region near H with the lone-pair density of the acceptor oxygen decreases the original lone-pair density as the O···O distance diminishes (Eisenstein & Hirshfeld, 1983; Lunell, 1984). For this reason the O(4) lone-pair density (Fig. 2) is depleted most in the direction of the shorter hydrogen bond [O···O distance = 2·566 (1) Å].

Financial support from the Swedish Natural Science Research Council is gratefully acknowledged. We are indebted to Mr H. Karlsson for assistance with the data collection. We wish to thank Professor Philip Coppens for making the computer program *DEBYE* available to us.

References

ABRAHAMS, S. C. & KEVE, E. T. (1971). *Acta Cryst.* A**27**, 157–165.

ALMLÖF, J., LINDGREN, J. & TEGENFELDT, J. (1972). *J. Mol. Struct.* **14**, 427–437.

BECKER, P. & COPPENS, P. (1974). *Acta Cryst.* A**30**, 129–147.

BECKER, P. & COPPENS, P. (1975). *Acta Cryst.* A**31**, 417–425.

BERKOVITCH-YELLIN, Z. & LEISEROWITZ, L. (1975). *J. Am. Chem. Soc.* **97**, 5627–5628.

COPPENS, P., BOEHME, R., PRICE, P. F. & STEVENS, E. D. (1981). *Acta Cryst.* A**37**, 857–863.

COPPENS, P., DAM, J., HARKEMA, S., FEIL, D., FELD, R., LEHMANN, M. S., GODDARD, R., KRUGER, C., HELLNER, E., JOHANSEN, H., LARSEN, F. K., KOETZLE, T. F., MCMULLAN, R. K., MASLEN, E. N. & STEVENS, E. D. (1984). *Acta Cryst.* A**40**, 184–195.

COPPENS, P., MOSS, G. & HANSEN, N. K. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 16.01–16.21. Bangalore: Indian Academy of Sciences.

DELAPLANE, R. G., TELLGREN, R. & OLOVSSON, I. (1984). *Acta Cryst.* C**40**, 1800–1803.

DENNE, W. A. (1977). *Acta Cryst.* A**33**, 438–440.

EISENSTEIN, M. & HIRSHFELD, F. L. (1983). *Acta Cryst.* B**39**, 61–75.

FERNANDES, N. G., TELLGREN, R. & OLOVSSON, I. (1988). *Acta Cryst.* C**44**, 1168–1172.

HERMANSSON, K. (1984). Thesis, Univ. of Uppsala, Sweden.

HERMANSSON, K. (1985). *Acta Cryst.* B**41**, 161–169.

HERMANSSON, K., OLOVSSON, I. & LUNELL, S. (1984). *Theor. Chim. Acta*, **64**, 265–276.

HERMANSSON, K. & THOMAS, J. (1982). *Acta Cryst.* B**38**, 2555–2563.

HIRSHFELD, F. L. (1971). *Acta Cryst.* B**27**, 769–781.

HIRSHFELD, F. L. (1976). *Acta Cryst.* A**32**, 239–244.

HIRSHFELD, F. L. (1977). *Isr. J. Chem.* **16**, 226–229.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

JOHANSEN, H. (1979). *Acta Cryst.* A**35**, 319–325.

JOHNSON, C. K. (1965). *ORTEP*. Report ORNL-3794. Oak Ridge National Laboratory, Tennessee, USA.

LEGROS, J.-P. & KVICK, Å. (1980). *Acta Cryst.* B**36**, 3052–3059.

LEHMANN, M. S. & LARSEN, F. K. (1974). *Acta Cryst.* A**30**, 580–584.

LUNDGREN, J.-O. (1979). *Acta Cryst.* B**35**, 1027–1033.

LUNDGREN, J.-O. (1982). *Crystallographic Computer Programs*. Report UUIC-B13-04-05. Institute of Chemistry, Univ. of Uppsala, Sweden.

LUNDGREN, J.-O. & LIMINGA, R. (1979). *Acta Cryst.* B**35**, 1023–1027.

LUNELL, S. (1984). *J. Chem. Phys.* **80**, 6185–6193.

MCCANDLISH, L. E., STOUT, G. H. & ANDREWS, L. C. (1975). *Acta Cryst.* A**31**, 245–249.

OLOVSSON, I. (1980). *Electron and Magnetization Densities in Molecules and Crystals*, edited by P. BECKER, pp. 883–890. New York: Plenum.

OLOVSSON, I. (1982). *Croat. Chem. Acta*, **55**, 171–190.

REES, B. (1977). *Isr. J. Chem.* **16**, 180–186.

STEVENS, E. D. (1980). *Acta Cryst.* B**36**, 1876–1886.

STEVENS, E. D. & COPPENS, P. (1980). *Acta Cryst.* B**36**, 1864–1876.

TELLGREN, R. & OLOVSSON, I. (1971). *J. Chem. Phys.* **54**, 127–134.

TELLGREN, R., THOMAS, J. O. & OLOVSSON, I. (1977). *Acta Cryst.* B**33**, 3500–3504.

Acta Cryst. (1990). B**46**, 370–377

Determination of the Crystal Structure of Recombinant Pig Myoglobin by Molecular Replacement and its Refinement

BY STEPHEN J. SMERDON, TOM J. OLDFIELD, ELEANOR J. DODSON, GUY G. DODSON, RODERICK E. HUBBARD AND ANTHONY J. WILKINSON*

Department of Chemistry, University of York, Heslington, York YO1 5DD, England

(*Received* 22 *August* 1989; *accepted* 6 *November* 1989)

Abstract

As part of a protein engineering study, the X-ray crystal structure of recombinant pig myoglobin, prepared and crystallized from *E. coli*, has been determined. Diffraction data were collected to 2·5 Å spacing using a synchrotron X-ray source. The structure was solved using the molecular-replacement method and refined using least-squares minimization procedures to a crystallographic *R* factor of 18·5% using 14 481 reflections between 10 and 2·5 Å. A preliminary comparison of the structure of pig myoglobin with other myoglobin structures is presented.

* To whom correspondence should be addressed.

0108-7681/90/030370-08$03.00

© 1990 International Union of Crystallography

Introduction

Myoglobin is amongst the most extensively studied proteins and has been examined by a panoply of biophysical and spectroscopic techniques. Sperm-whale myoglobin was the first protein crystal structure to be solved (Kendrew *et al.*, 1960) and since then high-resolution crystal structures of the protein from different sources (Scoloudi & Baker, 1978; Evans & Brayer, 1988) in various liganded states have been reported. These structures and detailed kinetic analyses have implicated residues that may be involved in the ligand-binding process and the determination of the ligand-binding affinity and stereochemistry. The contributions that individual side chains make to the function and properties of myoglobin can be probed by the introduction of amino-acid substitution mutations into the protein. Recently, a number of groups have embarked on such protein engineering studies of myoglobin focusing on sperm-whale (Springer & Sligar, 1987), human (Varadarajan, Szabo & Boxer, 1985) and pig (Dodson, Hubbard, Oldfield, Smerdon & Wilkinson, 1988) proteins.

The gene encoding pig myoglobin has been manipulated for high level expression of the protein in *E. coli* using the cII–factor X_a fusion expression system of Nagai & Thøgersen (1984). Intact porcine myoglobin was produced following reconstitution with exogenous haem and proteolytic cleavage with trypsin. This recombinant protein was used to grow crystals suitable for X-ray structural studies (Dodson *et al.*, 1988). Here we report the solution of the structure of pig myoglobin using the molecular-replacement method and the subsequent refinement of the structural model.

Crystallization and data collection

Recombinant pig apomyoglobin was expressed as a fusion protein with the *N*-terminal 31 amino acids of bacteriophage λcII protein and a short peptide linker that includes a cleavage site for the protease factor X_a. Haematin (Fe^{3+}) was added to the fusion protein and the complex was subsequently digested with trypsin to generate the intact protein. Following purification, this material was used for crystallization, with no attempt being made to alter the ligation/oxidation state of the haem iron. Crystals were grown in hanging drops by vapour diffusion from 80% ammonium sulfate in a 50 m*M* sodium phosphate buffer, pH 7·1, as described previously (Dodson *et al.*, 1988). The crystals have space group *C*2 and, assuming that the asymmetric unit contains two molecules, there are eight molecules per unit cell giving a solvent content of 45%. A single crystal was mounted in a glass capillary in the usual way and used for the collection of data on film using the synchrotron-radiation source at Daresbury, UK, station 7.2 ($\lambda = 1{\cdot}488$ Å). The crystal was aligned on an Arndt–Wonnacott oscillation film camera with the c^* axis parallel to the rotation axis to take advantage of the twofold symmetry. The crystal-to-film distance was 88 mm and data were collected to 2·5 Å spacing. 110° of data were collected using a 2° oscillation and a 50 s exposure.

Processing

Films were digitized using a Joyce–Loebl Scandig-3 scanning densitometer with a 50 μm raster. Data processing and reduction employed the *CCP*4 suite of programs for X-ray crystallography (SERC Daresbury Laboratory, 1986). Preliminary analysis of the recombinant pig myoglobin crystals by precession photography had shown systematic absences for $h + k + l = 2n + 1$ and apparent $I222$ symmetry (Fig. 1). However, one of the angles is not 90° (92·2°) and the space group is in fact $C2$ ($a = 156{\cdot}9$, $b = 42{\cdot}6$, $c = 92{\cdot}2$ Å, $\beta = 127{\cdot}9°$).

The data were processed in the non-standard space group $I2_1$ ($a = 124{\cdot}91$, $b = 42{\cdot}63$, $c = 92{\cdot}2$ Å, $\beta = 92{\cdot}16°$). 49 979 spots were fully recorded with 15 941 reflections recorded on more than one film. After rejection of 1453 (2·2% of the total) rogue reflections, the merging R factor for symmetry-related reflections was 5·9%. We obtained 14 481 independent intensity measurements (90% of the total data).

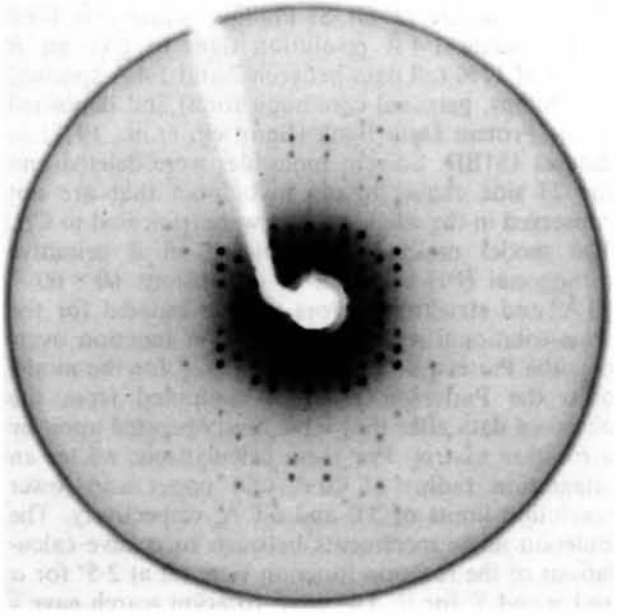

Fig. 1. Zero-layer ($0kl$) precession photograph of recombinant porcine myoglobin crystals showing the apparent systematic absences along the b^* (horizontal) and c^* (vertical) directions.

The data have a mean $I/\sigma(I)$ value of 29 and an overall B value of 33 Å^2 estimated from the Wilson plot.

Molecular replacement

Self-rotation search

The volume of the $C2$ unit cell suggested that there are two molecules in the crystallographic asymmetric unit. Precession photographs of recombinant pig myoglobin crystals exhibit pseudo $I222$ symmetry (Fig. 1), indicating that there is a non-crystallographic twofold axis of symmetry perpendicular to the crystallographic twofold symmetry axis. The relative orientation of the molecules within the asymmetric unit can be determined through a rotation search in Patterson space. Self-rotation searches of the Patterson function calculated using the observed intensities for data between 8·0 and 4·0 Å were made using an integration radius of 20 Å. The overlap function was calculated at 5° increments in the Eulerian angles. This revealed a strong positive peak with a height equivalent to 85% of that of the origin peak, corresponding to the expected twofold axis of symmetry along **a*** (Fig. 2*a*). The mean and root-mean-square deviation from the mean density represent 6 and 16%, respectively, of the origin peak density.

Cross-rotational search

The starting model for molecular replacement was the coordinates from the sperm-whale deoxymyoglobin structure of Dr S. Phillips, which has been refined using 1·4 Å resolution data to give an R factor of 17% (all data between 7 and 1·4 Å spacing) (S. Phillips, personal communication) and deposited in the Protein Data Bank (Bernstein *et al.*, 1977) as dataset 1MBD. Solvent molecules were deleted and the 23 side chains in pig myoglobin that are not conserved in the whale protein were truncated to Cβ. The model molecule was placed in a primitive orthogonal ($P1$) unit cell of dimensions 60 × 60 × 60 Å^3 and structure factors were calculated for the cross-rotational search. The rotation function overlaps the Patterson function calculated for the model onto the Patterson function computed from the observed data after they have been operated upon by a rotation matrix. For these calculations, we set an integration radius of 20 Å and upper and lower resolution limits of 3·0 and 6·0 Å, respectively. The Eulerian angle increments between successive calculations of the rotation function were set at 2·5° for α and γ and 5° for β. The cross-rotation search gave a clear set of solutions related by the expected symmetry operations (Figs. 2*b* and 2*c*). The mean and standard deviation from the mean represent 0·5% and 13% of the height of the solution peaks. The next highest peak in the rotation function was 49% of the solution peak height.

Translational search

The translational molecular-replacement parameters were determined using a search for a minimum in the R value between the observed structure-factor amplitudes and the different sets of calculated amplitudes as the molecule is translated through the unit cell using the program *SEARCH* (SERC Daresbury Laboratory, 1986). Two search molecules oriented according to the rotation-function solutions were

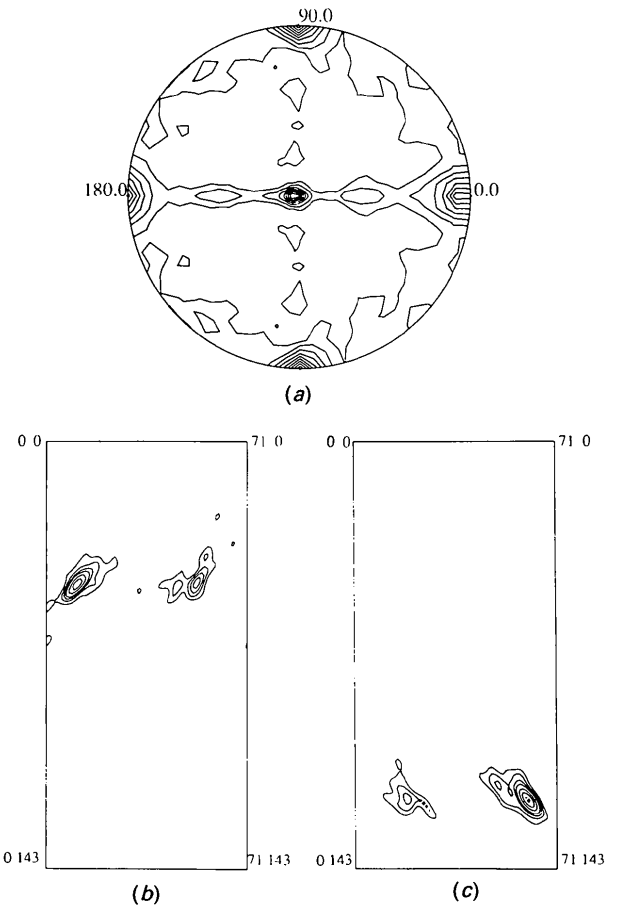

Fig. 2. (*a*) The spherical polar representation of the twofold rotational symmetry operations; stereographic plot of the κ = 180° (*i.e.* twofold rotations) section of the self-rotation search map. The map is contoured between 10 and 100% of the origin peak height at 10% intervals. The map shows the three perpendicular twofold symmetry axes; two are noncrystallographic along a^* and c^* with the third crystallographic axis along b. (*b*), (*c*) The two solutions to the cross-rotation function showing the match between the model structure factors and the observed ones. The two peaks, on the (*b*) β = 55° and (*c*) 125° sections, respectively, are related as expected. The map is contoured at intervals between 2·4 and 7·8 times the root-mean-square map density.

used independently in this procedure; *i.e.* the search was performed using only half the contents of the unit cell. The unit-cell directions, *a*, *b* and *c*, were divided into approximately 0·5, 1 and 0·5 Å equal spacings, respectively, and the *R* value was computed after positioning the molecule at each grid point. For monoclinic space groups, the *R* factor of a molecule is independent of the *y* coordinate. Furthermore, the *xz* origin can be placed at ($x = 0{\cdot}0$, $z = 0{\cdot}0$), ($x = 0{\cdot}5$, $z = 0{\cdot}0$), ($x = 0{\cdot}0$, $z = 0{\cdot}5$) or ($x = 0{\cdot}5$, $z = 0{\cdot}5$), so that it is only necessary to search from $x = 0{\cdot}0$ to $x = 0{\cdot}5$, and $z = 0{\cdot}0$ to $z = 0{\cdot}5$; the *R* factor will be the same for (X, y, Z), ($X + \frac{1}{2}$, y, Z), (X, y, $Z + \frac{1}{2}$) and ($X + \frac{1}{2}$, y, $Z + \frac{1}{2}$). But in this case, where the two molecules have been positioned independently, it is necessary to find the relative position of the second molecule to the first. So molecule *A* was assumed to be centred at ($X1$, 0, $Z1$), and the possible solutions for molecule *B* were explored: *viz.* ($X2$, y, $Z2$), ($X2 + \frac{1}{2}$, y, $Z2$), ($X2$, y, $Z2 + \frac{1}{2}$) and ($X2 + \frac{1}{2}$, y, $Z2 + \frac{1}{2}$), where y ranged along the *b* axis. The minimum *R* factor obtained was 47% for:

	α	β	γ	x	y	z
Molecule *A*	58·4	42·3	−20·2	53·3	0·0	10·1
Molecule *B*	125·3	−135·4	−15·2	52·8	9·6	34·8

Refinement of molecular-replacement parameters

A least-squares fitting procedure described by Derewenda (1989) was used to improve the initial fit of the model. Five cycles of least-squares refinement (Hendrickson & Konnert, 1979) were carried out. In these initial cycles the relative weight of the geometric terms to the X-ray terms was 1:1. At this weighting at 2·5 Å resolution, the X-ray gradient will be able to adjust the model towards the real molecule in the structure and in so doing will generate more precise molecular-replacement parameters. This can have real advantages in the subsequent refinement calculations. The central core of the initial model was refitted to the refined coordinates, and the refinement procedure begun again with these new coordinates. Convergence was achieved after repeating the procedure nine times. The final molecular replacement parameters were:

	α	β	γ	x	y	z
Molecule *A*	61·9	43·6	−21·6	55·3	0·5	11·0
Molecule *B*	122·0	−134·3	−16·6	54·3	9·0	34·5

This represents a root-mean-square shift of 0·09 and 1·2 Å for the *A* and *B* molecules, respectively, and the *R* factor had dropped to 27%.

Refinement

Thirty-three iterations of restrained least-squares refinement (Hendrickson & Konnert, 1979) were punctuated with three sessions of manual model fitting into Fourier maps ($2F_o - F_c$ and $F_o - F_c$) displayed on an Evans and Sutherland PS300 graphics device using the program *FRODO* (Jones, 1978). Structure factors were calculated using the fast Fourier transform package, *FFT* (SERC Daresbury Laboratory, 1986). The various geometric terms were weighted by $1/\sigma$ (Table 2). No weight was applied to the various X-ray terms. The X-ray matrix and the matrix derived from the geometric restraints were weighted in a 1:2 ratio. The course and progress of refinement are outlined in Table 1.

Difference Fourier maps ($F_o - F_c$) were used initially to position the 23 side chains that are not conserved between the sperm-whale and pig myoglobin sequences. Such maps were also used to identify and build in water molecules. In the course of refinement, we truncated side chains for which clear density was not discernible in $2F_o - F_c$ maps to eliminate contributions of wrongly positioned atoms to the overall phasing. This strategy was successful in improving the definition of a number of residues, particularly Phe150 in both molecules. The *R* factor obtained for all data between 10 and 2·5 Å is 18·5%

Table 1. *Course of the refinement*

Cycle	Resolution (Å)	*R**	Comments
1–10	10–2·5	0·27–0·23	*XYZ* refinement only, starting from molecular-replacement/least-squares model. Nonconserved side chains truncated to C^{β}. Porcine side chains added as appropriate, 22 waters inserted after cycle 10.
11–22	10–2·5	0·26–0·19	5 cycles *XYZ*, 1 cycle unrestrained isotropic *B*-factor refinement. 4 cycles *XYZ*, 2 cycles *B*-factor refinement. 35 waters added after cycle 22.
23–48	10–2·5	0·23–0·18	18 *XYZ* with 2 cycles *B* after each set of 5.

* $R = \sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$.

Table 2. *Restraints applied during refinement and the deviations from ideality of the final overall structural model*

Parameter restrained	R.m.s. Δ	σ
Distance restraint		
Bond	0·017	0·02
Angle	0·065	0·04
Planar 1–4	0·081	0·06
Metal coordination	0·304	0·03
Plane	0·016	0·02
Chiral centre	0·158	0·23
Non-bonded		
Single torsion	0·235	0·50
Multiple torsion	0·250	0·50
Possible H bond	0·199	0·50
Conformational torsion		
Planar	3·400	20·0
Staggered	25·10	20·0
Orthonormal	41·00	20·0

after 57 water molecules had been added.* Table 2 shows the stereochemical restraints that were imposed and the root-mean-square deviations from ideal values that were obtained. The clarity of the electron density map can be assessed from an examination of the superimposed contour sections in Fig. 3 and from the electron density associated with representative side chains (Figs. 4*a–c*).

Structure of pig myoglobin

There are two pig myoglobin molecules in the crystallographic asymmetric unit. As no noncrystallographic symmetry averaging was applied in the course of refinement, the two molecules represent independent structures. When the *A* and *B* molecules are overlapped, the root-mean-square deviation in the positions of the backbone atomic positions is 0·29 Å, in reasonable agreement with the positional errors in atomic coordinates estimated from a Luzzati (1952) plot (data not shown).

Pig myoglobin has close structural similarity to sperm-whale myoglobin as expected from the high level of amino-acid sequence conservation between the two proteins. When the *A* and *B* molecules are overlapped on the sperm-whale deoxymyoglobin model of Dr S. Phillips refined against 1·4 Å data (1MBD, Bernstein *et al.*, 1977), the root-mean-square deviations in backbone atomic positions are 0·53 and 0·50 Å, respectively, showing there exist small but distinct structural differences between the two globin species. The molecule itself has an ellipsoid form with overall dimensions of 45 × 35 × 25 Å^3. The positions and lengths of the eight α helices are strongly conserved as are the residues surrounding the haem cavity. The largest overall r.m.s. deviations in the backbone atomic coordinates occur in the *CD*, *EF*, *GH* and *HC** regions for both molecules, *i.e.* between and outside the helical segments.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1PMB, R1PMBSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centres at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37031 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

* The alphanumeric nomenclature referring to the residue positions in terms of loops and helices is according to Dickerson & Geis (1983).

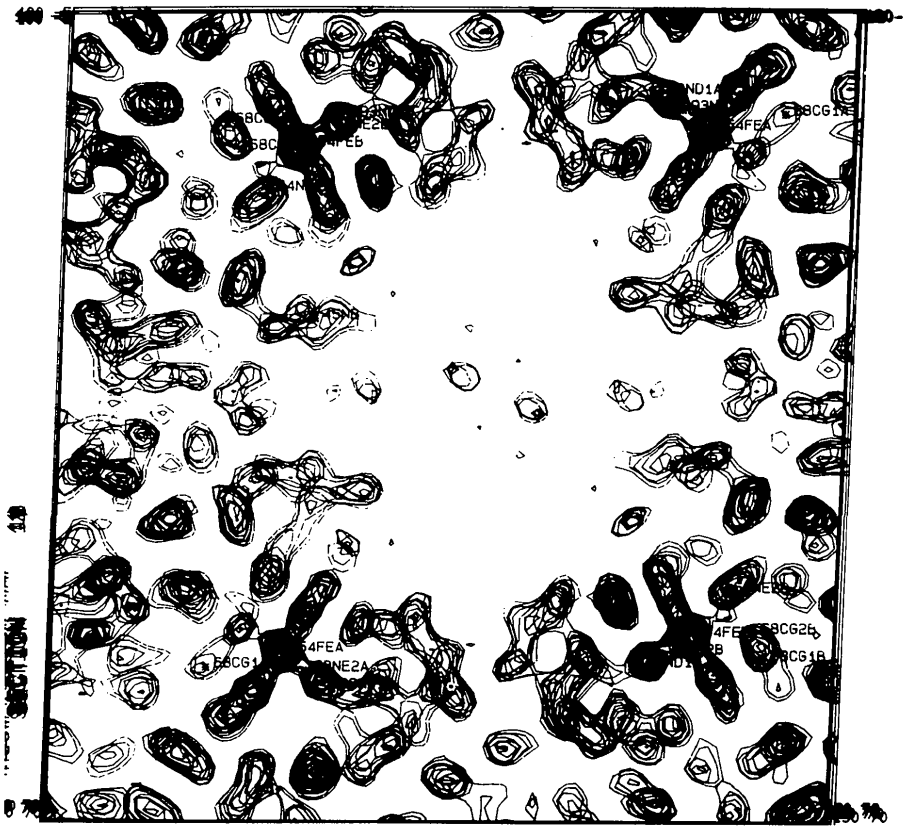

Fig. 3. Six sections (b = 7/72 to 12/72) of part of the $2F_o - F_c$ electron density map overlaid. The map shows four haem pockets from two asymmetric units arranged around a wide solvent channel; the *A* molecules are in the lower left and upper right corners. The map is contoured at intervals between 0·21 and 1·85 e Å^{-3}. The Fe ions represent the strongest features in the map.

Haem environment

A surprising observation from our maps is the absence of a peak in the electron density map situated at 2·1 Å from the iron atom in the distal haem pocket that would be consistent with a liganded water molecule (Figs. 3 and 5). This is true for both the *A* and the *B* molecules. Sperm-whale metmyoglobin crystallized at pH 5·75, has six-coordinated iron with a water molecule forming the distal ligand (Takano, 1977). Here, with crystals grown at pH 7·1, we observe for both of the pig myoglobin molecules a peak in the electron density map 3·5 Å from the iron and away from the normal to the haem plane that passes through the iron atom (Fig. 5). This density is more prominent in the *A* molecule than in the *B* molecule (Fig. 3); if this density is refined as a full occupancy water molecule, the *B* value associated with this molecule is average for water molecules in the structure as a whole. Fig. 5 shows a section of electron density through the haem pocket of the *A* molecule including the density associated with the distal pocket water molecule. The position of this water molecule, which is within hydrogen-bonding distance of histidine E7 Nε and in van der Waals contact distance of valine E11, is similar to that of the water molecule in the deoxymyoglobin model of Dr S. Phillips. This density is unlikely to be an artifact since it appeared in the first calculation of the map; a difference Fourier phased only on protein atoms. Omitting this water from the phasing and carrying out five subsequent cycles of refinement results in the reappearance of positive density in $F_o - F_c$ difference maps consistent with these observations. In contrast, inclusion of a water molecule at the sixth coordination site of the iron results in negative density appearing at this site in subsequent maps.

Recently, Bolognesi, Onesti, Gatti, Coda, Ascenzi & Brunori (1989) have reported the determination of the structure of *Aplysia limacina* metmyoglobin at high resolution. The coordination state of the iron in this protein (in which the distal E7 residue is a valine

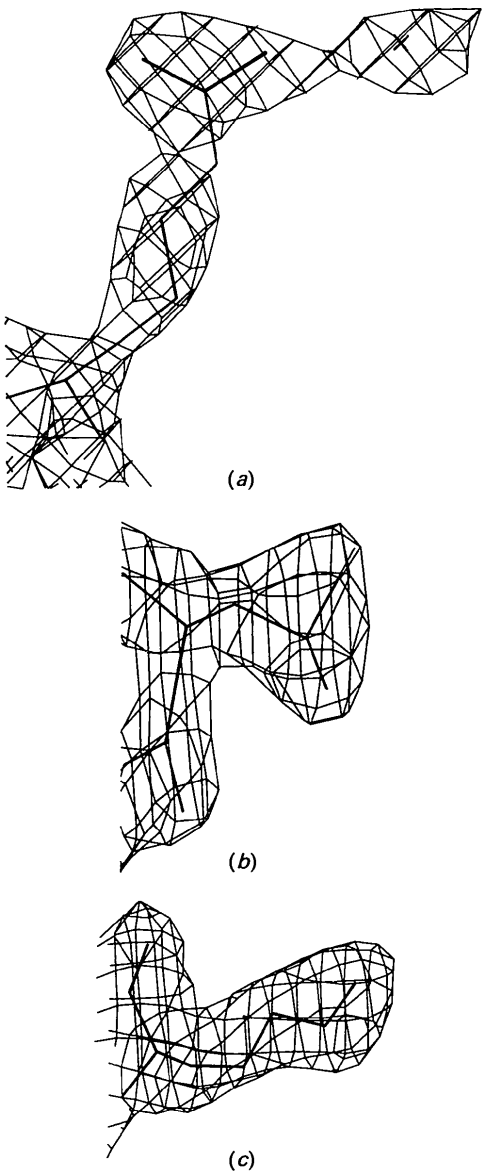

Fig. 4. Electron density associated with representative side chains in the *A* molecule of the structural model. (*a*) Arginine 31 contoured at 0·42 e Å^{-3}, (*b*) leucine 61 contoured at 0·42 e Å^{-3} and (*c*) lysine 45 contoured at 0·21 e Å^{-3}.

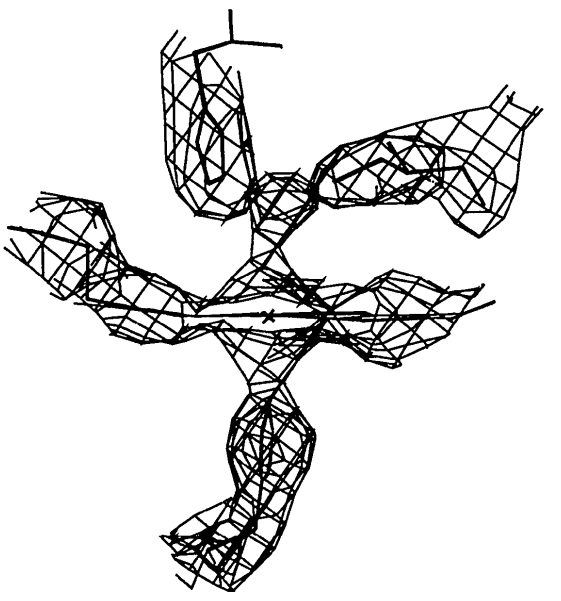

Fig. 5. A section through the haem pocket of the *A* molecule. The electron density is contoured at 0·4 e Å^{-3}. The haem group is seen in projection. In the distal pocket (above) the unliganded water molecule density is shown between residues histidine *E*7 and valine *E*11.

rather than a histidine) is pH dependent. At neutral pH no ligand is observed while at pH 9·0 there is a hydroxyl ligand at the sixth coordination position.* For *Chironomus thummi thummi* erythrocruorin, a monomeric haemoglobin, no bound ligand is observed in the met structure in crystals grown at pH 7·0 (Steigemann & Weber, 1979). Although this protein does have a distal histidine (E7), the crystal structure shows that its imidazole group is pointing out of the front of the haem pocket.

The iron–pyrrole nitrogen distances were not restrained in the course of refinement nor was the iron constrained to lie in the plane of the haem. The iron–proximal histidine distance was unrestrained up to the last two cycles of refinement. In our current model, the iron lies in the plane of the four pyrrole nitrogens (Fig. 5) though the iron–proximal histidine Nε distance (2·4 Å) is rather long. We have now collected higher-resolution (1·8 Å) data, and a more-detailed description of the structure and of the haem pocket will be presented when our structural model has been further refined.

When the sperm-whale and the pig myoglobin models are overlapped on the haem moieties, the amino-acid side chains in the neighbourhood of the haem are almost exactly superimposed. The major difference that is observed in this comparison results from the substitution of the arginine residue (R45), that spans the entrance to the distal haem pocket in sperm-whale myoglobin, for a lysine residue in pig myoglobin. This lysine is well defined in our structure with density extending to Nζ in both molecules (Fig. 4*a*). Kuriyan, Wilz, Karplus & Petsko (1986) have presented evidence for alternative conformations of the arginine 45 side chain in sperm-whale myoglobin and suggested that these may be associated with ligand entry and exit. For the pig protein, lysine 45 may perform the same role but the shorter side chain does not make the same bridging interactions as the guanidinium group of the arginine.

*Preliminary maps calculated on 1·8 Å data on a second crystal of pig myoglobin show a different ligation state; a water molecule is coordinated to the iron. It is possible that the pH of crystallization is close to the pH of the alkaline transition for pig myoglobin. Giacometti, Ascenzi, Brunori, Rigatti, Giacometti & Bolognesi (1981) have argued that the pH dependence of the ligation state in *Aplysia* myoglobin is directly linked to the alkaline transition of the protein. A variation in the pH of crystallization of a few tenths of a pH unit could therefore significantly influence the iron ligation chemistry.

Crystal contacts and crystal packing

Fig. 6 illustrates how the myoglobin molecules are packed in the crystal. The protein aggregates have 222 symmetry. The asymmetric units are stacked beside and on top of one another and do not interpenetrate. Consequently, there are very few contacts between the asymmetric units as is seen from Table 3. There is only a single contact between the *A* and *B* molecules within the same asymmetric unit, a probable hydrogen-bonding interaction between the imidazole rings of histidine 48 of each molecule. We have observed that the pH of the sodium phosphate buffer added to the crystallization solutions is critical, pH 6·8 to 7·2. A plausible explanation for this is that for the association of the two molecules of the asymmetric unit to take place, one of the histidine 48 residues must be ionized and the other must be uncharged. Thus, the probable charge–dipole interaction made between the two molecules may only be possible in a narrow pH range close to the pK_a of an imidazole group. Interestingly, as may be seen in Fig. 3, the haem and ligand sites are assembled around a *ca* 30 Å solvent channel in the crystal.

Concluding remarks

As these studies show (Fig. 2*b*), pig myoglobin is a remarkably good candidate for molecular replacement. Three properties of the crystals contribute to the quality of the solution obtained: (1) as is evident from our structural model, there is strong tertiary structure homology between the pig and sperm-whale myoglobins; (2) myoglobin has a high α-helical content giving rise to sets of parallel vectors in

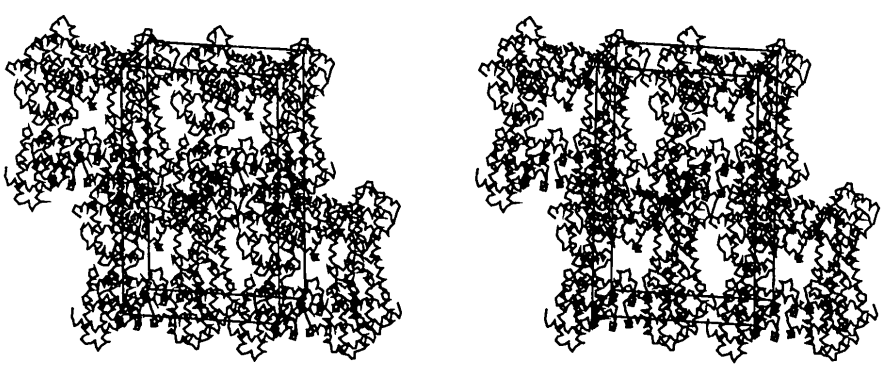

Fig. 6. Stereopair illustrating the packing of the myoglobin molecules within the $I2_1$ unit cell. The backbone of each monomer is represented as a ribbon using the program *FRODO* (Jones, 1978). The *c* and *a* cell directions are in the plane of the page with the *b* direction into the page.

Table 3. *Possible hydrogen-bonding and salt-bridging contacts between asymmetric units*

The table lists (*a*) interatomic distances between asymmetric units less than 3·0 Å and (*b*) possible ion/ion electrostatic interactions less than 3·5 Å. The symmetry operators 3 and 4 are $-x, y+\frac{1}{2}, -z$ and $\frac{1}{2}-x, y, \frac{1}{2}-z$, respectively.

Residue *I*		Residue *J*		Distance (Å)	Symmetry
(*a*) Interatomic distances					
Gly15*A*	O	Gln113*A*	Nε2	2·8	3
Asn66*A*	Nδ2	Gln128*A*	Oε2	2·7	3
Gln91*A*	Nε1	Glu148	O	2·8	4
Gly15*B*	O	Gln113*B*	Nε2	2·7	3
Asn66*B*	Nδ2	Gln128*B*	Oε2	2·9	3
Glu148*B*	O	Gln91*A*	Oε1	2·8	4
(*b*) Electrostatic interactions					
Lys16*A*	Nζ	Glu109*A*	Oε1	3·2	3
Glu18*A*	Oε2	Arg31*A*	Nη1	3·4	3
Glu18*A*	Oε2	Arg31*A*	Nη2	3·3	3
Lys63*A*	Nζ	Asp126*A*	Oδ2	3·3	3
Lys16*B*	Nζ	Glu109*B*	Oε2	3·2	3
Glu18*B*	Oε2	Arg31*B*	Nη1	3·2	3

the Patterson maps; (3) there is a paucity of close contacts between the molecules in the pig myoglobin crystal so that there are very few intermolecular vectors contributing to the observed Patterson function.

The manipulation of the gene encoding pig myoglobin for high-level expression of recombinant protein and the determination of an atomic resolution structure from crystals of this material provide a basis for detailed structure–function studies. This work is now being extended in two directions. Firstly, we have introduced a number of amino-acid substitutions into the protein by oligonucleotide-directed mutagenesis to probe the roles of individual side chains in the protein. The properties of the mutant proteins will be examined both kinetically and structurally. Secondly, we are extending the X-ray structural studies to higher resolution for both the wild-type and mutant proteins from which crystals have been grown.

We thank the SERC Protein Engineering Initiative for supporting this work.

References

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., ROGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. J. (1977). *J. Mol. Biol.* **112**, 535–542.

BOLOGNESI, M., ONESTI, S., GATTI, G., CODA, A., ASCENZI, P. & BRUNORI, M. (1989). *J. Mol. Biol.* **205**, 529–544.

DEREWENDA, Z. (1989). *Acta Cryst.* A**45**, 227–234.

DICKERSON, R. E. & GEIS, I. (1983). *Haemoglobin: Structure, Function, Evolution and Pathology.* Menlo Park: Benjamin/Cummings.

DODSON, G., HUBBARD, R. E., OLDFIELD, T. J., SMERDON, S. J. & WILKINSON, A. J. (1988). *Protein Eng.* **2**, 233–237.

EVANS, S. V. & BRAYER, G. D. (1988). *J. Biol. Chem.* **263**, 4263–4268.

GIACOMETTI, G. M., ASCENZI, P., BRUNORI, M., RIGATTI, R., GIACOMETTI, G. & BOLOGNESI, M. (1981). *J. Mol. Biol.* **151**, 315–319.

HENDRICKSON, W. A. & KONNERT, J. H. (1979). *Biomolecular Structure, Conformation, Function and Evolution,* Vol. 1, edited by R. SRINIVASAN, pp. 43–57. Pergamon: New York.

JONES, T. A. (1978). *J. Appl. Cryst.* **20**, 268–272.

KENDREW, J. C., DICKERSON, R. E., STRANDBERG, B. E., HART, R. G., DAVIES, D. R., PHILLIPS, D. C. & SHORE, V. C. (1960). *Nature (London)*, **185**, 422–427.

KURIYAN, J., WILZ, S., KARPLUS, M. & PETSKO, G. (1986). *J. Mol. Biol.* **192**, 133–154.

LUZZATI, V. (1952). *Acta Cryst.* A**25**, 712–713.

NAGAI, K. & THØGERSEN, H.-C. (1984). *Nature (London)*, **309**, 810–812.

SCOLOUDI, H. & BAKER, E. N. (1978). *J. Mol. Biol.* **126**, 637–660.

SERC Daresbury Laboratory (1986). *CCP4. A Suite of Programs for Protein Crystallography.* SERC Daresbury Laboratory, Warrington, England.

SPRINGER, B. A. & SLIGAR, S. G. (1987). *Proc. Natl Acad. Sci.* **84**, 8961–8965.

STEIGEMANN, W. & WEBER, E. (1979). *J. Mol. Biol.* **127**, 309–338.

TAKANO, T. (1977). *J. Mol. Biol.* **110**, 537–568.

VARADARAJAN, B., SZABO, A. & BOXER, S. G., (1985). *Proc. Natl Acad. Sci.* **82**, 5681–5684.

Acta Cryst. (1990). B**46**, 377–389

Structures of Six Terminally Substituted [*n*]Staffanes, *n* = 1–4

BY ANDRIENNE C. FRIEDLI, VINCENT M. LYNCH, PIOTR KASZYNSKI AND JOSEF MICHL

Center for Structure and Reactivity, Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712-1167, *USA*

(*Received* 1 *August* 1989; *accepted* 14 *December* 1989)

Abstract

1,3-Diacetylbicyclo[1.1.1]pentane (2), $C_9H_{12}O_2$, M_r = 152·19, monoclinic, $C2/m$ (No. 12), a = 11·691 (4), b = 6·624 (4), c = 5·807 (2) Å, β = 112·40 (2)°, V = 415·8 (3) Å^3, D_x = 1·22 g cm^{-3}, Z = 2, Mo $K\alpha$, λ = 0·71073 Å, μ = 0·7912 cm^{-1}, $F(000)$ = 164, T = 163 K, R = 0·0703 for 310 reflections $\{F_o \geq 4[\sigma(F_o)]\}$. Methyl [2]staffane-3-carboxylate (3), $C_{12}H_{16}O_2$, M_r = 192·26, monoclinic, $P2_1/m$ (No. 11), a = 5·793 (2),

0108-7681/90/030377-13$03.00

© 1990 International Union of Crystallography