PROTEINS: Structure, Function, and Genetics 8:133–155 (1990)

Comparison of the Structures of Globins and Phycocyanins: Evidence for Evolutionary Relationship

Annalisa Pastore[1] and Arthur M. Lesk[1,2]
[1]EMBL, Meyerhofstrasse 1, D-6900 Heidelberg, Federal Republic of Germany; and [2]MRC laboratory of Molecular Biology, Cambridge CB2 2QH, England

ABSTRACT **Globins and phycocyanins are two classes of proteins with different function, different ligands, and no substantial sequence similarity, yet the conformations of their polypeptide chains show very similar folding patterns. Does this arise from a genuine, albeit very distant, evolutionary relationship, or does it represent a common solution of a structural problem? We address this question by a very detailed comparison of the structures of the two protein families. An analysis of the helices and their interactions shows many features common to globins and phycocyanins, including some exceptional features of the globins such as a 3–10 *C* helix and the unusual "crossed-ridge" packing pattern at the *B/E* helix interfaces. We conclude that the evidence supports the hypothesis of distant evolutionary relationship between globins and phycocyanins.**

Key words: evolution, alignment, globins, phycocyanin, helix interfaces

INTRODUCTION

Phycocyanins, a class of proteins with covalently attached linear tetrapyrrole chromophores (bilins), are the major constituents of the phycobilisome, the supramolecular light-harvesting complexes of cyanobacteria and red algae, located in the outer surface of the thylakoid membrane. The crystal structures of phycocyanins from two species of bacteria, the cyanobacteria *Agmenellum quadruplicatum* (AQ) and *Mastigocladus laminosus* (ML), have recently been determined at 2.5 Å and 2.1 Å resolution, respectively.[1–3] Globins form another class of proteins, characterized by a noncovalently bound heme prosthetic group that binds oxygen reversibly. Globins are found in a wide variety of organisms, including bacteria, insects, vertebrates, and plants. They are one of the most widely studied classes of proteins, with over 400 known amino acid sequences and 12 crystal structures from distantly related species determined to atomic resolution (for reviews see Lesk and Chothia[4] and Dickerson and Geis[5]).

When the first phycocyanin structure was solved, it showed an entirely unsuspected, and quite surprising, close similarity to the globins in folding pattern (see Fig. 1), despite the difference in prosthetic group and function, and the absence of any detectable similarity between globin and phycocyanin sequences. What is the origin of the similarity in the fold? Do globins and phycocyanins share a common ancestor, or does their function dictate a similar three-dimensional structure?

A satisfactory answer to these questions is difficult to establish for two reasons: (1) because there is no detectable similarity in the sequences, there is no obvious way to align them with confidence, and (2) some of the structural differences between globins and phycocyanins are much larger than the differences between the most distantly related globins. The structures of the globins are constrained by the requirement of common function, and intercomparisons of globin structures cannot therefore show what structural changes might arise if this constraint is relaxed. The large structural differences between globins and phycocyanins are therefore not inconsistent with a true biological relationship, with divergence arising from differences in function. But if an evolutionary relationship exists it is a very distant one, lacking even a set of common functional residues to guide the comparison of the sequences and the structures. Superficially, the relationship between the structures is as close as that observed in some proteins genuinely related by evolution and no more distant than that of proteins believed to be unrelated.[6]

Work in recent years on the mechanism of protein evolution has revealed the kinds of similarities to be expected in distantly related members of a protein family.[4,7–9] One observes wide divergence of the sequences, but a retention of a *common core* of the structure: a central set of packed secondary structures present in all members of the family, surrounded by other structural elements that vary among members of the family. Even within the core there may be large shifts in the spatial relationships

Received September 11, 1989; revision accepted February 23, 1990.

Address reprint requests to Dr. Annalisa Pastore, European Molecular Biology Lab, Meyerhofstrasse 1, Postfach 1022.09, 6900 Heidelberg, Federal Republic of Germany.

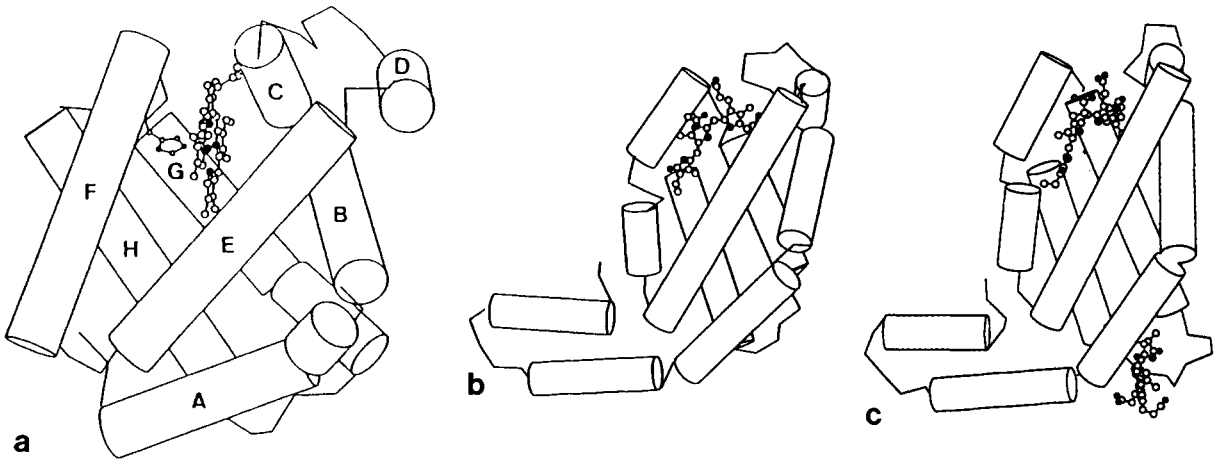


Fig. 1. A comparison between the *Aplysia* globin structure **(a)** and the α **(b)** and β **(c)** chains of phycocyanin from *Mastigocladus laminosus*. Cylinders represent the helical regions, and the prosthetic groups are shown in ball-and-stick representation.

between packed secondary structures—up to 7 Å in the globins. However, the patterns of residue–residue contacts in the interfaces between packed elements of secondary structure in the core are largely preserved—this appears to be a constraint imposed by the mechanism of evolution.[4]

The aim of the present work is to assess the similarity in structural detail of the phycocyanins and globins, and to ask whether the picture that emerges is consistent with, or rules out, an evolutionary relationship. An essential step in this investigation is to decide whether it is possible to determine an alignment of the sequences. In a previous paper we have discussed different approaches, including those dependent on structural comparisons, by which sequences of distantly related proteins may be aligned.[10] In this case, our seeking of an alignment does not assume—at least as anything more than a working hypothesis—that there is an evolutionary relationship; rather we are asking whether we can identify residues that play similar structural roles in both families and, provided we can, whether the relationship among them is consistent with observed patterns of evolution.

General Description of Phycocyanins and Globins

Phycocyanins are accessory light-harvesting pigments. They are associated with photosynthetic organelles, called phycobilisomes. The function of phycocyanins and other accessory pigments is to absorb energy in spectral regions in which the absorption by chlorophyll is weak, and to transfer the excitation energy to sites of photochemical activity.

The major components of phycobilisomes are biliproteins, that is, proteins with covalently attached tetrapyrrole (bilin) pigments and linker polypeptides.[11] Phycobilisomes can be divided morphologically into two regions: rods and core. The rods are arranged radially in a bundle around a core. Different phycobiliproteins are arranged in an ordered polar manner: the pathway of excitation energy transfer is from the tips of the rod to the core. Phycoerythrins and phycoerythrocyanins form the tips of the rod, and phycocyanins are located in the region proximal to the core. Allophycocyanins make up much of the core and are absent from the rods.

Phycocyanins exist in nature as stacks of hexamers, which contain two related polypeptide chains: α and β. The amino acid sequences of α and β chains have 20–30% identical amino acids (Table I), and the known structures show a high similarity in the folding pattern. Altogether, 20 sequences of biliproteins now appear in the PIR database[12] distributed into the three different subclasses of phycocyanins, allophycocyanins, and phycoerythrins. All these three classes are homologous, the most distantly related pair of known amino acid sequences retaining 30% identical residues.

Globins occur in a wide variety of organisms, with the common function of oxygen storage and transport. Globins bind oxygen reversibly to the iron of a heme group. Common structural features of all the globins are eight helices, labeled with the letters A, B, C, E, F, G, and H. In some of the structures there is also a short D-helix. The C-helix is a 3–10 helix; the others are α-helices. The helices assemble into a common folding pattern, enclosing the heme group in pockets of similar geometry made up from homologous portions of the molecules. In vertebrates, the presence of both monomeric globins (myoglobin) and tetramers (hemoglobin) allows a highly efficient physiological control over oxygen capture, transport, and delivery. The allosteric change in hemoglobin alters the oxygen affinity, producing more efficient absorption of oxygen in the lungs and more

TABLE I. Sequence Alignment of the Four Phycocyanins of Known Structure*

```
                  10        20        30        40
PCMLα     VKTPITDAIAAADTQGRFLSNTELQAVNGRYQRAAASLEA
PCAQα     MKTPLTEAVALADSQGRFLSNTELQYLYGRLRQGAFALEA
PCMLβ     AYDVFTKVVSQADSRGEFLSNEQLDALANVVKEGNKRLDV
PCAQβ     MFDIFTRVVSQADARGEFISSDKLEALKKVVAEGTKRSDA
                  10        20        30        40
            T     AD  G F S   L

                  50        60        70        80
PCMLα     ARALTANAQRLIDGAAQAVYQKFPYLIQTSGPNYAADARG
PCAQα     AQTLTAKADTLVNGAAQAVYSKFPYTTSTPGNNFAADQRG
PCMLβ     VNRITSNASTIVTNAARALFEEQPQLIAPGGSATRNGT--
PCAQβ     VSRMTNNASSIVTNAARQLFADQPQLIAPGGNAYTNRR--
                  50        60        70
            T  A       AA        P       G

                  90       100       110       120
PCMLα     KSKCARDIGHYLRIITYSLVAGGTGPLDEYLIAGLNEIND
PCAQα     KDKCARDIGYYLRMVTYCLVAGGTGPMDEYLIAGVDEINR
PCMLβ     MAACLRDMEIILRYITYAILAGDASILDDRCLNGLRETYQ
PCAQβ     MAACLRDMEIILRYVTYATFTGDASVLNDRCLNGLRETYV
           80        90       100       110
              C RD     LR  TY    G            G  E

                 130       140
PCMLα     AFELSPSWYIEALKYIKANHGLS----------GQAAN
PCAQα     TFDLSPSWYVEALKHIKANHGLT----------GDAAT
PCMLβ     ALGTPGSSVAVGIQKMKEAAINIANDPNGITKGDCSALIS
PCAQβ     ALGVPGASVAAGVRAMGKAAVAIVMDPAGVTSGDCSSLQQ
          120       130       140       150

          150       160
PCMLα      EANTYIDYVINALS
PCAQα      ETNNYIDYAINALS
PCMLβ      EVASYFDRAAAAVA
PCAQβ      EIELYFETAAKAVE
          160       170
           E   Y       A
```

	Sequence similarity: Number of identical residues			
	PCMLα	PCAQα	PCMLβ	PCAQβ
PCMLα	162	113	43	35
PCAQα	113	162	42	35
PCMLβ	43	42	172	112
PCAQβ	35	35	112	172
	Percent identical residues			
	PCMLα	PCAQα	PCMLβ	PCAQβ
PCMLα	100	70	27	22
PCAQα	70	100	26	22
PCMLβ	27	26	100	65
PCAQβ	22	22	65	100

*PCMLα and PCAQα stand for the α-chains from the *Mastigocladus laminosus* and *Agmenellum quadruplicatum*, respectively; PCMLβ and PCAQβ indicate the corresponding β-chains. The line beneath the sequences indicates residues conserved in all four sequences.

efficient release of oxygen in the tissues. However, there appears to be no relationship—neither structural nor functional—between the oligomer formation in globins and phycocyanins.

More than 400 globin sequences are currently known and about 12 crystal structures from different species have been determined at atomic resolution.

COORDINATES AND CALCULATIONS

Drs. W. Bode, R. Huber, and T. Schirmer provided coordinates of *Agmenellum quadruplicatum* and *Mastigocladus laminosus* phycocyanin.[1–3] These structures had been refined by the Jack–Levitt procedure, to R-factors of 21.7% at 2.1 Å for the *M. laminosus* structure and 18.4% at 2.5 Å for the *A. quadruplicatum* structure. We used globin coordi-

TABLE II. Phycocyanins and Globins of Known Atomic Structure From Distantly Related Species

Protein	Reference
Mastigocladus laminosus phycocyanin (α and β)	1,3
Agmenellum quadruplicatum phycocyanin (α and β)	2
Human hemoglobin (α- and β-chain)	20,21
Sperm whale metmyoglobin	22
Larval *Chironomus thummi* globin	23
Sea lamprey hemoglobin	24
Lupinus luteus L. leghemoglobin	25
Glycera dibranchata globin	26
Aplysia limacina myoglobin	14
Scapharca inaequivalvis hemoglobin	27

nates available from the Protein Data Bank,[13] and the *Aplysia limacina* myoglobin structure provided by Prof. M. Bolognesi and Dr. S. Onesti[14] (see Table II).

Structures were analyzed using the graphics program HYDRA installed on a Evans and Sutherland PS390 system. Other calculations were performed using programs written by A.M.L.[15]

Phycocyanin Crystal Structures

The minimum unit of a phycocyanin crystal is a dimer, with polypeptide chains of 162 amino acids (α-chain) and 171 amino acids (β-chain), respectively. The different lengths of the two phycocyanin chains arise from an insertion in the α-chain around residues 70–74 and an insertion in the β-chain around residues 145–154 (see Fig. 2). The α-chain carries one phycocyanobilin, the β-chain two, linked covalently to cysteine residues. These modified residues, including the prosthetic group, are denoted Cyc.

We will denote the α-chains from *Mastigocladus laminosus* and *Agmenellum quadruplicatum* as PCMLα and PCAQα, and the corresponding β-chains as PCMLβ and PCAQβ.

The crystals of phycocyanin from *A. quadruplicatum* (space-group *P*321) contain three (αβ)-units/asymmetric unit, i.e., three hexamers in the unit cell. The hexamer located at the origin has exact crystallographic 32 point symmetry. The other two hexamers of the unit cell have crystallographic 3-fold and local 2-fold symmetry. They are virtually identical to the hexamer at the origin.[2,3]

The unit cell of the crystals of phycocyanin from *M. laminosus* ($P6_3$) contains two trimers related by a crystallographic 2-fold screw axis. Each monomer has the overall shape of an oblate ellipsoid, with long axis 80 Å and short axis 20–25 Å. The three identical (αβ)-units are arranged around a 3-fold symmetry axis to form a disc of approximate dimensions 110 Å × 30 Å with a central channel 35 Å in diameter. This form of aggregation is believed to be the same as that in the rods of native phycobilisomes.[1,3]

RESULTS

Secondary Structure Assignments

Both α and β phycocyanin chains have a high proportion of helices (α-chain 65%, β-chain 62%), and no β-sheets. Each chain contains 10 helices: *X, Y, A, B, C, E, F′, F, G,* and *H*. The letters denoting them correspond to their position in the globin structures, except for the N-terminal *X* and *Y* helices which have no analog in globins. To avoid confusion with the one-letter codes for the amino acids, we use italic capital letters to denote helices. A number following a helix name refers to Kendrew notation (see Table VIII), whereas a number following an amino acid name always refers to the sequential number of the residue in the molecule under discussion; for example: *E8* refers to the eighth residue of the *E* helix, but E8 refers to the glutamate that is the eighth residue in the chain.

All helices are α-helices except for the *C* helix: residues 63 to 68 of both subunits form a short one-turn 3–10 *C* helix (Table III shows the hydrogen bonding in this region).

The limits of all the helices are shown in Table IV. We take each helix to begin at the first residue in which the oxygen is involved in the typical $(C{-}O)_i$ to $(NH)_{i+4}$ hydrogen bond pattern and to end at the last residue containing a similarly involved HN (the cutoff distance for the hydrogen bonds was 3.5 Å and the C=O...N angle cutoff 110°). With the exception of the *F* helix, the helices are straight and fairly regular. The *F′* and *F* helices are distinctly separated into independent helices at one residue (G114). In PCMLα the carbonyl of residue 112 forms hydrogen bonds to the amide groups of both residue 114 and 115.

The interaction between the *X* and *Y* helices of one subunit with the *A* and *E* helices of the other determines the α/β interaction. The loss or acquisition of the *X* and *Y* helices, which is correlated with the capacity to form an (αβ) monomer, may mark the branching point in the divergence of the phycocyanin and the globin families. (We note that *S. inaequivalvis* hemoglobin has a helix in the extension N-terminal to the A helix, but this is not involved in its dimer formation.[27])

Helix Packing

Secondary structure is usually well conserved in protein evolution, and the central pattern of residue contacts at the interface between two helices is largely invariant within the same family of proteins.[4] We therefore performed a detailed comparison of the helix packings common to globins and phycocyanins. Our initial goal was to test possible alignments of the sequences. If it were possible to establish a consistent alignment of the sequences,

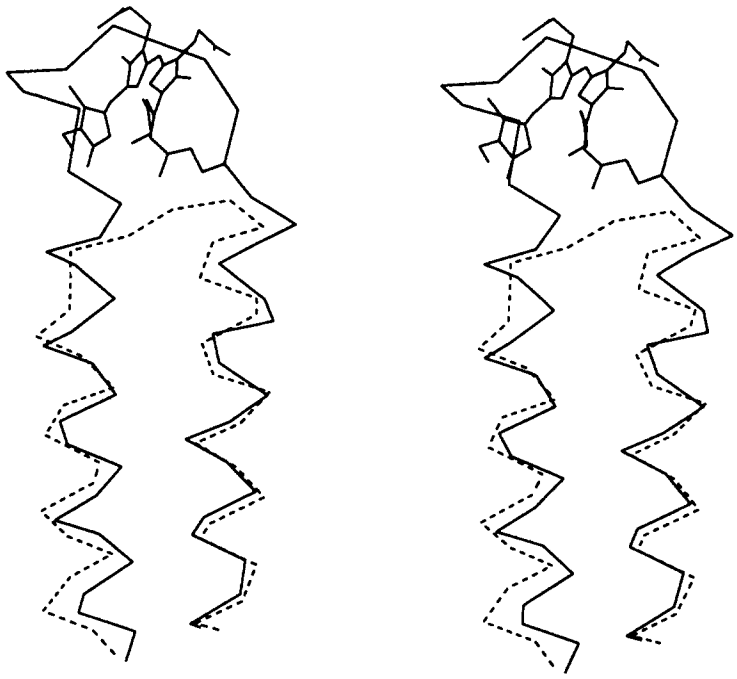


Fig. 2. Phycocyanin β-chains have a localized insertion, including residues linked to a second bilin group, between the *G* and *H* helices. This figure shows the *G/H* region of the α-chain (broken lines) and the β-chain plus its ligand (solid lines).

TABLE III. Hydrogen Bonds in the *C*-Helix

			PCAQα							PCMLα			
66	Thr	N	—	O	63	Phe	66	Leu	N	—	O	63	Phe
67	Thr	N	—	O	64	Pro	67	Ile	N	—	O	64	Pro
68	Ser	O_γ	—	O	64	Pro							
							69	Thr	N	—	O	65	Tyr
69	Thr	$O_\gamma 1$	—	O	65	Tyr							
			PCAQβ							PCMLβ			
66	Leu	N	—	O	63	Gln	66	Leu	N	—	O	63	Gln
67	Ile	N	—	O	64	Pro							
							68	Ala	N	—	O	65	Gln

we could then compare the patterns of residue contacts at helix interfaces, important evidence for or against evolutionary relationship.

To describe the patterns of residue contacts we use the "ridges into grooves" model introduced by Chothia et al.[16,17] Rows of side chains on the surface of each helix form irregular ridges separated by shallow grooves. Two helices pack together by means of a mutual intercalation of their ridges and grooves. The type of ridges occurring in the interactions determines, within a certain range, the interaxial angle.

Most commonly, ridges on the surface of helices are formed by residues separated by four in the sequence: we refer to these as "±4" ridges. In other cases ridges can be formed by residues separated by 3 in the sequence ("±3" ridges) or by residues consecutive in the sequence ("±1" ridges). The regions on the helix surface between the ridges form grooves parallel to the ridges. The different types of ridges and grooves make different angles with the helix axis, and therefore the packing of different pairs of types of ridges and grooves produces different characteristic interaxial angles.

The most common packing is given by intercalation of $\pm 4n$ ridges from both helices (typical angle around $-60°$ with a range of variability from $-20°$ to $-80°$), but patterns involving $\pm 1n$ ridges with $\pm 4n$ ridges (interhelix angle around $-110° \pm 30°$) or $\pm 3n$ ridges with $\pm 4n$ ridges (interhelix angle around $+19° \pm 15°$) are also frequently observed.[14,17] Sometimes ridges from two helices cross at a point, which

TABLE IV. Helical Regions in PCAQ and PCML*

	PCAQα	PCAQβ	PCMLα	PCMLβ
X	5–17	5–16	3–14	5–14
Y	20–34	25–33	20–32	20–33
A	36–48	33–48	34–48	33–48
B	47–60	47–63	50–62	47–63
C	63–68	63–68	63–68	63–68
E	77–102	75–100	78–102	75–100
F'	104–111	103–110	104–110	103–110
F	114–122	112–121	114–123	112–121
G	126–140	123–143	126–139	123–144
H	145–162	155–171	145–162	153–170

*The numbers refer to the sequential order of residues of each chain. Each helix was considered to begin at the first residue in which the oxygen is involved in the typical $(C{-}O)_i$ and $(H{-}N)_{i+4}$ for α-helices and $(H{-}N)_{i+3}$ for 3–10 helices and to end at the last residue containing a similarly involved H–N. The angle and the distance cutoffs for the hydrogen bonds were 110° and 3.5 Å, respectively.

provides a useful guideline for sequence alignment. Globins are unusually rich in such exceptional cases.

Helix packing in the globins has been described by Lesk and Chothia[4] and Pastore et al.[10] A summary of the types of the major interhelical contacts in the phycocyanins and in the globins is shown in Table V and VI.

Here we describe the helix packing in the phycocyanins, and compare the interfaces with the corresponding ones in the globins.

Helix packing in phycocyanins

Eight helix contacts occur in all four phycocyanin molecules: *A/E, A/H, B/E, B/G, E/G, E/H, F/H,* and *G/H*. All but one, *E/G,* are also common to all known globin structures, although an *E/G* contact is found in the β-strand of horse hemoglobin and in leghemoglobin. In globins, *A/E* and *E/H* are only minor contacts. The *X* and *Y* helices in phycocyanins are generally separated from the rest of the monomer, except for *A/Y* and *E/Y* interactions in the α structures and *A/Y* in the β structures. The interhelix axis angles and distances are shown in Table V, and a comparison with the corresponding values in globins is reported in Figure 3.

We now discuss the individual interhelix interfaces, and compare them with the corresponding ones in the globins.

The A/E helix interface. The *A* and *E* helices in both α- and β-chains of phycocyanins cross at an angle of approximately −170°. The packing of residues is fairly regular, with the intercalation of ±4*n* ridges of the *E* helix and ±3*n* ridges from the *A* helix. In PCMLα, the ridge formed by L92–T96–V100 intercalates between the two ridges formed by residues A41–L44 and T45–A48 (see Fig. 4). An *A/E* contact is found in some but not all globins. In *Aplysia* myoglobin, for instance, the interaxial angle is similar (−146°).

The A/H helix packing. The *H* helix meets the *A* helix at its very end. The *A/H* contact shows different types of packing in the α and β unit.

In the α-chains, the angle between helix axes is around −70°, and the interhelix axis distances are 8.4 Å and 7.0 Å for PCAQα and PCMLα, respectively. The two helices in the α chains pack together by the intercalation of ±1*n* ridges and grooves of helix *A* and ±4*n* ridges and grooves of helix *H* (Fig. 5).

In the α-subunits there are many small side chains (mainly Ala, Ser, and Gly) in both the *A* and the *H* helices. They are replaced by charged or polar groups (Arg, Lys, Gln, and Asn) in the β-chains as well as in the allophycocyanins and phycoerythrocyanins. The presence of small side chains in the contact region of the α-chain allows a short distance between the two helices. In PCMLα the ±1*n* ridges formed by A36–S37 and by E39–A40 intercalate between the ±4*n* ridges in the *H* helix running through Q145–E149–Y153–V157 and through A146–A150.

In the β-subunits the interaxis angle between helices *A* and *H* is around −50° with a larger interaxial distance (10.3 Å and 9.5 Å, in PCMLβ and PCAQβ respectively) and the packing is formed in this case by the intercalation of ±4*n* ridges from both helices, the most common form of helix packing (see Fig. 5). Residue 153 (Cyc) changes the orientation of the A chain by 22°, altering the nature of the interface, so that the ridge formed by K36–V40–I44 intercalates in the groove formed by L156–V160–F164 and Cyc153. The *A/H* contact region in PCMLβ contains much bulkier side chains than PCMLα at positions *A7, A14* and *H4* (see Figure 5).

In globins the *A/H* contact shows the same pattern as in the phycocyanin α-subunit, although the angle formed by the helix axes is around −110°, a more typical value for this packing (see Fig. 3). This may be related to the presence of large side chains such as F and W in position *H8* in globins which force the angle to be larger. In the α-phycocyanin chains the equivalent position is taken by Ala and the value of the angle is reduced to −70°.

B/E helix packing. In the *B/E* interface in phycocyanins ±4*n* ridges of the *B* helix cross a ±4*n* ridge of the *E* helix (see Fig. 6). The angle is around −50° for all four phycocyanin chains. In the α-chains, the ridges on the *E* helix formed by Cyc84–I88–L92–T96 and K81–A85–G89 intersect the ridge of the *B* helix formed by A48–V52–A56, with I88–G89 crossing over position 52. The residue at position 52 is V in PCAQα and I in PCMLα; the amino acid sequence in the *E* helix is very similar in the α-chains of both species.

Similarly, in the β-chain the ridges of the *E* helix E87–L83–M79 and T94–L90–M86–Cyc82 pack

TABLE V. Interaxial Angles (in Degrees) and Distances (in Å) for the Interhelical Contacts Made by the Four Phycocyanin Chains*,†

	PCAQα			PCMLα			PCAQb			PCMLβ		
A/E	7.9 Å	−174°	5/4	7.9 Å	−164°	6/4	10.6 Å	−173°	5/6	10.7 Å	−173°	6/5
A/H	8.4 Å	−66°	3/4	7.0 Å	−73°	5/4	10.3 Å	−51°	4/3	9.5 Å	−48°	6/4
B/E	9.6 Å	−52°	4/4	9.4 Å	−53°	4/5	10.1 Å	−51°	5/5	9.9 Å	−52°	6/6
B/G	8.4 Å	−48°	5/5	8.2 Å	−48°	7/6	8.9 Å	−43°	5/5	8.3 Å	−44°	8/8
E/H	9.9 Å	−53°	4/3	10.2 Å	−55°	4/4	10.0 Å	−40°	5/5	10.1 Å	−43°	3/4
E/G	11.0 Å	−57°	5/5	11.4 Å	−57°	4/4	11.4 Å	−57°	4/5	10.8 Å	−50°	6/5
F'/H	11.6 Å	−32°	2/3	11.4 Å	−31°	3/2	11.0 Å	−26°	2/4	10.0 Å	−23°	3/3
F/H	9.5 Å	−58°		9.1 Å	−60°	1/2	11.0 Å	−58°	3/2	10.3 Å	−68°	1/2
G/H	9.4 Å	−172°	8/6	8.7 Å	−174°	5/6	9.5 Å	170°	9/9	10.6 Å	166°	9/7

†No packing between the *X, Y* helices and the rest of the monomer.
*The values were calculated using the helical ranges in Table II excluding one or two terminal residues so that the rms deviation between the residues considered and an ideal α-helix of the same length was less than about 0.6 Å. The third column for each structure refers to the number of residues in contact in each helix pair.

TABLE VI. Summary of the Type of the Major Interhelical Contacts in Phycocyanins*

	PCα		PCβ		Globins	
A/E	3*n*/4*n*	(−169°)	3*n*/4*n*	(−173°)		
A/H	1*n*/4*n*	(−69°)	4*n*/4*n*	(−50°)	1*n*/4*n*	(−110°)
B/E	4*n*/4*n*	(−52°)	4*n*/4*n*	(−52°)	4*n*/4*n*	(−70°)
B/G	4*n*/4*n*	(−48°)	4*n*/4*n*	(−44°)	4*n*/4*n*	(−55°)
E/H	4*n*/4*n*	(−54°)	4*n*/4*n*	(−41°)		
E/G	4*n*/4*n*	(−57°)	4*n*/4*n*	(−54°)		
F/H	4*n*/4*n*	(−59°)	4*n*/4*n*	(−63°)	4*n*/4*n*	(−50°)
G/H	4*n*/4*n*	(−173°)	4*n*/4*n*	(168°)	4*n*/4*n*	(155°)

*Globins are shown here for comparison. In parentheses the interhelix angles (in degrees) are indicated.

against the ridge on the *B* helix formed by A48–V52–A56, with E87 crossing V52. The contacts are quite well conserved within the phycocyanins, as the sequence similarity is fairly high.

These observations can be compared directly to the similar structure of the *B/E* interface in globins, where the residue at position *B6* passes over position *E8*. In the globins, however, the interhelix axis angle tends to be somewhat larger (around −70°). It is of great importance to this investigation that both phycocyanins and globins show this exceptional "crossed-ridge" structure of the corresponding interface.

B/G helix contact. This contact is formed by intercalation of ±4*n* ridges from both helices and gives the expected value of the interhelix axis angle, around −45° (see Fig. 7). The values for the interhelix axis angle and distance are very similar in all four phycocyanin molecules, in spite of quite large changes in the amino acid sequence. In the α-chains, the inner ridge formed by the *B* helix (L51–A55–V59) runs almost parallel to the two ridges N139–Y135–E131 and I136–A132–W128. Similarly in the β-chains, the ridge formed by I51–A55–L59 packs between the two ridges A137–A(or K)133–A129 and M134–G130–S126.

In the globins the *B/G* contact is also of the regular ±4*n*/±4*n* type (Fig. 7).

E/G helix contact. The *G* and *H* helices are almost parallel to one another and to the *E* helix (the *G* helix is slightly closer to the *E* helix). The *E/G* contact involves a fairly regular ±4*n*/±4*n* packing in all the phycocyanins, with an interaxial angle of around −50°. The two ridges involving residues 129–133 and 128–132–136 of the *G* helix are intercalated between the ridges formed by residues 91–95–99 and 84–88–92 of the *E* helix.

E/H helix contacts. In the *E/H* contacts, the interhelix angles in the phycocyanin β-chains are around 10° smaller than in the α-chains. The two helices do not pack as intimately in PCMLα and PCAQα, where only one ridge (153–157) runs antiparallel to the ridge formed by residues 99–95–91 as in the β-chains. In the β-chains, the ridge formed by residues 93–89 packs between the two ridges formed by residues 163–167 and 160–164.

F'F/H helix contact. The *F'F* and the *H* regions are quite far apart in the phycocyanins (interaxial distances 11 Å, as compared to 9–10 Å in globins). In phycocyanins the *F* region is clearly broken into two helices between residues A113 and G114 (in PCMLα): we defined the *F'* helix in PCMLα as including residues 104–110, and the *F* helix as including residues 114–123. In phycocyanins the interaction with the *H* helix involves almost exclusively the *F'* helix.

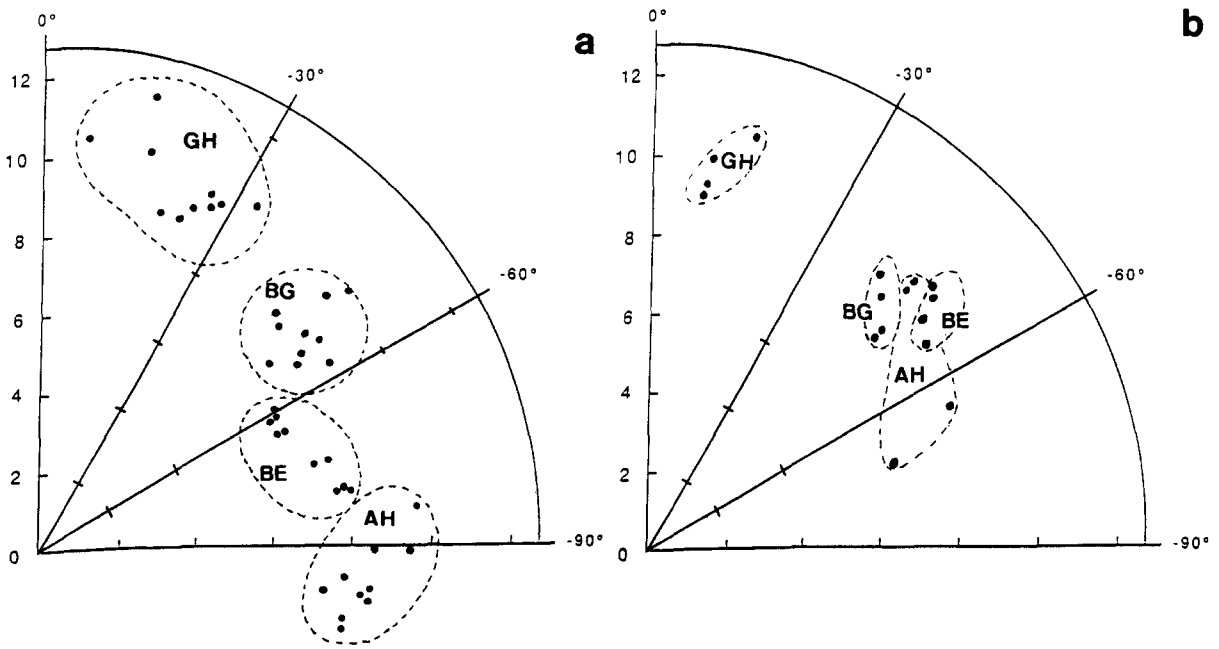


Fig. 3. Comparison of the distributions of interhelix distances and angles in the packed helices of globin (**a**) and phycocyanins (**b**).

The *F'* and *H* helices cross at an angle of −30°. The packing at the interface is of the most common type, ±4*n* ridges from both helices, with the ridges formed by 107–111–115 and 104–108–112 packing between 153–157 and 156–160 (e.g., L107–L111–L115 and T104–D108–I112 pack against Y153–V157 and Y156–A160 in PCMLα).

In the globins, the *F'F/H* contact shows a similar packing pattern, but the angle tends to be somewhat higher: around −50°.

G/H contact. These two helices are almost parallel, with interaxial angles of about +8° for the α-chains and −10° for the β-chains. The β-chains show a cross-ridged structure. The α-chains have an irregular interface permitting two possible interpretations of the packing: the interface can be considered either as similar to the one in the β-chain with a crossing point, or as formed by ±4*n*/±4*n* packing. In the second interpretation, two ridges running over residues P126–I130–K134 and L133–K137 intercalate between the two ridges formed in the *H* helix by I158–I154 and L161–V157–Y153. The expected value of the interaxial angle for such a packing is around +25°, reduced here to a smaller value (around 8°).

In the globins, the *G/H* contact is also formed by ±4*n* ridges from both helices, which cross each other (see Fig. 8).

Summary. The interaxial distances and angles in the helix interfaces in phycocyanins are summarized in Tables V and VI and Figure 3, which compares them with the corresponding values in the globins. We see in Figure 3 that for some contacts the clusters of interaxial distances and angles occupy the same region in the graph, but for others—notably *B/E* and *A/H*—they are very different.

Prosthetic Groups and Their Environment

The structural basis of the heme–globin interaction has been discussed extensively. The polypeptide chain of each globin accommodates the heme in a pocket, stabilized by hydrophobic interactions and by the coordination with two histidines (positions *E7* and *F8*, apart from a few exceptions, such as the globins from *Aplysia limacina* and *Glycera dibranchata*, which lack the distal His *E7*). A total of approximately 15 heme–protein atomic contacts is common to most of the known structures. These involve the carboxy-terminus of the *C* region, the central region of the *E* helix, and the *F* and the *G* helices. Because of their structural role, residues in contact with the heme are well-conserved in the globins.

The bilin group in phycocyanins is stabilized in position by packing contacts but also is covalently attached to the *E* chain. The separation and orientation of the chromophores are critical for the absorption and transfer of light in the photosystem. Because of its elongated shape, bilins in phycocyanins form contacts with more residues than does the heme in globins, but the distribution of the interactions involves the same helices (we are describing the contacts in the α-chains, which contain a single bilin bound in roughly the same region of the molecule as the heme in globins; in the β-chains there is

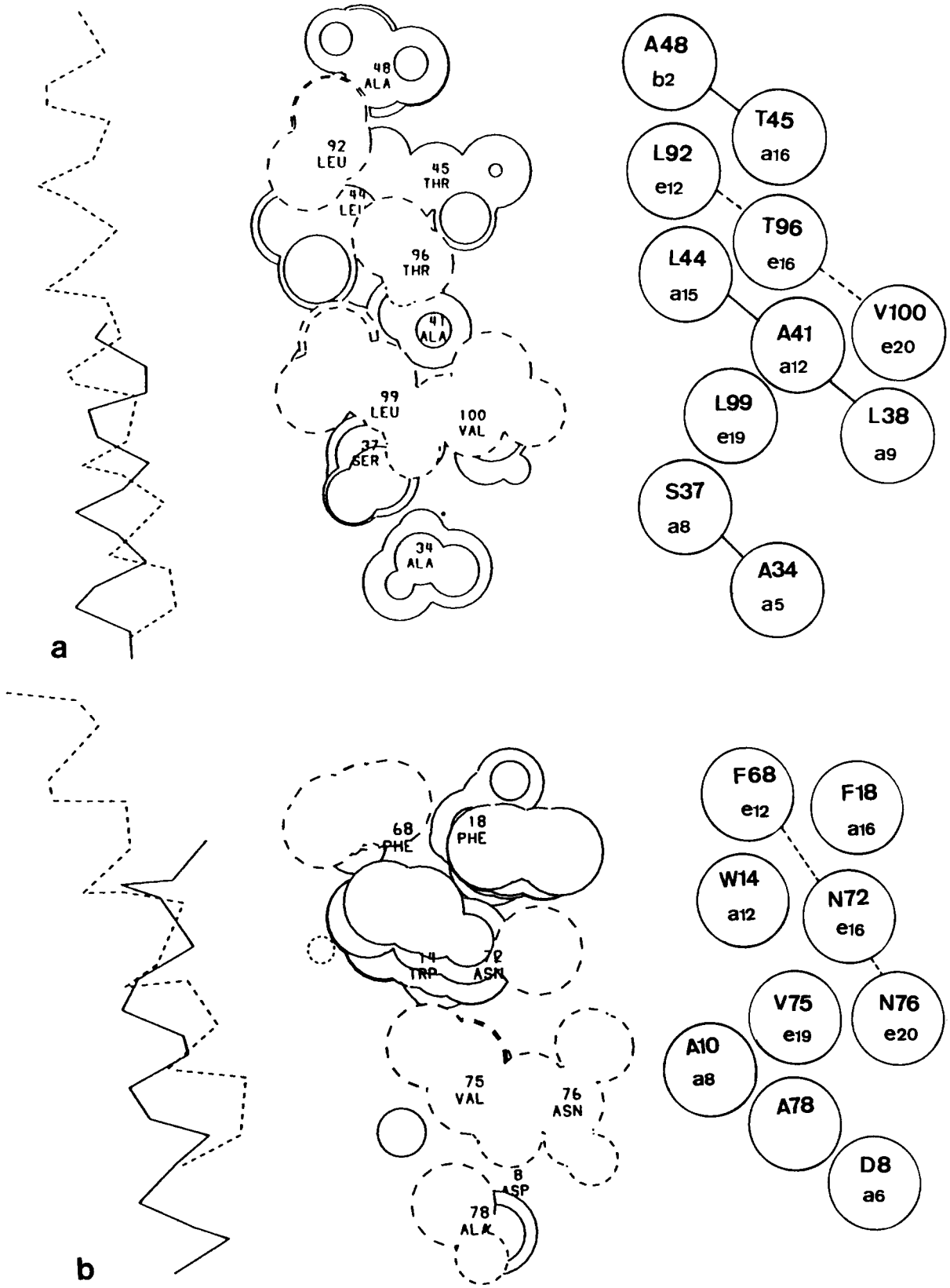
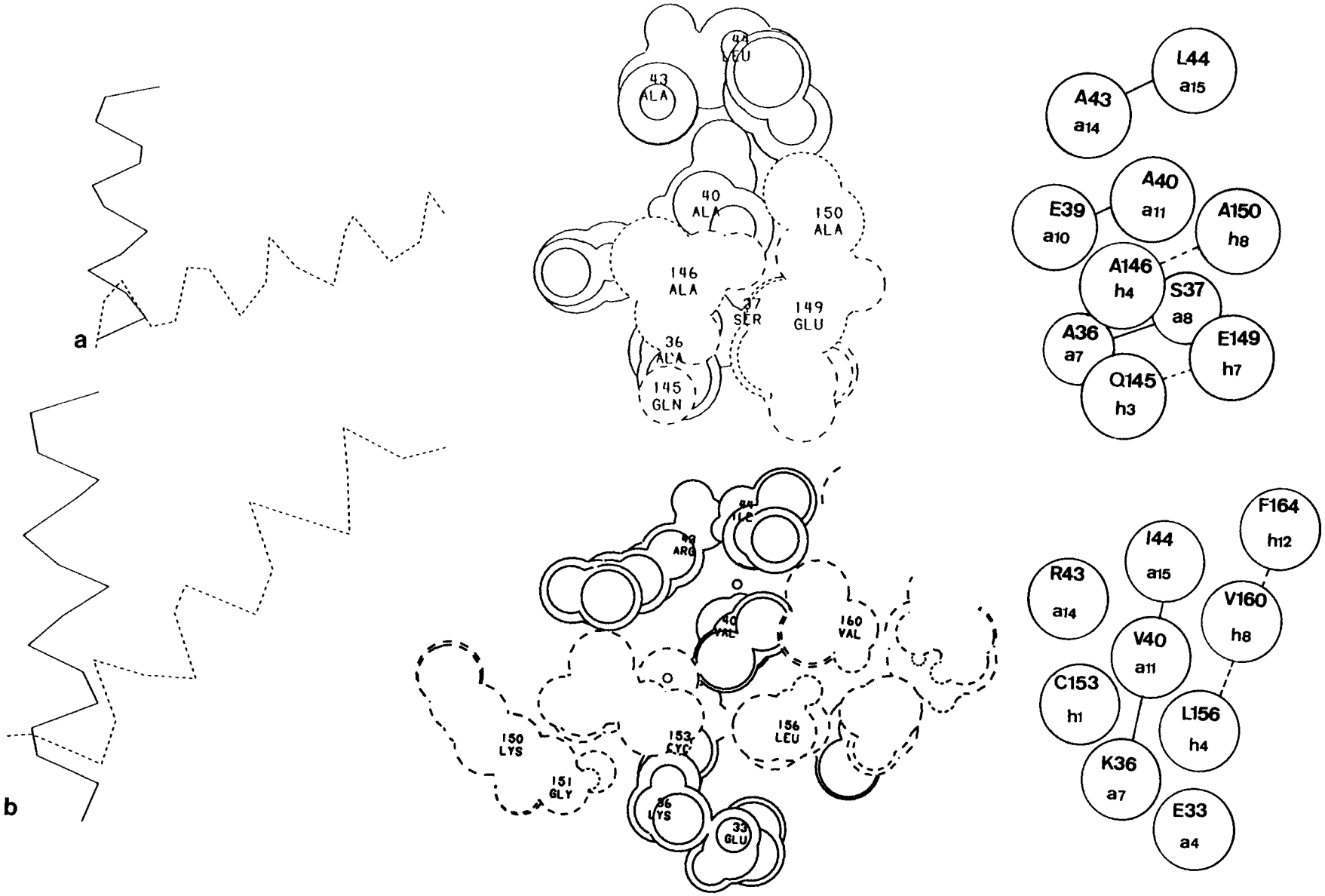


Fig. 4. Comparison of the geometry of the *A/E* interface in PCMLα **(a)** and in *Aplysia limacina* **(b)** and the corresponding residue packing. In this as in the following figures, we show, for each structure, a Cα trace of each of the pair of helices, the residue packing at the interfaces, and a schematic representation of the residue contacts. Residues from the *E* helix are indicated as broken lines.


a
b
43 ALA
40 ALA
150 ALA
146 ALA
37 SER
149 GLU
36 ALA
145 GLN
L44 a15
A43 a14
A40 a11
E39 a10
A150 h8
A146 h4
S37 a8
A36 a7
E149 h7
Q145 h3
160 VAL
150 LYS
151 GLY
153 LYS
156 LEU
33 GLU
F164 h12
I44 a15
R43 a14
V160 h8
V40 a11
C153 h1
L156 h4
K36 a7
E33 a4

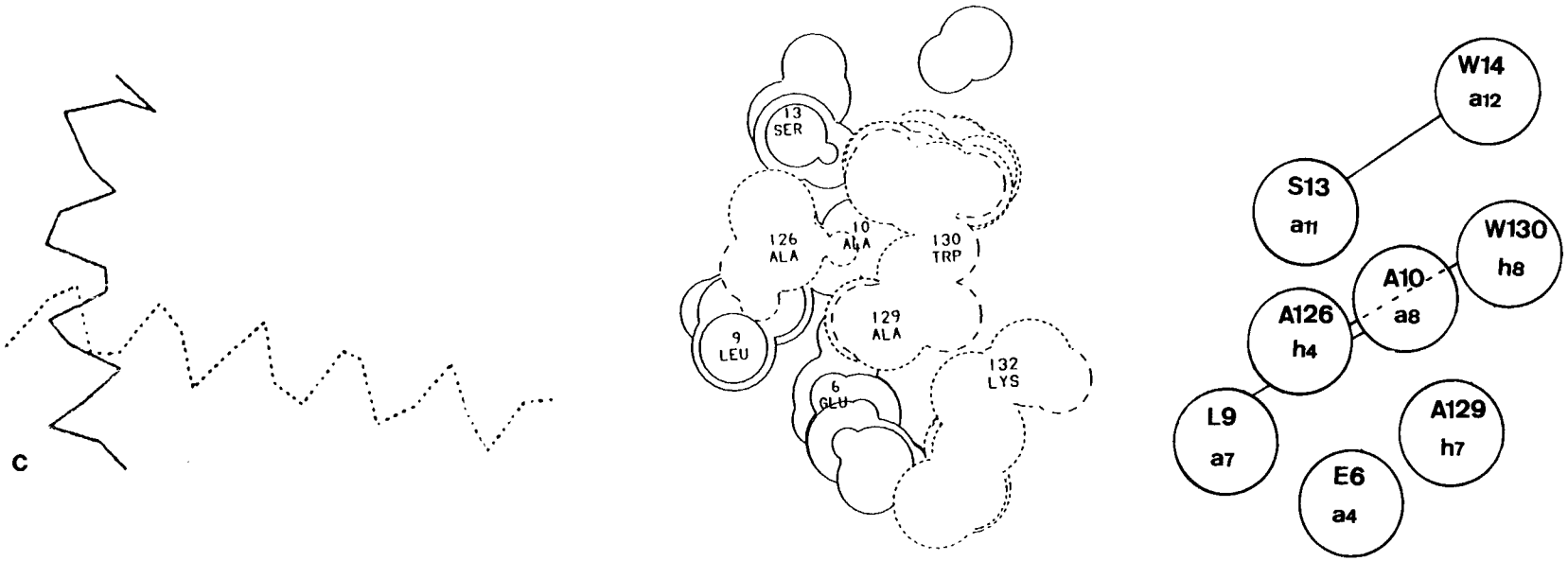


Fig. 5. Comparison of the *A/H* interface in PCMLα **(a),** in PCMLβ **(b)** and in *Aplysia limacina* **(c)**. Residues from the *H* helix are represented as broken lines.

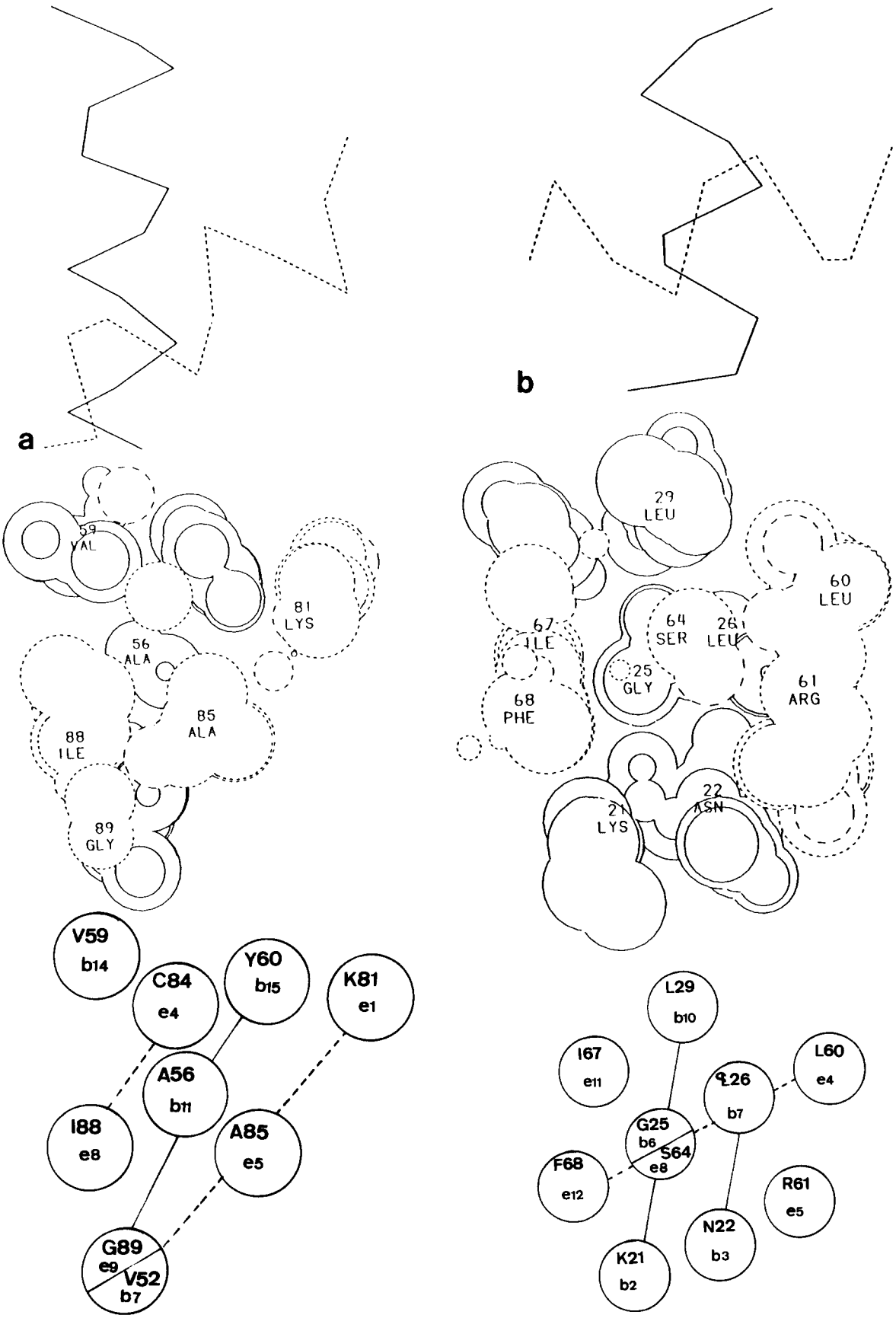


Fig. 6. Comparison of the *B/E* interface in PCMLα **(a)** and in *Aplysia limacina* **(b)** and the corresponding residue packing. Residues from the *E* helix are represented as broken lines.

an additional chromophore bound between the *G* and the *H* helices). Around the bilin the *E, F, G,* and *H* helices cluster together and the prosthetic group is almost completely sandwiched between the *E* and *F* helices (see Fig. 9). The angle between the axes of helices *E* and *F'* is 162°, while the angle between the *E* and the whole *F'F* helix is −161°, comparable to the same angle in Apl (−152°). The function of the *F* helix as a support would explain the break around the middle of this region to allow the *F* helix to get close enough to the bilin to form hydrophobic interactions involving I118, F122, and L124.

DISCUSSION

Method of Alignment of the Amino Acid Sequences

In previous work, we reviewed several possible approaches to the problem of the alignment of the amino acid sequences of distantly related proteins[10]:

1. According to the sequences, by procedures maximizing similarities of corresponding residues;
2. According to rules applicable to sequences, but derived from the general structural features of the class of proteins (templates);
3. A three-dimensional alignment obtained by observing similar spatial positions of the backbone atoms;
4. Using the patterns of residue packing at internal interfaces.

For the alignment of phycocyanins and globins, (1) is not useful, and (2), (3), and (4) are not straightforward to apply, because of the very great differences in sequence and structure. We regard (4) as the "court of last resort," and indeed we rely mainly on it.

The procedure we followed is based on the following idea. From each pair of helices in contact we may obtain one or more possible alignments—ambiguities arise because of the large structural differences. Most of the helices are involved in more than one interaction, so that for each helix we will have several possible alignments derived from the different contacts that the helix makes. We will then seek a unique global alignment consistent with the indications derived from each helix interface. The goal is to construct a self-consistent overall alignment.

In our analysis, we always used more than one structure, but, for simplicity, we will present here the alignment of PCMLα with *Aplysia* myoglobin (Apl).

Alignment Indications From Individual Interfaces

The following helix contacts occur in both phycocyanins and globins and therefore are sources of alignment information:

A: Helix *A* forms *A/E* and *A/H* interfaces
B: Helix *B* forms *B/E* and *B/G* interfaces
E: Helix *E* forms *A/E* and *B/E* interfaces
F: Helix *F* forms an *F/H* interface
G: Helix *G* forms *B/G* and *G/H* interfaces
H: Helix *H* forms *A/H* and *G/H* interfaces

By comparing the patterns of residue packing at corresponding helix interfaces in phycocyanins and globins, we can identify a set or in many cases several possible sets of corresponding residues. Using Figure 3, for each interface we tried to choose a phycocyanin and a globin with similar values of the interhelix axis distance and angle. In most but not all cases we compare PCMLα with *Aplysia* globin (Apl). In Table VII we list the possible alignment indicators derived from each of the helix contacts common to phycocyanins and globins, which are described in this section. In the next section we describe how we were able to resolve the ambiguities and choose from this table a single consistent alignment.

In reading the following paragraphs, the reader should examine the appropriate slices in Figure 4–8 and see how the entries in Table VII are derived from comparisons of the interfaces in phycocyanins and globins.

Alignment indications from the A/E interface

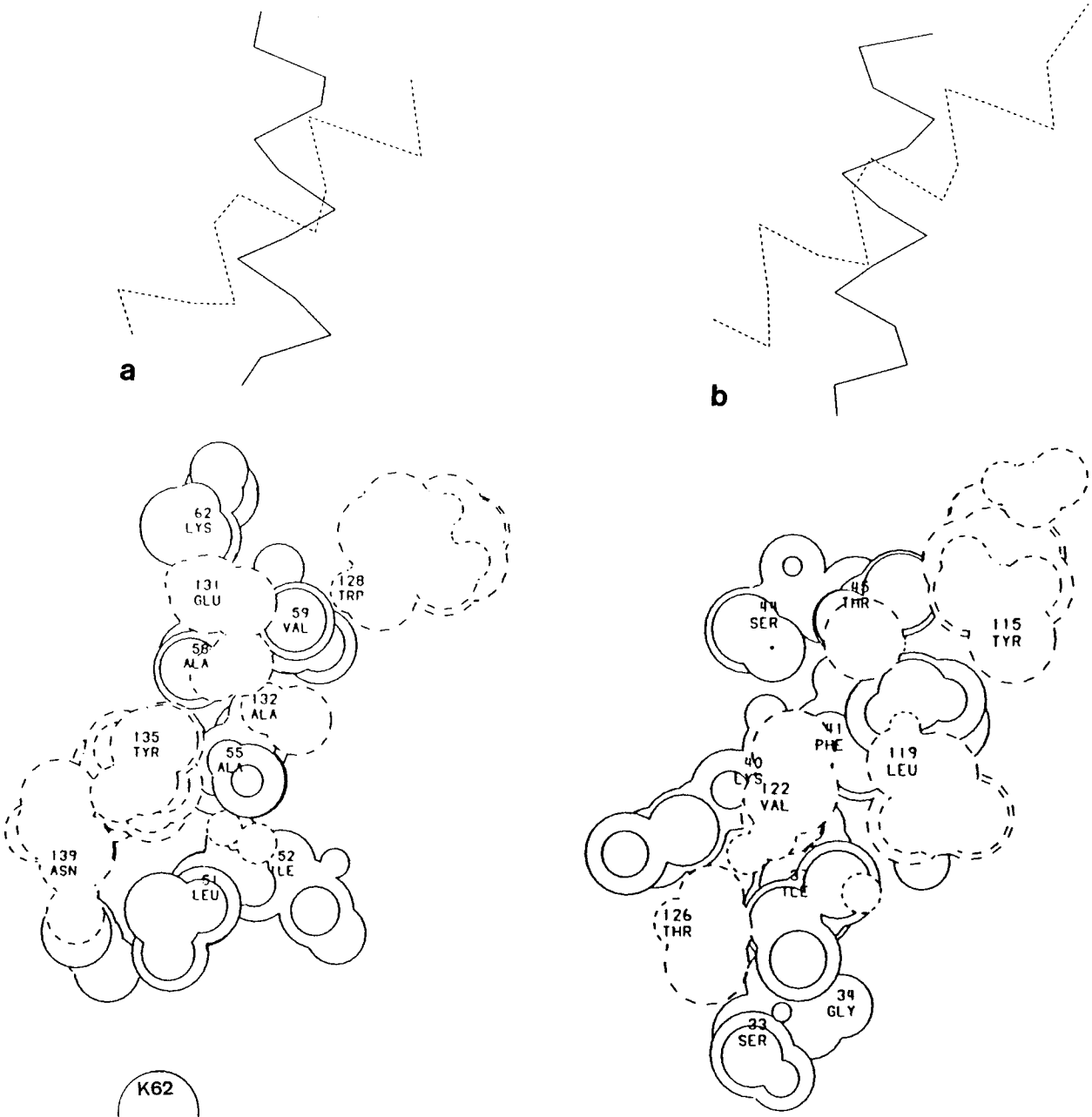
In the PCMLα *A/E* contact, the ridges formed by residues 92–96–100 pack between ridges formed by residues 44–41 and 48–45. This interface suggests the alignment of W14 of Apl with L44 of PCMLα. For the *E* region, the most plausible alignment is PCMLα96 = Apl72 (= position *E16* in globins).

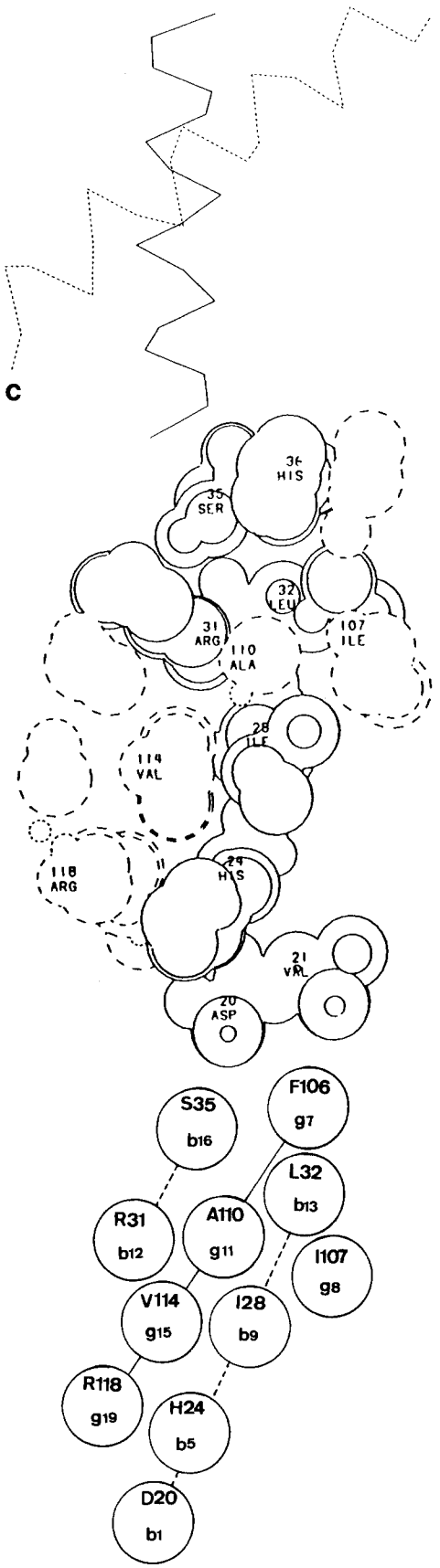
Alignment indications from the A/H interface

In the *A/H* contact in PCMLα the *A* helix shows parallel ±1*n* ridges: A36–S37, E39–A40, A43–L44. The first and the second pack against ±4*n* ridges of the *H* helix (Q145–E149 and A146–A150). They can be identified with the following *A* helix residues in Apl: L9–A10 and S13–W14, packed against ridges A126–W130 and A129. Plausible alignments of the *A* helix are therefore A40 in PCMLα with A10 in Apl or S37 in PCMLα with A10 in Apl. For the *H* region, the most plausible alignment is PCMLα150 = Apl130 (that is position *H6* in globins).

From both the *A/E* and *A/H* interfaces, the same two alignments of the *A* region are plausible:

```
          36  40      46
PCMLα     ASLEAARALTANA
Apl       LAGKSWAPVFA
          10          19
or
          36  40      46
PCMLα     ASLEAARALTANA
Apl         LAGKSWAPVFA
            10        17
```


a
b
62 LYS
131 GLU
128 TRP
59 VAL
58 ALA
132 ALA
135 TYR
55 ALA
139 ASN
52 LE
51 LEU
44 SER
45 THR
115 TYR
41 PHE
40 LYS
122 VAL
119 LEU
126 THR
34 GLY
33 SER
K62 b16
E131 g7
W128 g4
A58 b12
V59 b13
Y135 g11
A132 g8
A55 b9
N139 g15
I136 g12
L51 b5
I52 b6
S44 b16
V118 g7
Y115 g4
F41 b13
K40 b12
V122 g11
L119 g8
T126 g15
I37 b9
S33 b5



Alignment indications from the B/E interface

In the *B* helix, residue A56 in PCMLα could correspond with 25G or 29L in Apl. In the *E* helix, residues 85A or 88I in PCMLα could correspond with 64S in Apl (position *E8*).

Alignments of the *E* helix using the *E/H* and *E/G* contacts are not possible as these helices are too distant in globins (around 14 Å in Apl), and too few residues are in contact, to discern a pattern of interaction. In Apl, the *E/H* contact has only two pairs of residues and the *E/G* contact involves only one pair of residues with large side chains (F and L).

Alignment indications from the B/G interface

For the *B/G* interface we compared PCMLα with lamprey globin (Lam). In the *B* helix, residue A55 in PCMLα could correspond to F41 or I37 in Lam (that is, position *B13* or position *B9*). In the *G* helix, the most plausible correspondence is Y135 in PCMLα = V122 in Lam *(G11)*.

Alignment indications from the G/H interface

For the *G/H* interface we have compared PCMLβ with sperm whale myoglobin (SWM), because in this interface phycocyanin β-chains are more similar in structure to globins than α-chains. In the *G* helix, residue K135 or I139 in PCMLβ could correspond with I112 in SWM (position *G13*). If, in the *G* helix, K135 in PCMLβ = I112 in SWM, then in the *H* helix, F164 in PCMLβ should correspond to L135 or R139 in SWM (*H12* or *H16*). But if, in the *G* helix, I139 in PCMLβ = I112 in SWM, then in the *H* helix, V160 in PCMLβ should correspond to L135 or R139 in SWM.

Choosing a Consistent Alignment

Table VII shows that in many cases the helix interfaces do not suggest unambiguous residue correspondences. In those cases we have listed in Table VII several plausible alignments; indeed, because of the large structural differences we have been fairly generous about what we consider plausible.

What we must do next is to see whether it is possible to extract from Table VII a unique consistent alignment of all the helices.

For the *A* helix, both *A/E* and *A/H* interfaces suggest the *same* two alternatives. Therefore it is not possible to choose one of them from these two interfaces alone.

Consider next the *B* helix. The *B/E* and *B/G* interfaces each suggest two possible alignments of the *B* helix, one of which is common: PCMLα55 = Lam37 = Apl28 (position *B9*).

Fig. 7. Comparison of the *B/G* interface in PCMLα **(a)**, sea lamprey globin **(b)**, and sperm whale myoglobin **(c)** and the corresponding residue packing. Residues from the *G* helix are represented as broken lines.

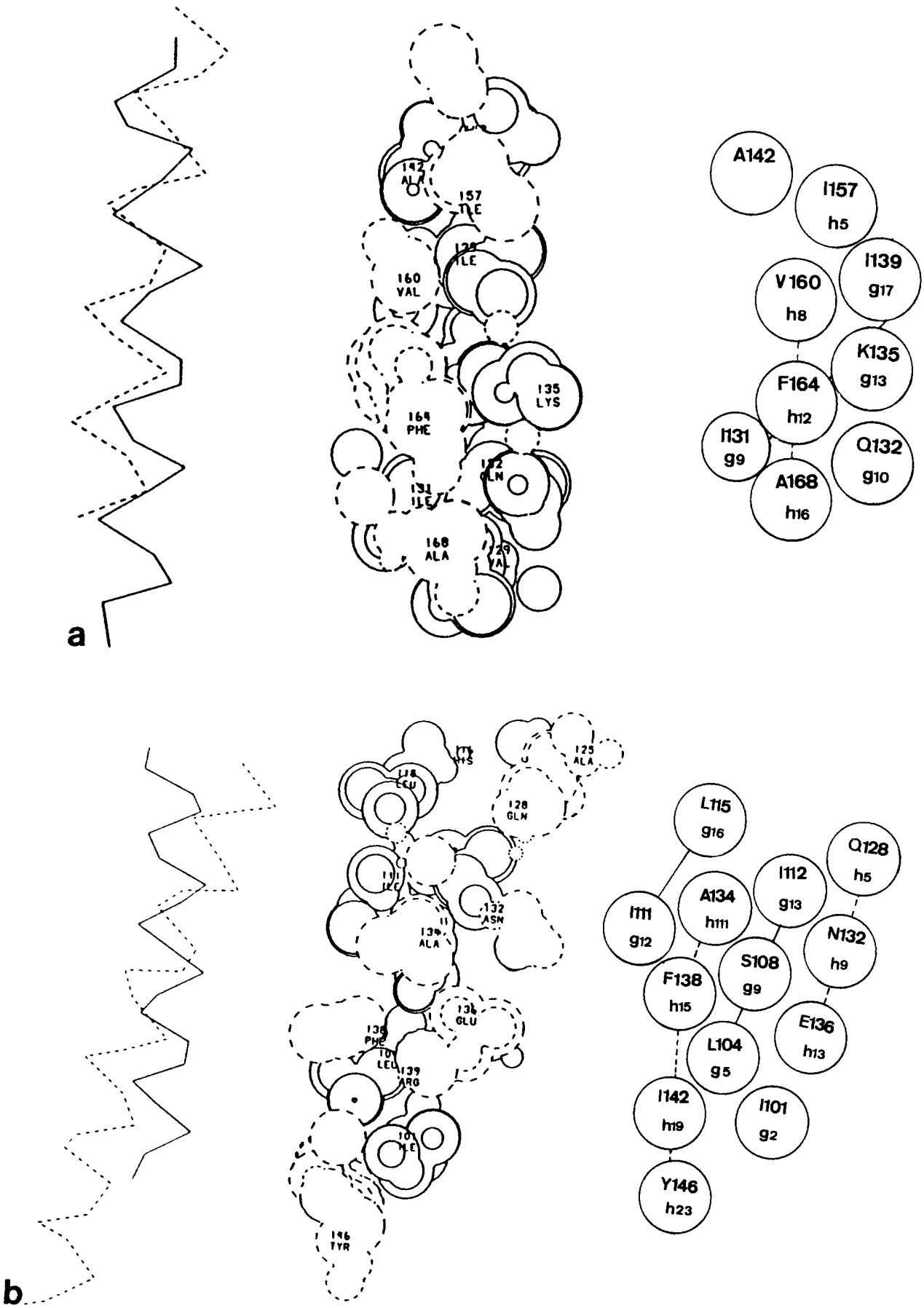

a
142 ALA
157 ILE
160 VAL
135 LYS
164 PHE
131 ILE
168 ALA
A142
I157 h5
I139 g17
V160 h8
K135 g13
F164 h12
I131 g9
Q132 g10
A168 h16
b
125 ALA
128 GLN
132 ASN
134 ALA
136 GLU
139 ARG
101 ILE
146 TYR
L115 g16
Q128 h5
I112 g13
I111 g12
A134 h11
N132 h9
S108 g9
F138 h15
E136 h13
L104 g5
I142 h19
I101 g2
Y146 h23

TABLE VII. Indications of Alignment of Residues from Packing Patterns at Interfaces Between Pairs of Helices*

	A	*B*	*E*	*G*	*H*
A			96α = 72Apl		150α = 130Apl
B			85α = 64Apl	135α = 122Lam	
			88α = 64Apl		
E	44α = 14Apl	56α = 25Apl			
	41α = 14Apl	56α = 29Apl			
G		55α = 41Lam			164β = 135SWM
		55α = 37Lam			160β = 135SWM
					160β = 139SWM
H	40α = 10Apl			135β = 112SWM	
	37α = 10Apl			139β = 112SWM	

*SWM stands for sperm whale myoglobin and Lam for lamprey globin. For the globin alignment see Table VIII.

Reexamining the *B/E* contact in light of this alignment of the *B* helix implies the alignment of the *E* helix: PCMLα88 = Apl64 *(E8)*. This is consistent with the indication from the *A/E* contact that PCMLα96 = Apl72 *(E6)*.

Reexamining the *A/E* contact, having fixed the alignment of the *E* helix, we are led to the alignment PCMLα41 = Apl14 *(A12)*.

We can now use the alignment of the *A* helix to fix the alignment of the *H* helix. Considering the *A/H* interface, the alignment of the *A* helix implies that for the *H* helix, PCMLα150 = Apl130 *(H6)*. This means that PCMLα154 = PCMLβ164 = SWM135 *(H12)*. This is one of the possibilities for the *H* helix, derived from comparing the *G/H* interfaces of PCMLβ and SWM).

Referring again to the *G/H* interface, the alignment of the *H* helix implies that, for the *G* helix, PCMLβ135 = SWM112 *(G13)*. But PCMLβ135 = PCMLα137, and SWM112 = Lam124 = Apl113, so that this alignment is consistent with the pattern observed in the *B/G* interface.

This discussion shows that it is possible to derive a self-consistent alignment of helices *A, B, E, G,* and *H,* with reasonable residue correspondences at each helix interface. This alignment is shown in Table VIII. It includes an alignment of the *C* helices that is most reasonable both from the sequences and from the relative spatial positions of the residues in the *C* helices as seen in Figure 10. The alignment of the *F* region was derived separately and is discussed in the next section.

The alignment derived from the individual helix interfaces, together with the *C* regions, includes the following portions of the helices of PCMLα and Apl:

	PCMLα		Apl	
A	34	45	7	18
B	50	62	23	35
C	63	68	36	41
E	81	100	57	76
G	126	139	102	115
H	145	162	125	142

A fit of the Cα atoms of these residues gave a root-mean-square deviation Δ = 3.84 Å. A superposition of these regions is shown in Figure 10. Because of shifts and rotations not all portions of the helices fit well. The C region in particular is highly displaced. What is significant is that this picture is consistent with an optimal alignment of two distorted structures; we do not see evidence that any of the regions has been misaligned; this would most likely appear as a shift along a helix axis by a turn.

By eliminating the most poorly fitting residues from the superposition, we can reduce Δ to 2.74 Å leaving the residues:

	PCMLα		Apl	
A	34	45	7	18
B	50	58	23	31
C		—		—
E	81	100	57	76
G	135	139	111	115
H	145	157	125	137

The superposition of these regions is shown in Figure 11.

In order to try to confirm this alignment further, we did the analog of crystallographers' "omit maps." For each of these five helices *A, B, E, G,* and *H,* we superposed four of the five remaining trimmed helices, and drew pictures of the remaining pair in the position and orientation derived from the superposition. Again the results did not suggest any misalignments.

Fig. 8. Comparison of the *G/H* interface in PCMLβ **(a)** and in sperm whale myoglobin **(b)** and the corresponding residue packing. Residues from the *H* helix are represented as broken lines.

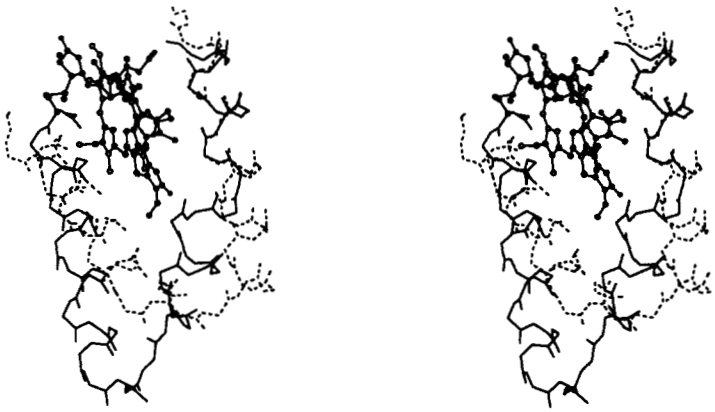


Fig. 9. Comparison between the regions around the prosthetic group of PCMLα (solid lines and filled circles) and Apl (broken lines and open circles). The two structures are superposed by fitting the *F'* helix of PCMLα with the corresponding region in Apl.

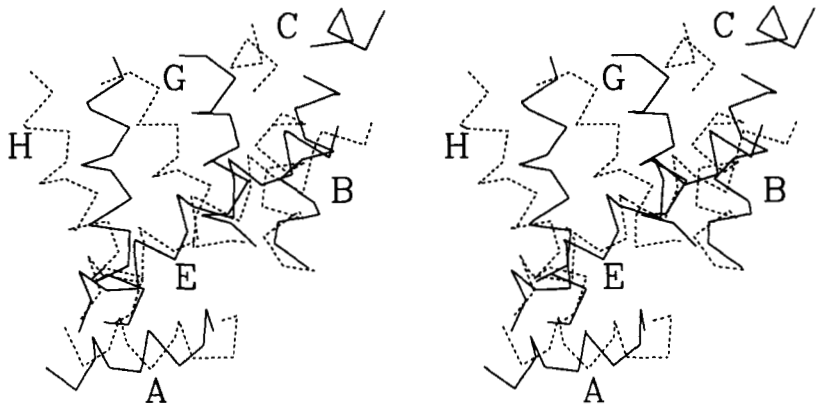


Fig. 10. Superposition of the residues common to helices *A, B, E, G,* and *H.* The Cα trace of PCMLα and of Apl are represented as solid lines and broken lines, respectively.

The *F'F* Helix Alignment

The alignment of the *F'* and *F* helices poses a problem that cannot be solved by appeal to packing patterns at helix interfaces. The *F'* and *F* helices in phycocyanins do not form well-developed interfaces with other helices. In this case we determined the best fit for this region by inspection of superposed structures by molecular graphics. We fitted together the *E* and the *G* helices, using the alignment previously obtained for them, and looked at the spatial relationships of the residues in the *F'F* helices. Both the different length of the *F* helix in the globin and the break in this helix in phycocyanins make the alignment of this region obscure. We fitted together the last two helix turns of the *F* helix (residues 115 to 119 in PCMLα with residues 91 to 95 in Apl). The result is shown in Figure 12. This clearly shows the one-residue insertion between the *F'* and *F* helices and the rather longer region between *E* and *F* helices in Apl. In the alignment we propose, residues 105 to 110 in PCMLα correspond to residues 82 to 87 in Apl (phycocyanin *F'* helix) and residues 115 to 119 in PCMLα correspond to 91 to 95.

This alignment seemed to us the most convincing one.

Fits and Shifts of the Helices

Using the alignment derived in the previous sections, it is possible to determine the differences in relative position and orientation of corresponding pairs of packed helices in globins and phycocyanins. This is done as follows. For each pair of packed helices one of the helices is superposed on the corresponding helix in the other molecule. Then the translation and rotation required to superpose the other pair of corresponding helices are calculated:

TABLE VIII. Alignment of the Four Phycocyanin Sequences, Lamprey Globin (Lam), Sperm Whale Myoglobin (SWM), and *Aplysia limacina* Globin (Apl)*

```
                    1         2         3         4         5
                    0         0         0         0         0
                  xxxxxxxxxxxx      yyyyyyyyyyyyyy aaaaaaaaaaaaaaa bbb
PCMLα       VKTPITDAIAAADTQGRFLSNTELQAVNGRYQRAAASLEAARALTANAQRLI
PCAQα       MKTPLTEAVALADSQGRFLSNTELQYLYGRLRQGAFALEAAQTLTAKADTLV
PCMLβ       AYDVFTKVVSQADSRGEFLSNEQLDALANVVKEGNKRLDVVNRITSNASTIV
PCAQβ       MFDIFTRVVSQADARGEFISSDKLEALKKVVAEGTKRSDAVSRMTNNASSIV
Lam                           PIVDTGSVAPLSAAEKTKIRSAWAPVYSTYETSG
SWM                                   VLSEGEWQLVLHVWAKVEADVAGHG
Apl                                   SLSAAEADLAGKSWAPVFANKNANG
                                        aaaaaaaaaaaaaaaa bbbbbb
                                                1         2
                                                0         0

                                                                1
                    6         7                   8         9         0
                    0         0                   0         0         0
                  bbbbbbbbbbccccccc               eeeeeeeeeeeeeeeeeeeeeee
PCMLα       DGAAQAVYQKFPYLIQTSGPNYAA----DARGKSKCARDIGHYLRIITYSLV
PCAQα       NGAAQAVYSKFPYTTSTPGNNFAA----DQRGKDKCARDIGYYLRMVTYCLV
PCMLβ       TNAARALFEEQPQLIAPGGSATRN----GT--MAACLRDMEIILRYITYAIL
PCAQβ       TNAARQLFADQPQLIAPGGNAYTN----RR--MAACLRDMEIILRYVTYATF
Lam         VDILVKFFTSTPAAQEFFPKFKGLTTADQLKKSADVRWHAERIINAVDDAVA
SWM         QDILIRLFKSHPETLEKFDRFKHLKTEAEMKASEDLKKHGVTVLTALGAILK
Apl         LDFLVALFEKFPDSANFFADFKGKSVA-DIKASPKLRDVSSRIFTRLNEFVN
            bbbbbbbbbbccccccc          ddddddeeeeeeeeeeeeeeeeeee
                3         4         5         6         7
                0         0         0         0         0

                      1         1         1         1
                      1         2         3         4
                      0         0         0         0
            ee  fffffff   fffffffff  999999999999999
PCMLα       AG-GTGPLDEYLIAGLNEINDAFELSPSWYINALKYIKANHGLS---------
PCAQα       AG-GTGPMDEYLIAGVDEINRTFDLSPSWYVEALKHIKANHGLT---------
PCMLβ       AG-DASILDDRCLNGLRETYQALGTPGSSVAVGIQKMKEAAINIANDPNGIT
PCAQβ       TG-DASVLNDRCLNGLRETYVALGVPGASVAAGVRAMGKAAVAIVMDPAGVT
Lam         SMDDTEKMSMKLRNLSGKHAKSFQVDPQYFKVLAAVIADTV----------
SWM         K K  GHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGNF---
Apl         NAANAGKMSAMLSQFAKEHV-GFGVGSANFENVRSMFPGFVASV---------
                      fffffffffff  99999999999999999
                                        1         1
                8         9         0         1
                0         0         0         0

                      1         1
                      5         6
                      0         0
                    hhhhhhhhhhhhhhhhhh
PCMLα       ----GQAANEANTYIDYVINALS
PCAQα       ----GDAATETNNYIDYAINALS
PCMLβ       KGDCSALISEVASYFDRAAAAVA
PCAQβ       SGDCSSLQQEIELYFETAAKAVE
Lam         ----AAGDAGFEKLMSMICILLRSAY
SWM         ---GADAQGAMNKALELFRKDIAAKYKELGYQG
Apl         AAPPAGADAAWTKLFGLIIDALKAAGK
              hhhhhhhhhhhhhhhhhhhhhhhhhh
            1         1         1
            2         3         4
            0         0         0
```

*The lower case letter above and below the sequences refer, respectively to the helix assignment in phycocyanins, and to Kendrew notation for globins.

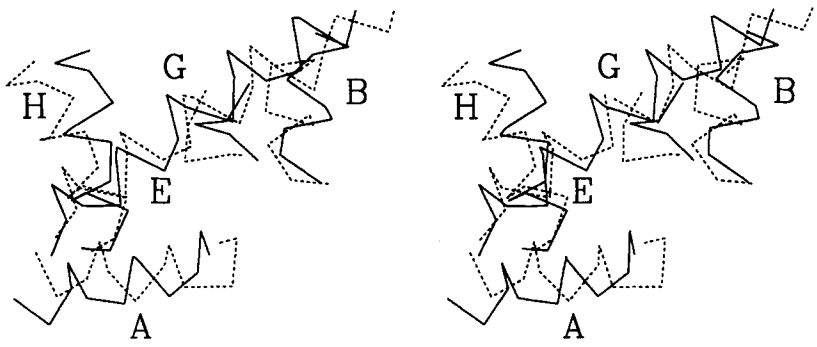


Fig. 11. Superposition of a subset of the residues common to helices *A*, *B*, *E*, *G*, and *H*, trimmed to an overall rms deviation of Cα atoms of 2.74 Å. The Cα trace of PCMLα and of Apl are represented as solid lines and broken lines, respectively.

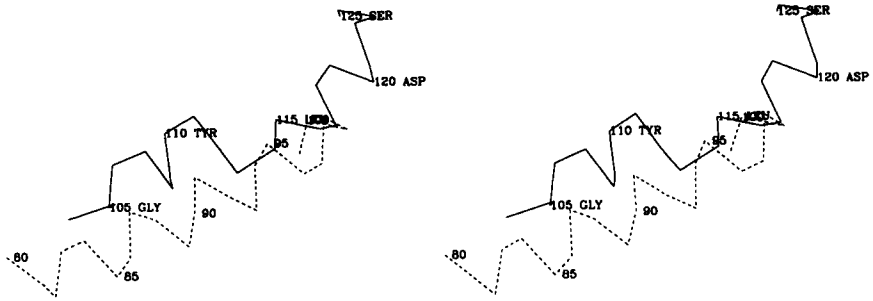


Fig. 12. Fit of the *F* region in PCMLα and Apl, after superposition shown in Figure 11.

the translation distance and rotation angle are measures of the changes in the relative geometry of the helices.

The results of these calculations appear in Table IX. The displacements are of the same order of magnitude (although slightly larger) than the largest displacements observed in distantly related globins, but the changes in angles are substantially larger than any observed in distantly related globins (cf. ref. 4, Figs. 7 and 12).

The alignment of phycocyanins and sperm whale myoglobin published by Schirmer et al.[1] is identical with ours for the *A*, *B*, *C*, and *H* helices. Their alignment of the *G* helix is shifted by 4 (it is consistent with one of our alternative alignments of this helix). Their alignment of the *E* helix is shifted by three residues, aligning position *E7* in globins with Cyc84 in phycocyanins.

TEMPLATES AND EXPOSURE TO THE SOLVENT

Bashford et al.[18] have identified a set of 66 conserved positions common to all the globins. These motifs constitute a set of rules or templates that identifies virtually all the proteins belonging to the globin class. Each motif, representing a unit of the secondary structure such as a helix, consists of an explicit list of allowable residues at each position. In applying a template to scan a sequence, each motif is allowed to slide along the sequence and a score between 0 and 1 is given to the violations from the set of allowed residues at each position. Insertions and deletions within motifs are not allowed. A low score corresponds to a good match.

Although the proposed templates have been shown to be satisfactory in selecting globin sequences from the sequence database, they do not also select the phycocyanins.[18] After an extension of the template removing some of the features considered stringent for globins, it has become possible to match all the phycocyanins present in the databank with a score comparable to the one reported for globins. However, in addition to the phycocyanins, 1330 other sequences were found to match the template. This shows the lack of specificity introduced by going to too general templates.

The alignment derived from the templates is consistent with the one we derived from the structure for all helices except the *G* helix. A detailed examination of the differences of the patterns of residue

TABLE IX. Differences in Relative Position and Orientation of Corresponding Pairs of Packed Helices in Globins and Phycocyanins Using the Alignment Derived in the Text (see also Table VIII)*

(a)

	A	*B*	*E*	*G*	*H*
A			6.5Å/12°		7.4 Å/46°
B			4.9Å/36°	7.1Å/11°	
E	4.5Å/11°	3.7Å/36°			
G		8.5Å/11°			7.0Å/43°
H	2.0Å/46°			4.8Å/43°	

(b)

	A	*B*	*E*	*G*	*H*
A			6.5Å/12°		10.0Å/47°
B			5.0Å/36°	6.6Å/19°	
E	4.5Å/12°	4.9Å/36°			
G		4.9Å/19°			5.9Å/26°
H	2.1Å/47°			4.5Å/26°	

*Each entry in the table corresponds to calculations on a pair of packed helices. For example, the entry corresponding to the *A/E* helices, 6.5 Å/12° was computed as follows: The *A* helices of PCML1 and Apl were superposed. Then the transformation required subsequently to superpose the *E* helices was calculated: this required a translation of 6.5 Å and a rotation of 12°. Note that doing these calculations in reverse order (first superposing the *E* helices and then finding the transformation required to superpose the *A* helices) gives a different result. In each case, the first superposition involved the helix named at the left of the row, and the second superposition involved the helix named at the top of the column. In (a) the calculation is done by using the helical regions which would give a fit of the Cα smaller or equal Δ = 3.84 Å. In (b) the regions were reduced to give a fitting of Δ = 2.74 Å.

conservation in globins and phycocyanins will be carried out in future work.

CONCLUSIONS

Having presented an alignment, and a description based on it of the similarities and the differences between globins and phycocyanins, we now return to the question of whether they are related by evolution or not.

A basic problem is the distinction between points of similarity and criteria for relationship. It is useful in this context to quote from W.E. Le Gros Clark: "While it may be broadly accepted that, as a general proposition, degrees of genetic relationship can be assessed by noting degrees of resemblance in anatomical details, it needs to be emphasized that morphological characters vary considerably in their significance for this assessment. Consequently it is of the utmost importance that particular attention should be given to those characters whose taxonomic relevance has been duly established by comparative anatomical and palaentological studies."[19] For the problem at hand, the implication is that we must weigh the structural similarities differentially, disregarding those features that are common to many other classes of proteins, and emphasizing the unusual or unique ones.

In light of this, we suggest that the following points argue for the hypothesis of evolutionary relationship and common ancestry:

1. One of the properties of a protein structure that tends to be the most well conserved in evolution is the reticulation, or pattern of interactions, of residues at the interfaces between packed secondary structures. Although in many cases the interhelix axis distances and angles are quite different, most of the packed helices show the same types of packing in their interfaces: *A/H, B/E, B/G, F/H, G/H* (see Table VI).

2. Two of these interfaces are particularly significant because they are unusual: The *B/E* and *G/H* packings. In the globins, these interfaces show the unusual "crossed-ridge" structure. For the *B/E* pair we find a similar crossing in phycocyanins: at the crossing point in Apl there are G25–S64, corresponding to I52–I88 in PCMLα (see Fig. 6). This is particularly significant, because in phycocyanins the interaxial angles in the *B/E* interfaces are all in the region of −50° to −60° that would be expected for the normal $\pm 4n$–$\pm 4n$ packing. By contrast, in the globins the interaxial angles of the *B/E* contacts are somewhat larger, in the range −60° to −80° (see Fig. 6). Therefore it cannot be argued that the phycocyanin structure requires the special crossed-ridge packing to achieve an unusual interaxial angle.

The *G/H* interface is quite irregular in the α- and β-chains of phycocyanins. This helix pair is quite variable in the globin family too. The crossing point in Apl at R109/F134 perfectly corresponds to the crossing point K135/F164 in PCMLβ, where not only the size is conserved but also the types of residues.

3. An unusual feature common to all the globins is the short 3–10 *C* helix between the *B* and *E* regions. In Apl, it involves five i-$i+3$ hydrogen bonds and spans the residues F6 and F43. In the same region of phycocyanins, we find a shorter but well defined 3–10 helix between residue F63 and I67 with two hydrogen bonds. It could not be argued that a 3–10 *C* helix is required as a necessary solution of a structural problem, because the length of the region between the *B* and *E* helices is quite long, and could take up a variety of conformations; indeed within the globins this region sometimes contains a *D* helix and sometimes does not.

4. If we examine the most conserved positions of buried residues of globins and we compare them with the corresponding residues in the entire known family of phycocyanins (eight sequences) we find the following list:

B14	LF	YF
C2	P	P
CD1	F/h	SPG
E11	VIL/h	YI
FG4	VI/h	WSL
G5	LF/h	LVI

where we indicate by "/h" the occurrence of interactions with the heme. We find thus two positions well conserved across phycocyanins and globins; although the others present a larger variability, they retain their hydrophobic character and, to a large extent, their range of side chain volumes. This would seem to be consistent with the fact that positions *B14* and *C2* are not involved in the interactions with the ligand, and therefore play a role in determining the structure or, perhaps, the folding pathway.

SUMMARY

We have described the similarities and differences between the globins and phycocyanins, to provide data to distinguish between the two possible hypotheses for their apparent similarity: common ancestry, or a similar solution to a structural problem. We have performed a detailed structural analysis of the helix packing in the phycocyanins and compared their features to those of globins. We have found that the helix packing, one of the best conserved structural features in proteins, is in most cases similar in the two classes, so that, on the basis of contact patterns, it is possible to trace a self-consistent alignment.

On the basis of this alignment, it is possible to describe structural changes between phycocyanins and globins. Previous work on evolution in protein families has treated sets of molecules of similar function. For such molecules, it has been observed that the pattern of residue packing at interfaces between secondary structures is conserved, although this allows large local structural changes in the relative geometry of packed secondary structures. However, in order to preserve function, these changes are coupled and integrated over all or most of the molecule. In the globin–phycocyanin case we observe the retention of the packing pattern of residues at interfaces and the large local changes characteristic of other families. However, the constraint of coupling the changes over the entire molecule has been released by the alteration in function, or, rather, has been replaced by an alternative set of global constraints.

In conclusion, we suggest that the results presented here support the idea that the globins and phycocyanins are distantly related members of a common family of proteins, that the major differences between them are the result of divergent evolution from a common ancestor, and that the major differences are in fact extremely large because of the requirements of the different functions that these proteins carry out.

ACKNOWLEDGMENTS

We thank Drs. W. Bode, R. Huber, and T. Schirmer for the coordinates of the phycocyanin structures. We are grateful to them and to Drs. C. Chothia and O.B. Ptitsyn for critical readings of the manuscript. We thank Dr. D. Bashford for his template-matching program.

REFERENCES

1. Schirmer, T., Bode, W., Huber, R., Sidler, W., Zuber, H. X-ray crystallographic structure of the light-harvesting biliprotein C-phycocyanin from the thermophilic cyanobacterium *Mastigocladus laminosus* and its resemblance to globin structures. J. Mol. Biol. 184:257–277, 1985.
2. Schirmer, T., Huber, R., Schneider, M., Bode W., Miller, M., Hackert, M.L. Crystal structure analysis and refinement at 2.5 Å of hexameric C-phycocyanin from the cyanobacterium *Agmenellum quadruplicatum*. J. Mol. Biol. 188: 651–676, 1986.
3. Schirmer, T., Bode, W., Huber, R. Refined three-dimensional structures of two cyanobacterial C-phycocyanins at 2.1 and 2.5 Å resolution. J. Mol. Biol. 196:677–695, 1987.
4. Lesk, A.M., Chothia, C. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. J. Mol. Biol. 136:225–270, 1980.
5. Dickerson, R.E., Geis, I. "Haemoglobin." Menlo Park, CA: Benjamin/Cummings, 1983.
6. Lesk, A.M., Brändén, C.-I., Chothia, C. Structural principles of α–β barrel proteins: The packing of the interior of the sheet. Proteins: Structure, Function Genet. 5:139–148, 1989.
7. Bajaj, M., Blundell, T. Evolution and the tertiary structure of proteins. Annu. Rev. Biophys. Bioeng. 13:453–492, 1984.
8. Lesk, A.M., Chothia, C. The response of protein structures to amino acid sequence changes. Phil. Trans. R. Soc. London 317:345–356, 1986.
9. Chothia, C., Lesk, A.M. Relationship between the divergence of sequence and structure in proteins. EMBO J. 5: 823–826, 1986.
10. Pastore, A., Lesk, A.M., Bolognesi, M., Onesti, S. Structural alignment and analysis of two distantly related pro-

teins: *Aplysia limacina* myoglobin and sea lamprey globin. Proteins: Structure. Function Genet. 4:240–250, 1988.
11. Glazer, A.N. Phycobilisome: A macromolecular complex optimised for light energy transfer. Biochim. Biophys. Acta 768:29–51, 1984.
12. George, D.G., Hunt, L.T., Barker, W.C. The Protein Identification Resource (PIR) software. In: "Computational Molecular Biology: Sources and Methods for Sequence Analysis," A.M. Lesk, (ed.). Oxford: Oxford University Press, 1988.
13. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
14. Bolognesi, M., Onesti, S., Gatti, G., Coda, A., Ascenzi, P., Brunori, M. *Aplysia limacina* myoglobin: Crystallographic analysis at 1.6 Å resolution. J. Mol. Biol. 205:529–544, 1989.
15. Lesk, A.M. Integrated access to sequence and structural data. In: "Biosequences: Perspectives and User Services in Europe," C. Saccone (ed.). Bruxelles: EEC, 1986:23–28, and references contained therein.
16. Chothia, C., Levitt, M., Richardson, D. Structure of proteins: Packing of α-helices and pleated sheets. Proc. Natl. Acad. Sci. U.S.A. 74:4130–4134, 1977.
17. Chothia, C., Levitt, M., Richardson, D. Helix to helix packing in proteins. J. Mol. Biol. 145:215–250, 1981.
18. Bashford, D., Chothia, C., Lesk, A.M. Determinants of protein fold: Unique features of the globin amino acid sequences. J. Mol. Biol. 196:199–216, 1987.
19. Clark, W.E. Le Gros. "The Antecedents of Man," 3rd ed. Chicago: Quadrangle Books, 1971: 11–12.
20. Shaanan, B. Structure of human oxyhaemoglobin at 2.1 angstroms resolution. J. Mol. Biol. 171:31–59, 1983.
21. Fermi, G., Perutz, M.F., Shaanan, B., Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74 angstroms resolution. J. Mol. Biol. 175:159–174, 1984.
22. Phillips, S.E.V. Structure and refinement of oxymyoglobin at 1.6 angstroms resolution. J. Mol. Biol. 142:531–554, 1980.
23. Steigemann, W., Weber, E. Structure of erythrocruorin in different ligand states refined at 1.4 angstroms resolution. J. Mol. Biol. 127:309–316, 1979.
24. Honzatko, R.B., Hendrickson, W.A., Love, W.E. Refinement of a molecular model for lamprey hemoglobin from *Petromyzon marinus*. J. Mol. Biol. 184:147–164, 1985.
25. Harutyunyan, E.H., Kuranova, I.P., Vainshtein, B.K. X-Ray structural investigation of leghemoglobin. Structure of acetate-ferrileghemoglobin at a resolution of 2.0 angstroms. Kristallografiya 25:80–92, 1980.
26. Arents, G., Love, W.E. *Glycera dibranchiata* hemoglobin: Structure refinement at 1.5 Å resolution. J. Mol. Biol. 210: 149–161, 1989.
27. Royer, W.E., Jr., Hendrickson, W.A., Chiancone, E. The 2.4-Å crystal structure of *Scapharca* dimeric hemoglobin. J. Biol. Chem. 264:21052–21061, 1989.