) and *R. capsulatus* (Rb. cap.), G4 from *A. thaliana* (A. thal.), chlorophyll synthase from *Synechocystis* sp. strain PCC6803 (Syn. 6803), a partial sequence of chlorophyll synthase from rice, BchG2 from *C. aurantiacus*, UbiA from *E. coli*, Coq2 from *S. cerevisiae*, and MenA from *E. coli*. The alignment was generated by using CLUSTALW 1.5 (33) from the *BCM Search Launcher* WWW site and SeqVu, a Macintosh shareware program from the Garvan Institute of Medical Research (Sydney, Australia). Regions of homology are shaded in gray, and regions of identity are boxed. Gaps in the alignment are indicated by dashes. There is a single black bar under each region containing a potential transmembrane helix. Helix predictions were made by using the hydrophobic predictions from Fig. 2, as well as the TMAP (24) and TMpred (19) WWW sites. Only seven helices are shown; the eighth is located in the C-terminal region not included in this figure. There is a double bar under the region homologous to polyprenyltransferase domain II.

identity of 46.6% (data not shown). There are several areas in this alignment that show especially high levels of identity. Amino acids 33 to 53, 90 to 106, and 194 to 216 from *C. aurantiacus* have percent identities of 61.9, 88.2, and 78.3%, respectively, with their aligned counterparts in *R. capsulatus*. These regions are underlined in Fig. 1.

Sequence comparisons. Database searches using the highly homologous regions of the *C. aurantiacus* BchG sequence were conducted. The searches were done with the BLASTP and TBLASTN programs of the BLAST server (2). These searches identified three proteins. The three proteins are a putative chlorophyll synthase (G4) from *Arabidopsis thaliana* (GenBank accession no. U19382) (16), a hypothetical protein translation from the genomic sequence of *Synechocystis* sp. strain PCC6803 (DDBJ accession no. D64001), and the product from an open reading frame upstream of the *csmM* and *csmN* genes of *C. aurantiacus* (EMBL accession no. Z34000) (23). All three have identities to the *R. capsulatus bchG* gene. Two partial cDNA-expressed sequence tags, one from *A. thaliana* (EMBL accession no. Z34566) and the other from rice (DDBJ accession no. D48639), were also identified. The *A. thaliana* sequence is identical to that of the G4 synthase gene.

The five protein sequences were aligned with the PILEUP program of the Genetics Computer Group package (17). The multiple sequence file produced was run through the PROFILEMAKE and PROFILESEARCH programs (17, 18). Three related protein sequences, MenA (Swiss-Prot accession no. P26601) and UbiA (Swiss-Prot accession no. P32166) from *E. coli* and Coq2 (Swiss-Prot accession no. P32378) from *Saccharomyces cerevisiae*, were identified. UbiA and Coq2 are homologs involved in ubiquinone biosynthesis (5, 22, 32), and MenA is involved in menaquinone biosynthesis (31). These enzymes are polyprenyltransferases that attach an aliphatic alcohol PP_i to an aromatic substrate. This suggests a broader relationship between chlorophyll synthases and other polyprenyltransferases.

Comparisons of the hydrophobic predictions of these amino acid sequences (Fig. 2) strongly suggest that all of these proteins are integral membrane proteins. The plots show an es-

sentially identical pattern with eight putative membrane-spanning regions.

The PILEUP (17) multiple sequence alignment of these eight proteins produces the dendrogram shown in Fig. 3. The dendrogram shows the pairwise clustering relationships of these eight proteins. The BchG sequences from *R. capsulatus* and *C. aurantiacus* and the chlorophyll synthases from *A. thaliana* and *Synechocystis* sp. strain PCC6803 cluster together, with BchG2 as the next nearest neighbor. This is interesting for two reasons. First, *C. aurantiacus* contains two bacteriochlorophylls with different esterifying alcohols. Second, BchG2 is clustered with two genes that code for proteins associated with the peripheral antenna system (and therefore bacteriochlorophyll *c*) and BchG is located near genes coding for proteins associated with the bacterial reaction center and core antenna (and therefore bacteriochlorophyll *a*). This suggests that *bchG* codes for the synthase that produces bacteriochlorophyll *a* and that the *bchG2* gene product produces bacteriochlorophyll *c*.

Figure 4 shows the multiple sequence alignment but does not contain the C- and N-terminal ends. The partial protein sequence from rice has also been included in this alignment. In this alignment, regions of homology are shaded and regions of identity are boxed.

Sequence comparisons of many polyprenyl synthase enzymes revealed three conserved domains (4, 11). Of these three, domain II (region above double bar in Fig. 4) has been found in a broad range of proteins that catalyze the condensation of a polyprenyl group to a variety of substrates. Domain II has been suggested to be the binding site of the polyprenyl PP_i (5, 22, 32). There are many regions throughout the sequence alignment that show surprising degrees of identity, considering the wide range of organisms and functions represented. There are 10 absolutely conserved residues; half of them are in or near domain II. Greater identity is shown only if the chlorophyll synthase proteins are compared. There are 45 identical residues; 15 are in or near domain II. This region also contains the DDXXD motif proposed to be responsible for the binding of the divalent cations (Mg^{2+} or Mn^{2+}) required for the catalytic activities of polyprenyltransferases (4, 11). Our sequence alignment suggests that this motif is actually DRXXD. The requirement of chlorophyll synthase for divalent cations has not been investigated. The conserved regions may be useful in identifying chlorophyll synthase genes from other organisms through PCR or as targets in site-directed mutagenesis experiments.

In *R. capsulatus*, the proteins involved in the biosynthetic pathway of bacteriochlorophyll *a* are encoded by genes located in a 45-kb region of the chromosome called the photosynthesis gene cluster (1). This cluster is bordered on one side by the L and M subunits of the reaction center and the structural proteins for the LHI antenna complex and on the other by the H subunit of the reaction center. *bchG* lies approximately halfway through the cluster. The positioning of *bchG* is very different in *C. aurantiacus. bchG* begins directly downstream of the genes encoding cytochrome *c*-554 and the structural proteins of the core antenna. The genes for the *C. aurantiacus* reaction center L and M subunits are a significant distance away (greater than 3 kb) (30). There is no H subunit in the *C. aurantiacus* reaction center, just as there is no tetraheme cytochrome (cytochrome *c*-554) in the *R. capsulatus* reaction center. Further studies may be very revealing, since most of the information collected from studies of other photosynthetic bacteria suggests a clustering of the photosynthetic genes, as seen in *R. capsulatus* (1, 7, 30).

Nucleotide sequence accession number. The sequence data presented here has been submitted to the GenBank database and assigned accession number U43963.

This work was supported by a grant from the Exobiology Program of NASA.

REFERENCES

1. **Alberti, M., D. H. Burke, and J. E. Hearst.** 1995. Structure and sequence of the photosynthesis gene cluster, p. 1083–1106. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), Anoxygenic photosynthetic bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Amesz, J.** 1991. Green photosynthetic bacteria and heliobacteria, p. 99–119. *In* J. M. Shively and L. L. Barton (ed.), Variations in autotrophic life. Academic Press, London.
4. **Ashby, M. N., and P. A. Edwards.** 1990. Elucidation of the deficiency in two yeast coenzyme Q mutants: characterization of the structural gene encoding hexaprenyl pyrophosphate synthetase. J. Biol. Chem. **265:**13157–13164.
5. **Ashby, M. N., S. Y. Kutsunai, S. Ackerman, A. Tzagoloff, and P. A. Edwards.** 1992. *COQ2* is a candidate for the structural gene encoding *para*-hydroxybenzoate:polyprenyltransferase. J. Biol. Chem. **267:**4128–4136.
6. **Beanland, T. J.** 1990. Evolutionary relationships between "Q-type" photosynthetic reaction centres: hypothesis-testing using parsimony. J. Theor. Biol. **145:**535–545.
7. **Beatty, J. T.** 1995. Organization of photosynthesis gene transcripts, p. 1209–1219. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), Anoxygenic photosynthetic bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.
8. **Blankenship, R. E.** 1992. Origin and early evolution of photosynthesis. Photosynth. Res. **33:**91–111.
9. **Blankenship, R. E., J. M. Olson, and M. Miller.** 1995. Antenna complexes from green photosynthetic bacteria, p. 399–435. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), Anoxygenic photosynthetic bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.
10. **Bollivar, D. W., S. Wang, J. P. Allen, and C. E. Bauer.** 1994. Molecular genetic analysis of terminal steps in bacteriochlorophyll *a* biosynthesis: characterization of a *Rhodobacter capsulatus* strain that synthesizes geranylgeraniol-esterified bacteriochlorophyll *a*. Biochemistry **33:**12763–12768.
11. **Carattoli, A., N. Romano, P. Ballario, G. Morelli, and G. Macino.** 1991. The *Neurospora crassa* carotenoid biosynthetic gene (Albino 3) reveals highly conserved regions among prenyltransferases. J. Biol. Chem. **266:**5854–5859.
12. **Del Sal, G., G. Manfioletti, and C. Schneider.** 1989. The CTAB-DNA precipitation method: a common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. BioTechniques **7:**514–519.
13. **Dracheva, S., J. C. Williams, G. Van Driessche, J. J. Van Beeumen, and R. E. Blankenship.** 1991. The primary structure of cytochrome *c*-554 from the green photosynthetic bacterium *Chloroflexus aurantiacus*. Biochemistry **30:** 11451–11458.
14. **Engelman, D. M., T. A. Steitz, and A. Goldman.** 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. **15:**321–353.
15. **Fages, F. N., N. Greibenow, K. Greibenow, A. R. Holzworth, and K. Schaffer.** 1980. Characterization of light-harvesting pigments of *Chloroflexus aurantiacus*: two new chlorophylls: oleyl (octadec-9-enyl) and cetyl (hexadecanyl) bacteriochlorophyllides-*c*. J. Chem. Soc. Perkin Trans. I **1990:**2791–2797.
16. **Gaubier, P., H.-J. Wu, M. Laudié, M. Delseny, and F. Grellet.** 1995. A chlorophyll synthetase gene from *Arabidopsis thaliana*. Mol. Gen. Genet. **249:**58–64.
17. **Genetics Computer Group.** 1995. Program manual for the Wisconsin package, version 8. Genetics Computer Group, Madison, Wis.
18. **Gribskov, M., A. D. McLachlan, and D. Eisenberg.** 1987. Profile analysis: detection of distantly related proteins. Proc. Natl. Acad. Sci. USA **84:**4355–4358.
19. **Hofmann, K., and W. Stoffel.** 1993. TMbase—a database of membrane spanning protein segments. Biol. Chem. Hoppe-Seyler **347:**166. [WWW site *TMpred-Prediction of Transmembrane Regions and Orientation* (http://ulrec3.unil.ch/software/TMPRED_form.html) from the Bioinformatics Group at the ISREC (Swiss Institute for Experimental Cancer Research) in Epalinges s/Lausanne.]
20. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
21. **Larsen, K. L., M. Miller, and R. P. Cox.** 1995. Incorporation of exogenous long-chain alcohols into bacteriochlorophyll *c* homologs by *Chloroflexus aurantiacus*. Arch. Microbiol. **163:**119–123.
22. **Melzer, M., and L. Heide.** 1994. Characterization of polyprenyldiphosphate: 4-hydroxybenzoate polyprenyltransferase from *Escherichia coli*. Biochim. Biophys. Acta **1212:**93–102.
23. **Niedermeier, G., J. A. Shiozawa, F. Lottspeich, and R. G. Feick.** 1994. The primary structure of two chlorosome proteins from *Chloroflexus aurantiacus*. FEBS Lett. **342:**61–65.
24. **Persson, B., and P. Argos.** 1994. Prediction of transmembrane segments in proteins utilizing multiple sequence alignments. J. Mol. Biol. **237:**182–192. [WWW site *TMAP Request form for multiple aligned sequences* (http://www

.embl-heidelberg.de/tmap/tmap_mul.html) from the European Molecular Biology Laboratory, Heidelberg, Germany.]
25. **Rüdiger, W.** 1987. Chlorophyll synthetase and its implication for regulation of chlorophyll biosynthesis, p. 461–467. *In* J. Biggins (ed.), Progress in photosynthesis research, vol. 4. Martinus Nijhoff Publishers, Dordrecht, The Netherlands.
26. **Rüdiger, W., J. Benz, U. Lempert, S. Schoch, and D. Steffens.** 1976. Inhibition of phytol accumulation with herbicides: geranylgeraniol and dihydrogeranylgeraniol-containing chlorophyll from wheat seedlings. Z. Pflanzenphysiol. **80:**131–143.
27. **Rüdiger, W., and S. Schoch.** 1991. The last steps of chlorophyll biosynthesis, p. 451–464. *In* H. Scheer (ed.), Chlorophylls. CRC Press, Boca Raton, Fla.
28. **Schoch, S., U. Lempert, and W. Rüdiger.** 1977. On the last steps of chlorophyll biosynthesis: intermediates between chlorophyllide and phytol-containing chlorophyll. Z. Pflanzenphysiol. **83:**427–436.
29. **Senge, S., and K. M. Smith.** 1995. Biosynthesis and structures of the bacteriochlorophylls, p. 137–151. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), Anoxygenic photosynthetic bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.
30. **Shiozawa, J. A.** 1995. A foundation for the analysis of green sulfur, green filamentous and heliobacteria, p. 1159–1173. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), Anoxygenic photosynthetic bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.
31. **Suvarna, K., R. Meganathan, and M. E. S. Hudspeth.** 1994. Menaquinone (vitamin K_2) biosynthesis: cloning and nucleotide sequence of the *menA* gene from *Escherichia coli*, abstr. K-159, p. 303. *In* Abstracts of the 94th General Meeting of the American Society for Microbiology 1994. American Society for Microbiology, Washington, D.C.
32. **Suziki, K., M. Ueda, T. Nakagawa, M. Kawamukai, and H. Matsuda.** 1994. Evidence that *Escherichia coli ubiA* product is a functional homolog of yeast COQ2, and the regulation of *ubiA* gene expression. Biosci. Biotechnol. Biochem. **58:**1814–1819.
33. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680. [WWW site *BCM Search Launcher: Multiple Sequence Alignments* (http://dot.imgen.bcm.tmc.edu:9331/multi-align/multi-align.html) at the Human Genome Center, Baylor College of Medicine, Houston, Tex.]
34. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
35. **Zuber, H., R. Brunisholz, and W. Sidler.** 1987. Structure and function of light-harvesting pigment-protein complexes, p. 233–271. *In* J. Amesz (ed.), Photosynthesis. Elsevier Science Publishers B. V. (Biomedical Division), Amsterdam.