Molecular structure design, search and energy calculations were performed using the *SYBYL* molecular modeling system (Tripos Associates Inc., 1983).

Concluding remarks

On the basis of the present study we are able to make the following inferences in relation to the description of the opiate receptor. The modulation of the activity by the C16 substituent should be interpreted by steric hindrance, incompatible with the receptor geometry, and not by a misalignment of the nitrogen N′ atom or a distortion of the fused ring system. Secondly, we can assume that the increased opiate activity related to the presence of the 6–14 etheno bridge along with the R^2 substituent must be correlated to the presence of the aforementioned hydrogen bond, which also accounts for the C19 stereoselectivity. Finally, the presence of this intramolecular hydrogen bond confers on the opioid alkaloids the conformational properties for adopting common conformations with cyclic opioid peptides. This argument constitutes a key feature in further molecular modeling studies to determine the receptor-bound conformation of opioid alkaloids and peptides. This major finding will lead to pertinent studies of the distinction between opioid receptor types that modulate physiological processes.

Financial support by the Natural Sciences and Engineering Research Council of Canada (NSERC) is gratefully acknowledged. We are grateful to Professor B. Belleau for fruitful discussions and his help in obtaining the compounds.

References

BECKETT, A. H. & CASY, A. F. (1954). *J. Pharm. Pharmacol.* **6**, 986–988.

DIMAIO, J., BAYLY, C. I., VILLENEUVE, G. & MICHEL, A. G. (1986). *J. Med. Chem.* **29**, 1658–1663.

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

LEWIS, J. W., BENTLEY, K. W. & COWAN, A. (1971). *Annu. Rev. Pharmacol.* **11**, 241–243.

LEWIS, J. W., MAYOR, P. A. & HADDLESEY, D. I. (1973). *J. Med. Chem.* **16**, 12–14.

LOEW, G. H. & BERKOWITZ, D. S. (1979). *J. Med. Chem.* **22**(6), 603–607.

MAIN, P., FISKE, S. J., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J.-P. & WOOLFSON, M. M. (1980). *MULTAN80. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.

MICHEL, A. G., EVRARD, G., NORBERG, B. & MILCHERT, E. (1988). *Can. J. Chem.* **66**, 1763–1769.

MICHEL, A. G., PROULX, M., EVRARD, G., NORBERG, B. & MILCHERT, E. (1988). *Can. J. Chem.* **66**, 2498–2505.

STEWART, J. M., HALL, S. R., ALDEN, R. A., OLTHOF-HAZEKAMP, R. & DOHERTY, R. M. (1983). *The XTAL System of Crystallographic Programs: User's Manual.* Tech. Rep. TR-1364. Univ. of Maryland, College Park, Maryland, USA.

Tripos Associates Inc. (1983). *SYBYL. Molecular Modeling Software.* Tripos Associates Inc., St Louis, Missouri, USA.

VAN DEN HENDE, J. H. & NELSON, N. R. (1967). *J. Am. Chem. Soc.* **89**, 2091–2905.

Acta Cryst. (1990). B**46**, 409–418

Refinement of a Partially Oxygenated T State Human Haemoglobin at 1·5 Å Resolution

BY D. A. WALLER* AND R. C. LIDDINGTON†

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

(*Received* 7 *August* 1989; *accepted* 2 *January* 1990)

Abstract

The degree of ligation of T state human haemoglobin crystals is reduced by inositol hexaphosphate (IHP). The structure of a partially ligated haemoglobin has been refined using fast Fourier restrained-least-squares techniques. Manual interventions were required to escape from local minima and introduce a large number of solvent molecules. Individual isotropic temperature factors were refined for all atoms and the final average atomic temperature factor is 32·3 Å^2. The final R factor is 19·6% for all data between 10 and 1·5 Å. The final model consists of 4560 protein atoms and 313 solvent molecules. The occupancies of the ligand atoms and the anisotropic behaviour of the iron atoms have been refined, demonstrating that the α haem groups are only partially ligated and that there is no ligation of the β haems. Density for the IHP indicates that it is not

* To whom correspondence should be addressed at his present address: Astbury Department of Biophysics, University of Leeds, Leeds LS2 9JT, England.

† Present address: Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts, USA.

0108-7681/90/030409-10$03.00

well ordered even though changes in the ligation and structure of the haemoglobin indicate its presence.

Introduction

The respiratory protein in vertebrate blood is haemoglobin which transports oxygen from the lungs to other tissues. This function depends on the cooperative behaviour of the protein; as one oxygen molecule is bound, haemoglobin's affinity for the next oxygen increases until four oxygen molecules are bound – one to each subunit. The cooperative behaviour of proteins has been described in terms of a two-state model in which the states have different quaternary structures with different affinities for ligands (Monod, Wyman & Changeux, 1965). These states are known as the T (tense) state which in the case of haemoglobin corresponds to the deoxygenated conformation and the R (relaxed) state which corresponds to the oxygenated conformation.

Haemoglobin has a molecular weight of 64 500 daltons. It is a tetramer consisting of two α-globin chains each containing 141 amino-acid residues and two β-globin chains containing 146 amino acids each. The α chains contain seven and the β chains eight helical segments separated by non-helical 'corners'. The helices of each subunit are folded to form a pocket for a haem group, protoporphyrin IX, which consists of an iron atom coordinated to the four pyrrole nitrogens of a porphyrin ring. One axial coordination position of the iron is occupied by a histidine side chain from the globin, forming the only covalent link between haem and globin, the other forms the ligand binding site.

The principal human haemoglobin structures presently available are, for the R state: 2·1 Å oxyhaemoglobin (Shaanan, 1983) and 1·9 Å cross-linked deoxyhaemoglobin (Luisi, 1986); for the T state: 1·7 Å deoxyhaemoglobin grown from $(NH_4)_2SO_4$ (Fermi, Perutz & Shaanan, 1984) and 2·1 Å deoxy-, semioxy- and aquomethaemoglobins grown from poly(ethylene glycol) (PEG) 8000 (Liddington, Derewenda, Dodson & Harris, 1988). It was observed (Liddington, 1986) that crystals of haemoglobin grown from PEG 1000 with inositol hexaphosphate (IHP), which is a powerful allosteric effector (similar to the avian allosteric effector inositol pentaphosphate), diffracted to higher resolution and were more stable in an X-ray beam than crystals grown without IHP. This observation led directly to the collection of 1·5 Å diffraction data from these IHP-containing T state crystals and the refinement which this paper describes.

Data collection and processing

Haemoglobin was extracted from human blood (obtained from the local hospital) by the process of Drabkin (1949) and crystals were grown in batch. Optimum conditions were 18–25% *v/v* PEG 1000 and pH 7·2–7·4 with a twofold excess of IHP present in the crystallization mixture (Liddington, 1986). The solution of oxyhaemoglobin deoxygenates over a number of days following the addition of PEG. The crystals produced are of tabular or wedge-shaped habit and generally large – typically 2 × 1 × 0·5 mm with some as large as 3 × 3·3 × 5 mm. They are of space group $P2_12_12$ and have cell dimensions: a = 95·8 (3), b = 97·8 (3), c = 65·5 (3) Å. They are typical of orthorhombic crystals with the longest crystal dimension corresponding to the smallest cell dimension. It appears that crystals removed from crystallization tubes are initially unliganded but partial ligation can occur with subsequent exposure to oxygen.

Data were collected on station 9.6 of the Synchrotron Radiation Source, Daresbury, England, using an Arndt–Wonacott oscillation camera with CEA Reflex-25 film. It was necessary to collect two data sets because the exposure times required for the high-resolution data lead to the saturation of lower-angle reflections. The wavelength used for both data collections was 1·00 Å.

The first data set extends from infinity to 1·9 Å. This data set was collected from two large crystals; one mounted about c^* was used to collect 84° of data with periodic translations of the crystal when necessary. The second crystal was mounted using a 'bridle' mount so that a crystal with its c^* axis along a capillary could be used to collect 30° of blind-region data rotating about a^*. Conventional flat cassettes were used with three films in a pack. A pair of orthogonal still photographs were taken to determine the orientation of each crystal.

The second data set only contains useful data between 3·0 and 1·5 Å. Five crystals mounted about c^* were used to collect 75·2° of data and a further 'bridle'-mounted crystal was used to collect 14·4° of blind-region data. 'V'-shaped cassettes (Arndt & Wonacott, 1977) with four films in a pack were used throughout. The intention was that the fourth film should be used for measurement of at least some of the low-resolution data which were saturated on the first three films; although this seemed feasible the necessary modifications to the software were never completed and only the first three films in each pack were used to obtain integrated intensities. A pair of orthogonal still photographs were taken to determine the orientation of each crystal.

The two data sets were treated independently to produce two sets of unique reflections. Films were digitized using a Joyce–Loebel Scandig-3-microdensitometer with a 50 × 50 μm raster and an optical density range of 0–2*D*. The initial orientation of each crystal was determined using still photo-

graphs and the program *IDXREF* running on an ICL Perq computer, which allowed the positions of reflections to be entered directly from the film using a data tablet. The subsequent refinement of orientation and integration of intensities was performed on a VAX 11/750 computer using the *MOSCO* suite of programs (Wonacott, 1980). Each crystal was treated in a similar fashion, starting from the orthogonal stills and proceeding pack by pack. When three films in eack pack were processed, corrections for oblique incidence, Lorentz and polarization effects were applied. Finally the intrapack scales were calculated and applied giving a set of integrated intensities ready for reduction to a unique set.

Data reduction employed the Fox & Holmes algorithm (1966) to calculate and apply scales. Post-refinement (Winkler, Schutt & Harrison, 1979) was used to improve the classification of partial reflections by refining cell dimensions, crystal orientation and reflection width. After a series of tests, partial reflections were summed where possible and scaled when summation was not possible but the partiality was greater than 0·4. On average each unique reflection in the 1·9 Å data set was measured five times and each unique reflection in the 1·5 Å data set was measured four times. The merging R_m factor ($R_m = \sum|I_i - \langle I\rangle|/\sum|I_i|$, where I_i is the intensity of an observation, $\langle I\rangle$ is the mean value for that reflection and the summations are over all reflections) for the 1·9 Å data set containing 48 943 unique reflections was 7·7%, whilst for the 1·5 Å data set with 73 013 unique reflections $R_m = 14{\cdot}0\%$. 78·6% of the reflections in the 1·9 Å data set and 70·5% of the reflections in the 1·5 Å data set have intensities greater than $3\sigma_I$. Merging the two merged data sets gave a value of $R_m = 10{\cdot}9\%$ for the 30 680 common reflections. Overall 90·3% of the unique data to 1·5 Å resolution were observed.

Refinement

The structure was refined using a version of the restrained-least-squares refinement program *PROLSQ* (Hendrickson, 1985) incorporating fast Fourier structure-factor and gradient calculation (Agarwal, 1978). A typical cycle of positional parameter refinement with 4560 atoms and 87 000 reflections took 15 min of CPU time on a VAX 8550. The stereochemistry was restrained using the weighting scheme shown in Table 1, structure-factor amplitudes being given unit weights rather than adopting a resolution-dependent weighting scheme. Stereochemical restraints and structure-factor terms were given equal weight. Temperature factors have been subjected to unrestrained refinement throughout this work. This contrasts with the conventional Konnert–Hendrickson approach which uses the temperature factors of covalently linked neighbours to restrain the temperature factor of a given atom. Large temperature factors were reduced before refinement was allowed to continue once again: in this way high temperature factors which resulted from incorrectly positioned atoms were eliminated after the atoms were repositioned. In general, temperature factors were only refined after positional parameter refinement had converged. Occupancies have not been refined since, even with 1·5 Å data, experience has shown that they are highly correlated to the temperature factors.

Table 1. *Weighting scheme for geometric restraints and standard deviations after the last cycle of refinement*

Equal weights were applied to X-ray and geometric terms.

	σ*	Standard deviations
Distances (Å)		
Bond lengths	0·02	0·025
Bond angles	0·04	0·064
Dihedral angles	0·06	0·075
Fe—Nε2	0·10	0·168
Planar groups	0·02	0·019
Chiral volumes (Å^3)	0·12	0·173
Non-bonded contacts (Å)		
Single torsion contacts	0·50	0·210
Multiple torsion contacts	0·50	0·222
Possible hydrogen bonds	0·50	0·189
Torsion angles (°)		
Peptide planes	20	4·7
Staggered	20	20·1
Orthonormal	20	34·9

* The weight in each class corresponds to $1/\sigma^2$.

The possible starting points for refinement were the structures of deoxy-, semioxy- and aquomet-haemoglobin grown from PEG 8000 (Liddington *et al.*, 1988). The semioxygenated form was not used because it seemed likely that the new data were from a partially oxygenated form and it was undesirable to bias the starting model. *R* factors were calculated for the 1·9 Å data with the deoxy and aquomet models ($R = \sum_h||F_o| - |F_c||/\sum_h|F_o|$ where $|F_o|$ and $|F_c|$ represent the observed and calculated structure-factor amplitudes and the summations are over all measurements). The aquomet model gave an *R* factor of 48·5% and the deoxy model 31·0%. The deoxy model, including 89 water molecules, was therefore selected as a starting point for refinement.

The first 13 cycles of refinement used all data between 10 and 1·9 Å. Starting with the deoxy model, three cycles of positional parameter refinement reduced the *R* factor from 31·0 to 24·8%. Next the deoxy model was rotated and translated into the new coordinate positions using a core of helices for a least-squares overlap. This procedure, described in detail by Derewenda (1989), provides a simple technique for the refinement of the rotation and translation parameters which need to be applied

to the starting model. Six further positional parameter refinement cycles and one cycle of temperature-factor refinement reduced the R factor from 28·6 to 22·0%. At this stage Fourier maps with coefficients $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ and phases α_c were calculated and inspected on an Evans and Sutherland PS330 using the program *FRODO* (Jones, 1978). This inspection showed that the general quality of the density was good and that both α haems appeared to be liganded. Dioxygen ligands were introduced to both α haems and two positional parameter and one temperature-factor refinement cycles performed. This reduced the R factor from 21·8 to 21·1%. At this stage the 1·5 Å data were introduced.

The initial R factor following the introduction of the 1·5 Å data (10–1·5 Å, all data) was 26·5%. A total of 69 refinement cycles were completed after the introduction of the 1·5 Å data; of these 47 cycles are positional parameter refinements and 22 temperature-factor refinements. The cycles were punctuated by a series of 12 manual interventions which were used for the adjustment of numerous side chains and the indentification of water molecules. Systematic examinations of every amino-acid residue and its immediate environment were performed twice, once working from the N terminus of the α_1 chain towards the C terminus of the β_2 chain and again in the reverse direction. The N termini of the β chains and all the haem groups were omitted from the phasing at various points and their appearance in difference maps used to adjust the model structure. In this way these regions can be examined without introducing bias from the model structure. The N termini were examined in this fashion because they are not well ordered and the haem groups because their geometry is crucial when the mechanism of ligand binding is being studied.

The refinement started with the 89 water molecules present in the deoxy model: of these 71 remain and a further 242 new water molecules have been added making 313 in total. New water molecules were inserted as the work proceeded on the basis of density appearing in both $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps (at a level of at least three times the r.m.s. deviation in the difference map). The majority of the new water sites were identified in the course of the examination of maps on the graphics system; however, some 85 possible sites were located using a peak-search program to examine a difference map. These sites were then examined on the graphics system and 65 of them included as water sites. After a total of 68 cycles of refinement all of the water molecules were removed and new Fourier maps calculated and examined. Even with no waters contributing to the phasing 272 water molecules appeared in these maps. When these water molecules were included in the phasing and maps recalculated, the remaining 41 water molecules reappeared. The positions of the water molecules were fixed and their contributions to the structure factors calculated for use in all further refinement cycles. Each of the water molecules was assigned unit occupancy; since occupancies were not refined the high temperature factors of many of them reflect the fact that the sites are only partially occupied.

The fluctuation of R factor during the refinement is presented in Fig. 1. The rotation and translation of the deoxy coordinates described above is marked LSQ and the introduction of high-resolution data is marked 1·5. The removal of all water molecules after cycle 68 is marked by WAT.

In a number of cases, examining the temperature factors for atoms of the side chains of histidine residues indicated that the side chain should be flipped. This effectively swaps atom Cδ1 with Nδ1 and Cε2 with Nε2. Such a flip was performed at a late stage in the refinement for five of the histidine groups. Fig. 2 shows the temperature factors for one of these histidines before and after flipping.

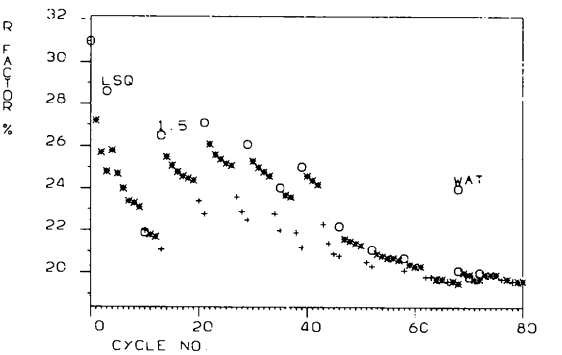

Fig. 1. Reliability index R plotted against cycle number. Temperature-factor refinement, +; positional-parameter refinement, *; and restart after manual intervention, o.

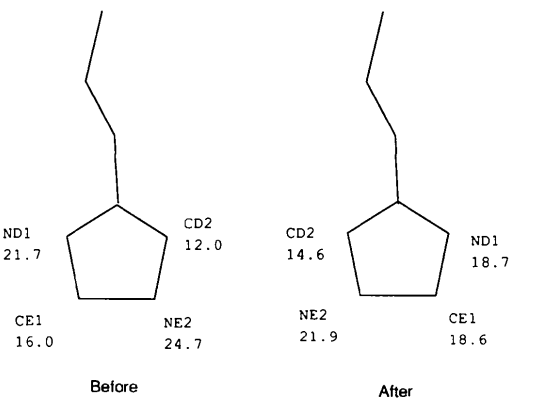

Fig. 2. Temperature factors (Å^2) of histidine α_1 G10 before and after rotation about the C_β—C_γ bond.

Hydrogen-bonding considerations should also indicate the correct orientation of the side chain, but it is interesting that, at least in the well defined parts of the structure, it is possible to distinguish between carbon and nitrogen atoms.

Final refined structure

The final R factor is 19·6% for all data between 10 and 1·5 Å. R.m.s. deviations of geometry for the last refinement cycle and their target σ's are shown in Table 1.* Out of the total number of first-, second- and third-neighbour distances only 4% deviate by more than three standard deviations from the corresponding ideal values. The r.m.s. coordinate shift in the last cycle of refinement was 0·006 Å. The only significant peaks in the final $|F_o| - |F_c|$ map are due to inositol hexaphosphate which was not included in the phasing at any time. The quality of the refined structure is illustrated by parts of the $2|F_o| - |F_c|$ electron density map shown in Fig. 3.

A Luzzati plot (Luzzati, 1952) is presented in Fig. 4. It shows an error of 0·2 Å between 4·0 and 2·0 Å, rising to 0·25 Å at 1·7 Å and 0·3 Å at 1·5 Å. This indicates that larger errors were introduced in the high-resolution stages of this work, and is a reflection of the lower quality of the 1·5 Å data set compared to the 1·9 Å set. However, the Luzzati plot is generally regarded as providing an upper limit on errors owing to its assumptions that differences between $|F_o|$ and $|F_c|$ only arise from coordinate errors and that all atoms are identical. The Luzzati plot also assumes that errors are random; any systematic error may of course increase the errors in specific regions. Alternative estimates of error obtained from least-squares overlaps of non-crystallographically related identical parts of the structure indicate that in the ordered helical regions the error in coordinates is 0·15 Å. Because of the greater scattering power of the iron atoms, the error in their positions is estimated to be no more than 0·05 Å.

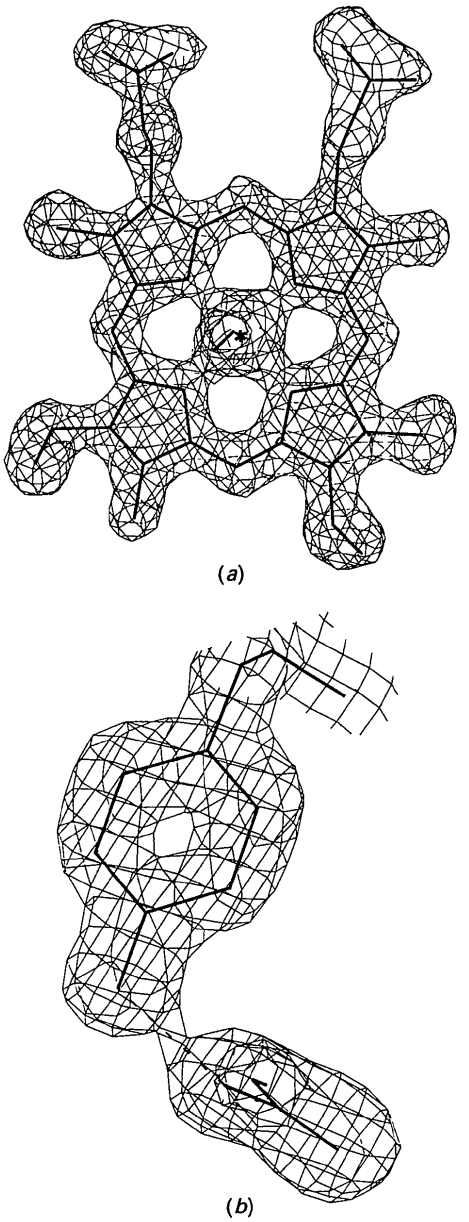

Fig. 3. Two parts of the final $2|F_o| - |F_c|$ electron density map calculated at 1·5 Å resolution showing: (*a*) α_1 haem and (*b*) α_1 Tyr42 to β_2 Asp99 short hydrogen bond (2·43 Å) both contoured at 0·23 e Å^{-3}.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1THB, R1THBSF), 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 37034 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

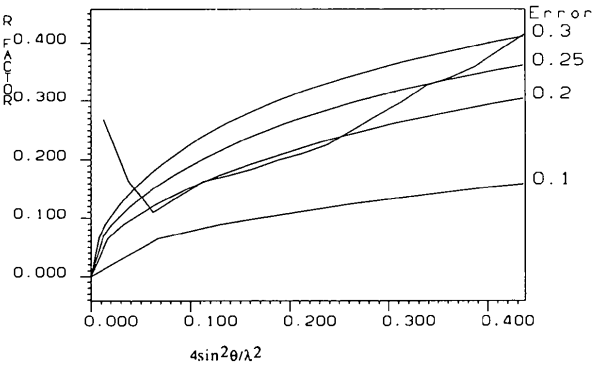

Fig. 4. Luzzati plot for the final refined structure.

In the final model atomic temperature factors for all atoms range from 10·7 to 77·3 Å^2 with a mean value of 32·3 Å^2. Removing the water molecules from the survey of temperature factors has no effect on the maximum and minimum values, the average being reduced by just 1·3 Å^2 to 31·0 Å^2. These values are somewhat higher than the temperature factor of 23·1 Å^2 obtained from a Wilson plot for the data. The temperature factors exhibit a physically sensible gradual increase along side chains. The refinement was halted as the largest temperature factors approached 79·0 Å^2 – which is equivalent to an r.m.s. displacement of 1 Å. Fig. 5 shows plots of temperature factors averaged over main-chain and side-chain atoms for the four subunits of haemoglobin. The helices are represented by blocks along the bottom of the plots lettered *A–H*. The lower temperature factors generally correlate well with the helical segments of the protein.

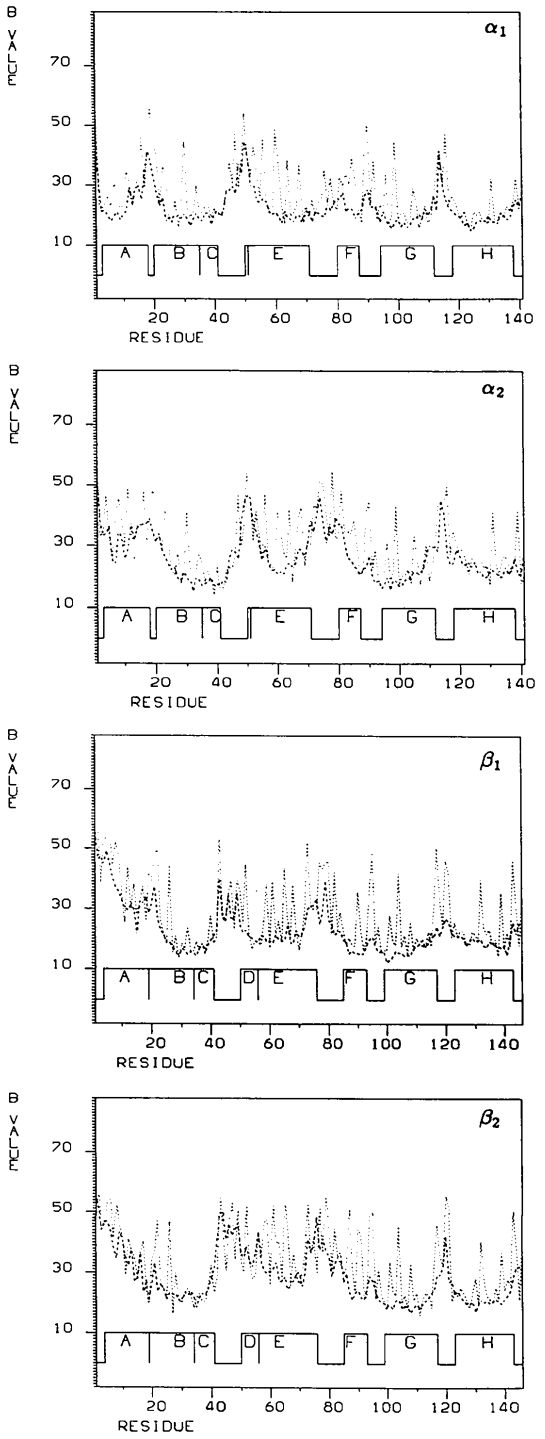

Fig. 5. Averaged main-chain (dashed line) and side-chain (dotted line) temperature factors plotted against residue number for the four subunits of haemoglobin.

A Ramachandran plot (Ramachandran, Ramakrishnan & Sasisekharan, 1963) for the whole molecule is shown in Fig. 6. This indicates that the geometry of the main chain is well behaved with the vast majority of the φ, ψ angle pairs within the 'hard-sphere' allowed regions. Those residues which lie outside the allowed regions can be classified in one of three groups. First there are quite a large number of residues in the so called 'bridge' region between the allowed α-helix and β-sheet regions. This reflects the number of residues involved in β turns in the non-helical segments of the protein. Second there are eight glycine residues – six in the lower right-hand corner of the map and two close to the bottom right-hand corner of the left-handed α-helix region. These are all reasonable because large areas of conformational space which are not accessible to ordinary amino-acid residues are accessible to a glycine residue owing to its lack of a C_β atom.

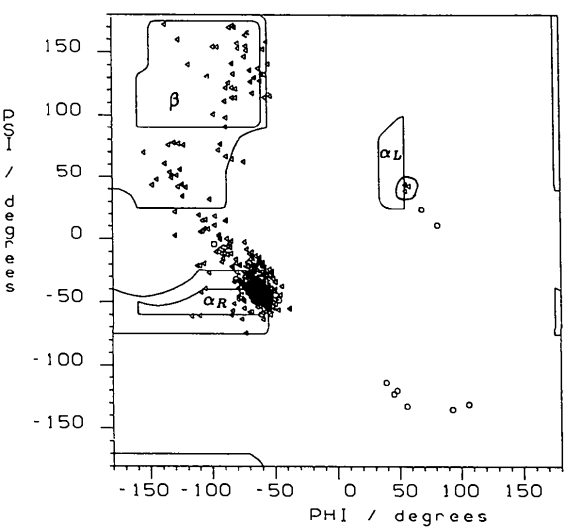

Fig. 6. Ramachandran map for the final refined structure. Dihedral-angle regions for α-helices, β-pleated sheets and left-handed α-helices are indicated, respectively, by α_R, β and α_L. Glycines are indicated by circles and other residues by triangles.

These eight residues are in fact only four independent residues because of the duplication in the pairs of chains. Third there are four residues on the edge of the left-handed α-helix region (they are ringed in Fig. 6). These four are in fact equivalent residues in the four subunits, being arginine *FG*4 in the α chains and histidine *FG*4 in the β chains. This unusual conformation of residue *FG*4 is observed in both haemoglobin and myoglobin structures and may be significant in maintaining the geometry of the *FG* corner which forms part of the functionally significant $\alpha_1\beta_2$ interface.

Ligand binding was considered first by examining contoured sections of electron density maps calculated in such a way that the ligand atom closest to the iron (named *OL*1), the iron atom itself, two of the porphyrin nitrogens and the Nε atom of the proximal histidine all lie in the plane of the section. Fig. 7 shows sections of the $2|F_o| - |F_c|$ map including the ligand in the phasing and the $|F_o| - |F_c|$

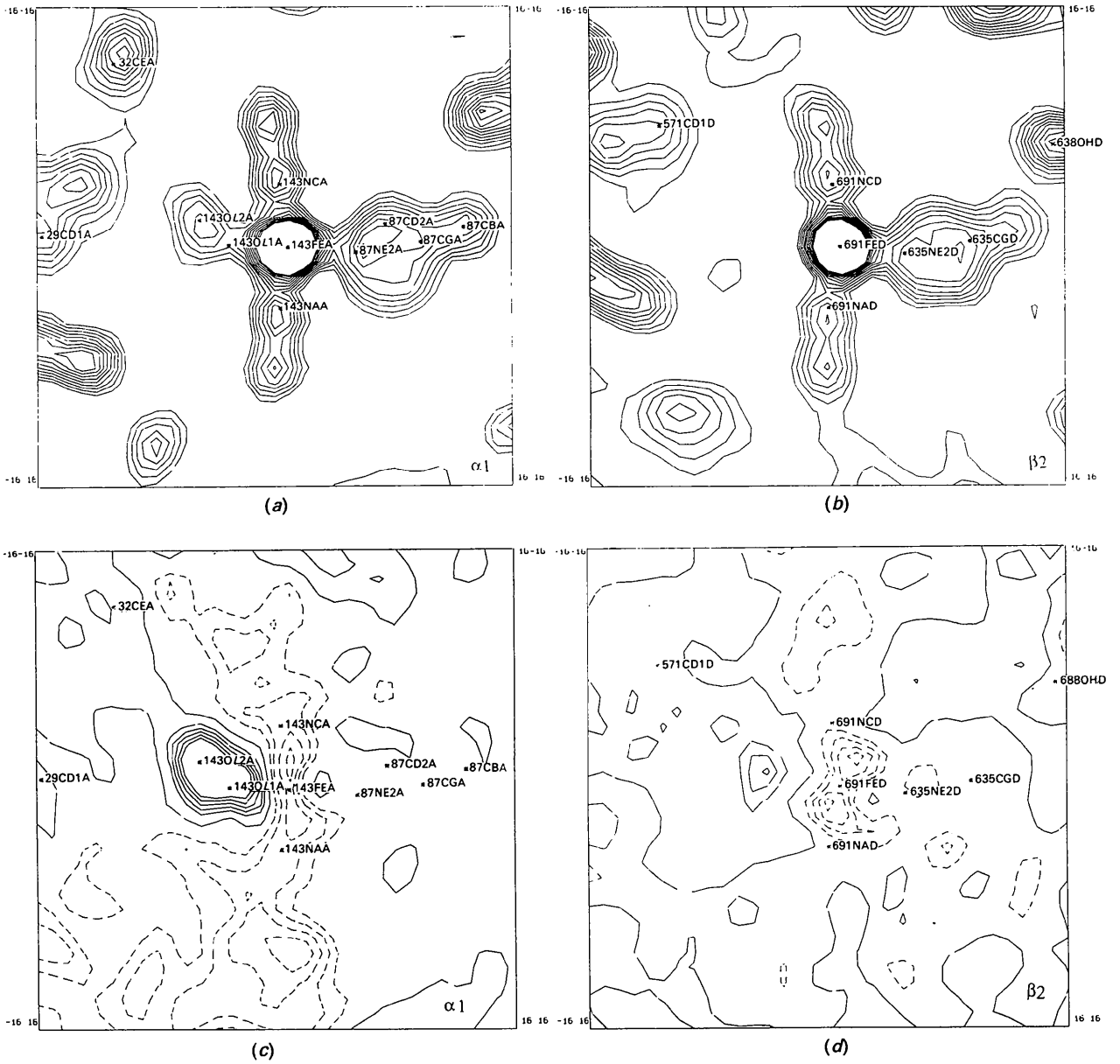

Fig. 7. (*a*) $2|F_o| - |F_c|$ map section through α_1 haem, (*b*) $2|F_o| - |F_c|$ map section through β_2 haem; both maps are contoured from zero in steps of 0·15 e Å^{-3}. (*c*) $|F_o| - |F_c|$ map section through α_1 haem with ligand omitted, (*d*) $|F_o| - |F_c|$ map section through β_2 haem; both maps are contoured from $-0{\cdot}3$ to 0·3 e Å^{-3} in steps of 0·05 e Å^{-3}. Solid contours are positive, dashed contours are negative. ($\pm 0{\cdot}05$ e Å^{-3} contours are omitted but zero level is included.)

Table 2. *Parameters* (Å, °) *of haem geometry and environment*

	Distances of Fe from the plane*			NA—Cε1	NC—Cδ2	Nδ1—C=O	OL1–Nε2	OL2–Nε2			Fe—Nε2
	Np	Pyr	Cp	(F8)	(F8)	(F8) (F4)	(E7)	(E7)	Fe—O	Fe—O—O	(F8)
Haemoglobin with IHP											
α_1	0·28	0·44	0·59	3·16	3·77	2·81	2·59	2·52	1·98	140·9	2·28
α_2	0·32	0·49	0·62	3·15	3·77	2·79	2·77	2·70	2·05	144·6	2·26
β_1	0·30	0·43	0·53	3·24	3·64	2·79	—	—	—	—	2·25
β_2	0·29	0·41	0·52	3·28	3·65	2·78	—	—	—	—	2·18
Deoxygenated haemoglobin											
α_1	0·32	0·49	0·63	2·96	3·84	2·55	—	—	—	—	2·25
α_2	0·35	0·51	0·64	3·24	3·73	2·66	—	—	—	—	2·27
β_1	0·32	0·39	0·45	3·36	3·82	2·84	—	—	—	—	2·26
β_2	0·52	0·54	0·54	3·27	3·78	2·90	—	—	—	—	2·17
Semioxygenated haemoglobin											
α_1	0·17	0·35	0·49	3·02	3·66	2·83	2·86	2·63	1·78	149·0	2·18
α_2	0·20	0·35	0·48	3·28	3·72	2·88	2·88	2·89	1·86	156·0	2·32
β_1	0·21	0·29	0·35	3·20	3·80	2·91	—	—	—	—	2·24
β_2	0·23	0·23	0·21	3·10	3·47	2·99	—	—	—	—	2·11

* Np = plane through pyrrole N atoms, Pyr = plane through all pyrrole atoms, Cp = plane through outer two pyrrole C atoms.

map with the ligand omitted from the phasing for the α_1 and β_2 haems. For the α_1 haem the $|F_o| - |F_c|$ map shows clear indications of ligand present at the haem group; however, the $2|F_o| - |F_c|$ synthesis suggests that the ligand site may not be fully occupied. The β_2 haem shows no sign of ligation.

The occupancy of the ligand atoms and the possible anisotropic behaviour of the iron atoms were refined using a program originally designed for the refinement of heavy-atom parameters in isomorphous replacement. Tests using carbonyl oxygen atoms of peptide groups showed that the refinement of the occupancy of such light atoms was feasible and gave reasonable results. The anisotropic refinement of the iron atoms was straightforward; however, determining the occupancy of the ligand sites was not, eventually the following procedure yielded useful results. The temperature factors of the putative ligands were set to zero and their occupancies refined in shells of $4\sin^2\theta/\lambda^2$. These occupancies were then plotted against $4\sin^2\theta/\lambda^2$ along with $\exp(-B\sin^2\theta/\lambda^2)$ for various values of B. Comparison of occupancy and $\exp(-B\sin^2\theta/\lambda^2)$ plots provided an estimate of the occupancy and temperature factor of each ligand atom; these estimated values were then used in a further round of refinement where temperature factors alone were refined to correspond to the estimated occupancies. The result of this refinement was that both α haems were estimated to have an occupancy of 0·4 for the OL1 atom and 1·0 for the OL2 atom. The occupancy of the OL1 atoms indicate that the α haems are not fully ligated. It is thought that when the ligand is not present a water molecule occupies the position of the second oxygen atom, hydrogen bonding to the side chain of the distal histidine and giving rise to the unit occupancy of this site. The OL1 and OL2 atoms have temperature factors of 25·2 and 32·7 Å^2 respectively for the α_1 haem, and 21·6 and 34·3 Å^2 for the α_2 haem. The β haems show no sign of ligation. This contrasts with the previously studied semioxygenated haemoglobin which had fully occupied oxygen molecules bound to the α haems and a low level (25%) of ligation at the β haems. The iron atoms in the α haems exhibit a distinct anisotropy in the direction of the iron–oxygen bond, supporting the idea that they are at least partially ligated. There is no sign of anisotropy in this direction for the β haem groups.

Various parameters which describe the geometry of the haem groups of this structure and both the semioxygenated and deoxygenated forms grown from PEG 8000 are listed in Table 2. The difference in ligation of the haem groups correlates well with differences in the haem geometries. Most notably the α-haem iron atoms have not moved into the plane of the haem groups as they do in the semioxy structure. The lack of movement of the iron atom leads to longer apparent iron-to-oxygen distances than expected. The β subunit haem groups appear to be distorted in such a way that they approach a stable five-coordinate geometry for the iron atoms which means that there is neither space nor necessity for the binding of ligand to the β haems, see Fig. 8.

Comparison of this structure with semioxy- and deoxyhaemoglobin structures reveals numerous small (<0·4 Å) differences between the structures but very few large ones (>0·6 Å). What large differences there are occur at the N termini of both β_1 and β_2 subunits and at the *EF* corner of the β_2 subunit. Both β-chain N termini and *EF* corners form part of the IHP binding site first described by Arnone & Perutz (1974). The lack of a substantial difference at the β_1 *EF* corner is probably due to packing constraints. The $|F_o| - |F_c|$ map at the entrance to the central cavity between the N termini of the β subunits is shown in Fig. 9 along with the coordinates of

IHP from Arnone & Perutz (1974). Contouring the map at three times it's r.m.s. deviation (0·23 e $Å^{-3}$ – solid mesh) reveals just one substantial peak that corresponds to the phosphate group which is coordinated to histidines β_1 2 and β_2 143. The lack of any other density at this level suggests that the IHP is disordered and this is confirmed by inspection of a lower-level contour (0·15 e $Å^{-3}$ – dashed mesh) which reveals poorly defined density around the rest of the IHP molecule.

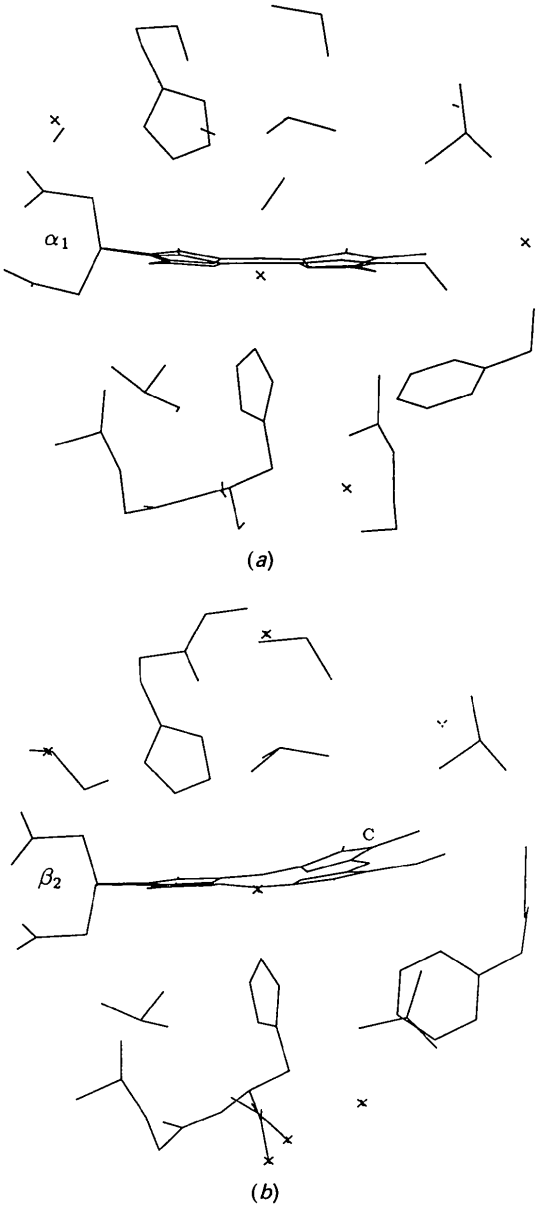

Fig. 8. The α_1 (*a*) and β_2 (*b*) haem groups and their environments, illustrating the distortion of the haem group in the β subunits. Note that a dioxygen ligand is included in the α_1 haem and that pyrrole ring *C* shows the greatest displacement.

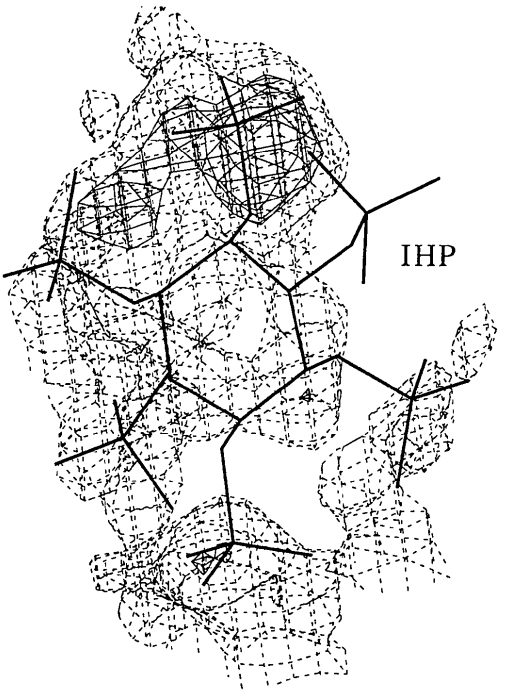

Fig. 9. $|F_o| - |F_c|$ map in the IHP binding site contoured at 0·23 e $Å^{-3}$ (solid mesh) and 0·15 e $Å^{-3}$ (dashed mesh). IHP coordinates from Arnone & Perutz (1974).

Concluding remarks

The structure of human haemoglobin grown from PEG 1000 with IHP has been refined at 1·5 Å resolution. The ligation of this structure differs from that of the previously studied semioxygenated haemoglobin. The partial ligation of the α haems provides the first crystallographic observation of an allosteric inhibitor at work. Further details of the IHP binding site, comparison of this structure with the semioxy and deoxy structures and the effects of IHP on the structure will be reported elsewhere when the detailed analysis is complete. This work has prompted the examination of crystals of human haemoglobin grown from PEG with diphosphoglycerate which is the mammalian allosteric effector.

The authors thank Mrs E. J. Dodson for her assistance with computing, Mr P. H. Holden for his assistance with data processing, and Dr J. R. Helliwell and the Daresbury staff for X-ray facilities. The Institut Laue–Langevin, Grenoble, France, is thanked for financial support for DAW, and the Medical Research Council is thanked for financial support for RCL.

References

AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

ARNDT, U. W. & WONACOTT, A. J. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: Biomedical Press.

ARNONE, A. & PERUTZ, M. F. (1974). *Nature (London)*, **249**, 34–36.
DEREWENDA, Z. S. (1989). *Acta Cryst.* A**45**, 227–234.
DRABKIN, J. (1949). *Arch. Biochem. Biophys.* **21**, 224–236.
FERMI, G., PERUTZ, M. F. & SHAANAN, B. (1984). *J. Mol. Biol.* **175**, 159–174.
FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886–891.
HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
LIDDINGTON, R. C. (1986). DPhil Thesis, Univ. of York, England.
LIDDINGTON, R. C., DEREWENDA, Z., DODSON, G. & HARRIS, D. (1988). *Nature (London)*, **331**, 725–728.
LUISI, B. F. (1986). PhD Thesis, Univ. of Cambridge, England.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
MONOD, J., WYMAN, J. & CHANGEUX, J. P. (1965). *J. Mol. Biol.* **12**, 88–118.
RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. (1963). *J. Mol. Biol.* **7**, 95–99.
SHAANAN, B. (1983). *J. Mol. Biol.* **171**, 31–59.
WINKLER, F. K., SCHUTT, C. E. & HARRISON, S. C. (1979). *Acta Cryst.* A**35**, 901–911.
WONACOTT, A. J. (1980). *MOSCO. A Suite of Programs for On-Line Evaluation of Integrated Intensities on Small-Angle Rotation/Oscillation Photographs.* Unpublished.

Acta Cryst. (1990). B**46**, 418–425

Molecular-Replacement Structure Determination of Two Different Antibody:Antigen Complexes

BY STEVEN SHERIFF,* EDUARDO A. PADLAN, GERSON H. COHEN AND DAVID R. DAVIES

National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

(*Received* 12 *July* 1989; *accepted* 9 *January* 1990)

Abstract

We have used molecular replacement to determine the structures of two antibody:antigen complexes, HyHEL-5 Fab:lysozyme and HyHEL-10 Fab:lysozyme by orienting and locating the C_L:C_H1 domains, Fv and lysozyme. We used the model of McPC603 as our probe for the C_L:C_H1 domains and Fv. In the case of HyHEL-5 Fab:lysozyme, there were two closely related crystal forms and the top peak in the rotation function was correct in five out of six cases (two crystals and three probes). The top peak in the Crowther–Blow translation function was also correct in five out of six cases. We used the program *BRUTE* [Fujinaga & Read (1987). *J. Appl. Cryst.* **20**, 517–521] to put together the three pieces and thereby solve the relative origin problem in space group $P2_1$. In the case of HyHEL-10 Fab:lysozyme the top peak in the rotation function was correct for the Fv and lysozyme, but even with an appropriate model (the C_L:C_H1 domains of HyHEL-5) it was no better than the seventh peak (72% of the top peak) for C_L:C_H1. In the Crowther–Blow translation function, the top peaks were correct on two of three Harker sections for the Fv domains and lysozyme. The Crowther–Blow translation function was unable to locate C_L:C_H1, when using McPC603 C_L:C_H1 as the probe; however, when HyHEL-5 C_L:C_H1 was used, the top peak in all three sections was correct.

* Current address: Squibb Institute for Medical Research, PO Box 4000, Princeton, NJ 08543-4000, USA.

Introduction

Molecular replacement can be used to determine the location of a known structure in an unknown crystal. This use of molecular replacement was first demonstrated at low resolution by the determination of seal myoglobin using the structure of sperm-whale myoglobin (Tollin, 1969) and at intermediate resolution by the determination of carboxypeptidase B using the structure of carboxypeptidase A (Schmid, Lattman & Herriott, 1974; Schmid & Herriott, 1976). Since that time molecular replacement has become a routine technique for determining the structure of closely related macromolecules or the same macromolecule in a different unit cell. Several workers have used models which account for only part of the macromolecule in the crystal, *e.g.* the use of *Streptomyces griseus* protease B in the determination of the SGPB-ovomucoid inhibitor complex (Fujinaga, Read, Sielecki, Ardelt, Laskowski & James, 1982) and the use of trypsin in the determination of a trypsinogen–Kazal-type inhibitor complex (Bolognesi *et al.*, 1982).

More recently, Cygler, Boodhoo, Lee & Anderson (1987) and Cygler & Anderson (1988*a*,*b*) have used the C_L:C_H1 domains and the Fv (V_L:V_H) domains of the McPC603 Fab (Satow, Cohen, Padlan & Davies, 1986) to independently determine the location of these two pieces of structure for an unknown Fab, HED10. Building on this result we have determined the structures of two different antibody:antigen complexes, HyHEL-5 Fab:lysozyme and HyHEL-10 Fab:lysozyme. We determined the orientation and