J. Mol. Biol. (1986) **188**, 443–454

Structure and X-ray Amino Acid Sequence of a Bacteriochlorophyll *a* Protein from *Prosthecochloris aestuarii* Refined at 1·9 Å Resolution

D. E. Tronrud, M. F. Schmid† and B. W. Matthews

Departments of Physics and Chemistry and Institute of Molecular Biology
University of Oregon
Eugene, OR 97403, U.S.A.

(Received 21 August 1985, and in revised form 14 November 1985)

The structure of the water-soluble bacteriochlorophyll *a* protein (Bchl protein) from the green photosynthetic bacterium *Prosthecochloris aestuarii* has been refined at 1·9 Å resolution to a crystallographic residual of 18·9%. The refinement was carried out without knowledge of the amino acid sequence and has led to an "X-ray sequence". The structure consists of seven Bchl molecules enclosed within a protein "bag" and the refinement supports the general conformation of the molecule described previously. The refinement also supports the previous suggestion that the ligands to the seven Bchl magnesiums are, respectively, five histidines, a carbonyl oxygen from the polypeptide backbone of the protein, and a bound water molecule.

The conformations of the seven Bchl head-groups are described in detail. In two cases the magnesium atoms are approximately 0·48 Å "below" the plane of the conjugated macrocycle while in the other five cases the atoms are, on average, 0·48 Å "above" the plane. The acetyl ring substituents are more-or-less coplanar with the dihydrophorbin macrocycle, consistent with a previous resonance Raman study. The conjugated atoms in each of the seven macrocycles have significant departures from strict planarity. These deviations are similar for Bchls 1, 2 and 3 (class I) and are also similar for Bchls 4, 5, 6 and 7 (class II). Ethylchlorophillide also belongs to class II. The out-of-plane deformations for the class I and class II bacteriochlorophylls appear to correspond to two distinct modes of bending or curvature of the dihydrophorbin macrocycle.

1. Introduction

The light-gathering apparatus of the green photosynthetic bacteria consists of three chlorophyll-containing entities. First, there are the "chlorophyll bodies", which reside beneath the cytoplasmic membrane. These incorporate 95% of the total chlorophyll of the organism and constitute the principal light-gathering pigment. Second, there is the "bacteriochlorophyll *a* protein", which includes most of the remaining chlorophyll, and resides between the chlorophyll bodies and the cytoplasmic membrane. Third, there are the reaction centers. The bacteriochlorophyll *a* protein is thought to transmit excitation energy to a special pair of chlorophyll molecules in the reaction center where the first step in energy transduction, namely a charge separation, occurs.

The bacteriochlorophyll *a* protein of *Prosthecochloris aestuarii* (Bchl protein), first isolated and crystallized by Olson (1971, 1978), has been subject to crystallographic analysis, and the structure determined to a nominal resolution of 2·8 Å (Fenna & Matthews, 1975; Matthews *et al.*, 1979). The molecule, of molecular weight 150,000, consists of three identical subunits related by 120° rotations about a 3-fold axis of symmetry. In each subunit the polypeptide backbone forms a large, twisted β-sheet of 16 strands that forms the "outside" of the protein molecule (i.e. the part exposed to solvent), and encloses a central core of seven bacteriochlorophyll *a* molecules (Fig. 1). The available evidence suggests that the trimeric structure seen in the crystals occurs *in vivo* and is not an artifact of crystallization.

† Present address: Department of Biochemistry, University of Arizona, Tucson, AZ 85721, U.S.A.

0022-2836/86/070443-12 $03.00/0

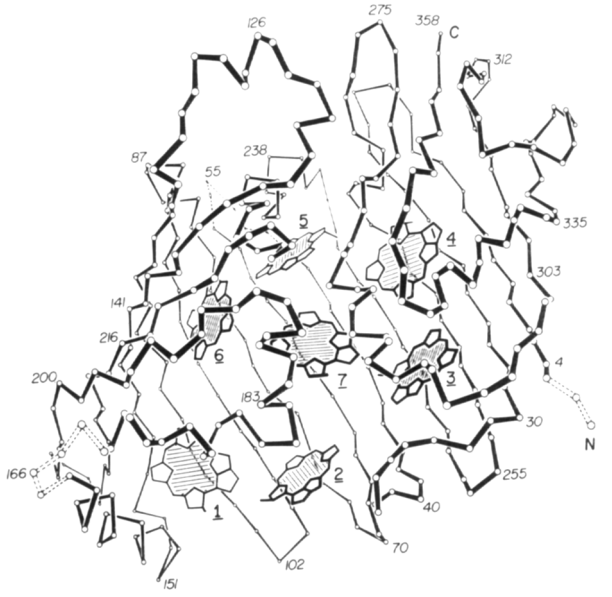

Figure 1. Simplified view of one subunit of the Bchl protein, viewed from the center of the trimer, with the 3-fold axis horizontal. For clarity the phytyl chains and other bacteriochlorophyll ring substituents have been omitted. The numbering of the amino acid sequence corresponds to that used in the text, although it should be noted that there are uncertainties in the exact number of residues at the N and C termini and at the gaps in the electron density between residues 51–57, 164–169 and 205–208, marked in the Figure by broken connections. (After Fenna *et al.* (1977).)

Several different aspects of this structure are of interest. The interactions between the phytyl chains of the bacteriochlorophyll *a* molecules and the amino acid side-chains provide a good model for lipid–protein interactions. Also, the packing of the bacteriochlorophyll *a* molecules gives insight into other protein–chlorophyll interactions. In particular, it will be of interest to compare the chlorophyll–chlorophyll and chlorophyll–protein interactions in the Bchl protein with the recently determined structure of the reaction center from *Rhodopseudomonas viridis* (Deisenhofer *et al.*, 1984). Theoretical calculations of the absorption and circular dichroism spectra of the bacteriochlorophyll core of this protein have been attempted but the agreement between the observed and calculated spectra is poor (Perlstein & Heminger, 1978).

In this paper we describe the refinement of the Bchl protein structure at a resolution of 1·9 Å. The refinement was carried out without knowledge of the amino acid sequence and has led to an "X-ray sequence". The X-ray sequence at an intermediate stage of refinement has been reported by Schmid *et al.* (1983). In this paper we report the X-ray sequence and describe those aspects of the structure that do not depend critically on the correct identification of individual amino acids, i.e. the backbone conformation and the configuration of the bacteriochlorophylls.

2. Methods

(a) *Data collection*

Crystals of the bacteriochlorophyll *a* protein have space group $P6_3$ with cell dimensions $a = b = 111{\cdot}9$ Å, $c = 98{\cdot}6$ Å (Fenna *et al.*, 1974; Matthews *et al.*, 1979). In the initial X-ray analysis, precession photography was used to obtain an electron density map at 2·8 Å resolution. This resolution is essentially the highest that could be achieved using conventional precession photographs, a limitation that can be attributed in part to the relatively large volume of the unit cell ($1{\cdot}08 \times 10^6$ Å^3) and in part to the high (70%) solvent content of the crystals. By using oscillation photography and monochromatic radiation (Schmid *et al.*, 1981) it has been possible to extend the resolution of the data to 1·9 Å. A total of 9 crystals was used to obtain 76 oscillation photographs. These were processed using a program based on that of Rossmann (1979). A total of 126,000 reflections was measured and reduced to a unique set of 46,700 intensities. The average agreement between symmetry-related intensities on the same film was 4·6% and the overall agreement between reflections measured on different films was 11·6%. (The latter value is much higher than the former partly because the reflections come from different crystals in different orientations, but also because the number of measurements per reflection is larger.) For the 15,000 reflections common to the precession and oscillation data sets the agreement was 9·8%.

(b) *Starting co-ordinates*

A model of the structure was built in an optical comparator (Richards, 1968; Colman *et al.*, 1972). Raw co-ordinates were then obtained by placing markers in the electron density map to correspond with the model and the density. These co-ordinates were then adjusted to have acceptable stereochemistry and were the starting point for the structure refinement. The initial crystallographic *R*-value was 44·2% for data between 6·0 Å and 1·9 Å resolution.

(c) *Refinement*

The protein was refined using a new, fast, flexible, restrained least-squares refinement package developed in collaboration with L. F. TenEyck (unpublished results). The program restrains bond lengths and bond angles to be within acceptable limits. Also, groups of atoms known to be planar (e.g. aromatic side-chains, carboxyl, peptide and amide groups) are likewise restrained. Distances between non-bonded atoms within the protein, between neighboring molecules in the crystal, and between protein and solvent atoms are routinely checked and can be restrained if desired.

Although it was expected that the 22 conjugated atoms in each of the dihydrophorbin macrocycles would be essentially coplanar, it was still necessary to allow for the possibility that the Bchl head-groups might be distorted in one way or another. Therefore, instead of constraining all the conjugated atoms to be near a single plane, we divided the atoms into sets of smaller subplanes that had common overlapping atoms but were otherwise independent. This procedure ensured that planarity was maintained locally but permitted larger-scale deviations. The subgroups of atoms were as follows: (C1C, CHC, C4B, C3B, C2B, C1B, CHB, C4A, NA, C1A, CHA, C4D), (C1A, CHA, C4D, CBD, C2D, C1D, CHD, C4C, NC, C1C, CHC,

C4B), (NB, C1B, C2B, C3B, C4B), (ND, C1D, C2D, C3D, C4D) and (CAD, OBD, CBD, C3D).

Because the bacteriochlorophyll *a* protein was used as a test case for development of the refinement package the refinement proceeded in a somewhat irregular manner. Also, in the early stages a graphics system was not available and the model was checked by direct inspection of plotted contour maps. By these methods the crystallographic residual was reduced to about 23%. In the latter stages an MMS-X graphics system (Molnar *et al.*, 1976) running the program FRODO (Jones, 1982) was used to facilitate the inspection of both $(2F_o-F_c)$ and (F_o-F_c) maps. Also $(2F_o-F_c)$ "omit" maps were calculated in which 10 amino acids at a time were omitted from the phase calculation and their density checked for consistency with the model. (Our experience was that such omit maps are not obviously superior to conventional difference maps.) We adopted the conservative policy of deleting atoms from the model rather than having them present in doubtful positions. Also we only permitted side-chains that corresponded to the standard amino acids. Thermal factors were refined in cycles independent of co-ordinate refinement. The presence of atoms with large thermal factors was taken as an indication that the identification of the amino acid in question should be changed. Solvent atoms with thermal factors greater than 50 Å^2 were deleted.

As additional checks, the refinement program provides a list of the worst occurrences of each type of stereochemistry including bond lengths, bond angles, torsion angles, inverted chiral centers and, if desired, a list of the atoms with the largest crystallographic gradients. These lists were routinely used to check for possible errors in the model.

Altogether, 5 rounds of manual model revision and automated refinement were completed before convergence was achieved. The final R-value is 18·9% to 1·9 Å resolution. In the final model the bond lengths and angles depart from their ideal values by root mean square values of 0·02 Å and 3·2°, respectively. The overall shifts in the polypeptide backbone during refinement are shown in Fig. 2. The refined co-ordinates have been deposited in the Brookhaven Data Bank.

3. Results and Conclusions

(a) *Amino acid sequence*

There are several difficulties in attempting to determine an amino acid sequence from crystallographic data. In other instances it has proved possible to identify correctly about 50 to 75% of the amino acids by X-ray methods (e.g. see Matthews, 1977). Several pairs of amino acids are chemically different but isostructural, namely (Val, Thr), (Glu, Gln) and (Asp, Asn). Members of these pairs are virtually impossible to differentiate in the electron density map. In addition, some amino acid side-chains, especially on the surface of the protein, are mobile and hard to identify. In extreme cases the backbone atoms also undergo large "thermal" displacements and cannot be seen. Therefore, we have given a confidence rating, ranging from 1 to 4, for each amino acid identification (Table 1), Very badly determined residues were given a rating of 4. An amino acid in this category did not fit the existing electron density well, had questionable stereochemistry and its atoms usually had high B values; 36 amino acids were determined as belonging in this category. If a residue did not agree with the electron density but the disagreement was not so bad as to eliminate the possibility of the identity being correct it was rated as 3; 48 amino acids fell into this class. If an amino acid had some disagreement with the electron density but was still considered likely to be correct it was rated 2; 42 amino acids fit this description. The final class (rating 1) was reserved for amino acids that agreed well with all the criteria; 218 amino acids were so classified.

The sequence confidence rating scheme is unquestionably subjective. In the case of the isostructural pair, Thr and Val, the confidence rating does not take into account the ambiguity between these residues, i.e. if a residue appeared to be a good Val or Thr it was considered to be well determined even though the two alternatives could not be reliably distinguished. In such cases an attempt was made to differentiate between the two alternatives by consideration of the local environment, but this is, at best, an indirect assessment. As far as possible we kept the numbering of the amino acid sequence in register with that inferred from the original isomorphous replacement electron-density map. Therefore, no particular significance should be placed on the exact numbering of residues in the "gaps" or at the amino terminus. For example, the X-ray sequence has a gap between residues 51 and 57, but this is not meant to imply that there are thought to be exactly five intervening residues. Also there is no residue corresponding to sequence number 83. Compared to the preliminary X-ray amino acid sequence obtained from inspection of

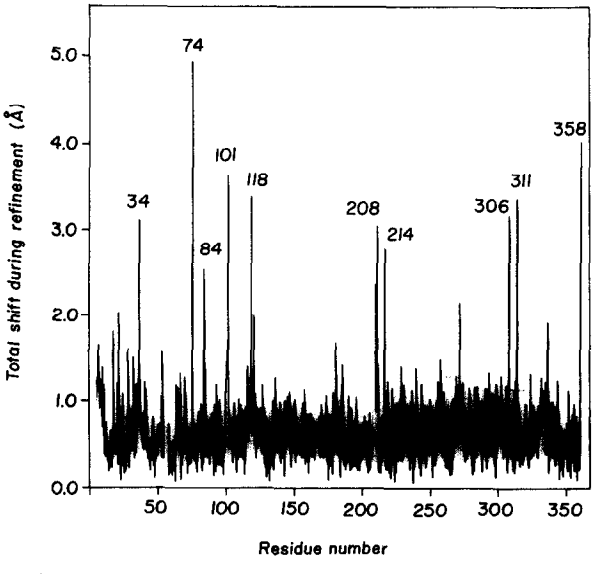

Figure 2. Co-ordinate shifts of the backbone atoms of the Bchl protein during refinement. The graph shows, for each backbone atom (N, CA, C, O), the difference between the position at the start of refinement and in the final model. Atoms that have undergone the largest shifts are numbered by their residue number.

Table 1

X-ray amino acid sequence

								1	10									20
-	-	-	-	-	-	Val	Ser	Ala	Asx	Ser	Ala	Tyr	Lys	Ile	Ile	Leu	Glx	Gly	Gly	Ala	Ser	Ser	Trp
						2	3	1	4	1	4	1	4	1	1	1	3	1	2	4	1	1	1
21													30										40
Gly	Glx	Val	Ala	Gly	Ala	Ala	Ser	Val	Ser	Val	Pro	Ala	Ser	Ile	Pro	Leu	Asx	Pro	Thr
1	3	1	3	1	3	2	3	1	4	1	1	1	3	1	1	1	3	1	1
41											50												60
Asx	Cys	Ser	Ile	Lys	Ile	Asx	Ala	Ser	Pro	Ser	-	-	-	-	-	-	-	-	-	-	Thr	Val	Lys	Phe
3	1	3	1	2	1	1	1	4	1	4											4	1	1	1
61											70												80
Thr	Val	Ala	Ile	Ala	Ser	Thr	Ile	Asx	Ala	Thr	Ala	Asx	Thr	Leu	Ser	Val	Ala	Thr	Ser
4	1	3	1	4	1	1	3	1	3	1	2	1	2	3	4	1	4	2	3
81											90												100
Ile	Ala		Asx	Ser	Ala	Ala	Ser	Lys	Arg	Ile	Ala	Thr	Gly	Ala	Gly	Ser	Val	Ala	Val
1	1		1	4	3	3	3	1	1	1	1	1	1	4	1	2	4	3	1
101											110												120
Gly	Ser	Phe	Ala	His	Ala	Phe	Ser	Phe	Met	Gly	Ser	Thr	Thr	Asx	Met	Tyr	Tyr	Ser	Gly
1	4	1	3	1	1	2	1	2	3	1	2	1	1	1	4	1	1	4	4
121											130												140
Ser	Ser	Ala	Thr	Ala	Arg	Asx	Ile	Pro	Asx	Pro	Ile	Tyr	Met	Glx	Gly	Arg	Glx	Phe	His
4	4	1	1	2	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1
141											150												160
Asx	Ile	Asx	Met	Lys	Val	Pro	Leu	Asx	Asx	Gly	Asx	Leu	Thr	Ser	Thr	Trp	Lys	Gly	Phe
1	1	1	1	1	1	1	1	3	2	3	2	1	1	4	1	1	3	1	2
161											170												180
Ser	Ala	Ala	Ser	-	-	-	-	-	-	-	-	Asx	Phe	Gly	Asx	Trp	Ile	Arg	Asx	Phe	Trp	Phe	Ile
4	4	1	4									1	1	1	1	1	1	1	3	1	1	1	1
181											190												200
Gly	Pro	Ala	Phe	Ala	Ala	Ile	Asx	Glx	Gly	Gly	Glx	Arg	Ile	Ser	Pro	Val	Thr	Thr	Asx
1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	3	1	2	1
201											210												220
Ser	Ala	Ser	Thr	Glx	-	-	-	-	Gly	Pro	Asx	Gly	Thr	Thr	Arg	Trp	Ser	Phe	Ser	His	Ala
2	2	3	2	4					4	3	4	1	1	1	1	1	4	1	2	1	1
221											230												240
Gly	Ser	Gly	Val	Val	Asx	Ser	Ile	Ser	Arg	Trp	Thr	Glx	Leu	Phe	Pro	Thr	Ala	Lys	Leu
1	1	1	2	1	1	4	1	2	1	1	1	1	1	1	1	2	4	1	1
241											250												260
Ser	Lys	Pro	Ala	Ala	Ile	Glx	Gly	Gly	Phe	Ser	Ser	Asx	Ser	Ala	Gly	Ile	Ser	Val	Ala
3	1	1	1	2	1	4	1	1	1	3	1	2	2	2	1	1	4	1	1
261											270												280
Val	Ala	Gly	Ser	Leu	Pro	Gly	Val	Ser	Lys	Ser	Ala	Gly	Gly	Gly	Asx	Lys	Lys	Ile	Leu
1	4	1	3	1	1	1	1	1	2	3	2	3	3	3	2	2	1	1	1
281											290												300
Asx	His	Pro	Asx	Ile	Pro	Leu	Thr	His	His	Gly	Met	Thr	Gly	Lys	Phe	Asx	Ser	Phe	Ser
1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	3
301											310												320
Ser	Asx	Thr	Ala	Asx	Lys	Ile	Thr	Leu	Pro	Lys	Gly	Tyr	Ala	Ile	Ser	Tyr	Ala	Ala	Pro
3	1	1	3	2	1	1	1	1	1	3	1	1	2	2	3	1	1	1	1
321											330												340
Ala	His	Ser	Ser	Lys	Asx	Asx	Glx	Ala	Tyr	Lys	Trp	Ala	Gly	Gly	Ala	Tyr	Ala	Arg	Trp
2	2	3	2	3	2	1	2	4	1	1	1	2	1	2	3	1	1	1	1
341											350							358					
Val	Glx	His	Val	Cys	Lys	Gly	Gly	Thr	Gly	Glx	His	Glx	Thr	Leu	Tyr	Ala	Ala						
1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	3	4						

The amino acid sequence is inferred from the crystallographic refinement. It is not possible to distinguish Thr and Val directly. The identifications given here are inferred indirectly from the environment. Numbers under each amino acid indicate the level of confidence in the identification (1, most confident; 4, least confident). The exact number of amino acids at the N and C termini and in the gaps is uncertain.

Table 2
Amino acid composition

Amino acid	Experimental†	X-ray model	Discrepancy
Gly	41	36	−5
Asx	40	31	−9
Val	34–35	18	−16–17
Glx	34	13	−21
Ser	28	47	19
Ile	24	23	−1
Ala	22	48	26
Leu	20	12	−8
Arg	20	8	−12
Pro	18	17	−1
Lys	18	17	−1
Phe	17	15	−2
Thr	14	26	12
Tyr	10	9	−1
His	8	9	1
Trp	7	8	1
Met	3–4	5	1–2
Cys	2	2	0
Total	360	344	−16

† From Olson *et al.* (1976), scaled so that there is a total of 360 amino acids per polypeptide.

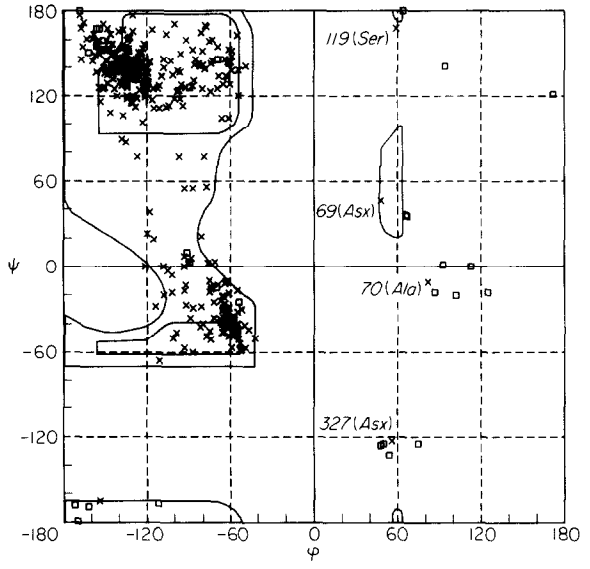

Figure 3. Ramachandran diagram for the Bchl protein. (□) Presumed glycine residues; (×) non-glycine residues.

the isomorphous replacement electron density map (Matthews *et al.*, 1979), 179 amino acids remain the same and 165 were altered during refinement.

(b) *Amino acid composition*

The amino acid composition of the Bchl *a* protein is known (Olson *et al.*, 1976) and is compared with the amino acid composition of the X-ray model in Table 2. The X-ray model contains 344 residues and is presumed to be missing several additional residues (e.g. at gaps in the electron density and possibly at the ends of the polypeptide chain). Therefore, the amino acid composition determined chemically and given in Table 2 has been put on a scale so that it corresponds to a total composition of 360 residues.

Table 2 indicates that there are too few glycine residues in the X-ray model. This is somewhat surprising since one tends to put glycine in regions of the model that are poorly determined. It is possible that there are additional glycine residues, not included in the model, that are in flexible bend or terminal regions. It is not surprising that the model has too many serine and alanine residues since these could be surface residues for which the distal part of the side-chain is mobile and not seen in the electron density map. In particular, the 45 "extra" serine and alanine residues in the model could well correspond to many of the 42 "missing" Asx, Glx and Arg residues, which would be expected to be on the surface of the molecule.

(c) *Backbone conformation*

The refinement supports the overall conformation of the protein backbone reported previously (e.g. see Figs 5, 6 and 7 of Matthews *et al.*, 1979). Figure 3 shows the Ramachandran diagram for the backbone dihedral angles. These angles were not monitored or manually adjusted during the refinement to conform to acceptable limits. There are four non-glycine residues that occur in regions where residues with beta-carbons are not normally expected. Residues 69 and 70 are the central two residues in a type I′ turn (Venkatachalam, 1968). Amino acid 69 has been identified with reasonable confidence as Asx and while the identity of residue 70 is not clear it seems to be at least as large as alanine. The turn is distorted with the carbonyl oxygen of residue 68 accepting the bridging turn hydrogen bond as well as a second hydrogen bond from a water molecule, as is often associated with these hairpin bends (Rose *et al.*, 1983). Residue 327 is also an apparent non-glycine residue (Asx) in a turn (type II′). Residue 119 is within a poorly defined stretch of chain and the conformational angles for residues in this vicinity are unreliable.

In the present numbering scheme for the amino acid sequence, residues that form α-helices are 121–128, 150–164, 169–179, 183–190, 225–230, 230–234, 284–294 and 335–346. One of the characteristics of the Bchl protein is the large number of residues in extended conformations. These residues form two β-sheets, one of 16 strands and the other of four strands (Matthews *et al.*, 1979). The β-sheet strands

Table 3
1—3 Hydrogen bonds

Residue	φ (°)	ψ (°)	*d* (Å)	Angle (°) C—O . . . N
30 (Ser)	−88	58	3·1	84
148 (Leu)	−94	55	3·1	82
179 (Phe)	−98	80	3·2	88
240 (Leu)	−80	76	2·8	98

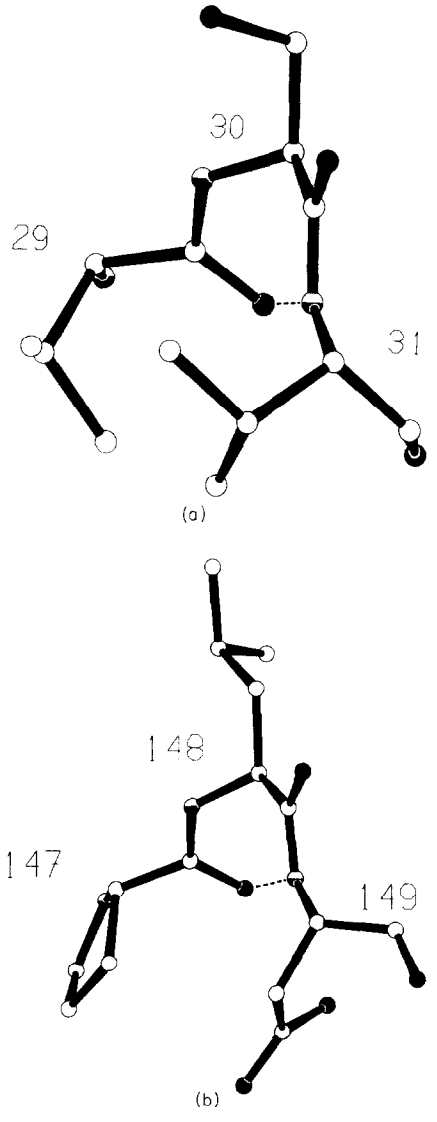

Figure 4. Two examples of 1—3 hydrogen bonds between successive peptide groups. (a) The presumed sequence Val29-Ser30-Val31. (b) The presumed sequence Pro147-Leu148-Asx149.

include residues 3–14, 20–29, 39–49, 57–68, 71–85, 88–101, 103–118, 135–146, 191–205, 209–222, 244–253, 256–265, 268–272, 276–280, 299–309, 314–319, 324–326, 328–333 and 352–356. (Residues are included in α-helices and β-sheets if they contribute one or more hydrogen bonds to the secondary structure element.) There is an extended six-residue segment (278–283) that begins within one of the β-sheet strands and continues beyond in a conformation similar to that of collagen (Soman & Ramakrishnan, 1983).

In four instances the main chain forms 1—3 hydrogen bonds between the carbonyl oxygen of one peptide and the nitrogen of the succeeding peptide (e.g. see Ramachandran & Sasisekharan, 1968; Baker & Hubbard, 1984). Two examples are shown in Figure 4(a) and (b). In all cases the Ramachandran angles for the α-carbon enclosed within the hydrogen-bonded peptides are in the vicinity of $\phi \sim -90°$, $\psi \sim 70°$ (Table 3).

(d) *Dynamics*

The apparent motion of the polypeptide backbone as estimated from the refined crystallographic thermal factors is shown in Figure 5. The three places where there are gaps in the backbone (i.e. residues 52–56, 165–168 and 206–207) are, as expected, regions of large apparent thermal motion or disorder. Often the regions of high mobility are at the ends of hairpin loops between successive β-strands, but there is also a case (near residue 164) where one end of an α-helix is very mobile. This α-helix is at the left edge of the subunit shown in Figure 1. It lies against the surface of the protein but is largely exposed to solvent.

The residues of the protein that form ligands to the magnesium atoms of the bacteriochlorophylls tend to have relatively low mobility (Fig. 5). This is presumably due to the fact that such residues are internal, and should not be taken to imply that the interactions with the metals anchor the protein.

The thermal factors for both the bacteriochlorophyll head-groups and the phytyl chains are relatively low, showing that these atoms are as well ordered as the interior protein atoms. In particular, the refinement confirms that the phytyl chains have well-defined conformations (Matthews *et al.*, 1979; Matthews, 1982) and are not disordered as in a fluid lipid bilayer.

(e) *Bacteriochlorophyll a conformations*

The atom nomenclature used here for the bacteriochlorophyll *a* molecules is illustrated in Figure 6. This scheme follows the conventions used by the Brookhaven Data Bank, and replaces the nomenclature used previously (Fenna *et al.*, 1977; cf. Strouse, 1974).

There are seven bacteriochlorophyll *a* molecules in the Bchl protein. This is a reasonable sample within which to compare variables such as the out-of-plane distances of the magnesium atoms, the pucker of rings II and IV, possible deformations of the chlorin rings and variability in the conformations of the ring substituents.

In Figure 7 the porphine head-groups of all seven Bchls have been superimposed by a least-squares process (Kabsch, 1976). The Figure gives an overall impression of the variability in the respective conformations. One of the striking features is the distribution of the magnesium positions. In two cases (Bchls 3 and 7) the metal ions are below the ring plane whereas in the other five cases they are above. There are no intermediate states.

Table 4 gives, for each Bchl, the distances of all the magnesium atoms from the planes with best least-squares fit to the 22 conjugated atoms. The Table indicates that the magnesium atoms of Bchls 3 and 7 have a mean displacement of 0·48 Å below the plane of the ring whereas the other five

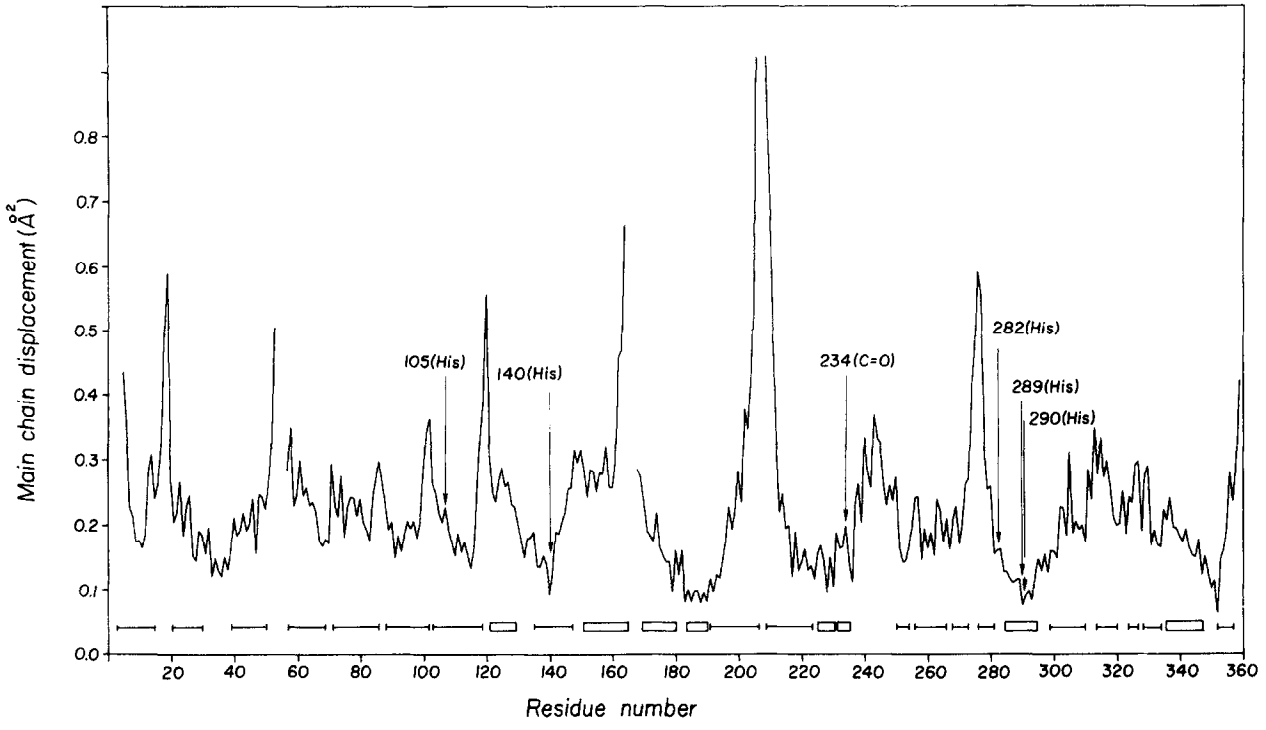

Figure 5. Apparent thermal motion of the backbone of the Bchl protein. There are 3 breaks in the chain at residues 52–56, 165–168 and 206–207 where the electron density is too weak to follow. Locations of α-helices are indicated by open rectangles and β-sheet strands by single bars. The residues that are liganded to the bacteriochlorophyll magnesium atoms are indicated.

magnesium atoms are, on average, 0·48 Å above. These distances agree reasonably well with the value of 0·4 Å observed in the structure of ethylchlorophyllide *a* (Strouse, 1974; Chow *et al.*, 1975; Kratky & Dunitz, 1975).

The displacement of each magnesium is always towards the side of the ring plane from which the metal is liganded to the protein. (Each magnesium is pentaco-ordinate.) Therefore the protein environment determines the direction of the out-of-plane magnesium displacement; the displacement direction is not an intrinsic property of the bacteriochlorophyll molecule itself.

The protein ligands for the two Bchls with their magnesium atoms below the ring plane appear to be histidines although one of these (Bchl 7) is unique in that it is the ND1 imidazole nitrogen that binds to the metal. For the other four Bchls with presumed histidine ligands the magnesium is liganded by nitrogen NE2. Table 4 lists the presumed ligands for each of the Bchls together with the ligand distances.

For the ring I acetyl substituent, the oxygen and the methyl group are essentially equivalent in the X-ray refinement. Therefore, these moieties have to be differentiated indirectly from their presumed

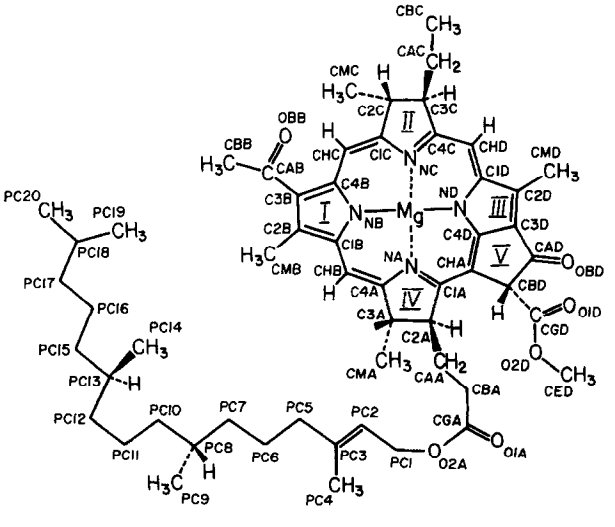

Figure 6. Schematic drawing of the structure of bacteriochlorophyll *a* showing the scheme used for atom identification. It should be noted that the scheme is somewhat anomalous in that atoms associated with rings I, II, III, IV or V have the letter B, C, D, A or D, respectively, as the last character in the atom name. We have changed to this scheme because it is the one used by the Brookhaven Data Bank.

Table 4
Magnesium ligands

Bchl	Presumed Mg ligand	Mg–ligand distance (Å)	Mg out-of-plane distance (Å)
1	His105 : NE2	2·2	0·44
2	Solvent (SOL1 : 14S)	2·0	0·45
3	His290 : NE2	2·1	−0·54
4	His282 : NE2	2·1	0·46
5	Leu234 : O (backbone)	2·0	0·52
6	His140 : NE2	2·2	0·54
7	His289 : ND1	2·1	−0·43

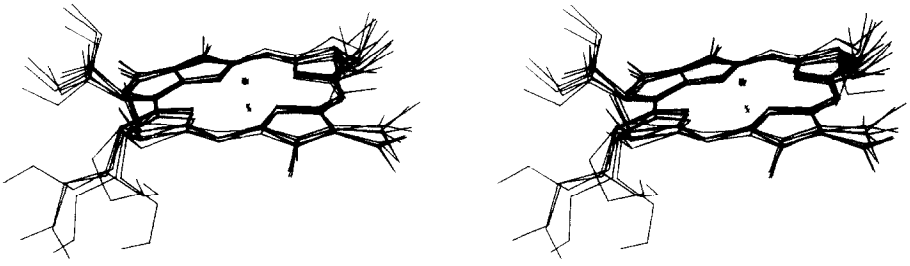

Figure 7. Stereo drawing showing the head-groups of all 7 Bchls brought into the same orientation. × indicates the position of a magnesium atom.

interactions with the protein. These are summarized in Table 5. During refinement, the acetyl group was constrained to be planar but was permitted free rotation about the bond between C3B and CAB. On the basis of a resonance Raman study, Lutz *et al.* (1982) suggested that at least four and possibly all seven of the acetyl groups should be within 18° of coplanar with the conjugated ring system. Allowing for experimental error, this is as observed (Table 5; Fig. 8(a)). Five of the acetyl groups are within 24° of coplanarity and the greatest departure (Bchl 5) is 36°. The torsion angles (Table 5) fall into two classes differing by approximately 180°. In two cases the ester oxygen does not make any apparent hydrogen bond to the protein and so cannot be distinguished from the methyl group. In such cases the torsion angle could be χ or $\chi-180°$. It is worth noting that Lutz *et al.* (1982) also proposed from their Raman study that up to two of the acetyl carbonyl groups were free of hydrogen bonding, in agreement with Table 5. The acetyl group with greatest departure from coplanarity with the chlorin system is not hydrogen bonded but has its methyl group "wedged" between hydrophobic groups of the protein. These close contacts appear to prevent the acetyl group relaxing to the coplanar configuration.

The torsion angles for the ring II ethyl substituents are listed in Table 6 and shown in Figure 8(b). The conformations are invariably staggered with one value in the 60° region, a second in the 180° region but a strong preference for torsion angles near −60° (five cases out of seven).

In refining the ring V carbomethoxy groups, atoms CBD, CGD, OD1 and OD2 were constrained to lie in a common plane but free rotation was permitted about the CBD—CGD bond. Even though the methyl group CED was not restrained to be coplanar with the remainder of the carbomethoxy moiety, it does lie close to the common plane in six of the seven cases (Fig. 7). The exception is Bchl 6 where the methyl group appears to be held away from coplanar by steric interference with neighboring protein side-chains. The plane of the carbomethoxy group is, in all cases, essentially perpendicular to the plane of the conjugated chlorin ring system (Fig. 8(c); Table 7). In no case does the ester oxygen O2D appear to accept a hydrogen bond from the protein. (This is also the case for the ester oxygen O2A in each of the phytyl chains.)

In each of the seven Bchls the keto oxygen (OBD) is essentially coplanar with the remainder of the conjugated system. Five of the keto oxygens appear to accept hydrogen bonds from the protein or a bound solvent molecule, while two seem not to be hydrogen bonded (Table 8). The resonance Raman spectra suggest that up to three of the keto oxygens are free from hydrogen bonding (Lutz *et al.*, 1982).

Each Bchl head group is connected to its phytyl tail *via* a propionic acid linkage. The conformations of the phytyl chains are all different and the linking groups also have varied conformations. As mentioned, none of the linking oxygens, O2A, accept hydrogen bonds. Also, in only one case (Bchl 1) does oxygen O1A appear to make a

Table 5

Configurations of the ring I acetyl groups

Bacterio-chlorophyll	Dihedral angle† (°)	Atom H-bonded to OBB‡	H-bonded distance (Å)
1	171	Solvent (SOL3 : 410)	2·6
2	−151	Ser66 : OG	2·8
3	24	Tyr9 : OH	2·7
4	176 or −4	—	—
5	144 or −36	—	—
6	174	Trp176 : NE1	3·0
7	158	Solvent (SOL1 : 43S)	2·7

† The dihedral angle is the angle between the planes defined by (C2B, C3B, CAB) and (C3B, CAB, OBB) and is 0° when C2B and OBB are *cis*. The calculation assumes that atoms OBB and CBB can be differentiated (see the text).

‡ The identification of these atoms is based only on the X-ray results and should be regarded as very tentative, pending confirmation with the chemically determined amino acid sequence.

Table 6

Configurations of the ring II ethyl groups

Bacteriochlorophyll	Torsion angle† (°)
1	−75
2	−45
3	−159
4	−83
5	−67
6	82
7	−75

† The torsion angle is zero when C2C and CBC are *cis*.

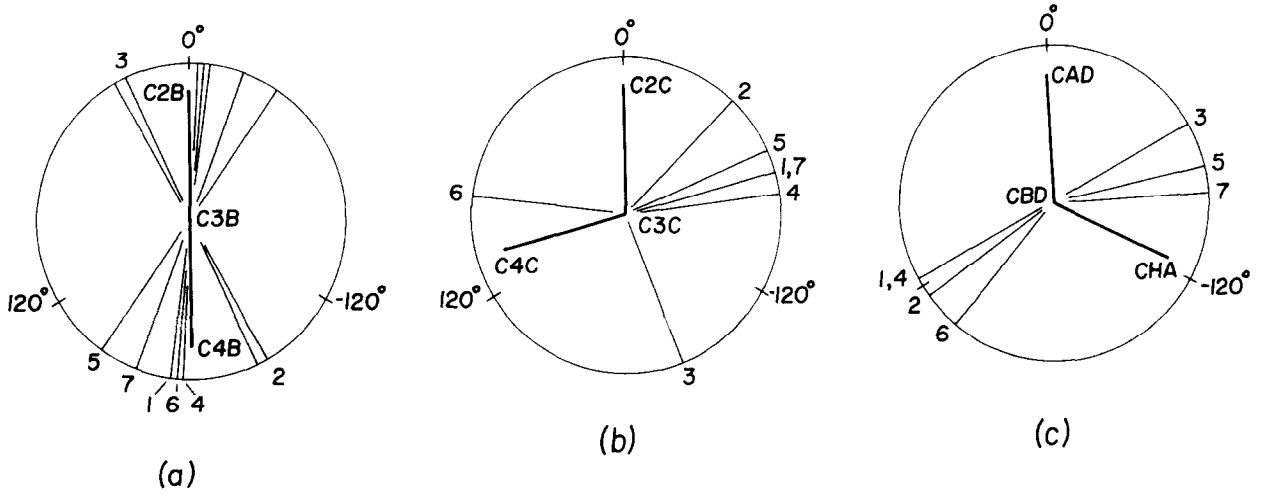

Figure 8. Newman projections illustrating the torsion angles of the ring substituents for the 7 Bchls. In each case the direction of view is from the substituent toward the Bchl ring. (a) Ring I acetyl groups. The direction of view is along the CAB—C3B bond and the lines labeled 1 to 7 indicate the planes of the acetyl groups of the 7 Bchls. In this Figure no attempt is made to distinguish between the acetyl oxygen (OBB) and the acetyl methyl group (CBB) (cf. Table 5). (b) Ring II ethyl groups. The direction of view is along the CAC—C3C bond and the numbers 1 to 7 indicate, for the 7 Bchls, the torsion angles about this bond (cf. Table 6). (c) Ring V carbomethozy groups. The view direction is along the CGD—CBD bond and the numbers 1 to 7 indicate the apparent torsion angles (defined by atom O2D) for the 7 Bchls (cf. Table 7).

hydrogen bond and this is to an atom in one of the symmetry-related monomers of the Bchl protein trimer.

There is a series of close contacts between the oxygen-containing ring substituents of Bchls 2 and 7. In the carbomethoxy group of Bchl 7, oxygen O1D is 3·0 Å from oxygen O2A in the same Bchl. At the same time the other oxygen (O2D) in the carboxymethyl group of Bchl 7 is 3·0 Å from oxygen O2A in Bchl 2. None of these close approaches are hydrogen bonds since no protons are involved.

Although the 22 atoms that constitute the conjugated part of each of the Bchl head-groups are essentially coplanar (Table 9) it can be asked if there is any evidence for systematic distortions away from planarity. Such an analysis is complicated because the relatively large uncertainty in the co-ordinates of the conjugated atoms (approx. 0·15 Å) could well mask small but systematic departures from planarity. Our first step, therefore, was to see if there were systematic departures from planarity that were common to two or more of the Bchls.

Table 10 shows the correlation coefficients between the out-of-plane displacements of the 22 conjugated atoms calculated for all possible pairs of Bchls. The striking result is that some pairs of Bchls (e.g. 2 and 3; 4 and 6; 5 and 6) have correlation coefficients close to unity, indicating that they depart from planar in virtually the same way. There are some instances of negative correlation (e.g. 1 and 4; 3 and 6) indicating that these pairs tend to depart from planarity in opposite directions, but these correlations are less striking than the positive ones. Taken together, the entries in Table 10 show clearly that Bchls 1, 2 and 3 correlate strongly with each other and therefore depart from planarity in the same way. We call this

Table 7
Configuration of the ring V carbomethoxy groups

Bacterio-chlorophyll	Dihedral angle† (°)	Atom H-bonded to O1D‡	H-bond distance (Å)
1	118	—	—
2	124	—	—
3	−62	—	—
4	118	Solvent (SOL1 : 26S)	2·7
5	−78	—	—
6	139	Thr113 : OG1	3·1
7	−88	—	—

† The dihedral angle is 0° when CAD and O2D are *cis*.
‡ Identifications provisional, pending confirmation with the chemically determined amino acid sequence.

Table 8
Interactions of the ring V keto oxygens

Bacteriochlorophyll	Interacting atom†	Distance (Å)
1	—	—
2	Peptide NH : Ile132‡	2·9
	Peptide NH : Tyr133‡	3·1
3	Peptide NH : Ser34	2·7
	OG : Ser34	2·9
4	OH : Tyr337	2·9
5	Solvent (SOL1 : 41S)	3·0
6	NH2 : Arg90	3·0
7	—	—

† All identifications tentative, pending confirmation with the chemically determined amino acid sequence.
‡ Symmetry-related molecule.

Table 9

Departures from planarity of the seven Bchl head-groups

	Bchl 1	Bchl 2	Bchl 3	Bchl 4	Bchl 5	Bchl 6	Bchl 7
Mg	0·44	0·45	−0·54	0·46	0·52	0·54	−0·43
C1A*	−0·05	−0·24	−0·19	0·05	0·03	0·05	−0·05
C2A	−0·01	−0·27	−0·20	0·03	0·08	−0·31	−0·01
CAA	−1·09	−1·71	−1·57	−1·34	−1·35	−1·77	−1·42
CBA	−2·59	−1·80	−1·20	−2·31	−1·17	−2·69	−1·57
CGA	−3·00	−3·21	−2·51	−3·01	−2·48	−4·12	−0·71
O1A	−3·05	−3·88	−3·18	−3·04	−2·54	−4·39	−0·94
C3A	−0·14	0·08	0·24	0·62	0·44	0·14	0·39
CMA	0·94	1·42	1·72	2·09	1·88	1·51	1·84
C4A*	−0·02	0·01	−0·02	0·16	0·27	0·24	0·17
NA*	0·00	−0·18	−0·18	0·11	0·01	0·19	0·04
CHB*	−0·06	0·09	0·13	0·26	0·01	0·02	0·07
C1B*	0·10	0·22	0·15	0·04	−0·06	−0·02	0·07
C2B*	0·06	0·11	0·18	−0·10	−0·11	−0·24	−0·21
CMB	−0·26	−0·05	0·23	−0·33	−0·03	−0·29	−0·35
C3B*	−0·02	−0·03	0·13	−0·24	−0·07	−0·28	−0·20
CAB	−0·04	0·04	0·21	−0·56	0·25	−0·65	−0·46
OBB	−0·28	0·42	−0·08	−0·71	−0·28	−0·61	−0·80
CBB	0·36	−0·23	0·54	−0·70	1·14	−0·77	−0·32
C4B*	0·00	−0·03	0·00	−0·20	−0·02	−0·05	0·05
NB*	0·04	0·17	0·03	0·01	0·01	0·14	0·22
CHC*	0·02	−0·13	−0·24	−0·14	−0·17	−0·14	−0·07
C1C*	0·04	−0·07	−0·14	−0·03	0·04	0·05	0·07
C2C	0·16	−0·33	−0·32	−0·13	0·24	0·01	0·51
CMC	1·59	0·70	0·46	0·78	1·61	1·10	1·96
C3C	−0·24	−0·09	0·30	0·24	0·06	0·15	0·63
CAC	−1·72	−1·15	−0·51	−0·85	−1·28	−1·03	−1·14
CBC	−1·97	−1·45	0·17	−0·45	−1·24	−2·37	0·46
C4C*	−0·09	−0·09	0·03	0·21	0·06	0·08	0·10
NC*	0·02	0·00	−0·05	0·08	0·07	0·18	0·01
CHD*	−0·06	0·12	0·17	0·29	0·03	0·08	0·02
C1D*	0·07	0·04	−0·02	−0·03	0·04	−0·01	−0·07
C2D*	0·00	0·04	0·12	−0·07	0·00	−0·11	0·01
CMD	−0·07	0·35	0·30	−0·05	0·04	−0·16	−0·05
C3D*	−0·08	0·03	0·07	−0·03	0·03	−0·02	−0·01
CAD*	0·04	−0·02	−0·05	−0·19	−0·14	−0·25	−0·08
OBD*	0·20	0·22	0·24	−0·18	−0·16	−0·21	0·03
CBD*	−0·25	−0·33	−0·28	−0·21	−0·06	−0·32	0·00
CGD	0·87	0·65	0·75	1·04	1·09	0·79	1·18
O1D	2·04	1·82	0·60	2·17	0·93	1·90	1·04
O2D	0·49	0·06	1·97	0·76	2·30	0·55	2·44
CED	1·67	0·99	2·96	1·97	3·48	1·89	3·51
CHA	−0·13	−0·18	−0·19	0·01	−0·08	0·04	−0·06
C4D*	−0·08	−0·03	−0·07	−0·01	0·08	0·12	−0·02
ND*	0·01	−0·03	−0·11	0·00	0·11	0·15	−0·09

The entries give the distance in ångström units from the best least-squares plane through the 22 atoms (asterisks) constituting the conjugated part of the dihydrophorbin macrocycle.

Table 10

Correlations between non-planarity of the conjugated Bchl macrocycles

		Class I			Class II				
	Bchl	1	2	3	4	5	6	7	EC†
Class I	1	1·0	0·56	0·32	−0·43	−0·37	−0·36	0·01	−0·35
	2		1·0	0·84	−0·02	−0·12	−0·22	0·23	−0·19
	3			1·0	0·01	−0·15	−0·40	0·03	−0·36
Class II	4				1·0	0·58	0·72	0·50	0·76
	5					1·0	0·80	0·46	0·80
	6						1·0	0·62	0·92
	7							1·0	0·59
	EC								1·0

† Ethylchlorophillide; co-ordinates from Kratky & Dunitz (1975).

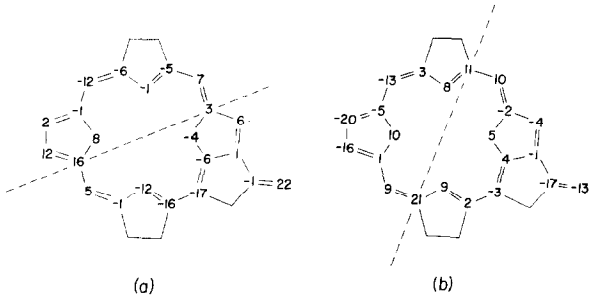

Figure 9. Schematic drawing illustrating the out-of-plane distances of the atoms in the conjugated chlorin system. The Bchls are in the same alignment as in Fig. 6 and the numbers give the out-of-plane distances in 0·01 Å units. (a) Average distances for conjugated atoms in Bchls 1, 2 and 3. (b) Average distances for conjugated atoms in Bchls 4, 5, 6 and 7.

class I. Similarly, Bchls 4, 5, 6 and 7 form a second distinct, highly correlated set (class II). There is little if any (perhaps slightly negative) correlation between the 1, 2, 3 set and the 4, 5, 6, 7 set. As a further test we have included in Table 10 the correlations between Bchls 1 to 7 and the corresponding 22 atoms in the structure of ethylchlorophyllide (Chow *et al.*, 1975; Kratky & Dunitz, 1975). Clearly, the non-planarity of ethylchlorophyllide corresponds to the Bchl 4, 5, 6, 7 class and not to the Bchl 1, 2, 3 class. The high positive correlations between the non-planarity of ethylchlorophyllide and Bchls 4, 5 and 6, in particular, confirm that these deviations are physically meaningful. In particular, the deviations cannot be attributed to the restraints that were placed on the Bchl head-groups during refinement since no such restraints were applied to ethylchlorophyllide.

Figure 9 illustrates the out-of-plane distances averaged over the class I and class II Bchls. With the exception of the keto group, the class I deformation appears to be, in large part, a bending or curvature of the conjugated atoms about a line through atoms C1B and C1D. For the class II Bchls there also seems to be overall bending or curvature of the conjugated macrocycle, in this case through a line connecting atoms C4A and C4C. Figure 10 shows the class I and class II Bchls superimposed in stereo.

(f) *Phytol conformations*

The phytyl chains of the seven Bchls and their interactions with the Bchl protein provide a model for lipid–protein interactions. Inspection of the phytols and their environments suggests the following general features (see Matthews, 1982, for a more detailed discussion).

(i) *Lipid conformations are well-defined but irregular*

In contrast to the fluid-like flexibility envisaged for lipids in a bilayer, the phytyl chains in the Bchl protein occupy well-defined positions in space with thermal factors comparable to the surrounding protein. In some cases the hydrocarbon chains are quite extended but they are not in idealized all-*trans* conformations. In this respect the Bchl phytols also differ markedly from crystal structures of isolated lipids in which very regular extended conformations are the rule (Pascher, 1976; Pascher & Sondell, 1977; Pearson & Pascher, 1979).

(ii) *Lipid tails prefer parallel interactions*

In a membrane bilayer, the amphipathic nature of membrane lipids results in the well-known bilayer structure with polar head-groups at the

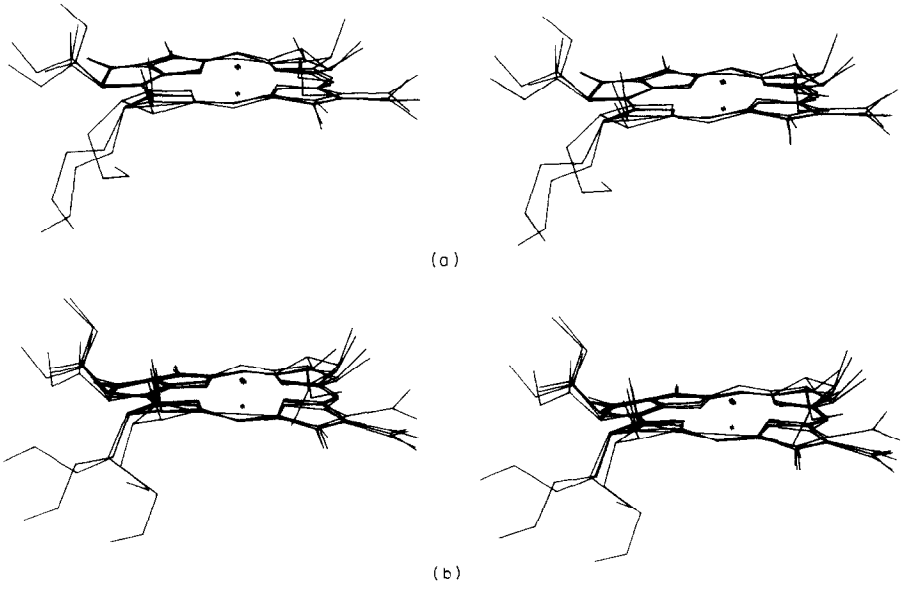

Figure 10. Superimposed stereo drawings of the class I and class II Bchl head-groups. (a) Bchls 1, 2 and 3 (class I). (b) Bchls 4, 5, 6 and 7 (class II).

surface and non-polar tails below the surface, interacting with each other. Because of these constraints the lipid tails are forced to lie more or less parallel to each other. In the Bchl protein, the constraint on the phytol packing is very different from that in a membrane bilayer, but where the hydrocarbon tails do interact they clearly prefer to do so in a parallel manner. There are seven cases in which two phytol chains from different Bchls remain in contact for five or more contiguous carbon atoms and in six of the seven cases the phytyl chains run in the same sense. The ester linkage to the chlorin head-group gives each phytol a distinct polarity, and, in addition, the asymmetric centers at PC8 and PC13 also imply polarity within the phytyl chain itself. Perhaps this inherent asymmetry favors parallel rather than antiparallel packing of the hydrocarbon chains.

(iii) *Bends occur at substituted positions*

The individual phytyl chains have unique conformations and often make sharp bends. When such bends occur they are invariably at, or adjacent to, substituted positions (including CAA and CBA, between the chlorin ring and the phytol ester linkage). This is not to suggest that the hydrocarbon chains are fully extended except at or near substituted positions. Rather, there are local departures from all-*trans* conformation but gross changes in direction occur at or near the sites of substitution.

(iv) *Lipid–lipid interactions do not dominate over lipid–protein interactions*

Although there are extensive interactions between the phytyl chains of neighboring Bchls, these interactions do not dominate over interactions between the phytols and the surrounding protein. Inspection of the structure suggests that there is roughly the same contact area occupied in phytol–phytol, phytol–chlorin and phytol–protein interactions.

We are grateful to Dr John Olson for generous gifts of Bchl protein, Dr Larry Weaver for help with data collection, Dr Lynn TenEyck for the initial development of the refinement package, Dr Martino Bolognesi and Giovanna Bolognesi for help with measuring the starting co-ordinates and Dr Roger Fenna for helpful discussions concerning the Bchl protein.

This work was supported in part by grants from the National Science Foundation (PCM8312151), the National Institutes of Health (GM20066; GM35114) and the Murdock Charitable Trust.

References

Baker, E. N. & Hubbard, R. E. (1984). *Progr. Biophys. Mol. Biol.* **44**, 97–179.

Chow, H.-C., Serlin, R. & Strouse, C. E. (1975). *J. Amer. Chem. Soc.* **97**, 7230–7237.

Colman, P. M., Jansonius, J. N. & Matthews, B. W. (1972). *J. Mol. Biol.* **70**, 701–724.

Deisenhofer, J., Epp, O., Huber, R. & Michel, H. (1984). *J. Mol. Biol.* **180**, 385–398.

Fenna, R. E. & Matthews, B. W. (1975). *Nature (London)*, **258**, 573–577.

Fenna, R. E., Matthews, B. W., Olson, J. M. & Shaw, E. K. (1974). *J. Mol. Biol.* **84**, 231–240.

Fenna, R. E., TenEyck, L. F. & Matthews, B. W. (1977). *Biochem. Biophys. Res. Commun.* **75**, 751–756.

Jones, T. A. (1982). In *Crystallographic Computing* (Sayre, D., ed.), pp. 303–317, Oxford University Press.

Kabsch, W. (1976). *Acta Crystallogr. sect. A*, **32**, 922–923.

Kratky, C. & Dunitz, J. D. (1975). *Acta Crystallogr. sect. B*, **31**, 1586–1589.

Lutz, M., Hoff, A. J. & Brehamet, L. (1982). *Biochim. Biophys. Acta*, **679**, 331–341.

Matthews, B. W. (1977). In *The Proteins* (Neurath, H. & Hill, R. L., eds), vol. 3, 3rd edit., pp. 403–590, Academic Press, New York.

Matthews, B. W. (1982). In *Lipid–Protein Interactions* (Jost, P. & Griffith, O. H., eds), vol. 1, pp. 1–23, John Wiley, New York.

Matthews, B. W., Fenna, R. E., Bolognesi, M. C., Schmid, M. F. & Olson, J. M. (1979). *J. Mol. Biol.* **131**, 259–285.

Molnar, C. E., Barry, C. D. & Rosenberger, F. U. (1976). *Tech. Memo. no. 229*, Computer Systems Laboratory, Washington University, St Louis, Miss.

Olson, J. M. (1971). In *Photosynthesis Part A* (San Pietro, A., ed.) (vol. 23, *Methods Enzymol.*), pp. 636–644, Academic Press, New York.

Olson, J. M. (1978). In *The Photosynthetic Bacteria* (Clayton, R. K. & Sistrom, W. R., eds), pp. 161–197, Plenum Press, New York.

Olson, J. M., Shaw, E. K. & Engleberger, F. M. (1976). *Biochem. J.* **159**, 769–774.

Pascher, I. (1976). *Biochim. Biophys. Acta*, **455**, 433–451.

Pascher, I. & Sondell, S. (1977). *Chem. Phys. Lipids*, **20**, 175–191.

Pearson, R. H. & Pascher, I. (1979). *Nature (London)*, **281**, 499–501.

Perlstein, R. M. & Hemenger, R. P. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 4920–4924.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Advan. Protein Chem.* **23**, 283–438.

Richards, F. M. (1968). *J. Mol. Biol.* **37**, 225–230.

Rose, G. D., Young, W. B. & Gierasch, L. M. (1983). *Nature (London)*, **304**, 655–657.

Rossmann, M. G. (1979). *J. Appl. Crystallogr.* **12**, 225–238.

Schmid, M. F., Weaver, L. H., Holmes, M. A., Grutter, M. G., Ohlendorf, D. H., Reynolds, R. A., Remington, S. J. & Matthews, B. W. (1981). *Acta Crystallogr. sect. A*, **37**, 701–710.

Schmid, M. F., Tronrud, D. E. & Matthews, B. W. (1983). *Chemica Scripta*, **21**, 67–69.

Soman, K. V. & Ramakrishnan, C. (1983). *J. Mol. Biol.* **170**, 1045–1048.

Strouse, C. E. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 325–328.

Venkatachalam, C. M. (1968). *Biopolymers*, **6**, 1425–1436.

Edited by R. Huber