Acta Cryst. (1997). D53, 302–310

Structure of Ferric Soybean Leghemoglobin *a* Nicotinate at 2.3 Å Resolution

PAUL J. ELLIS,[a] CYRIL A. APPLEBY,[b] J. MITCHELL GUSS,[a]† WILLIAM N. HUNTER,[a]‡ DAVID L. OLLIS[a]§ AND HANS C. FREEMAN[a]*

[a]*School of Chemistry, University of Sydney, NSW 2006, Australia, and* [b]*Division of Plant Industry, CSIRO, Canberra, ACT 2601, Australia. E-mail: freemanh@chem.usyd.edu.au*

(*Received 21 October 1996; accepted 7 January 1997*)

Abstract

Soybean leghemoglobin *a* is a small (16 kDa) protein facilitating the transport of O_2 to respiring N_2-fixing bacteria at low free-O_2 tension. The crystal structure of soybean ferric leghemoglobin *a* nicotinate has been refined at 2.3 Å resolution. The final *R* factor is 15.8% for 6877 reflections between 6.0 and 2.3 Å. The structure of soybean leghemoglobin *a* (143 residues) is closely similar to that of lupin leghemoglobin II (153 residues), the proteins having 82 identical residues when the sequences are aligned. The new structure provides support for the conclusion that the unique properties of leghemoglobin arise principally from a heme pocket considerably larger and more flexible than that of myoglobin, a strongly ruffled heme group, and a proximal histidine orientation more favourable to ligand binding.

1. Introduction

The monomeric leghemoglobins (Lb's) of the symbiotic tissue of nitrogen-fixing legume root nodules have a function analogous to that of animal cytoplasmic myoglobins: the reversible binding and transport of oxygen by facilitated diffusion (Appleby, 1984, 1992). In animal cells this oxygen is used for respiration and energy generation by oxygen-tolerant mitochondria. Within nodule symbiotic tissue it is consumed by the rapidly respiring yet oxygen-intolerant *Rhizobium*, *Bradyrhizobium* or *Azorhizobium* bacteroids to generate the large amount of ATP necessary for the reduction of nitrogen to ammonia. The leghemoglobins have very high oxygen affinities (Gibson, Wittenberg, Wittenberg, Bogusz & Appleby, 1989), so that a sufficient oxygen flux is achieved at a free concentration (~5–20 n*M*) which is too low to inhibit, damage or prevent biosynthesis of the oxygen-sensitive nitrogen-reducing enzymes within these bacteroids.

The leghemoglobins of the various legume symbioses consist of a 16–17 kDa polypeptide chain synthesized in the host cytoplasm of bacteroid-infected plant cells, and a protoheme moiety that is also a plant product (O'Brian, 1996). The assembled leghemoglobin remains in the host cell cytoplasm, being separated from intimate contact with bacteroid surfaces by the presence of peribacteroid membranes that surround each small group of bacteroids (Appleby, 1992). Kinetics measurements (Gibson *et al.*, 1989) reveal that the high oxygen affinity of leghemoglobins is due to association ('on') rate constants 10–20 times faster than for myoglobins, with dissociation ('off') rate constants of the same order as for myoglobins. The normal 'off' rate is essential for the biological function of leghemoglobins in facilitated diffusion. Other unusual properties include the ability to bind bulky ligands such as nicotinic acid and carboxylic acids from acetic to valeric (C1 to C5) (Ellfolk, 1961; Appleby, 1984; Atanasov, Dimitrova, Kudryatseva, Zhiznevskaya & Appleby, 1989).

Most of the studies of Lb in solution have dealt with Lb *a* from soybean (*Glycine max*), while most of the X-ray crystallographic studies have centered on Lb II from yellow lupin (*Lupinus luteus*) (see, for example, Narula, Dalvit, Appleby & Wright, 1988; Harutyunyan *et al.*, 1995). The soybean and lupin proteins are distant on the phylogenetic tree (Andersson, Jensen, Llewellyn, Dennis & Peacock, 1997), and differ significantly in length: soybean Lb *a* has 143 residues while lupin Lb II has 153 residues. When the two sequences are aligned, they have identical residues at 82 positions.

To help understand how Lb's function we present here the crystal structure of soybean ferric Lb *a* nicotinate refined at 2.3 Å resolution. Nicotinate (pyridine 3-carboxylate) binds to ferric Lb about 50 000-fold more strongly than to ferric myoglobin. Although the nicotinate anion complex is not of physiological significance (Klucas & Appleby, 1991), the crystal structure analysis has revealed features that may help to explain the fast oxygenation rate constants of the leghemoglobins.

2. Experimental

The structure of soybean Lb *a* nicotinate was solved by the method of multiple isomorphous replacement (MIR) at 3.3 Å resolution, as described previously (Ollis *et al.*, 1983). We here report the refinement at 2.3 Å resolution.

† Present address: Department of Biochemistry, University of Sydney, NSW 2006, Australia.

‡ Present address: Department of Biochemistry, The University, Dundee DD1 4HN, Scotland.

§ Present address: Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia.

Table 1. *Data collection and refinement*

Space group	$P2_1$
Unit-cell dimensions (Å, °)	$a=52.69, b=56.55, c=57.80$ $\beta=109.9$
Molecules/asymmetric unit	2
No. of crystals	6
No. of measurements	18416
No. of unique reflections	14153
R_{merge} (%)	6.1
No. of reflections used in refinement	6877
Criterion for inclusion	$F \geq 2\sigma(F)$
Resolution (Å)	6.0–2.3
Final R factor	0.158
No. of non-H atoms	2332
No. of solvent atoms	70
Range of B values (Å^2)	2.0–35.0
Mean B values (Å^2)	
All atoms	12.3
Protein atoms	12.2
Water molecules	15.1
R.m.s. deviations from ideal values	
Bonds (Å)	0.009
Bond angles (°)	1.5

Selected unit-cell properties, data-collection statistics, and results of the refinement are shown in Table 1.

The initial coordinates used in the present work were obtained from a partial refinement of the MIR structure without non-crystallographic symmetry restraints (W. N. Hunter, unpublished work). These coordinates were modified by imposing the non-crystallographic symmetry on the two independent molecules, setting the B factors of all atoms to 20 Å^2, manually optimizing the fit to F_{obs} maps, and deleting about 100 atoms that could not be accommodated in good density. A similar starting model would have been obtained if the same modifications had been applied to the MIR structure.

Refinement was carried out by restrained least-squares, using *PROLSQ* (Hendrickson & Konnert, 1980; Hendrickson, 1985) and later *PROFFT* (Finzel, 1987), alternating with manual intervention. The program used for manual adjustments of the model was *FRODO* (Jones, 1978). Electron-density maps were calculated with coefficients F_{obs}, $F_{obs}-F_{calc}$ and $2Fobs-F_{calc}$. All regions of the structure were subjected to detailed examination using omit maps. During the refinement, the data range was extended in stages from 5.0–2.5 Å (5584 reflections) to 6.0–2.3 Å (6877 reflections), the initially tight non-crystallographic symmetry restraints were relaxed and finally removed, and H atoms were added. The Fe—ligand bonds were weakly restrained during the early stages of the refinement, and unrestrained during the final stages. Only restraints to keep each heme pyrrole ring and the two adjacent methene C atoms coplanar were applied throughout the refinement. Electron-density peaks with amplitude $\geq 4\sigma$ in $F_{obs}-F_{calc}$ or omit maps, and with one or more plausible hydrogen-bonding contacts, were assigned as water molecules. Water molecules whose amplitude fell below 4σ, or whose B value exceeded 35 Å^2, were deleted. The refinement was completed after 24 rounds of automatic refinement and manual rebuilding. The atomic coordinates have been deposited with the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).*

3. Results

3.1. *Assessment of accuracy and precision*

The final model comprises 2332 non-H and 2258 H atoms, representing two Lb molecules plus 70 solvent molecules [treated as O(water) atoms]. Each Lb molecule consists of the polypeptide, a heme group, and a coordinated nicotinate. The structure has a residual $R=15.8\%$ for 6877 reflections with $F_{obs} \geq 2\sigma(F)$ in the range 6.0–2.3 Å. The r.m.s. deviations from the dictionary geometry are 0.009 Å for bond distances and 1.5° for bond angles, respectively. A number of indicators suggest that the uncertainty in the light-atom positions is about 0.2 Å. (i) The maximum average error in the atomic positions estimated from a Luzzati plot (Luzzati, 1952) is 0.18 Å. (ii) The r.m.s. difference between corresponding Fe—N bond distances in the two independent molecules is 0.17 Å. We assume that the contribution of the Fe-atom positions to this difference is negligible in comparison with the N-atom positions. (iii) When the two independent molecules are superposed, the r.m.s. difference between the positions of corresponding backbone atoms is 0.36 Å. If we make the conservative assumption that this difference is due entirely to errors, then the r.m.s. error in the individual atomic positions is 0.25 Å (*i.e.*, $0.36/2^{1/2}$).

In Ramachandran plots for the final structure (Ramachandran & Sasisekharan, 1968), only one non-glycine residue lies significantly outside the allowed regions. The aberrant residue is Lys142, the residue before the C terminus, in molecule *B*. The torsion angles for this residue in molecule *A* are normal. No importance is attached to this discrepancy, as the two C-terminal residues (Lys142 and Ala143) of both molecules are in weak density and have high temperature factors (25–31 Å^2).

3.2. *Description of the molecule*

Soybean Lb *a* (Fig. 1) has the same 'globin fold' as other monomeric hemoglobins (Bashford, Chothia & Lesk, 1987). The seven helices (*A–C*, *E–H*) common to all globin structures account for approximately 75% of the amino acids. The refinement of the structure and a rigorous application of the Kabsch & Sander (1983) criteria have resulted in several minor changes from the original helix assignment of Ollis *et al.* (1983) (Fig. 2).

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1FSL, R1FSLSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: gr0686).

The temperature factors B averaged over the backbone atoms of each residue have similar patterns in both molecules (Fig. 3). The temperature factors are lowest in the well ordered α-helical regions, and become higher in the 3_{10}-helix C and in the interhelical regions. The highest temperature factors are observed in the carboxy-terminal region.

3.3. *The heme environment*

As in other hemoglobins, the heme in soybean Lb *a* is located in a hydrophobic pocket between the E and F helices. The heme is oriented with the propionate groups projecting away from the hydrophobic pocket and into the solvent (Fig. 4). The side chains which line the heme pocket belong to residues in the B, C, E, F, G and H helices, and in the CD and FG interhelical regions. The amino acids in the immediate vicinity of the heme are shown in Fig. 5. The heme is strongly ruffled (see below).

The heme divides the pocket into two sides, proximal and distal. On the proximal side, the heme is linked to the polypeptide by a bond between the Fe atom and the $N^{\epsilon 2}$(imidazole) atom of the conserved proximal histidine, His$F7$. On the distal side, the Fe atom may be coordinated by an exogenous ligand, resulting in a quasi-octahedral geometry. The binding of exogenous ligands is influenced by steric and hydrogen-bonding interactions with residues inside the heme pocket. Prominent

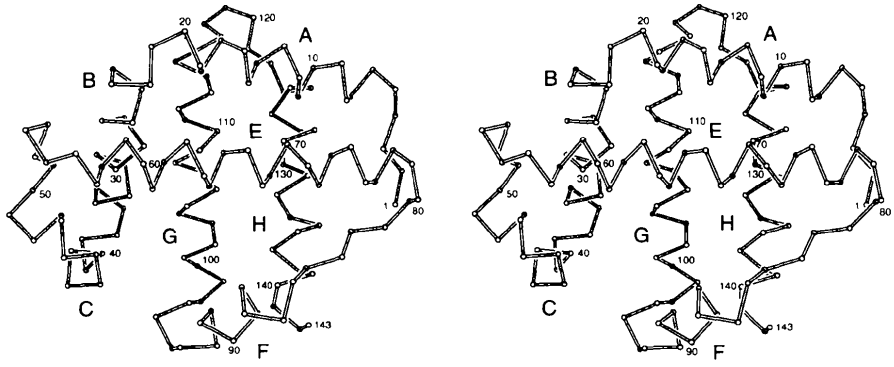

Fig. 1. C^{α} diagram of soybean leghemoglobin *a*, showing the conventional helix labels. This diagram and Figs. 3–10 are based on the coordinates of molecule *A*.

1				5				10	A				15							20
Val	Ala	Phe	Thr	Glu	Lys	Gln	Asp	Ala	Leu	Val	Ser	Ser	Ser	Phe	Glu	Ala	Phe	Lys	Ala	
NA1			NA4	A1				A5					A10					A15	A16	
				25				B	30					35						40
Asn	Ile	Pro	Gln	Tyr	Ser	Val	Val	Phe	Tyr	Thr	Ser	Ile	Leu	Glu	Lys	Ala	Pro	Ala	Ala	
AB1	B1				B5					B10					B15	BC1	C1			
C				45					50					55						60
Lys	Asp	Leu	Phe	Ser	Phe	Leu	Ala	Asn	Gly	Val	Asp	Pro	Thr	Asn	Pro	Lys	Leu	Thr	Gly	
	C5	C6	CD1				CD5					CD10		CD12	E1				E5	
				65		E			70					75						80
His	Ala	Glu	Lys	Leu	Phe	Ala	Leu	Val	Arg	Asp	Ser	Ala	Gly	Gln	Leu	Lys	Ala	Ser	Gly	
E6				E10					E15					E20				E24	EF1	
DISTAL																				
				85				F	90					95						100
Thr	Val	Val	Ala	Asp	Ala	Ala	Leu	Gly	Ser	Val	His	Ala	Gln	Lys	Ala	Val	Thr	Asp	Pro	
			EF5	EF6	F1				F5		F7	F8	FG1				FG5	G1		
											PROXIMAL									
				105				G	110					115						120
Gln	Phe	Val	Val	Val	Lys	Glu	Ala	Leu	Leu	Lys	Thr	Ile	Lys	Ala	Ala	Val	Gly	Asp	Lys	
		G5					G10					G15				G19	GH1			
				125					130		H			135					140	143
Trp	Ser	Asp	Glu	Leu	Ser	Arg	Ala	Trp	Glu	Val	Ala	Tyr	Asp	Glu	Leu	Ala	Ala	Ala	Ile	Lys	Lys	Ala
	GH5	H1				H5					H10					H15			H19	HC1	HC2

Fig. 2. The amino-acid sequence of soybean leghemoglobin *a*. The helical segments are indicated by boxes. The helix assignments in the refined structure differ from those of Ollis *et al.* (1983) as follows: Helix *A*, residues 5–20 (instead of 4–20); helix *B*, residues 22–36 (21–36); helix *C*, 38–43 (37–43); helix *E*, 56–79 (55–79); helix *F*, 86–93 (86–94); helix *H*, 123–141 (122–141). A sequence error at the C terminus has been corrected. The helical segments are assigned according to Kabsch & Sander (1983), without distinguishing between the criteria for 3_{10}- and α-helices.

among these is another highly conserved residue, the distal histidine (His*E*6 in the present structure) (Bashford *et al.*, 1987). The effect of mutations at His*E*6 on dioxygen binding and other properties of Lb is under investigation (Hargrove *et al.*, 1997). The only other residue that is totally conserved in all hemoglobins is a phenylalanine, Phe*CD*1, which is also located on the distal side of the heme pocket and close to the exogenous ligand position. The presence of a water molecule on the distal side of the heme pocket should be noted (Figs. 4–6).

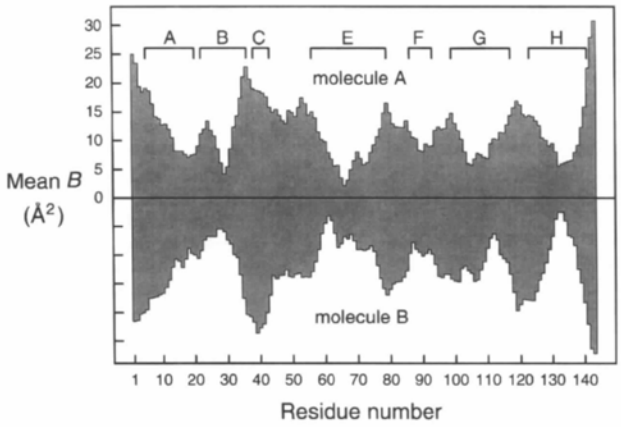

Fig. 3. Temperature factors *B* averaged over the backbone atoms of each residue. The seven helices are indicated at the top of the figure.

4. Discussion

4.1. *The heme group and heme cavity*

For a comparison between soybean Lb *a* nicotinate and myoglobin (Mb) we use a crystallographically characterized derivative of Mb with a bulky exogenous ligand, the sperm whale Mb:imidazole complex (Lionetti, Guanziroli, Frigerio, Ascenzi & Bolognesi, 1991). A superposition of their C^{α} atoms shows that Lb has a significantly larger distal heme cavity than Mb (Fig. 7). The increase in volume is associated with a displacement of the heme. The *A*, *B*, *C*, *G* and *H* helices of the two proteins align with good agreement, the *E* helices have closely similar conformations but are out of register, and the *CD* loops and *F* helices have quite different conformations, positions and orientations. The shift of the *F* helix is critical because it permits the change in the position of the heme and the consequent expansion of the heme cavity.

The fact that Lb has a larger distal heme cavity than Mb has been reported previously for both soybean Lb *a* (Ollis *et al.*, 1983) and lupin Lb II (Arutyunyan, Kuranova, Vainshtein & Steigemann, 1980; Arutyunyan, Kuranova, Tovbis *et al.*, 1980; Harutyunyan *et al.*, 1995). The present work enables us to make a semi-quantitative comparison between the volumes of the distal heme cavities: the volume traced by the centre of a 1.4 Å rolling sphere is ~15/~18 Å^3 for the independent molecules

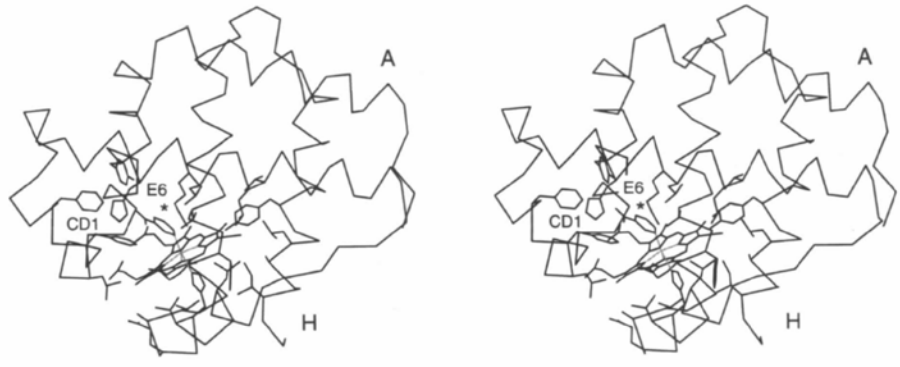

Fig. 4. The heme pocket of soybean leghemoglobin *a* nicotinate. The distal and proximal cavities lie above and below the heme, respectively. In the distal cavity, the invariant residues His*E*6 (the distal histidine) and Phe*CD*1 are seen to the left and slightly in front of the nicotinate ligand and an internal water molecule (denoted by the asterisk). The heme propionate groups protrude into the solvent.

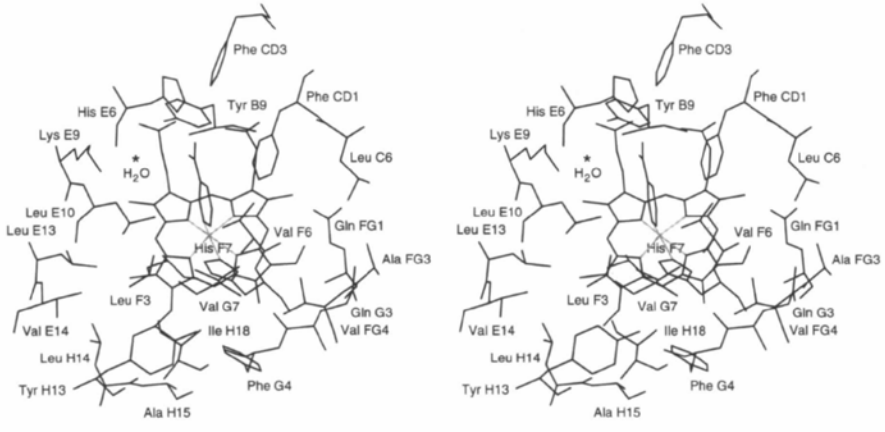

Fig. 5. The heme pocket of soybean leghemoglobin *a* nicotinate, showing all residues which make a contact ≤ 4 Å with the heme or nicotinate (Tyr*B*9, Leu*C*6, Phe*CD*1, Phe*CD*3, His*E*6, Lys*E*9, Leu*E*10, Leu*E*13, Val*E*14, Leu*F*3, Val*F*6, His*F*7, Gln*FG*1, Ala*FG*3, Val*FG*4, Gln*G*3, Phe*G*4, Val*G*7, Tyr*H*11, Leu*H*14, Ala*H*15, Ile*H*18). The asterisk denotes the internal H_2O molecule.

of soybean Lb *a* nicotinate, ~14 Å^3 for lupin Lb II nicotinate, and only ~5 Å^3 for the imidazole complex of sperm whale Mb. [The comparison is based on models from which the exogenous ligand has been deleted, and in which the conformation of the distal histidine is adjusted to have $\chi_1 = 120°$.] Three other differences from Mb, noted previously for lupin Lb II and now confirmed in soybean Lb *a*, are a proximal histidine orientation more favorable for ligand binding, a strongly ruffled heme group, and different orientations of the heme vinyl groups.

Despite the significant difference between the polypeptide chain lengths of soybean Lb *a* and lupin Lb II (see *Introduction*), the two proteins exhibit 56% sequence identity (Fig. 4 of Ollis *et al.*, 1983) and strong structural homology, as shown in Fig. 8. The *A*,

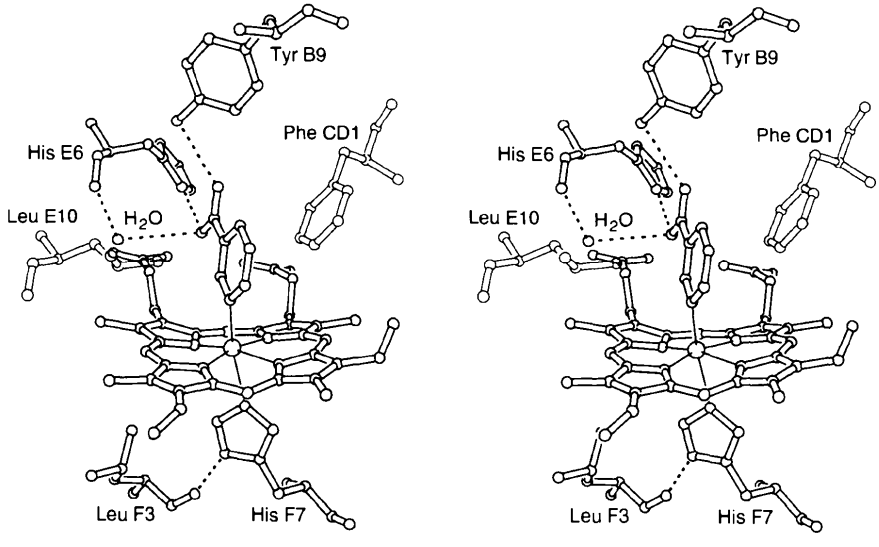

Fig. 6. The heme group, including hydrogen bonds involving the heme ligands. The Fe—ligand bond lengths, averaged over the two molecules in the asymmetric unit, are Fe—N(porphyrin) = 2.0(1), Fe—N(His*F7*) = 2.2(1), Fe—N(nicotinate) = 2.2(1) Å. (The e.s.d.'s of the averaged bond lengths are assigned as half the estimated uncertainty in the light-atom positions, see text.)

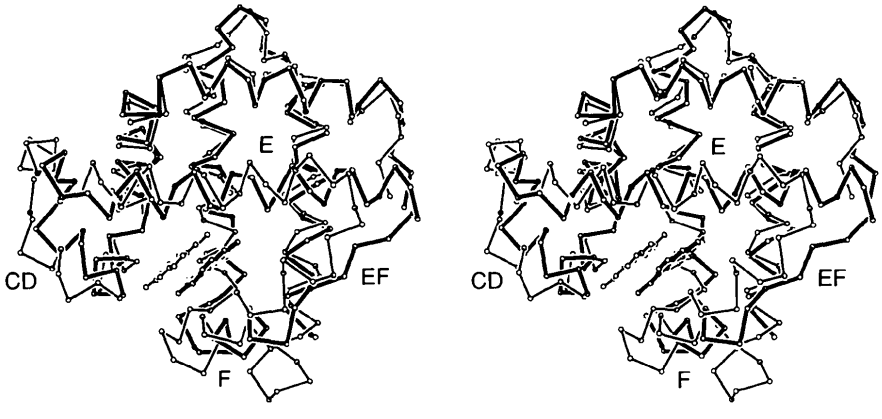

Fig. 7. Superposition of soybean leghemoglobin *a* nicotinate (heavy lines) and sperm whale myoglobin (light lines), based on the C^α and porphyrin atoms. The protein side chains and heme substituents have been omitted. (Myoglobin coordinates from Protein Data Bank entry 1MBI for sperm whale myoglobin:imidazole.)

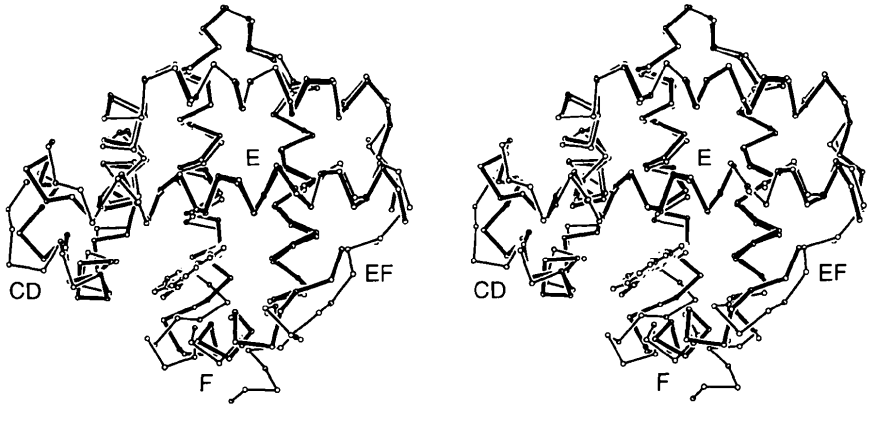

Fig. 8. Superposition of soybean leghemoglobin *a* nicotinate (heavy lines) and lupin leghemoglobin II nicotinate (light lines), based on the C^α and porphyrin atoms. The protein side chains and heme substituents have been omitted. (Lupin leghemoglobin coordinates from Protein Data Bank entry 1LH6 for lupin leghemoglobin II nicotinate.)

B, *C*, *E*, *G* and *H* helices all align closely. The main differences again occur in the *CD* region where soybean Lb has two residues fewer than lupin Lb, and the *F* helix which is three residues shorter in soybean Lb. The position of the *F* helix is influenced by its contacts with the *H* helix, which in soybean Lb is two turns (six residues) shorter at the C-terminal end. The differences at the *F* helix are associated with a significantly changed conformation at the *FG* bend. [The helix lengths used in these comparisons are derived by applying the Kabsch & Sander (1983) criteria rigorously to both Lb's.]

It is remarkable that the differences between the soybean and lupin Lb polypeptides leave the position and orientation of the heme in relation to the rest of the molecule almost unchanged (Fig. 8). As mentioned above, both proteins have a large distal heme cavity. The height of the cavity is defined largely by the *E* helix. The *E* helices of the two proteins superpose almost exactly.

The heme groups in soybean and lupin Lb's are similar not only in position and orientation, but also in the degree of non-planarity ('ruffling') (Fig. 9). The ruffling may be described by a parameter Γ.* Ellis (1995) has calculated values of Γ for all the crystallographically determined Lb and Mb structures. The values of Γ for the two independent molecules in soybean Lb *a* nicotinate are 50 and 41°, respectively. The corresponding value for lupin Lb II nicotinate is 48°. The smallest Γ among other lupin Lb II structures is 27°. In comparison, the heme in Mb is much less ruffled. The largest value of Γ found in any Mb structure is 20°. It occurs in the structure of the ferric Mb:imidazole complex (Lionetti *et al.*, 1991).

Fig. 9 includes a comparison between the vinyl substituents in soybean and lupin Lb's. Consistent with the conservation of the protein structure in the vicinity of the heme, the vinyl orientations in the two Lb's are closely similar. The orientations found in the crystals are consistent with those found by ^{1}H NMR for soybean carbonmonoxy- and oxy-Lb *a*, and for lupin carbonmonoxy-Lb I and II (Narula *et al.*, 1988; Mabbutt & Wright, 1983). The environments of the vinyl groups in soybean Lb *a* are shown in Fig. 10.

4.2. *The proximal histidine*

The hypothesis that oxygen binding in heme proteins is influenced by the proximal histidine is currently receiving attention (see, *e.g.*, Harutyunyan *et al.*, 1995). Within the limits of precision, the orientation of the imidazole group of His*F*7 in soybean Lb is the same as in lupin Lb, but different from that in myoglobin. The dihedral angle between the imidazole plane and the heme N_a—Fe—N_c direction is ~22° in soybean Lb *a* nicotinate, ~32° in lupin Lb II nicotinate, but only 0–7° in various forms of sperm whale myoglobin. Steric hindrance between the heme and the imidazole H(C^δ) and H(C^ε) atoms is greatest when the imidazole is eclipsed with N_a and N_c (as in Mb), and becomes smaller as the

* $\Gamma = -\angle_{AB} + \angle_{BC} - \angle_{CD} + \angle_{DA}$, where $\angle_{AB}$ is the angle between the normals to the planes of the heme *A* and *B* rings. $\angle_{AB}$ is positive if the normals to the planes incline towards each other on the distal side of the heme, and negative if they incline towards each other on the proximal side of the heme.

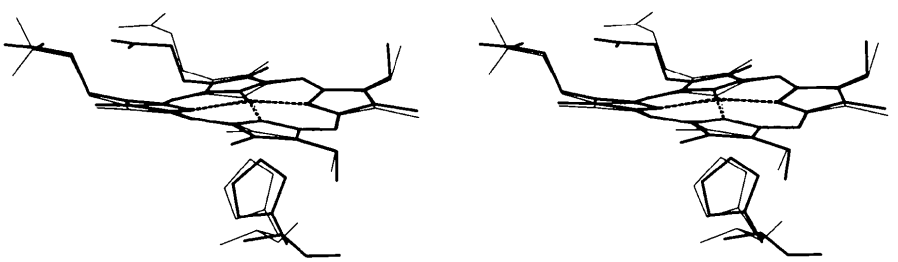

Fig. 9. Superposition of the heme groups in soybean leghemoglobin *a* nicotinate (heavy lines) and lupin leghemoglobin II nicotinate (light lines). The ruffling of the porphyrin, the orientations of the propionate groups, and the configurations of the vinyl groups are closely similar. The differences between the configurations of the heme propionate groups are not considered to be significant. These groups protrude into the solvent and interact only with solvent molecules in the crystals.

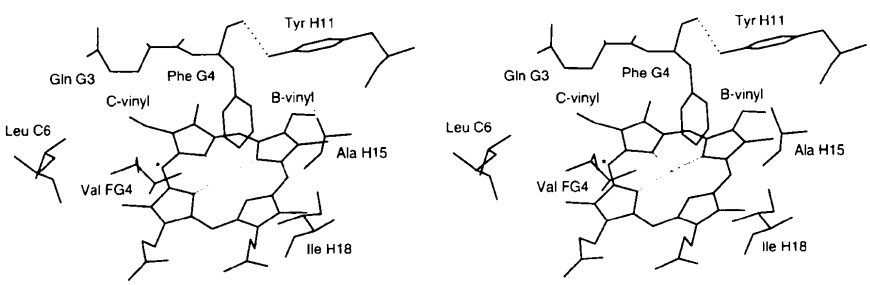

Fig. 10. Environment of the heme vinyl groups in soybean leghemoglobin *a* nicotinate.

imidazole ring rotates towards a staggered configuration (as in the Lb nicotinates). In comparison with Mb, the staggered configuration of the proximal histidine found in the present structure should make it easier for the imidazole group (and hence the Fe) to move towards the distal side when an exogenous ligand approaches, thus reducing the activation energy for ligation. In this regard, our description and conclusions concur with those of Harutyunyan *et al.* (1995). However, we note that the present work and previous reports of a staggered proximal histidine configuration refer to Lb derivatives with a 6-coordinate Fe atom: the exogenous ligand is already in place. The present work provides no new evidence concerning the type of torsional disorder said to occur in the 5-coordinate deoxyferrous form of lupin Lb (Harutyunyan *et al.*, 1995).

4.3. *Heme accessibility*

It is well known that the static X-ray structures of sperm whale Mb indicate no unhindered route of access for an O_2 molecule to the distal side of the heme pocket (Case & Karplus, 1979; Kottalam & Case, 1988; Elber & Karplus, 1990). The passage of ligands between the solvent and the heme pocket relies on dynamic variations in the protein. A pathway beginning between His*E*7 and Thr*E*10 has been proposed repeatedly. In complexes of Mb with bulky ligands such as alkyl isocyanides and imidazole, the side chain of His*E*7 is in an 'open' conformation which should permit O_2 to enter the heme pocket relatively freely.

The same '*E*7–*E*10' pathway is potentially available in soybean and lupin Lb's. However, to be consistent with the helix assignment in Fig. 2, the pathway in Lb's should be described as '*E*6–*E*9': the residues in soybean Lb *a* which correspond to His*E*7 and Thr*E*10 in Mb are His*E*6 and Lys*E*9. In soybean Lb *a* nicotinate the side chain of His*E*6 is held in the 'open' conformation by the bulky nicotinate ligand (Fig. 6). If the model of the structure is modified by removing the nicotinate and solvent molecules from the heme pocket, and rotating the heme propionate-A to the proximal side of the heme, a 1.5 Å radius probe sphere can move between His*E*6 and Lys*E*9 from the solvent to the heme without experiencing steric hindrance from any non-H atom (Fig. 11, upper part).*

During calculations of the distal heme cavity volume by the 'rolling sphere' method it was observed that a probe sphere can also escape from the protein through a gap between the *BC* junction and the *FG* loop. It was clear that the channel through which the probe left the protein is a second pathway by which ligands may enter. In fact, it corresponds to the alternative pathway previously identified in soybean Lb by Ollis *et al.* (1983) and in lupin Lb by Vainshtein *et al.* (1977). In soybean Lb *a*, the pathway passes between the side chains of Ala*BC*1, Ala*C*2, Leu*C*6 and Gln*G*3 (Fig. 11, lower part).

4.4. *Hydrogen bonds at the Fe-binding groups*

The orientations of the Fe-binding groups on both sides of the heme are influenced by hydrogen bonds to the polypeptide. On the proximal side, the orientation of the imidazole ring of His*F*7 is stabilized by a hydrogen bond from the $N^{\delta 1}$(imidazole) atom to the O(peptide) atom of Leu*F*3 (His*F*7 $N^{\delta 1}\cdots$O Leu*F*3 = 2.6/2.4 Å). (Here and elsewhere, two distances separated by a slash refer to the two independent molecules.) If the *F* helix were an ideal α-helix, the O(peptide) atom of Leu*F*3 would be hydrogen bonded to N His*F*7, but the interatomic distance is too long to be a hydrogen bond (Leu*F*3 O$\cdots$N His*F*7 = 3.7/3.9 Å). The hydrogen bond from the N1(imidazole) atom of His*F*7 to the O(peptide) atom of Leu*F*3 may affect dioxygen binding, since it restrains the imidazole ring of the proximal histidine towards the staggered configuration. The possible significance of this configuration in relation to dioxygen binding has been discussed in §4.2. Similar interactions have been observed in other hemoglobins, including sperm whale myoglobin (see, *e.g.*, Phillips, 1980) and lupin Lb II (Harutyunyan *et al.*, 1995).

On the distal side of the heme, the carboxylate group of the nicotinate ligand interacts strongly with the imidazole group of the distal histidine His*E*6 (Nic $O^1\cdots N^{\delta 1}$ His*E*6 = 2.8/2.7 Å). The effect of the interaction is to hold the side chain of His*E*6 in the 'open' conformation noted earlier, *i.e.*, rotated about C^{α}—C^{β} away from the Fe atom, as shown in Fig. 6. The interaction between the negatively charged nicotinate carboxylate group and the imidazole group of His*E*6 can be described either as a salt bridge or as a hydrogen bond with a strong electrostatic component: since the pK_a of His*E*6 in ferric Lb *a* nicotinate is 7.7–8.3 (Trewhella, Appleby & Wright, 1986; Atanasov *et al.*, 1989), it follows that the imidazole group is protonated under the conditions of X-ray data collection (pH 6.0).

The nicotinate carboxylate group forms two other hydrogen bonds, one to the phenol O atom of Tyr*B*9 (Nic $O^2\cdots O^{\eta}$ Tyr*B*9 = 3.2/2.7 Å), and the other to the heme-pocket water molecule (Nic $O^1\cdots O$ H_2O = 3.2/3.2 Å). This water molecule in turn is hydrogen bonded to the carbonyl O atom of the distal histidine (H_2O O$\cdots$O His*E*6 = 2.6/2.5 Å).

4.5. *TyrB9: a functional role?*

The existence of a hydrogen bond between the nicotinate carboxyl group and the phenolic O atom of Tyr*B*9 was predicted by Arutyunyan (1981), who noted that

* The orientation of the propionate-A side chain in Lb crystals is determined by intermolecular and solvent interactions. In the soybean Lb *a* nicotinate structure, the propionate side chain makes contact only with solvent atoms, and lies on the distal side of the heme. In seven of 11 lupin Lb structures, it lies on the proximal side. We conclude that the side chain is sufficiently flexible not to create a barrier against the *E*6–*E*9 pathway, and that, for the purposes of the rolling sphere calculation, it can be rotated out of the way.

such an interaction cannot occur in lupin Lb II where the residue corresponding to Tyr*B*9 is a Phe. Details of the interaction between the nicotinate and Tyr*B*9 are shown in Figs. 5 and 6. The location of Tyr*B*9 in relation to the rest of the molecule can be seen in Fig. 4, where the phenolic side chain points towards the reader just above and behind the label '*E*6'. The ability to form a tyrosine/nicotinate hydrogen bond explains why soybean Lb *a* and kidney bean Lb, which have a tyrosine at *B*9, have higher affinities for nicotinate than lupin Lb I, which (like lupin Lb II) has phenylalanine at *B*9. The relevant stability constants pK_L at the acid limit are 8.55 (soybean), 6.50 (kidney bean) and 4.45 (lupin), respectively (Atanasov *et al.*, 1989).

While nicotinate is not a physiological ligand of leghemoglobin, a tyrosine which can form a hydrogen bond with nicotinate may well be able to interact with other Fe ligands. It is possible that this hydrogen-bonding capability has a physiological significance. The hydrogen bond need not necessarily involve Tyr*B*9: if Phe*CD*3 in soybean Lb *a* were a tyrosine, its phenolic O^{η} atom would be just as favourably located for hydrogen bonding to the nicotinate carboxyl group as the O^{η} atom of Tyr*B*9 (Fig. 5). In a 1987 survey of 226 globin sequences, it was noted that only six have Tyr at the position corresponding to *B*9 in the present work, and only four have Tyr at a position corresponding to *CD*3 (Bashford *et al.*, 1987). In a more recent survey of 22 *plant* hemoglobin sequences, Appleby (unpublished work) has observed that (i) the residues corresponding to *B*9 and *CD*3 are always Tyr or Phe; (ii) 12 sequences have Tyr at the residue corresponding to *B*9, and two more have Tyr at the residue corresponding to Phe*CD*3; and (iii) no sequence has Tyr at both positions. Thus, the tendency of plant hemoglobins to have a Tyr residue within a hydrogen-bonding distance of a Fe-ligand appears to be above average.

The speculation that a Tyr at *B*9 or *CD*3 may have a physiological role is supported by the recent observation that a hydrogen bond from Tyr*B*9 stabilizes the heme-bound dioxygen molecule in the oxy-hemoglobin of the nematode *Ascaris suum* (Yang, Kloek, Goldberg

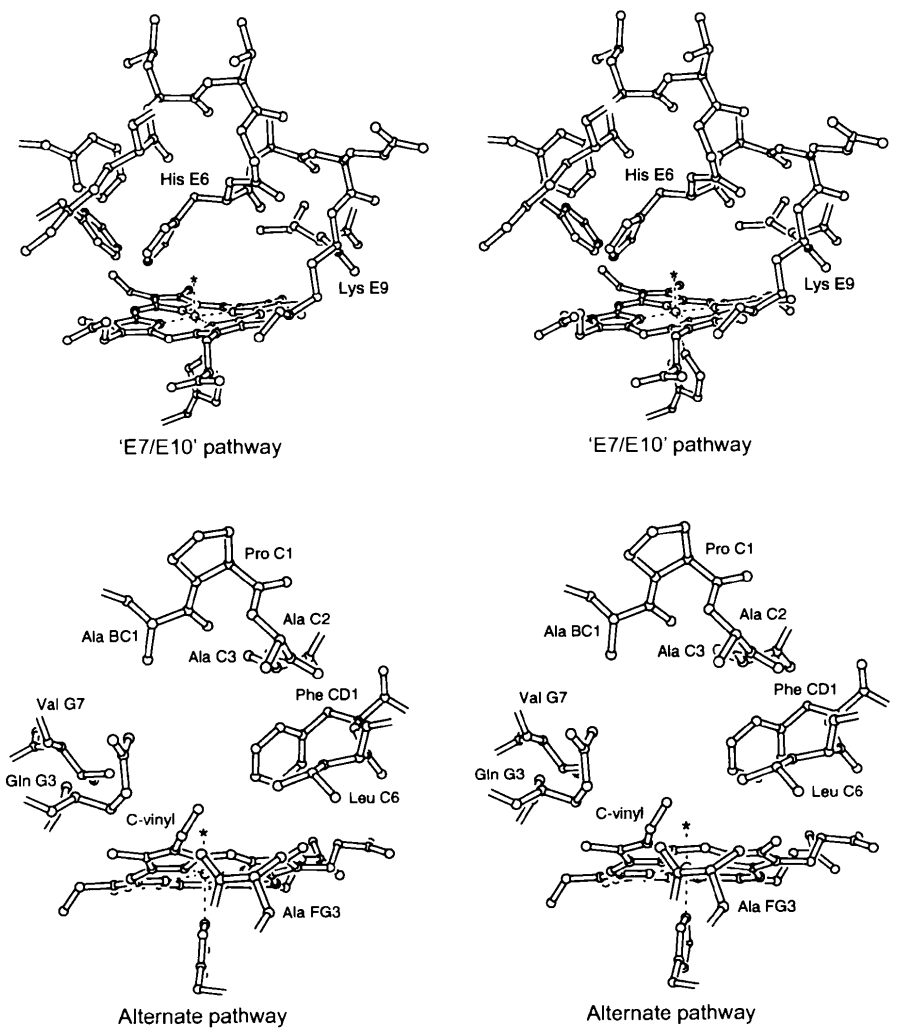

Fig. 11. The '*E*7–*E*10' and alternate pathways between the molecular surface and the heme in soybean leghemoglobin *a* nicotinate. (In soybean Lb *a*, the '*E*7–*E*10' pathway has to be described as '*E*6–*E*9', since the residues corresponding to *E*7 and *E*10 in Mb are His*E*6 and Lys*E*9.) The nicotinate ligand and the internal water molecule are omitted. The binding site for exogenous ligands is marked with an asterisk. In the upper figure, the heme-A propionate (pointing towards the reader) has been rotated to the proximal side of the heme.

& Mathews, 1995; Huang, Huang, Kloek, Goldberg & Friedman, 1996). On the other hand, Tyr*B*9 in the present structure would have to move ~3 Å closer to the heme if it were to interact with a ligand as small as dioxygen (the Tyr*B*9 $O^{\eta}\cdots$Fe distance is 8.8 Å). Further, there appears to be no correlation between the kinetic and thermodynamic parameters for ligand binding to plant hemoglobins listed by Gibson *et al.* (1989), and the presence or absence of Tyr residues at positions *B*9 and *CD*3. The possible significance of these observations remains to be explored.

This work was supported by grants C80/15377 and A29230677 from the Australian Research Council, a University of Sydney Gritton Senior Research Fellowship (WHN), and an Australian Postgraduate Research Award (PJE).

References

Andersson, C. R., Jensen, E. Ö, Llewellyn, D. J., Dennis, E. S. & Peacock, W. J. (1997). *Proc. Natl Acad. Sci. USA*. In the press.

Appleby, C. A. (1984). *Ann. Rev. Plant Physiol.* **35**, 443–478.

Appleby, C. A. (1992). *Sci. Progress (Oxford)*, **76**, 365–398.

Arutyunyan, E. G. (1981). *Mol. Biol. (Moscow)*, **15**, 27–44.

Arutyunyan, E. G., Kuranova, I. P., Tovbis, A. B., Grebenko, A. I., Voronova, A. A., Nekrasov, Yu. V. & Vainshtein, B. K. (1980). *Sov. Phys. Crystallogr.* **25**, 302–306.

Arutyunyan, E. G., Kuranova, I. P., Vainshtein, B. K. & Steigemann, W. (1980). *Sov. Phys. Crystallogr.* **25**, 43–58.

Atanasov, B. P., Dimitrova, E. A., Kudryatseva, N. N., Zhiznevskaya, G. Y. & Appleby, C. A. (1989). *Biochim. Biophys. Acta*, **998**, 80–84.

Bashford, D., Chothia, C. & Lesk, A. M. (1987). *J. Mol. Biol.* **196**, 199–216.

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. (1977). *J. Mol. Biol.* **112**, 535–542.

Case, D. A. & Karplus, M. (1979). *J. Mol. Biol.* **132**, 343–368.

Elber, R. & Karplus, M. (1990). *J. Am. Chem. Soc.* **112**, 9161–9175.

Ellfolk, N. (1961). *Acta Chem. Scand.* **15**, 975–984.

Ellis, P. J. (1995). PhD thesis, University of Sydney, Australia.

Finzel, B. C. (1987). *J. Appl. Cryst.* **20**, 53–55.

Gibson, Q. H., Wittenberg, J. B., Wittenberg, B. A., Bogusz, D. & Appleby, C. A. (1989). *J. Biol. Chem.* **264**, 100–107.

Hargrove, M. S., Barry, J. K., Brucker, E. A., Berry, M. B., Phillips, G. N. Jr, Olson, J. S., Arredondo-Peter, R., Dean, J. M., Klucas, R. V. & Sarath, G. (1997). *J. Mol. Biol.* Submitted.

Harutyunyan, E. H., Safonova, T. N, Kuranova, I. P., Popov, A. N., Teplyakov, A. V., Oblomova, G. V., Rusakov, A. A., Vainshtein, B. K., Dodson, G. G., Wilson, J. C. & Perutz, M. F. (1995). *J. Mol. Biol.* **251**, 104–115.

Hendrickson, W. A. (1985). *Methods Enzymol.* **115**, 252–270.

Hendrickson, W. A. & Konnert, J. H. (1980). In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

Huang, S., Huang, J., Kloek, A. P., Goldberg, D. E. & Friedman, J. M. (1996). *J. Biol. Chem.* **271**, 958–962.

Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Klucas, R. V. & Appleby, C. A. (1991). *Plant Physiol.* **91**, 551–555.

Kottalam, J. & Case, D. A. (1988). *J. Am. Chem. Soc.* **110**, 7690–7697.

Lionetti, C., Guanziroli, M. G., Frigerio, F., Ascenzi, P. & Bolognesi, M. (1991). *J. Mol. Biol.* **217**, 409–412.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Narula, S. S., Dalvit, C., Appleby, C. A. & Wright, P. E. (1988). *Eur. J. Biochem.* **178**, 419–435.

O'Brian, M. R. (1996). *J. Bacteriol.* **178**, 2471–2478.

Ollis, D. L., Appleby, C. A., Colman, P. M., Cutten, A. E., Guss, J. M., Venkatappa, M. P. & Freeman, H. C. (1983). *Aust. J. Chem.* **36**, 451–468.

Phillips, S. E. V. (1980). *J. Mol. Biol.* **142**, 531–554.

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

Trewhella, J., Appleby, C. A. & Wright, P. E. (1986). *Aust. J. Chem.* **39**, 317–324.

Vainshtein, B. K., Arutyunyan, E. G., Kuranova, I. P., Borisov, V. V., Sosfenov, N. I., Pavlovskii, A. G., Grebenko, A. I., Konareva, N. V. & Nekrasov, Yu. V. (1977). *Dokl. Akad. Nauk SSSR*, **233**, 238–241.

Yang, J., Kloek, A. P., Goldberg, D. E. & Mathews, F. S. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 4224–4228.