FRAASE STORM, G. M. & TUINSTRA, F. (1986). *J. Appl. Cryst.* Submitted.

GOEDE, R. DE (1986). To be published.

JOHNSON, C. K. (1965). *ORTEP.* Report ORNL-3794, revised June 1970. Oak Ridge National Laboratory, Tennessee.

STEWART, J. M., KRUGER, G. J., AMMON, H. L., DICKINSON, C. W. & HALL, S. R. (1972). The *XRAY72* system. Tech. Rep. TR-192. Computer Science Center, Univ. of Maryland, College Park, Maryland.

WANG, J. L., BERKOVITCH-YELLIN, Z. & LEISEROWITZ, L. (1985). *Acta Cryst.* B41, 341-348.

*Acta Cryst.* (1986). B42, 497-515

# The Refined Structure of Beef Liver Catalase at 2·5 Å Resolution

BY IGNACIO FITA,* ABELARDO M. SILVA,† MATHUR R. N. MURTHY‡ AND MICHAEL G. ROSSMANN

*Department of Biological Sciences, Purdue University, West Lafayette, Indiana* 47907, *USA*

(*Received* 23 *September* 1985; *accepted* 3 *April* 1986)

## Abstract

The crystal structure of beef liver catalase [Murthy, Reid, Sicignano, Tanaka & Rossmann (1981). *J. Mol. Biol.* **152**, 465-499] has now been refined by a restrained parameter least-squares method [Konnert & Hendrickson (1980). *Acta Cryst.* A**36**, 344-350] with respect to 2·5 Å data. Some extra density was discovered during the refinement process. This was interpreted in terms of a bound NADP molecule [Kirkman & Gaetani (1984). *Proc. Natl Acad. Sci. USA*, **81**, 4343-4348; Fita & Rossmann (1985). *Proc. Natl Acad. Sci. USA*, **82**, 1604-1608]. When the non-crystallographic symmetry was imposed as a constraint, the *R* factor was reduced to 21·2%. However, refinement of the two crystallographic independent subunits gave a final *R* factor of 19·1%. The refined coordinates have been re-analyzed for main-chain and side-chain hydrogen bonding, charge distribution, secondary structural element interactions, subunit contacts and molecular packing. The fractional accessibility and the temperature-factor variation are also discussed. The oligomerization process is considered in terms of the unusual quaternary structure. The organization of the heme channel and its relation to the enzyme's catalytic properties have been discussed elsewhere [Fita & Rossmann (1985). *J. Mol. Biol.* **185**, 21-37].

## Introduction

The biological function of catalase (EC 1.11.1.6) is still unclear, as is also the *in vivo* role of the bound NADP molecule found in some mammalian catalases (Kirkman & Gaetani, 1984; Fita & Rossmann, 1985*a*) and the extra 'flavodoxin-like' domain in fungal catalases (Vainshtein, Melik-Adamyan, Barynin, Vagin & Grebenko, 1981). However, the presence of catalase in most aerobic organisms has been related to the protection of cells from the toxic effects of small peroxides (*cf.* Deisseroth & Dounce, 1970; Schonbaum & Chance, 1976).

The structure of beef liver catalase (BLC) had been determined (Reid, Murthy, Sicignano, Tanaka, Musick & Rossmann, 1981; Murthy, Reid, Sicignano, Tanaka & Rossmann, 1981) and bears a close resemblance to fungal catalase (Vainshtein *et al.*, 1981). In this paper we describe the refinement of the atomic parameters and analyze the molecular structure. Elsewhere (Fita & Rossmann, 1985*b*) we have discussed the heme environment in relation to the catalytic properties of the enzyme.

## Refinement of the structure

### (*a*) *Crystal and molecular data*

The molecular weight of BLC is 235 000. There are four polypeptide chains per molecule, each with one heme group and one NADP moiety (Kirkman & Gaetani, 1984; Fita & Rossmann, 1985*a*). Each subunit has usually 506 amino acids, although some beef liver extracts have been reported with 10 to 15 additional residues (Schroeder, Shelton, Shelton, Robberson, Apell, Fang & Bonaventura, 1982; Schroeder, Shelton, Shelton, Apell, Evans, Bonaventura & Fang, 1982). The crystals are trigonal, have space group $P3_221$ with $a = 142{\cdot}0$ and $c = 103{\cdot}7$ Å and were first reported by Eventoff, Tanaka & Rossmann (1976). The molecule is positioned on a crystallographic diad with two subunits per crystallographic asymmetric unit. For most of the refinement, the coordinates were

* Present address: Departamento de Quimica Macromolecular, ETSIIB, Diagonal 647, 08028 Barcelona-28, Spain.

† Present address: Departamento de Fisica, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CC No. 67, 1900 La Plata, Argentina.

‡ Present address: Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India.

0108-7681/86/050497-19$01.50 

constrained by the previously determined (Murthy *et al.*, 1981) non-crystallographic symmetry relating the two crystallographically independent subunits. Only in the final stages of refinement was this constraint removed to permit entirely independent adjustments of the atomic parameters in the two subunits.

The three mutually perpendicular molecular diads serve to define the right-handed $P,Q,R$ coordinate system. The $Q$ axis is coincident with the crystallographic twofold axis at $z=\frac{5}{6}c$. The $R$ axis is very roughly parallel to $c$, or more precisely there is an angle of $\theta=-13{\cdot}65°$ between $R$ and $c$. The relationship between the molecular and crystallographic axes is given by:

$$\begin{aligned} P &= a\frac{\sqrt{3}}{2}\cos\theta x + 0y - c\sin\theta\left(z-\frac{5}{6}\right) \\ Q &= -\frac{a}{2}x + b(y-y_0) + 0z \\ R &= a\frac{\sqrt{3}}{2}\sin\theta x + 0y + c\cos\theta\left(z-\frac{5}{6}\right), \end{aligned} \tag{1}$$

where $(0, y_0, 0)$ is the position of the molecular center in fractional coordinates. Alternatively, the relationship between the two non-crystallographic asymmetric units can be expressed as:

$$\begin{aligned} x' &= -\cos 2\theta x + 0y + \frac{2c\sin 2\theta}{\sqrt{3}a}z - \left[\frac{5c}{3\sqrt{3}a}\sin 2\theta\right] \\ y' &= \frac{1-\cos 2\theta}{2}x - y + \frac{c}{\sqrt{3}a}\sin 2\theta z \\ &\quad + \left[2y_0 - \frac{5c}{6\sqrt{3}a}\sin 2\theta\right] \\ z' &= \frac{a}{\sqrt{3}c}\sin 2\theta x + 0y + \cos 2\theta z + \left[\frac{5}{6}(1-\cos 2\theta)\right]. \end{aligned} \tag{2}$$

Equations (2) were used to constrain the refinement to the non-crystallographic symmetry. The reference subunit, $S_1$, is defined as having all positive coordinates for its heme Fe position. Subunit $S_2$ is related to $S_1$ by the $R$ axis.

The X-ray diffraction data were collected with oscillation photographs as described by Murthy *et al.* (1981). The final native data extended to 2·5 Å resolution. Although 88% of the data were observed with amplitudes greater than one estimated standard error, yet only 67% of the data was found to be similarly significant in the 2·58–2·50 Å resolution shell. The overall $R$ factor was 10·3%, where $R$ factor $=\sum_h\sum_i|(\bar{I}_h - I_{hi})|/\sum_h\sum_i \bar{I}_h$ ($\bar{I}_h$ is the mean of $i$ observations $I_{hi}$).

### (*b*) *The starting model*

The original multiple-isomorphous-replacement phases were obtained from the native and four heavy-atom-derivative data sets. The map was then subjected to ten cycles of molecular-replacement density averaging. The resultant electron density, averaged over the two crystallographically independent molecular subunits, was interpreted with respect to the amino acid sequence in a Richards optical comparator (Murthy *et al.*, 1981). This model was used to measure $C_\alpha$ coordinates and details of some selected residues. These coordinates were the starting point of the work reported here.

The averaged electron density map, calculated with respect to the molecular $P,Q,R$ system, was then displayed in the MMS-X computer-graphics system (Barry, Bosshard, Ellis & Marshall, 1975) with A. Jones's (1978) *FRODO* program. The initial model building experienced some difficulties in a few regions. In particular, the amino-terminal (1–4) and carboxy-terminal (501–506) residues were not visible in the map. Apart from residues 3 and 4, these residues remained undetermined throughout the refinement.

### (*c*) *Refinement strategy*

The molecular coordinates were transferred to the crystal system and used to construct an electron density map. Each atom was represented as a sum of two Gaussians (Forsyth & Wells, 1959; Agarwal, 1978) modified by individual temperature factors and an artificial overall temperature factor (25 Å$^2$) related to resolution and grid sampling (Ten Eyck, 1977). The electron density generated by any given atom was limited to a radius of 3 Å. No density was placed in the solvent region of the crystal cell. The resultant electron density was back-transformed with a $P3_2$ specific fast Fourier program (Ten Eyck, 1973; E. Dodson, private communication) to give calculated structure factors, $F_c$. These structure factors were placed on a relative scale to the observed amplitudes $F_o$ by deriving independent linear scale factors ($\sum F_oF_c/\sum F_c^2$) within 15 spherical shells of equal volume. Actual scale factors were determined by interpolation between the mean value for each shell. The scaled values were used to compute $(2F_o - F_c)$ and $(F_o - F_c)$ difference maps without any figure of merit weighting. The maps were twofold averaged and skewed (Johnson, 1978) into the molecular $P,Q,R$ system for display in the MMS-X computer-graphics system. The $(2F_o - F_c)$ maps were used for checking and correcting the current model, while the $(F_o - F_c)$ map was used primarily for finding major errors and solvent molecules.

After all necessary atomic adjustments had been made by use of the *FRODO* program, the coordinates were returned from the MMS-X graphics system to the Purdue University Computer system by way of a 9800 baud telephone line. The atomic parameters were then subjected to least-squares refinement. The least-squares program had been written and con-

ceived by Hendrickson & Konnert (1980) and adapted for vector processing on a Cyber 205 computer. This procedure improves the agreement between $|F_o|$ and $|F_c|$ amplitudes while simultaneously maintaining reasonable stereochemical features. The objective is to minimize an observational function by setting up a set of normal equations in terms of the atomic parameter shifts.

The stereochemical restraints applied to the model included interatomic distances and angles, planarity, chiral volume, non-bonded contacts and torsion angles. The general form of the contribution to the residual for a given type of restraint is

$$\sum_i [w(d_T - d_M)/\sigma_T]^2,$$

where the sum is taken over all restraint parameters of a given type. The weight for each term in the sum has the form $w/\sigma_T$, where $\sigma_T$ is the target standard deviation for a particular restraint and $w$ is the weight given to that term in the residual; $d_T$ is the target value for the restraint and $d_M$ is its actual value derived from the model. The values of $w$, $\sigma_T$ and $R$ are given in Table 1 for the final stages of the refinement.

Two main modifications were introduced to the Hendrickson-Konnert refinement package. The first is the use of non-crystallographic symmetry as a constraint and the second is the vectorization of the structure factor and structure factor derivative calculations. In its original version, the program imposed non-crystallographic symmetry as a restraint by minimizing $\sum (x - [C]x')^2$, where $x$ and $x'$ are non-crystallographically related position vectors of atoms and [C] is the known non-crystallographic operator. This means that, while the number of observational equations would have increased, the number of parameters would not have decreased. By imposing the non-crystallographic symmetry as an algebraic constraint, the number of independent atomic parameters is halved (Rossmann, 1976).

The specific configuration of the Purdue University Cyber 205 computer at the time of this work included one vector processor, one scaler processor, $10^6$ 64-bit words of main memory and $3 \times 10^9$ bytes of disk space. The special feature of such a computer is the vector-processing option. This permits the same operation to be performed almost simultaneously on each set of paired elements in two vectors. The operation on the second pair of elements can commence very shortly after the operation on the first pair has started and so forth. Hence, the multiplication (say) of two vectors can be done in only a little extra time than the multiplication of two individual numbers. The calculation of structure factors and their derivatives was, therefore, performed by computing all atomic contributions for a given reflection at once. That is, the $x$, $y$ and $z$ coordinates of all atoms in the unit cell were initially expanded into three vectors. The length of each vector was about 50 000 elements. Similarly, the cosine and sine tables were placed into vectors and accessed by a special instruction which allows the gathering of all required values in the look-up table. All multiplications of scattering factors were also vectorized. In this manner the execution time was reduced by approximately a factor of 10 when compared with the unvectorized version of the program.

Table 1. *Refinement restraints*

| Model: | One-subunit refinement | Two-subunit refinement |
|---|---|---|
| Distances | | |
| $\omega$ | 1·5 | 1·4 |
| No.* | 11531 | 23062 |
| $>4\sigma$† | 17 | 60 |
| $\sigma_1$ (0·030) | 0·027 | 0·029 |
| $\sigma_2$ (0·040) | 0·044 | 0·048 |
| $\sigma_3$ (0·050) | 0·047 | 0·051 |
| Chiral volume (Å$^3$) | | |
| $\omega$ | 1·3 | 1·3 |
| No. | 573 | 1146 |
| $>3\sigma$ | 5 | 18 |
| $\sigma$ (0·150) | 0·180 | 0·182 |
| Planes (Å) | | |
| $\omega$ | 1·0 | 1·0 |
| No. | 723 | 1446 |
| $>2\sigma$ | 8 | 14 |
| $\sigma$ (0·020) | 0·017 | 0·017 |
| Non-bonded contacts (Å) | | |
| $\omega$ | 1·0 | 1·0 |
| $\sigma_1$ (0·500) | 0·223 | 0·232 |
| $\sigma_2$ (0·500) | 0·258 | 0·279 |
| $\sigma_3$ (0·500) | 0·233 | 0·249 |
| Torsion angles (°) | | |
| $\omega$ | 1·3 | 1·3 |
| No. | 1285 | 2570 |
| $\sigma_1$ (3·0) | 5·1 | 6·9 |
| $\sigma_2$ (15·0) | 23·5 | 24·0 |
| $\sigma_3$ (20·0) | 35·0 | 34·5 |
| Thermal factors (Å$^2$) | | |
| $\omega$ | 1·0 | 1·0 |
| $\sigma_1$ (1.0) | 1·34 | 1·13 |
| $\sigma_2$ (1.0) | 1·38 | 1·14 |
| $\sigma_3$ (1·5) | 2·11 | 1·81 |
| $\sigma_4$ (1·5) | 2·23 | 1·84 |
| R.m.s. shifts (Å) | 0·012 | 0·015 |
| $R$(%) | 21·2 | 19·1 |
| No. of reflections | 33255 | 33255 |
| Resolution (Å) | 8·5–2·5 | 8·5–2·5 |

Parenthetic values for the $\sigma$ values show target values.
Distances: $\sigma_1$ is the r.m.s. deviation (Å) for bond lengths, between atoms $i$ and $i+1$, from their idealized values.
$\sigma_2$ is the r.m.s. deviation (Å) for bond-angle distances, between atoms $i$ and $i+2$, from their idealized values.
$\sigma_3$ is the r.m.s. deviation (Å) in planar groups, between atoms $i$ and $i+3$, from their idealized values.
Planes: $\sigma$ is the r.m.s. deviation (Å) from the mean plane.
Chiral volume: $\sigma$ is the r.m.s. deviation (Å$^3$) from the idealized volume inscribed by the bonded atoms.
Non-bonded contacts: $\sigma_1$ is the r.m.s. deviation (Å) of single torsion contacts between atoms $i$ and $i+3$.
$\sigma_2$ is the r.m.s. deviation (Å) of multiple torsion contacts between atoms $i$ and $j > i+3$.
$\sigma_3$ is the r.m.s. deviation (Å) of possible H bonds.
Torsion angles: $\sigma_1$ is the r.m.s. deviation (°) from planar (0°, 180°) conformations.
$\sigma_2$ is the r.m.s. deviation (°) from staggered (±60°, 180°) conformations.
$\sigma_3$ is the r.m.s. deviation (°) from orthonormal (±90°) conformations.
Thermal factors: $\sigma_1$ is the r.m.s. deviation (Å$^2$) for main-chain, bonded (1-2) neighbors.
$\sigma_2$ is the r.m.s. deviation (Å$^2$) for side-chain, bonded (1-2) neighbors.
$\sigma_3$ is the r.m.s. deviation (Å$^2$) for main-chain, (1-3) neighbors.
$\sigma_4$ is the r.m.s. deviation (Å$^2$) for side-chain, (1-3) neighbors.

* The number of distances, planes, *etc.*, which were restrained.
† $>n\sigma$ indicates the number of restraints which remained larger than $n\sigma_T$ (target) at the end of the refinement stage.

Table 2. *Progress of the refinement*

| Cycle number (inclusive) | Atoms | Residues | Water molecules | Parameters refined* | Resolution range (Å) | Comments |
|---|---|---|---|---|---|---|
| 1–4 | 4032 | (5–500) Heme | — | *XYZ* | 10·0–3·0 | Initial model from *FRODO* |
| 5–7 | 4032 | (5–500) Heme | — | *XYZ* | 5·0–2·5 | |
| 8–11 | 4032 | (5–500) Heme | — | *XYZB* | 5·0–2·5 | |
| 12–14 | 4107 | (5–500) Heme | 75 | *XYZB* | 7·0–2·5 | $(2F_o-F_c)\alpha_c$ averaged |
| 15–17 | 4107 | (5–500) Heme | 75 | *XYZB* | 5·0–2·5 | |
| 18–20 | 4101 | (5–500) Heme | 69 | *XYZ* | 8·5–2·5 | |
| 21–29 | 4131 | (2–500) Heme | 75 | *XYZB* | 8·5–2·5 | $(2F_o-F_c)\alpha_c$ unaveraged; Pro 405 as *cis*; 17–23, 289–293 omitted from one subunit |
| 30–34 | 4145 | (2–500) Heme polyA peptide | 69 | *XYZB* | 8·5–2·5 | $(2F_o-F_c)\alpha_c$ and $(F_o-F_c)\alpha_c$ averaged; checked extra density |
| 35,38 | 4097 | (4–500) Heme | 54 | *XYZB* | 8·5–2·5 | |
| 39–42 | 4148 | (3–500) Heme AVHTYA | 52 | *XYZB* | 8·5–2·5 | $(2F_o-F_c)\alpha_c$ unaveraged |
| 43–51 | 4145 | (3–500) Heme | 49 | *XYZB* | 8·5–2·5 | |
| 52–57 | 4150 | (3–500) Heme | 54 | *XYZB* | 8·5–2·5 | $(F_o-F_c)\alpha_c$ averaged |
| 58–69 | 4150 | (3–500) Heme | 54 | *XYZB* | 8·5–2·5 | $(F_o-F_c)\alpha_c$ with and without averaging |
| 70–76 | 8300 | (3–500) Heme AVHTYA 2 subunits | 108 | *XYZB* | 8·5–2·5 | $(F_o-F_c)\alpha_c$ with and without averaging |
| 77–83 | 4149 | (3–500) Heme NADP | 50 | *XYZB* | 8·5–2·5 | $(2F_o-F_c)\alpha_c$ and $(F_o-F_c)\alpha_c$ with and without averaging |
| 84–93 | 8296 | (3–500) Heme NADP 2 subunits | 98 | *XYZB* | 8·5–2·5 | $(2F_o-F_c)\alpha_c$ averaged<br>$(2F_o-F_c)\alpha_c$ without averaging |

* *X*, *Y*, *Z* refer to molecular coordinates, *B* refers to temperature factors.

The Cyber 205 computer was not available when the refinement of BLC was initiated. Hence, the real-space transformation described above was used for structure factor calculations. This would not have been necessary had the Cyber been available. Thus, in the end, the reciprocal-space method of structure factor calculation was used in the least-squares procedure, whereas the real-space transformation was used in the construction of difference maps.

### (*d*) *Refinement progress*

The initial *R* factor for the model that had been fitted to the molecular-replacement map in the MMS-X graphics system by means of the *FRODO* program was 37·9% for data between resolution limits of 10 to 3 Å. The subsequent progress of refinement is detailed in Table 2. Explicit hydrogen bonds were used as restraints during the first 11 cycles while weighting heavily the X-ray contribution to the normal equations. An overall temperature factor was used only in the first seven cycles of refinement. Discontinuities in the overall *R*-factor convergence (Fig. 1) correspond to the incorporation of additional data into the least-squares process and the re-examination of the current model in the MMS-X system. Water molecules were introduced after the 11th cycle and refined with full occupancies. These were placed only where there was good density and no steric overcrowding. Those water molecules which exhibited shifts greater than 2 Å from their initial positions or refined to temperature factors larger than 50 Å$^2$ were removed. No restraints were put on water molecules, but short contacts were not permitted. Proline 405 was changed to a *cis* conformation after the 20th cycle. Residues 17–23 and 289–293 were omitted for one subunit only in the difference map calculations after cycle 20 in order to check on their conformation since they previously had poor density. The weight of the stereochemical constraints was

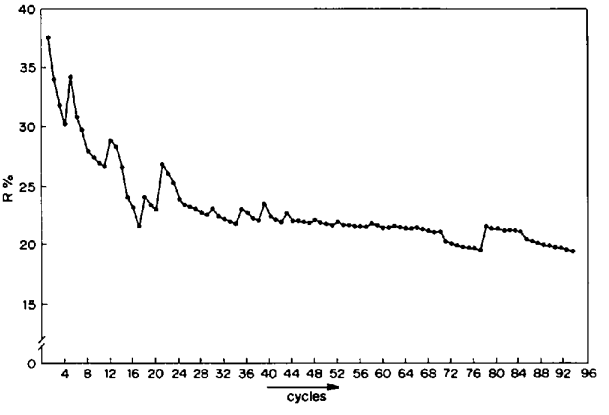



Fig. 1. *R*-factor variation during the process of refinement.

Table 3. *Alignment of carboxy-terminal peptides*

The generally accepted one-letter amino acid code used here is: A = Ala, C = Cys, D = Asp, E = Glu, F = Phe, G = Gly, H = His, I = Ile, K = Lys, L = Leu, M = Met, N = Asn, P = Pro, Q = Gln, R = Arg, S = Ser, T = Thr, V = Val, W = Trp, Y = Tyr.

```
                  500        505        510        515
BLC (Schroeder)   -N-E-E-K-P-K-N(V,E,H,G,S,H) (V-H-T-Y)
                                 <-- Th10 -->  <-Th11->

BLC (X-ray)       -N <------ not visible ------> H- V-H-T-Y-A

BEC (Schroeder)   -N-E-E-K-P-K-N(V-E-H-G-S-H) (V-H-T-Y,A)

HEC (Schroeder)   -N-A-E-K-P-K-N-A
                                 .
                                  (F-V-E-S-G-S-H) (I,H,T)      + (V-A-V)

                  492        495        500        505
PVC (Vainshtein,  -A - A-A-A-A-A-A-A-A-A-A-A- A-A-A-A-A-
  X-ray)
```

Beef liver catalase (BLC) crystals show additional electron density. The latter was initially interpreted as the above alignment of peptides Th 10 and Th 11 found in occasional preparations of BLC (Schroeder, Shelton, Shelton, Robberson *et al.*, 1982). Similar peptides are invariably found in beef erythrocyte catalase (BEC) which shows no other differences to BLC. Similar peptides are also found in human erythrocyte catalase (HEC) although the above alignment shows a conflict of residues in positions 506 and 507. The carboxy-terminal extension to the flavodoxin-like domain in *Penicillium vitale* catalase (PVC) showed spatial alignment of residues 512–517 (BLC) with 503–508 (PVC). Hence, the initial interpretation of the extra density in BLC was in terms of the hexapeptide H-V-H-T-Y-A. However, after the refinement had already been terminated, this density was reinterpreted in terms of the NADP recently reported (Kirkman & Gaetani, 1984) to be associated with BLC.

increased after cycle 29 so that the contribution per parameter of the X-ray data was roughly equal to the contribution per constraint in the overall sum that was to be minimized.

The constraint between the two independent crystallographic subunits was removed in the final cycles (70–76) giving a further significant decrease in $R$ factor. The refinement of residues 499 and 500 in subunit 2 was then found to be unstable. Hence, atoms $C_\gamma$ and the carbonyl O of Tyr 499 as well as $C_\gamma$ of Asn 500 were fixed in subunit 2. This cured the problem

A significant amount of uninterpreted electron density was assumed to be a carboxy-terminal extension after cycle 52. The density corresponded to some of the amino acids of peptides Th 10 and Th 11 found by Schroeder, Shelton, Shelton, Robberson *et al.* (1982) in occasional preparations and assigned by them to the carboxy-terminal extension of bovine and human erythrocyte catalase (Table 3). Furthermore, the extra density in BLC was very close to the superimposed polypeptide chain of *Penicillium vitale* catalase (PVC) corresponding to the hinge linking the final flavodoxin-like domain (absent in beef liver catalase). These additional residues 512–517 were included in refinements after cycle 38. The sequence was determined by homology with human and erythrocyte catalase and by inspection of the X-ray map. After the refinement was essentially ter-

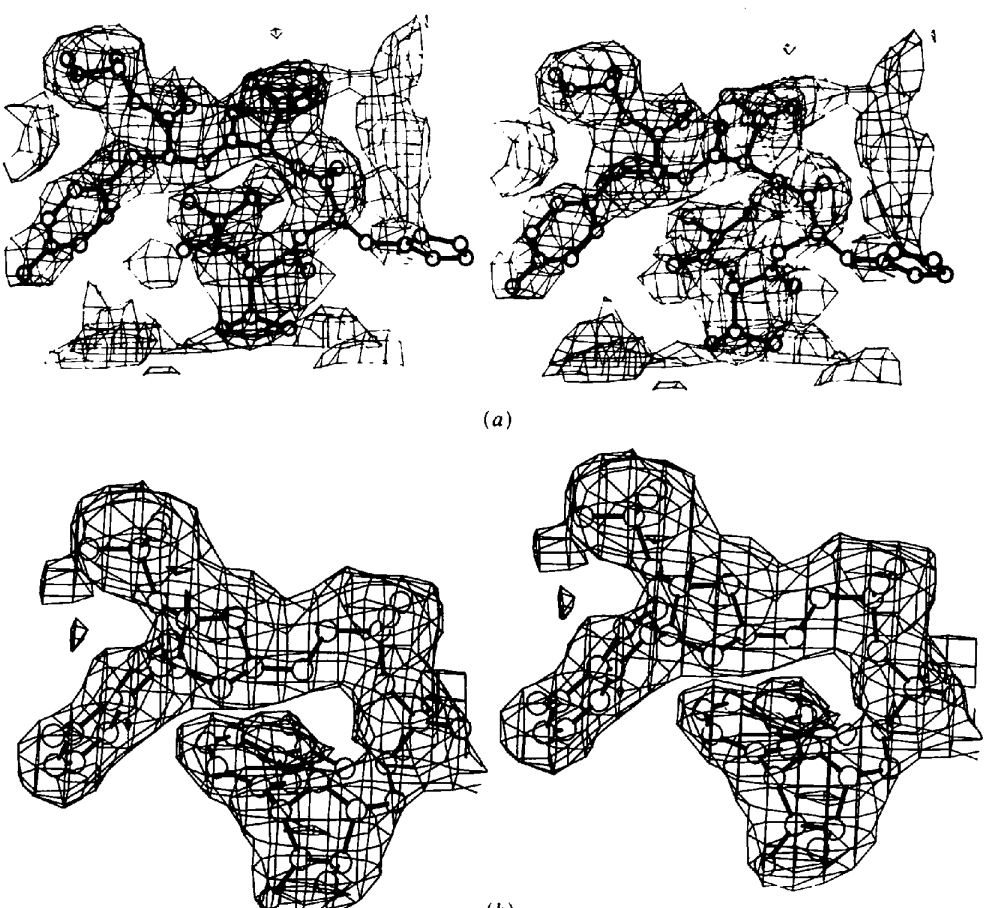


Fig. 2. Fit of erroneous hexapeptide (*a*) to the extra density compared to the correct fit of NADP (*b*).

minated, it was reported that BLC contained 4 molecules of tightly bound NADP (Kirkman & Gaetani, 1984). Reinspection of the extra density clearly showed a far better fit of NADP to the density, even after refinement with respect to the hexapeptide. The correspondence of fit can be described as:

hexapeptide: H V H T Y A

NADP: N R P P R A P

where the imidazole of the second histidine occupied only very low and poor density (Fig. 2). The NADP structure was refined in cycles 77 to 87 using the non-crystallographic constraint and starting with the averaged coordinates found at the end of cycle 69.

Final plots of the $R$ factor expressed as a function of the resolution for the final models are shown in Fig. 3. A plot of the $C_\alpha$ atoms in one subunit is given in Fig. 4 which is compared with a diagrammatic view of the molecule. Final coordinates of both subunits relative to the molecular $P,Q,R$ axes have been deposited with the Brookhaven Protein Data Bank as have also the structure amplitudes.*

## Convergence and accuracy

Although it is difficult to assess that convergence has been reached in the refinement process, the r.m.s. shifts of atoms at the end of cycle 76 were only 0·018 Å and the r.m.s. thermal adjustments were 0·37 Å$^2$. Such relatively small values do not assure that all parts of the structure have the correct conformation, but only that the present structure has been refined to a minimum, given the current weight of the conformational restraints and imposition of non-crystallographic constraints.

No easy method is available for determining the diagonal terms of the inverted matrix representing the normal equations of the least-squares procedure. Hence, it is not easily possible to determine a direct estimate of the errors. However, a maximum limit on the errors can be assessed by observing the difference between the independently refined subunits (Table 4). The larger positional differences are mostly associated with surface residues that encounter different external environments (Table 5). Thus, an outside error limit of 0·21 Å for main-chain and 0·37 Å for

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (References 7CAT, 8CAT and R1CATSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centers at Melbourne or Osaka. The data have also been deposited with the British Library Lending Division as Supplementary Publication No. SUP 37017 (5 microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

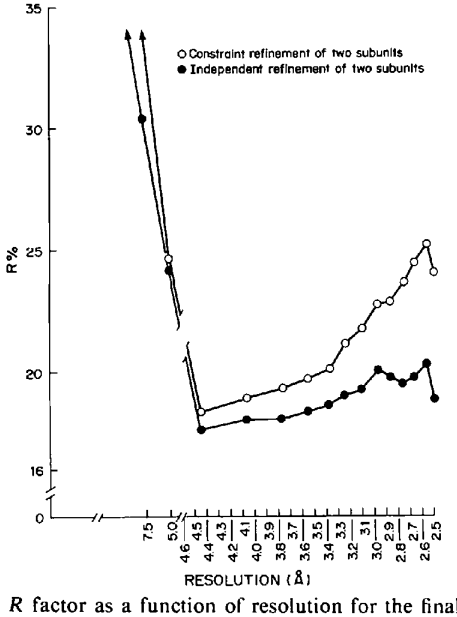



Fig. 3. $R$ factor as a function of resolution for the final models.

Table 4. *Variation between atomic parameters of subunits* 1 *and* 2

| | Distance (Å) | Temperature factor (Å$^2$) |
|---|---|---|
| Mean (r.m.s.) | | |
| All atoms | 0·31 | 5·0 |
| Main chain only | 0·21 | 4·0 |
| Side chain only | 0·37 | 5·3 |
| Maximum | | |
| Position | | |
| Side chain: Glu 453 | 1·20 | — |
| Main chain: Arg 421 | 0·46 | — |
| Temperature | | |
| Side chain: Glu 172 | — | 14·1 |
| Main chain: Arg 319 | — | 9·4 |

Table 5. *R.m.s. deviation of atoms in residues involved in molecular contacts, corresponding to different environments in the two crystallographically independent molecules*

| | Main chain | | Side chain | |
|---|---|---|---|---|
| | Position (Å) | Temperature factor (Å$^2$) | Position (Å) | Temperature factor (Å$^2$) |
| Pro 45 | 0·18 | 2·3 | 0·38* | 4·3 |
| Arg 46 | 0·17 | 1·3 | 0·32 | 2·1 |
| Glu 85 | 0·28* | 5·1* | 0·40* | 9·0* |
| Val 86 | 0·22* | 1·1 | 0·27 | 1·3 |
| Thr 87 | 0·23* | 1·1 | 0·37 | 2·7 |
| His 88 | 0·37* | 6·3* | 0·23 | 6·6* |
| His 101 | 0·18 | 3·8 | 0·35 | 1·6 |
| Ile 102 | 0·23* | 5·0* | 0·41* | 4·0 |
| Gly 103 | 0·20 | 4·6 | — | — |
| Arg 105 | 0·15 | 9·0* | 0·29 | 11·2* |
| Asp 238 | 0·33* | 2·8 | 0·33 | 6·6* |
| Tyr 273 | 0·34* | 2·6 | 0·34 | 3·0 |
| Ser 275 | 0·11 | 2·2 | 0·22 | 1·3 |
| Thr 277 | 0·12 | 8·0* | 0·13 | 10·1* |
| Tyr 279 | 0·14 | 7·5* | 0·24 | 10·2* |
| Glu 289 | 0·17 | 1·9 | 0·29 | 3·0 |
| Ile 290 | 0·17 | 0·56 | 0·48* | 1·1 |
| Pro 292 | 0·25* | 0·76 | 0·43* | 0·4 |
| Lys 314 | 0·20 | 7·6* | 0·45* | 11·5* |
| Arg 379 | 0·15 | 6·1* | 0·24 | 4·9 |
| Arg 421 | 0·46* | 1·1 | 0·71* | 1·4 |
| His 423 | 0·26* | 1·5 | 0·53* | 2·7 |
| Asp 427 | 0·15 | 4·2* | 0·29 | 4·5 |
| Gln 429 | 0·13 | 2·8 | 0·55* | 3·0 |
| Phe 431 | 0·12 | 2·6 | 0·38* | 6·5* |
| Asn 432 | 0·12 | 4·5* | 0·53* | 4·6 |

* Indicates values larger than the global average (Table 4).

side-chain atoms (Table 4) would seem to be very reasonable estimates.

The final $(2F_o - F_c)$ averaged map is compared with the starting map [multiple isomorphous replacement (MIR) improved by cyclic molecular replacement averaging] in Fig. 5. The poorest parts in the final electron density map are: (i) the amino-terminal (1–2) and carboxy-terminal (501–506) parts of the molecule where some residues could not be found even in the final map; (ii) the residues Ala 19-Ala 20-Glu 21-Lys 22 where only the tracing of the main chain was visible in the final map; and (iii) Leu 450-Asn 451-Glu 452-Glu 453 where the main chain is quite well defined but most of the side chains are completely unobservable. These three regions correspond to the highest temperature factors in the molecule (see below).

Table 6. *Hamilton significance test*

| | One-subunit refinement | Two-subunit refinement |
|---|---|---|
| No. of reflections | 33255 | 33255 |
| No. of parameters | 16597 | 33185 |
| No. of restraints | 29551 | 59169 |
| No. of data = No. of reflections + No. of restraints | 62806 | 92424 |
| No. of degrees of freedom = No. of data − No. of parameters | 46209 | 59239 |
| $R$ | 21·2 | 19·1 |
| $R'$ [$= (\sum w\Delta F^2/\sum wF_o^2)^{1/2} \times 10^2$] | 24·3 | 20·9 |

Therefore, $R'$(one-subunit)/$R'$(two-subunit) = 1·162. Hamilton's test predicts a value of 1·02 for this ratio at the 5% significance level. Since the value of $R'$(one-subunit)/$R'$(two-subunit) > 1·02, the two-subunit refinement should be a significant improvement. Note, however, that the $R$ factor is dependent only on X-ray data, whereas the number of degrees of freedom was augmented by the stereochemical restraints.

### Differences between the two subunits

Independent refinement of the two subunits permits:

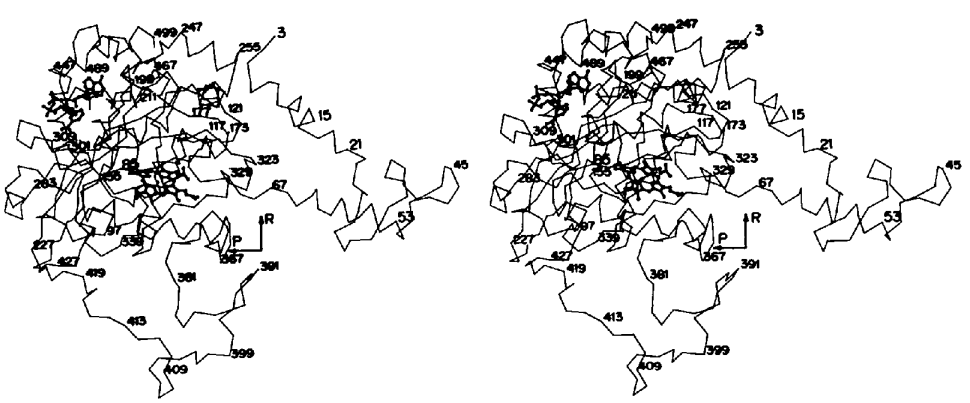
(1) redetermination of the position and orientation of the molecule,

(2) assessment of the accuracy of the molecular 222 symmetry, and

(3) determination of the local influence of the different crystallographic environment on the two independent subunits.

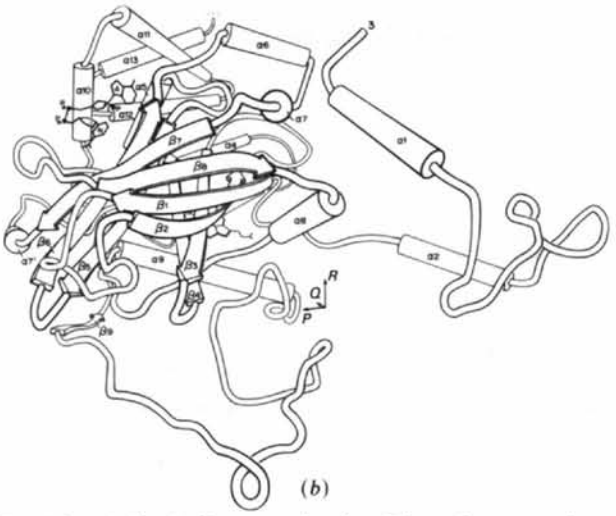
However, refinement of the two subunits requires doubling the number of refinable parameters without increase of the number of observed amplitudes, although the number of observational equations restraining the molecular stereochemistry is also doubled. Hamilton (1964, 1965) suggested a test to determine whether the increase of parameters made a significant improvement in the agreement of the observational equations or was due entirely to the increase in the number of parameters. Application of the test (Table 6) shows that a significant improvement occurs in the two-subunit refinement if it is assumed that the number of reflections (information) is to be augmented by the number of restraints. However, the test was not originally conceived for the combination of X-ray and stereochemical data. As Hamilton's test is not truly applicable in the case of a stereochemically restrained least-squares procedure, an alternative method has been applied which shows that the differences between the subunits may be significant and physically meaningful. The largest differences between the non-crystallographically related subunits may not be due to error in the determination of atomic positions, but be real differences due to differing molecular environments (Table 5).

All the $C_\alpha$ atoms were superimposed by a least-squares process (Rao & Rossmann, 1973; Rossmann & Argos, 1975) with varying constraints (Table 7). In (1) of Table 7, it was assumed that the $R$ axis was a true diad and intersected the molecular $Q$ axis coincident with the crystallographic twofold axis, while solution (2) referred to the free superposition of the two $R$-related subunits. The results show that the size of $\Delta$, $\kappa$ and $t$ (Table 7) were very close to 0·0 Å, 180° and 0·0 Å, respectively, making it reasonably certain




(*a*)

Fig. 4. Final $C_\alpha$ atoms in one subunit. (*a*) Stereoview.



(*b*)

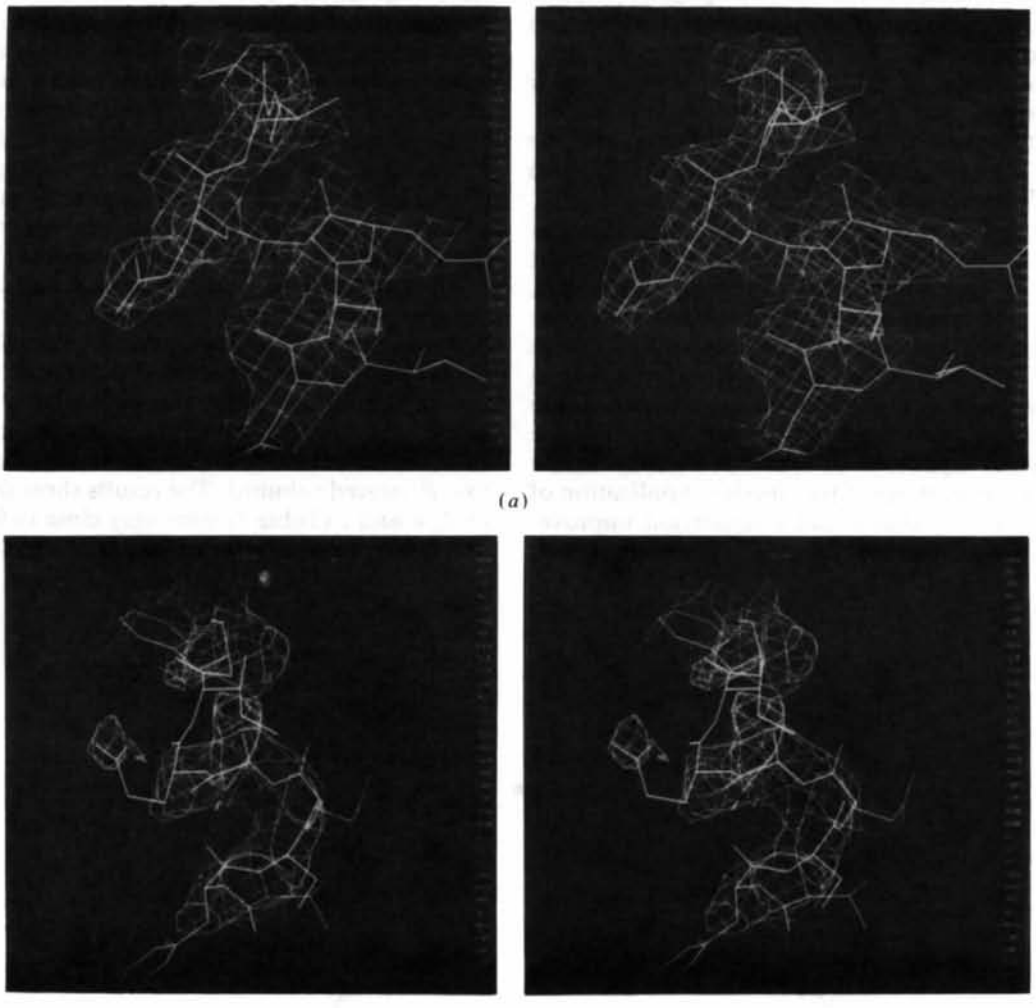
Fig. 4 (*cont.*) (*b*) A diagrammatic view. The various secondary structural elements $\alpha_1$, $\alpha_2$, ..., $\alpha_{13}$ and $\beta_1$, $\beta_2$, ..., $\beta_9$ are also identified.

Table 7. *Analysis of the non-crystallographic symmetry operator along the R axis*

| Parameter | Murthy *et al.* (1981) | Present results (1) | Present results (2) |
|---|---|---|---|
| $y_0$ (Å) | 92·59 | 92·57 | 92·58 |
| $\theta$ (°) | −13·65 | −13·64 | −13·64 |
| $\Delta$ (Å) | 0·000 | 0·000 | 0·004 |
| $\kappa$ (°) | 180·00 | 180·00 | 180·00 |
| $t$ (Å) | 0·00 | 0·00 | 0·015 |

$y_0$ is the translation of the molecular center along the $Q$ axis away from the $3_2$ axis.
$\theta$ is the angle between the molecular $R$ axis and crystallographic $c$-axis directions.
$\Delta$ is the perpendicular distance of the $R$ axis from the crystallographic diad.
$\kappa$ is the angle of rotation between the two subunits about $R$.
$t$ is the translation between two $R$-axis-related subunits parallel to $R$.
Solution (1) assumes exact molecular 222 symmetry.
Solution (2) permits free superposition of the $R$-related subunits.

that the molecule had exact 222 symmetry. The differences between the final values of $y_0$ and $\theta$ between those used in cycles 0–69 and those found subsequently were 0·02 Å and 0·01°, too small to be significant.



(*a*)

(*b*)

Fig. 5. Comparison of (*a*) the final $(2F_o - F_c)$ electron density map and (*b*) the starting MIR (improved by MR cycling) map in the region of the $\alpha_7'$ helix.

### Analysis of the refined molecule

#### (*a*) *Conformation of main-chain angles*

The main-chain torsional angles, $\varphi$, $\psi$ and $\omega$, are good indicators of the quality of the model. In their analysis, we have used the results of the averaged subunit refinement. Glycyl and non-glycyl residues were treated as two classes because of the greater freedom of the former. The Ramachandran plots for each class are shown in Fig. 6 which also show superimposed the conformational-energy map (Brant & Schimmel, 1967).

Fifteen glycyl residues, representing 44% of all glycines, as opposed to only twelve other residues (3% of all such residues) have positive $\varphi$ angles. Among the non-glycyl residues in the $0 < \varphi < 180°$ range, there is the catalytically essential His 74 which is stabilized both by the interaction of its carbonyl oxygen with guanidinium of Arg 111 and by a hydrogen bond between its amino group and the carbonyl oxygen of Val 115. However, with the exception of Ser 216, none of the residues depart significantly from the minima shown on the conformational-energy map (Fig. 6). Ser 216 has similarly unusual conformational angles in PVC (B. K. Vainshtein & W. R. Melik-Adamyan, private communication), and plays a part in closing the $\beta$-barrel between strands $\beta_4$ and $\beta_5$. There is a cluster of four residues (388, 391, 396 and 397) which also have positive $\varphi$ values. These form part of the $Q$-axis-generated subunit contacts and have an unusually high sulfur content, the sequence being Met 391-Cys 392-Met 393-Met 394.

The conformational angles $\omega$ were restrained toward the *trans* ($\omega = 180°$) position, except for Pro 405 which required a *cis* conformation. Table 8 lists those residues that differ by more than 10° from 180°. The largest discrepancy was 18° for the essential Tyr 357. Indeed, it is of some interest that many of the essential residues or those within the heme pocket (Table 8) have unusual conformational angles.

#### (*b*) *Secondary structure*

Although the general folding of BLC presented here (Fig. 4) does not differ from that presented previously (Murthy *et al.*, 1981), there are many changes in detail. Particularly large changes relate to the main-chain hydrogen-bonding arrangements of $\beta_5$ and $\beta_6$. Hydrogen bonds were assumed (Fig. 7) whenever the donor–acceptor distance was less than 3·5 Å and the precursor-donor–acceptor angle was between 110 and 180° (Baker & Hubbard, 1984). The main-chain hydrogen-bonding interactions between subunits are also shown in Fig. 7.

In BLC there are 14 helices ranging in size from one ($\alpha_7$, and $\alpha_8$) to $4\frac{1}{2}$ turns ($\alpha_{11}$, $\alpha_{12}$ and $\alpha_{13}$). Helix $\alpha_{7'}$ has longer than average hydrogen bonds (ranging from 2·7 to 3·3 Å) and had not been previously characterized. Seven helices ($\alpha_1$, $\alpha_2$, $\alpha_3$, $\alpha_5$, $\alpha_6$, $\alpha_{11}$, $\alpha_{12}$) are initiated or terminated by fragments of a $3_{10}$ helix. Helix $\alpha_8$ resembles more a $\pi$- than an $\alpha$-helix (see hydrogen bonding in Fig. 7), a situation which also occurs in PVC (Vainshtein, Melik-Adamyan, Barynin, Vagin, Grebenko, Borisov, Bartels, Fita & Rossmann, 1986).

The central feature of the catalase structure is an eight-stranded anti-parallel $\beta$-barrel. The first four strands ($\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$) are consecutive with each other and these have a fold similar to the second set of four strands ($\beta_5$, $\beta_6$, $\beta_7$, $\beta_8$). A similar $\beta$-barrel

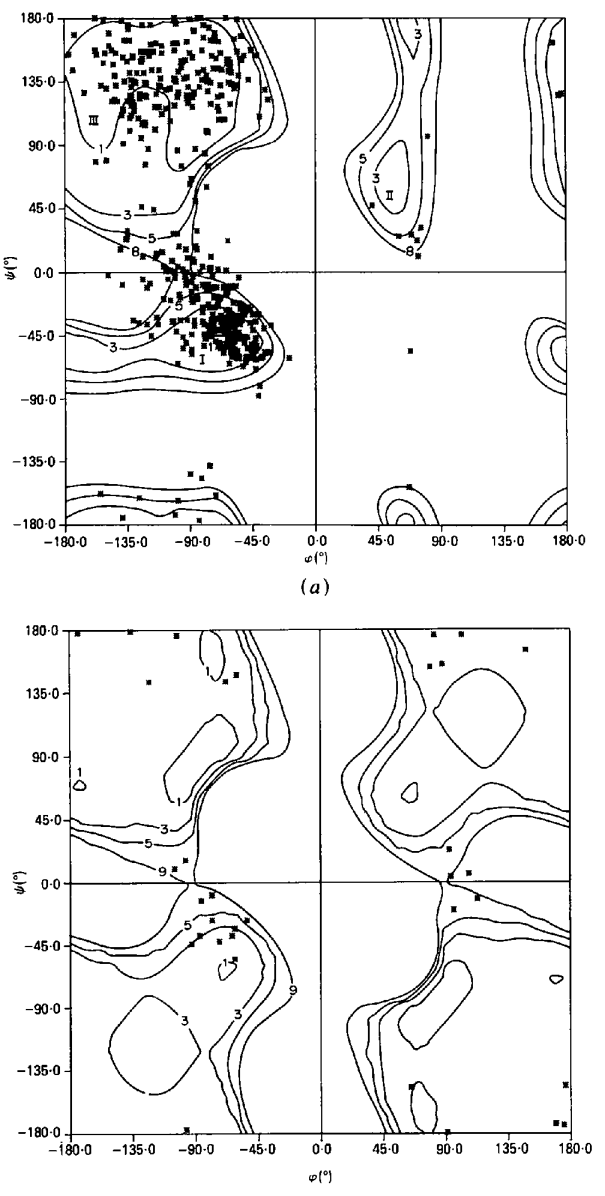
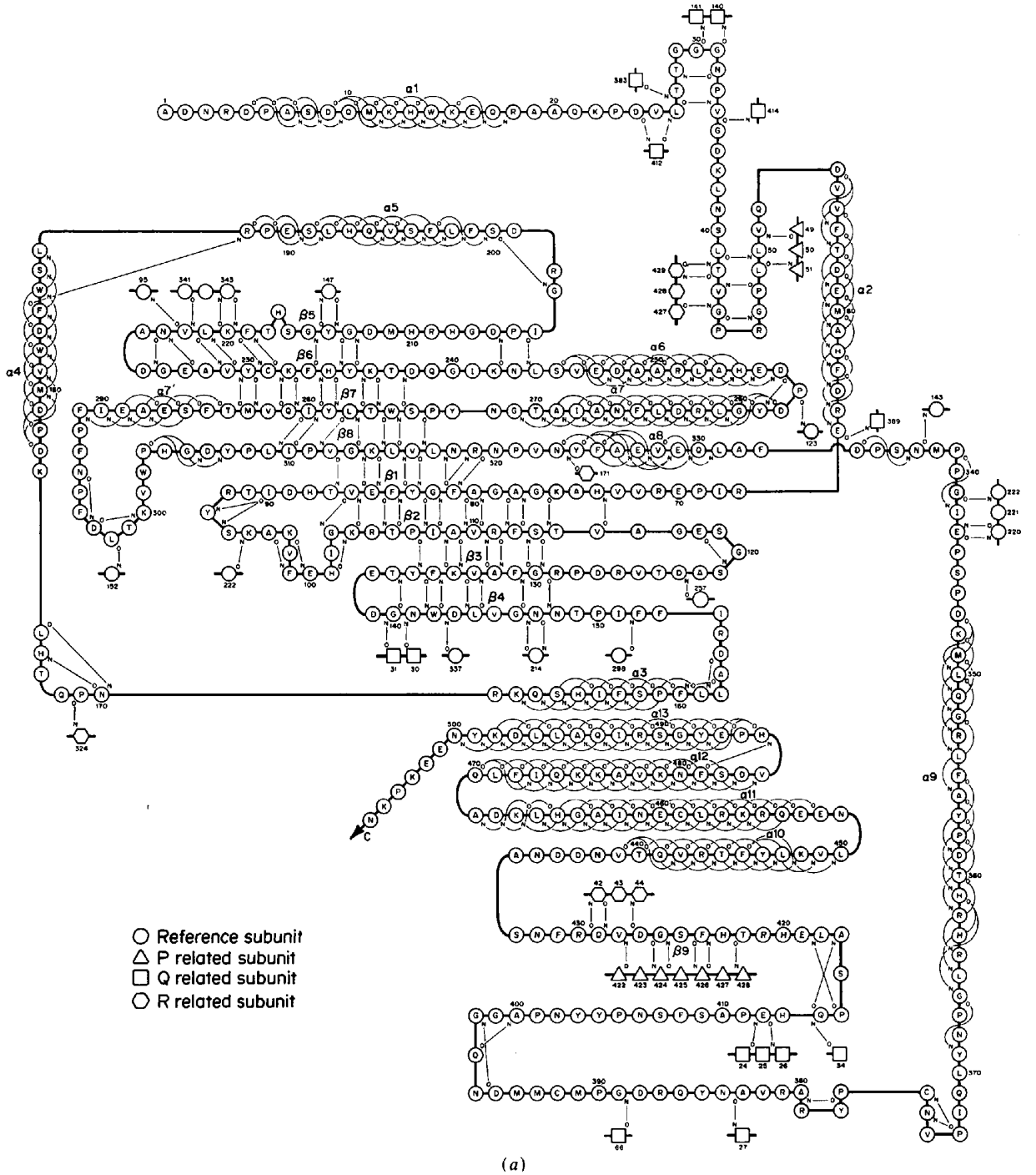



Fig. 6. Ramachandran plots of the main-chain conformational angles superimposed on energy maps for (*a*) poly L-alanine and (*b*) for poly Gly.

topology has been observed in retinol binding protein (Newcomer, Jones, Åqvist, Sundelin, Eriksson, Rask & Peterson, 1984). The continuity of the hydrogen bonding between strands within the $\beta$-barrel is interrupted only between strands $\beta_4$ and $\beta_5$ where there are only two well defined hydrogen bonds involving the essential Asn 147. Strand $\beta_5$ itself is irregular due to the insertion of residues 216–218 which are not involved in the $\beta$-sheet. This permits the amino-terminal part of $\beta_5$ to bind simultaneously to its neighboring strands $\beta_4$ and $\beta_6$, thus closing the $\beta$-barrel. The first half of the $\beta$-barrel contains the essential residues in the distal side of the heme pocket (His 74, Asn 147, Phe 152, Phe 160), whereas the

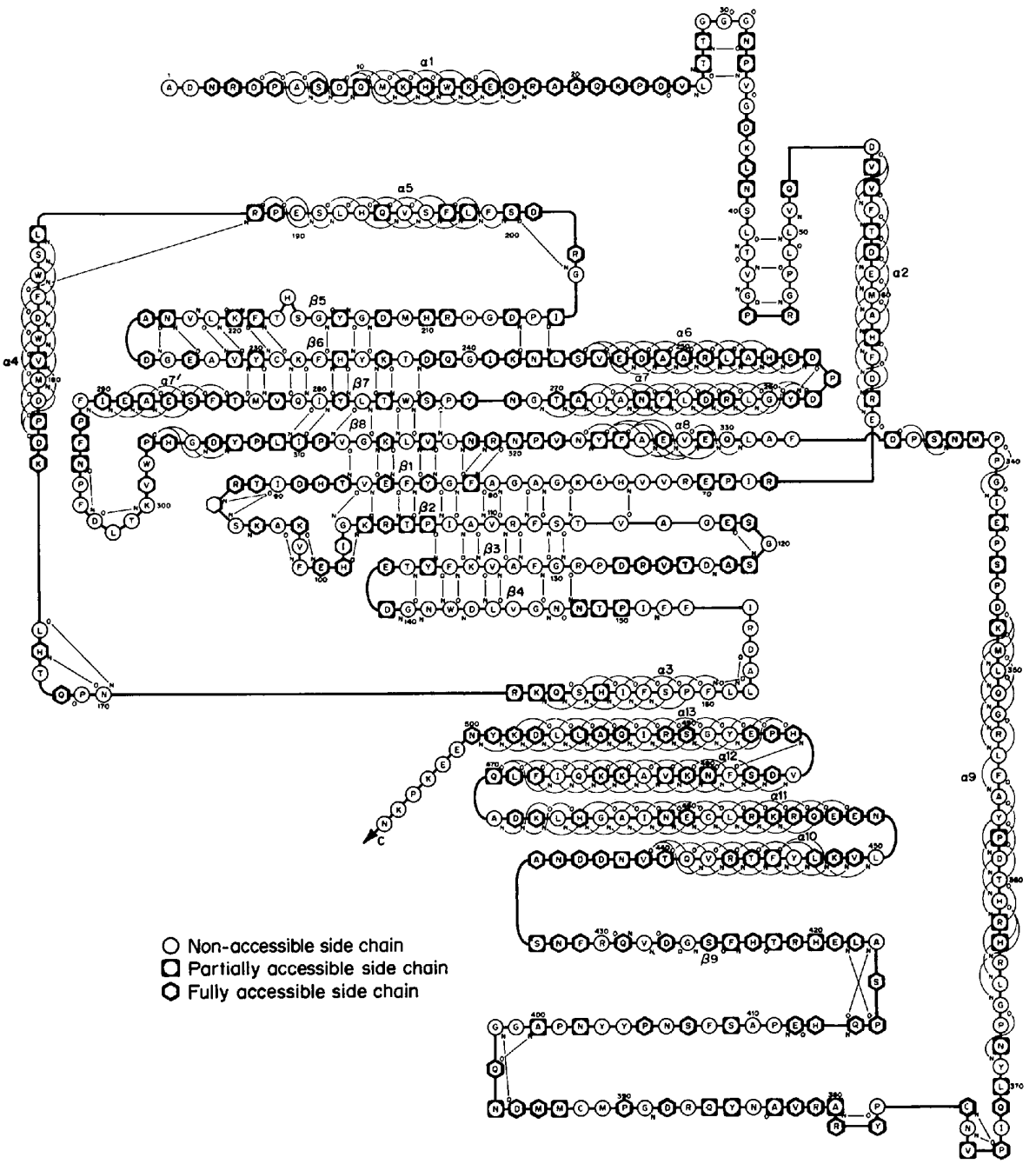



(*a*)

Fig. 7. (*a*) Main-chain hydrogen-bonding network for one BLC subunit. Nomenclature of the secondary structural elements is also shown. The main-chain hydrogen bonding between subunits is indicated.

second half of the $\beta$-barrel functions primarily to bind the NADP moiety (Ser 200, Arg 202, Asp 212, Lys 236, His 304, Val 301, Trp 302, Tyr 214, His 234). The amino-end extension of the essential helix (containing the heme liganding residue Tyr 357) is anchored between $\beta_4$ and $\beta_5$, on either half of the $\beta$-barrel, and forms three hydrogen bonds with $\beta_5$. The essential helix, which provides most of the residues for the heme's proximal side, consists of two segments on either side of the essential Tyr 357.

The $\beta$-barrel contains four $\beta$-bulges (Richardson, Getzoff & Richardson, 1978). Two of these are 'classic' (residues 278, 312, 313; and 81, 318, 319) with one having a forked hydrogen bond in the *X* position, and the other with a glycine in position 2. The other two are 'G1' bends with glycine in position




(*b*)

Fig. 7 (*cont.*) (*b*) Accessibility of side chains. Side chains which have more than 15% of their surface area accessible to water in the tetramer are marked ○ and those with 2-15% accessible area are marked □.

Table 8. *Residues with the peptide bond departing from planarity by more than* 10°

| Residue | $\varphi$ (°) | $\psi$ (°) | | $\omega$ (°) | Comment |
|---|---|---|---|---|---|
| Lys 22 | 81 | 96 | < | −167<br>−177 | Poor density |
| Phe 112 | −101 | 157 | < | 169<br>−177 | In the vicinity of the heme pocket |
| Thr 114 | −99 | −163 | < | 169<br>−179 | |
| Ala 116 | −103 | −35 | < | −168<br>178 | |
| Asp 183 | −52 | −67 | < | 168<br>179 | |
| Phe 197 | −59 | −44 | < | 169<br>−171 | |
| Asn 294 | −69 | 108 | < | 169<br>176 | |
| Ala 326 | −73 | −25 | < | −169<br>180 | |
| Glu 343 | −174 | 143 | < | −166<br>−179 | |
| Arg 353 | −57 | −36 | < | 166<br>173 | Essential residue |
| Tyr 357 | −57 | −34 | < | −162<br>−179 | Essential residue |
| Tyr 378 | −81 | −25 | < | −166<br>175 | |
| Asp 388 | 67 | −153 | < | 165<br>180 | |
| Met 394 | −72 | −159 | < | −169<br>171 | |
| Gly 398 | 65 | −147 | < | 165<br>−178 | |
| Phe 408 | −136 | 23 | < | −167<br>171 | |
| His 423 | −121 | 133 | < | −168<br>178 | |
| Asn 500 | −54 | 33 | ~ | 168 | Last fitted residue |

1 [residues 223, 226(G), 227; and 86, 103(G), 104]. While one of these bends has the usual associated type II tight turn, the other has a type II turn connecting the two $\beta$-strands involved in the $\beta$-bulge formation.

There is an unusual 6-stranded anti-parallel structure formed between subunits involving the wrapping domain and amino-terminal arm. This is shown as a hydrogen-bonding diagram in Fig. 8. The amino-terminal arm of the standard 'red' subunit itself has been slipped through a loop, composed of residues 380–420 of the *Q*-related 'yellow' subunit.

### (*c*) *Temperature-factor variation and solvent accessibility*

As in other proteins, catalase shows reasonable correlation between atomic temperature factors and molecular surface area. In Fig. 9 are shown graphs relating the average main-chain and side-chain temperature factors per residue with surface accessibility. The latter was determined with the Lee & Richards (1971) algorithm. The exposed fraction of the total surface area (Fig. 7*b*) was normalized with respect to the accessible surface area of each residue *X* as found in a Gly-*X*-Gly tripeptide in the same conformation as in the native protein (Shrake & Rupley, 1973). Although the greater surface flexibility (expressed as a 'temperature factor') in proteins has long been recognized, it has only occasionally been placed on a quantitative basis (Sheriff, Hendrickson, Stenkamp, Sieker & Jensen, 1985).

The variation of mean temperature factor per residue with position in the polypeptide chain is shown in Fig. 10 for both main- and side-chain atoms. While the main-chain temperature factors vary relatively smoothly, there are far greater differences in mean side-chain temperature factors between successive residues. The highest temperature factors (58·5 Å$^2$ for the main chain) occur for residues Ala 19-Ala 20-Gln 21 at the hinge between the amino-terminal arm helix $\alpha_1$ and the loop through which the arm is inserted in the *Q*-axis-related subunit. Possibly the arm requires this flexibility in order to form the 'knot'. Other regions of particularly high temperature factors are at the amino-terminal (49·6 Å$^2$ main chain) and carboxy-terminal (57·9 Å$^2$ main chain) regions where the first two and last six residues are entirely disordered. Other regions of high temperature factors are near the beginning of helix

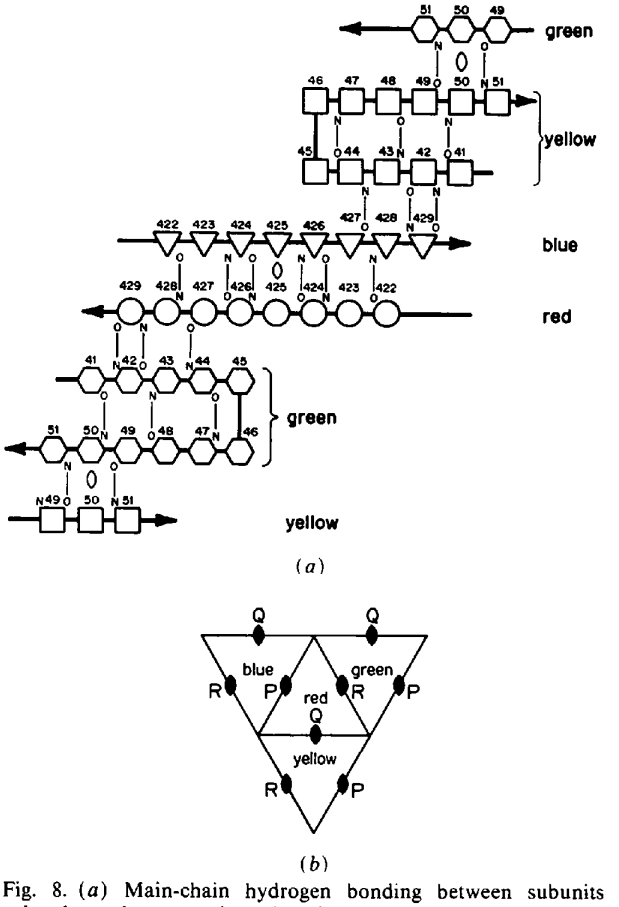



Fig. 8. (*a*) Main-chain hydrogen bonding between subunits involves the wrapping domain and amino-terminal arm. (*b*) Color code for naming of subunits in the tetramer.

$\alpha_{11}$ ($\simeq$49·0 Å$^2$) and in the wrapping domain Met 391-Met 394 ($\simeq$37.6 Å$^2$). Both these fragments are particularly accessible to water.

The most rigid parts of the molecule (60-75, 113-117, 159-163, 324-326, 350-368) are found in the heme pocket ($8 < B < 10$ Å$^2$ for main-chain atoms). The mean temperature factor of the heme groups themselves is only 17 Å$^2$ [see Fita & Rossmann (1985*b*) for a discussion of the variation of the temperature factors of the heme atoms]. Indeed, a number of other enzymes have their most rigid portions near their active sites [*e.g.* papain (Kamphuis, Kalk, Swarte & Drenth, 1984) and penicillopepsin (James & Sielecki, 1983; James, Hsu, Hofmann & Sielecki, 1981)]. In contrast, the NADP binding pocket is far more flexible and the NADP molecule itself has a mean temperature factor of 50 Å$^2$ (Table 9). This is particularly surprising in light of its strong affinity to the catalase molecule (binding constant $<10^{-8}$ *M*; Kirkman & Gaetani, 1984). The O2′ phosphate group has a particularly high temperature factor (57·1 Å$^2$).

Residues have been divided into three groups according to their percentage normalized surface accessibility: those with more than 15% of their area

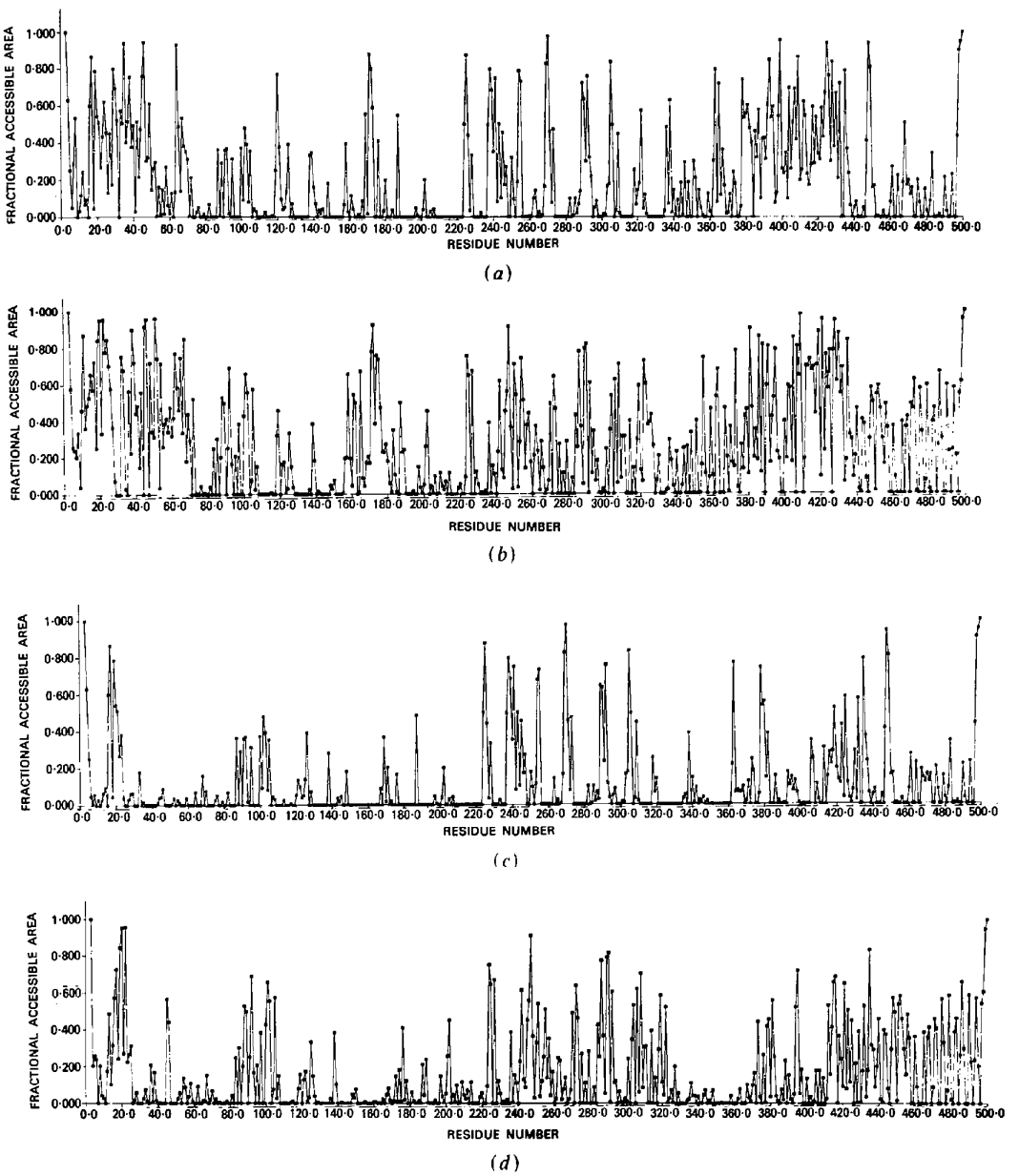



Fig. 9. Fractional accessibility surface areas for the main and side chains, respectively, of an isolated subunit (*a*, *b*) and the tetrameric catalase (*c*, *d*). Comparison with Fig. 10 shows the better correlation of the *B*-value distribution with the accessible surface in the tetramer.

Table 9. *Temperature factors for the bound NADP molecule*

| Moiety | Mean $B$ (Å$^2$) |
|---|---|
| Nicotinamide | 48 |
| Nicotinamide ribose | 47 |
| Pyrophosphate | 50 |
| Adenine ribose | 50 |
| O2′ phosphate | 56 |
| Adenine | 51 |

Table 10. *Side-chain accessibility*

| Residue | Number | Number accessible | Number partially accessible | Number inaccessible | Per cent inaccessible |
|---|---|---|---|---|---|
| Ala | 35 | 7 | 7 | 20 | 57* |
| Cys | 4 | 1 | — | 3 | 75 |
| Asp | 39 | 19 | 11 | 9 | 23 |
| Glu | 25 | 16 | 4 | 3 | 12* |
| Phe | 31 | 3 | 10 | 18 | 58 |
| His | 21 | 10 | 5 | 5 | 24* |
| Ile | 18 | 4 | 1 | 13 | 72 |
| Lys | 27 | 14 | 4 | 5 | 19* |
| Leu | 35 | 6 | 6 | 23 | 66 |
| Met | 10 | 1 | 2 | 7 | 70 |
| Asn | 30 | 8 | 14 | 7 | 23* |
| Pro | 38 | 11 | 11 | 15 | 39* |
| Gln | 21 | 9 | 7 | 5 | 23 |
| Arg | 31 | 18 | 6 | 7 | 22 |
| Ser | 25 | 12 | 4 | 9 | 36 |
| Thr | 22 | 6 | 9 | 7 | 31 |
| Val | 34 | 7 | 10 | 17 | 50 |
| Trp | 6 | — | 1 | 5 | 83 |
| Tyr | 20 | 7 | 6 | 7 | 35 |

* Residues 1–2 and 501–506, not included in the model, were considered accessible in this calculation.

accessible, those with between 15–2% accessible, and those with less than 2% accessible. Table 10 shows relative side-chain surface accessibility of the different kinds of residues and their variance. The results for catalase are very similar to those for penicillopepsin (James & Sielecki, 1983). The main chain has 106 'fully' accessible residues and 112 partially accessible residues. Their position (Fig. 7*b*) correlates with absence of observed main-chain hydrogen bonds, suggesting that most of these residues may interact with non-detected water molecules.

### (*d*) *Side-chain and solvent interactions*

Table 11 lists all possible hydrogen bonds between main-chain and side-chain atoms. Within one subunit, there are 48 such bonds with respect to main-chain amino nitrogens and 80 such bonds with main-chain carbonyls. There are 5 and 13 possible hydrogen bonds, respectively, from the main chain of one subunit to all other subunits. The main-chain to side-chain hydrogen bonds are divided between the 3 possible interfaces, generated by the *P*, *Q* and *R* axes, in the ratio 1:6:11, respectively. A total of 43% of the main-chain carbonyls were involved in main-chain interactions and 20% in other interactions, while the main-chain amino groups had 45% and 13% such interactions, respectively. Baker & Hubbard (1984) report that average values for other proteins are 46% and 11% for carbonyls and 68% and 11% for amino nitrogens.

Each heme group makes 7 hydrogen bonds with the tetramer (all within the same subunit) and interacts with two to four water molecules. Each NADP molecule makes 12 hydrogen bonds within a subunit and interacts with two water molecules. A total of 49 solvent molecules per subunit were introduced, representing only a small fraction of all the water molecules in the cell. Their average temperature factor was 22 Å$^2$, ranging from 3 to 45 Å$^2$. Of these, 20

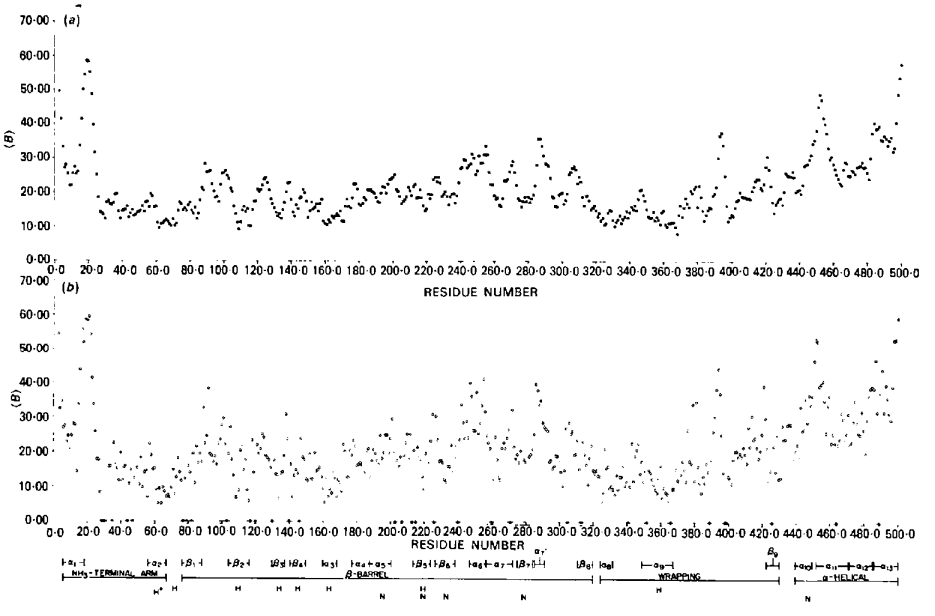



Fig. 10. Variation of mean temperature factor per residue along the polypeptide chain. (*a*) Main chain, (*b*) side chain. Shown also are the secondary structural elements, the residues involved in forming the heme pocket (H) and those which bind the NADP moiety (N).

form hydrogen bonds with main-chain amino nitrogens and 41 with main-chain carbonyls. Their disposition with respect to a subunit is shown in Fig. 11, while Table 12 counts the number of protein and water ligands for each water molecule. Of particular interest was one water molecule (identified as *W*52 with *P,Q,R* coordinates 19·6, 4·3, 28·3 Å). It had a temperature factor of 19 Å$^2$ and was associated with especially bulky density. It is positioned on the heme's distal side, near Asn 147, and at the entrance to the possible secondary heme channel (Fita & Rossmann, 1985*b*).

Possible ionic interactions, defined as distances less than 3·8 Å between groups that are probably charged at neutral pH (Arg, Lys, Glu, Asp), are listed in Table 13. There were 29 probable salt bridges (or 30 if the Arg 353 interaction with the essential tyrosine 357 is counted) within a subunit and 7 between subunits. The intersubunit salt bridges involve 6 arginines and 6 aspartic acids but only 1 glutamic acid and no lysines. Ionic subunit interactions in southern bean mosaic virus are also dominated by arginines (Rossmann, Abad-Zapatero, Hermodson & Erickson, 1983). Of some concern are residues Arg 387 and Lys 232 which are completely buried and apparently not neutralized directly by any other residues, although they are in polar environments.

### (*e*) *Tertiary and quaternary structure analysis*

The tertiary and quaternary structure has been outlined previously (Murthy *et al.*, 1981). Our aim here is to examine the relationship between secondary structural elements in a quantitative manner. The area of contact between any pair of secondary structural elements was examined by determining the reduction of surface area for the free elements when positioned in the molecule, using the Lee & Richards (1971) algorithm with a probe radius of 0·75 Å. The surface area concealed to the probe will be slightly different for each of two secondary structural elements. Hence, the average value of these two determinations was used. Finally, a distinction can be made between carbon and non-carbon atoms made inaccessible by contact between two secondary structural elements. The carbon-atom area can be considered as a function of the hydrophobic van der Waals interactions. These calculations can be represented as a matrix where the matrix elements represent interactions between any two secondary structural elements. The matrix can be extended by including the prosthetic heme and NADP groups (Table 14).

Analysis of the matrix for one subunit of catalase shows in a quantitative manner that (i) the amino-terminal arm is bound only weakly to its own subunit by way of small interactions between $\alpha_1$ and $\alpha_7$ and the primarily hydrophobic contact of $\alpha_2$ with $\alpha_8$; (ii) while there is extensive contact between the adjacent strands of the $\beta$-barrel, the contact area between $\beta_4$ and $\beta_5$ is much less; (iii) there are seven $\alpha$-helices ($\alpha_3$, $\alpha_4$, $\alpha_5$, $\alpha_{10}$, $\alpha_{11}$, $\alpha_{12}$, $\alpha_{13}$) which form a cluster; and (iv) NADP is in contact with $\alpha_5$, $\beta_5$, $\beta_6$, $\beta_7$ and $\alpha_{10}$ and the heme group with $\beta_2$, $\beta_3$, $\beta_4$, $\alpha_3$, $\beta_5$, $\alpha_8$ and $\alpha_9$.

A similar analysis can be made for the interactions between secondary structural elements in different subunits. These results are summarized in Tables 15 and 16. By far the smallest subunit surface contact area is between '*P*-related' dimers (that is, between dimers which themselves still contain subunits related by the *P* axis), while the *Q*- and *R*-related dimer contact surfaces are roughly equal. The *P*-related dimer contact surface is also dependent on a larger percentage of polar interactions. The wrapping domain provides the largest area of contact for subunit interactions, particularly for the *Q*-related dimers. The amino-terminal arm also provides a large amount of contact especially for *R*-related dimers.

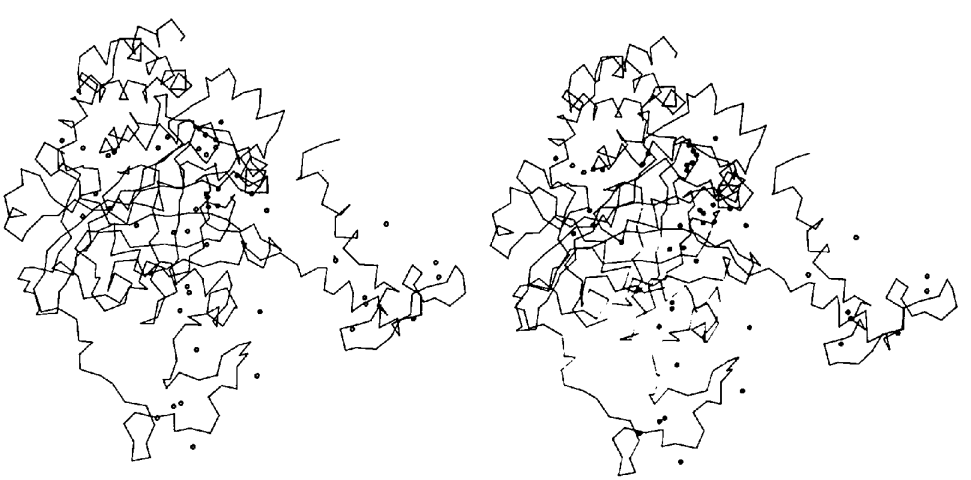


Fig. 11. Disposition of the water molecules that had definitely been located during the refinement around one subunit.

## Table 11. *Main-chain–side-chain hydrogen bonds*

(*a*) Main chain as H donors within a subunit

| Donor | | Acceptor | | Distance |
|---|---|---|---|---|
| Atom | Residue | Atom | Residue | (Å) |
| N | Ala 7 | OD1 | Asn 266 | 2·8 |
| N | Gly 29 | OD1 | Asp 58 | 3·1 |
| N | Gly 30 | OG1 | Thr 28 | 3·3 |
| N | Leu 38 | OD2 | Asp 36 | 3·2 |
| N | Leu 41 | OD1 | Asp 53 | 2·7 |
| N | Phe 56 | OD2 | Asp 53 | 3·2 |
| N | Val 115 | OD1 | Asp 127 | 2·8 |
| N | Ala 116 | OD1 | Asp 127 | 2·8 |
| N | Ser 119 | OE1 | Glu 70 | 2·7 |
| N | Arg 129 | OD1 | Asn 148 | 2·8 |
| N | Asp 139 | OG1 | Thr 137 | 2·6 |
| N | Ile 151 | OG1 | Thr 149 | 3·0 |
| N | Leu 158 | OD2 | Asp 156 | 3·2 |
| N | Gln 172 | OD1 | Asn 170 | 2·7 |
| N | Thr 173 | OD1 | Asn 170 | 3·1 |
| N | Asp 179 | OD1 | Asp 177 | 3·1 |
| N | Arg 202 | OG | Ser 200 | 3·3 |
| N | Asp 225 | OD1 | Asn 223 | 2·6 |
| N | Gly 226 | OD1 | Asn 223 | 3·1 |
| N | Glu 227 | OD1 | Asn 223 | 3·1 |
| N | Thr 237 | OD1 | Asp 212 | 3·0 |
| N | Glu 239 | OG1 | Thr 237 | 3·1 |
| N | Asp 248 | OG | Ser 245 | 3·1 |
| N | Gly 260 | OD2 | Asp 123 | 2·8 |
| N | Leu 261 | OD2 | Asp 123 | 3·2 |
| N | Ser 275 | OE1 | Gln 239 | 2·4 |
| N | Thr 284 | OE2 | Glu 287 | 3·1 |
| N | Phe 296 | OD1 | Asn 294 | 2·9 |
| N | Leu 298 | OD2 | Asp 297 | 3·1 |
| N | Thr 299 | OD2 | Asp 297 | 3·0 |
| N | Val 301 | OE1 | Gln 441 | 2·9 |
| N | Asn 323 | OE2 | Glu 327 | 3·2 |
| N | Ala 326 | OD1 | Asn 323 | 3·3 |
| N | Ala 332 | OD1 | Asn 375 | 2·9 |
| N | Met 349 | OD1 | Asp 347 | 3·1 |
| N | Gln 386 | OD1 | Asn 384 | 2·8 |
| N | Met 393 | OE1 | Gln 371 | 2·7 |
| N | Tyr 403 | OE1 | Asn 402 | 3·2 |
| N | Phe 408 | OD1 | Asn 406 | 2·6 |
| N | His 413 | OH | Tyr 404 | 3·2 |
| N | His 420 | OE1 | Glu 343 | 2·7 |
| N | Ala 434 | OD1 | Asn 432 | 3·2 |
| N | Asn 438 | OD2 | Asp 437 | 2·9 |
| N | Val 439 | OD2 | Asp 437 | 2·9 |
| N | Asn 451 | OE1 | Gln 454 | 3·0 |
| N | Glu 453 | OD1 | Asn 451 | 2·8 |
| N | Gln 454 | OD1 | Asn 451 | 3·0 |
| N | Gln 470 | OD2 | Asp 179 | 3·2 |

(*b*) Main chain as acceptor

| Acceptor | | Donor | | Distance |
|---|---|---|---|---|
| Atom | Residue | Atom | Residue | (Å) |
| O | Thr 28 | OG1 | Thr 27 | 2·8 |
| O | Gly 44 | OG1 | Thr 42 | 2·9 |
| O | Gly 140 | OG1 | Thr 137 | 2·8 |
| O | Arg 209 | OG1 | Thr 237 | 2·8 |
| O | Pro 295 | OG1 | Thr 218 | 3·2 |
| O | Gly 240 | OG1 | Thr 237 | 3·3 |
| O | Asn 266 | OG1 | Thr 270 | 3·2 |
| O | Ala 267 | OG1 | Thr 270 | 2·8 |
| O | Thr 114 | OG | Ser 113 | 3·4 |
| O | Glu 118 | OG | Ser 121 | 2·7 |
| O | Phe 163 | OG | Ser 166 | 2·4 |
| O | Trp 182 | OG | Ser 186 | 3·0 |
| O | Asp 183 | OG | Ser 186 | 3·0 |
| O | Arg 188 | OG | Ser 191 | 3·0 |
| O | Leu 192 | OG | Ser 196 | 2·6 |
| O | Phe 197 | OG | Ser 200 | 2·8 |
| O | Thr 149 | OG | Ser 216 | 2·5 |
| O | Thr 218 | OG | Ser 345 | 2·7 |
| O | Val 478 | OG | Ser 482 | 3·4 |
| O | Pro 486 | OG | Ser 490 | 3·0 |
| O | Ser 433 | NE | Arg 155 | 2·4 |
| O | His 174 | NH1 | Arg 169 | 2·4 |
| O | Arg 209 | NE2 | Gln 239 | 3·0 |
| O | Ser 275 | NE2 | Gln 239 | 2·7 |
| O | Trp 302 | NE2 | Gln 281 | 2·6 |
| O | Tyr 307 | NE2 | Gln 281 | 2·7 |
| O | Ser 345 | NE2 | Gln 351 | 2·8 |

(*b*) (*cont.*)

| Acceptor | | Donor | | Distance |
|---|---|---|---|---|
| Atom | Residue | Atom | Residue | (Å) |
| O | Val 145 | NH2 | Arg 353 | 3·1 |
| O | Glu 329 | NH2 | Arg 364 | 3·3 |
| O | Pro 390 | ND2 | Asn 368 | 2·7 |
| O | Gln 330 | ND2 | Asn 375 | 2·7 |
| O | Tyr 369 | ND2 | Asn 375 | 3·2 |
| O | Asn 384 | NH2 | Arg 387 | 3·2 |
| O | Ala 400 | NH2 | Arg 387 | 3·4 |
| O | Gln 386 | NH1 | Arg 387 | 2·5 |
| O | Tyr 403 | ND2 | Asn 406 | 3·2 |
| O | Asp 436 | ND2 | Asn 438 | 2·9 |
| O | Ala 434 | ND2 | Asn 438 | 3·0 |
| O | Val 301 | NE2 | Gln 441 | 2·8 |
| O | Val 484 | NE | Arg 443 | 3·1 |
| O | Leu 447 | NH2 | Arg 455 | 3·0 |
| O | Leu 450 | NH2 | Arg 455 | 2·9 |
| O | Asn 451 | NH2 | Arg 455 | 3·0 |
| O | Asn 451 | NH1 | Arg 455 | 3·2 |
| O | Tyr 446 | NE | Arg 455 | 2·6 |
| O | Leu 450 | NE | Arg 455 | 3·4 |
| O | Lys 37 | NZ | Lys 37 | 2·8 |
| O | Val 72 | NZ | Lys 168 | 3·4 |
| O | Lys 97 | NZ | Lys 104 | 2·9 |
| O | Ala 116 | NZ | Lys 168 | 2·8 |
| O | Ser 121 | NZ | Lys 76 | 3·5 |
| O | Ser 186 | NZ | Lys 476 | 2·3 |
| O | Ser 216 | NZ | Lys 232 | 2·9 |
| O | Lys 300 | NZ | Lys 232 | 3·2 |
| O | Ala 332 | NZ | Lys 134 | 3·3 |
| O | Ser 416 | NZ | Lys 95 | 3·3 |
| O | Phe 431 | NZ | Lys 348 | 3·3 |
| O | Asp 437 | NZ | Lys 300 | 2·5 |
| O | Gly 464 | NZ | Lys 467 | 3·1 |
| O | Ser 8 | NH1 | Arg 4 | 3·0 |
| O | Arg 18 | NE2 | Gln 21 | 2·3 |
| O | Ala 61 | NH1 | Arg 65 | 3·1 |
| O | Val 328 | NH2 | Arg 71 | 3·3 |
| O | Leu 331 | NH2 | Arg 71 | 2·6 |
| O | Glu 329 | NH1 | Arg 71 | 2·7 |
| O | His 74 | NH2 | Arg 111 | 2·2 |
| O | Ser 200 | NH1 | Arg 126 | 3·3 |
| O | Leu 178 | NE | Arg 126 | 2·9 |
| O | Asp 123 | NH2 | Arg 129 | 2·9 |
| O | Val 125 | NH2 | Arg 129 | 2·5 |
| O | Val 125 | NH1 | Arg 129 | 2·5 |
| O | Asp 127 | NH1 | Arg 129 | 2·8 |
| O | Ser 433 | NH2 | Arg 155 | 3·3 |
| O | Gly 77 | OH | Tyr 324 | 2·2 |
| O | Phe 131 | OH | Tyr 235 | 2·9 |
| O | Val 374 | OH | Tyr 136 | 2·6 |
| O | Pro 411 | OH | Tyr 385 | 2·2 |
| O | Asn 451 | NE | Arg 455 | 3·0 |
| O | Phe 199 | ND2 | Asn 461 | 2·9 |
| O | Ala 469 | NE2 | Gln 474 | 3·2 |

(*c*) Main chain as H donor

| Donor | | Acceptor | | Distance |
|---|---|---|---|---|
| Atom | Residue | Atom | Residue | (Å) |
| N | Ser 40 | OD1 | Asp 156 (*R*) | 2·7 |
| N | Leu 49 | OE1 | Gln 351 (*R*) | 3·4 |
| N | Val 322 | OE1 | Gln 172 (*R*) | 3·4 |
| N | Ala 383 | OG1 | Thr 27 (*Q*) | 3·4 |
| N | Cys 392 | OE1 | Gln 330 (*Q*) | 3·2 |
| N | Ser 407 | OD2 | Asp 183 (*P*) | 2·7 |

(*d*) Main chain as H acceptor

| Acceptor | | Donor | | Distance |
|---|---|---|---|---|
| Atom | Residue | Atom | Residue | (Å) |
| O | Asp 139 | ND2 | Asn 32 (*Q*) | 2·9 |
| O | Asn 141 | ND2 | Asn 32 (*Q*) | 2·9 |
| O | Gly 398 | ND2 | Asn 323 (*Q*) | 2·6 |
| O | Gln 397 | NE2 | Gln 10 (*Q*) | 2·8 |
| O | Asp 36 | NE | Arg 430 (*R*) | 3·4 |
| O | Asn 39 | NE | Arg 430 (*R*) | 2·9 |
| O | Asp 64 | OG1 | Thr 360 (*R*) | 2·7 |
| O | Val 72 | NH2 | Arg 67 (*R*) | 2·8 |
| O | Leu 158 | NZ | Lys 37 (*R*) | 3·2 |
| O | Pro 321 | NE2 | Gln 172 (*R*) | 2·4 |
| O | Phe 431 | ND2 | Asn 39 (*R*) | 2·7 |
| O | Asp 468 | NH2 | Arg 4 (*R*) | 3·3 |

Table 12. *Interaction of water molecules*

| Number of main-chain hydrogen bonds | Number of hydrogen bonds with other water molecules |
|---|---|
| 0 1 2 3 4 5 / 0: 4 6 1 1 1 / 1: 4 4 1 2 1 / 2: 1 3 1 | 0 |
| 0 1 2 3 4 5 / 0: 3 2 / 1: 3 3 2 / 2: 3 | 1 |
| 0 1 2 3 / 0: 1 / 1: / 2: 1 | 2 |
| 0 1 2 3 / 0: / 1: 1 / 2: | 3 |

The table shows the number of water molecules with different types of hydrogen-bonding interactions.

Table 13. *Ionic interactions*

| Positive | Negative | Positive | Negative |
|---|---|---|---|
| Lys 12 | Asp 9 | Arg 111 | Heme |
| Lys 37 | Glu 59 | Arg 155 | Asp 438, Asp 297 |
| Lys 76 | Asp 123, Asp 258 | Arg 169 | Glu 118, Asp 258 (*R*) |
| Lys 97 | Asp 139 | Arg 188 | Glu 190 |
| Lys 104 | Glu 100, Glu 138 | Arg 202 | NADP |
| Lys 134 | Asp 143 | Arg 209 | Asp 263 |
| Lys 236 | NADP | Arg 251 | Asp 248 |
| Lys 300 | Asp 297 | Arg 262 | Asp 5, Asp 256 |
| Lys 314 | Asp 238 | Arg 353 | Tyr 357 (deprotonated?) |
| Lys 456 | Glu 453 | Arg 364 | Glu 329, Heme |
| Lys 467 | Asp 468 | Arg 381 | Asp 24 (*Q*) |
| | | Arg 421 | Asp 427 (*P*) |
| Arg 4 | Asp 9, Asp 179 (*R*) | Arg 430 | Asp 53 (*R*), Glu 419 (*P*) |
| Arg 65 | Asp 359 (*R*) | Arg 455 | Glu 452 |
| Arg 67 | Glu 66 | Arg 457 | Asp 201, Glu 453 |
| Arg 71 | Glu 329, Heme | Arg 491 | Glu 452, Glu 487 |

Thus, the most flexible portions of the molecule are responsible for most of the quaternary structural interactions.

Included in Table 15 are the non-secondary structural elements. These were not included in the intra-subunit interactions (Table 14) on account of the end effects which relate to the surface area between adjacent elements and thus represent covalent interactions. However, these non-secondary structural elements can be introduced in considering intersubunit contacts without any such problems and are seen to contribute significantly (Table 15) to the interactions.

### (*f*) *Crystal packing*

In order to describe the molecular contacts in the crystal, the reference molecule will be considered to be positioned at $0,y_0,\frac{5}{6}$. Subunit 1 has its Fe atom such that all its $P,Q,R$ coordinates are positive, while the crystallographically different subunit 2 has its Fe atom at $P,Q,R$ (−16·9, 3·4, −15·2). Three different molecular interfaces can then be described (Table 17). The tetramer itself, therefore, contains a total of 6 different molecular contacts since each of the three types of interfaces will occur twice (Fig. 12).

Interface 1 involves at least five well defined hydrogen bonds; interface 2 involves primarily weak hydrophobic interactions and only two possible hydrogen bonds; and interface 3 has at least five hydrogen bonds, ionic interactions and larger hydrophobic contacts. No water molecules have been found in the space between the molecules in the contact region. This may, however, merely reflect the absence of well characterized water molecules. As observed earlier (Table 5), the molecular contact regions also correspond to maximum departure from non-crystallo-

Table 14. *Interactions between secondary structural elements*

| | $\alpha_1$ | $\alpha_2$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ | $\beta_5$ | $\beta_6$ | $\alpha_6$ | $\alpha_7$ | $\beta_7$ | $\alpha_{7'}$ | $\beta_8$ | $\alpha_8$ | $\alpha_9$ | $w$ | $\beta_9$ | $\alpha_{10}$ | $\alpha_{11}$ | $\alpha_{12}$ | $\alpha_{13}$ | Heme | NADP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha_1$ | 1530 | — | — | — | — | — | — | — | — | — | — | — | 110 | — | — | — | — | — | — | — | — | — | — | — | — | — |
| $\alpha_2$ | — | 1340 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | 3 | — | — | — | — | — | — | — | — | — |
| $\beta_1$ | — | — | 1450 | 327 | 22 | — | — | — | — | — | — | — | 81 | 20 | — | 268 | 72 | — | 11 | — | — | — | — | — | — | — |
| $\beta_2$ | — | — | 0·96 | 1540 | 281 | — | — | — | — | — | — | — | 14 | 16 | — | 63 | 49 | — | 22 | — | — | — | — | — | 35 | — |
| $\beta_3$ | — | — | 21·10 | 2·06 | 1240 | 293 | — | — | — | 15 | 44 | — | — | 27 | — | 18 | 20 | — | 106 | — | — | — | — | — | 25 | — |
| $\beta_4$ | — | — | — | — | 0·96 | 1130 | — | — | — | 68 | 55 | — | — | 5 | — | — | — | 32 | 44 | — | — | — | — | — | 81 | — |
| $\alpha_3$ | — | — | — | — | — | — | 1100 | 136 | 26 | — | — | — | — | — | — | — | — | 4 | — | — | — | — | — | — | 49 | — |
| $\alpha_4$ | — | — | — | — | — | — | 2·31 | 1230 | 150 | — | — | — | — | — | — | — | — | — | — | — | 145 | 173 | — | — | — | — |
| $\alpha_5$ | — | — | — | — | — | — | 4·30 | 2·33 | 1390 | — | — | — | — | — | — | — | — | — | — | — | 114 | 88 | 79 | 19 | — | 91 |
| $\beta_5$ | — | — | — | — | 31·00 | 1·43 | — | — | — | 1350 | 367 | — | — | 14 | 14 | — | — | 84 | — | 2 | — | — | — | — | 31 | 14 |
| $\beta_6$ | — | — | — | — | 3·58 | 5·63 | — | — | — | 1·24 | 1863 | — | — | 354 | 92 | 36 | — | — | — | — | — | — | — | — | — | 27 |
| $\alpha_6$ | — | — | — | — | — | — | — | — | — | — | — | 1150 | 89 | — | — | — | — | — | — | — | — | 8 | — | — | — | — |
| $\alpha_7$ | 1·31 | — | 2·56 | ∞ | — | — | — | — | — | — | — | 1·25 | 1450 | 22 | — | 49 | 19 | — | — | — | — | — | — | — | — | — |
| $\beta_7$ | — | — | ∞ | ∞ | 20·76 | ∞ | — | — | — | 14·47 | 1·31 | — | ∞ | 1467 | 79 | 290 | — | — | — | — | — | — | — | — | — | 9 |
| $\alpha_{7'}$ | — | — | — | — | — | — | — | — | — | ∞ | 9·76 | — | — | 1·91 | 990 | — | — | — | — | 3 | 4 | — | — | — | — | — |
| $\beta_8$ | — | — | 1·44 | 22·05 | ∞ | — | — | — | — | — | 1·00 | — | 4·71 | 1·72 | — | 1390 | 2 | — | — | — | — | — | — | — | — | — |
| $\alpha_8$ | — | ∞ | 1·91 | 5·96 | ∞ | — | — | — | — | — | — | — | ∞ | — | — | 0·00 | 990 | 52 | 90 | — | — | — | — | — | 2 | — |
| $\alpha_9$ | — | — | — | — | — | 1·00 | ∞ | — | — | 2·12 | — | — | — | — | — | — | 0·81 | 2060 | 91 | — | — | — | — | — | 207 | — |
| $w$ | — | — | ∞ | 2·07 | 1·51 | 1·51 | — | — | — | — | — | — | — | — | — | — | 1·89 | 3·02 | 4930 | — | — | — | — | — | — | — |
| $\beta_9$ | — | — | — | — | — | — | — | — | — | 1·00 | — | — | — | — | 1·00 | — | — | — | — | 1002 | — | — | — | — | — | — |
| $\alpha_{10}$ | — | — | — | — | — | — | — | — | 6·32 | — | — | — | — | — | 1·00 | — | — | — | — | — | 1260 | 74 | 103 | 101 | — | 116 |
| $\alpha_{11}$ | — | — | — | — | — | — | — | 2·61 | 3·25 | — | — | 1·81 | — | — | — | — | — | — | — | — | 1·64 | 1800 | 279 | 270 | — | — |
| $\alpha_{12}$ | — | — | — | — | — | — | — | 1·60 | 8·29 | — | — | — | — | — | — | — | — | — | — | — | 2·40 | 2·73 | 1600 | 247 | — | — |
| $\alpha_{13}$ | — | — | — | — | — | — | — | — | 0·19 | — | — | — | — | — | — | — | — | — | — | — | 3·36 | 2·84 | 2·14 | 660 | — | — |
| Heme | — | — | — | 0·53 | 12·25 | 2·68 | 11·12 | — | — | 3·19 | — | — | — | — | — | — | 0 | 1·92 | — | — | — | — | — | — | 715 | — |
| NADP | — | — | — | — | — | — | — | — | 0·97 | 0·93 | 0·71 | — | — | 0·12 | — | — | — | — | — | — | 1·38 | — | — | — | — | 740 |

This matrix shows areas (Å$^2$) in contact between two structural elements as determined by a probe atom of 1·5 Å diameter. The top half of the matrix gives total areas. The bottom half of the matrix gives the ratio of carbon atoms to non-carbon atoms in contact. The diagonal gives the total area of secondary structural elements.

Table 15. *Area of contact* (Å$^2$) *of the secondary elements in one subunit of BLC with, respectively, the P,Q,R molecular symmetry related subunits*

| Element | Residue (initial-final) | Target subunit: P-related Total | P-related C/≠C | Q-related Total | Q-related C/≠C | R-related Total | R-related C/≠C |
|---|---|---|---|---|---|---|---|
| $\alpha_1$ | 5–18 | | — | 220 | 4·5 | 115 | 1·5 |
| $\alpha_2$ | 55–66 | 10 | 0 | 228 | 2·0 | 447 | 2·0 |
| $\beta_1$ | 76–87 | | — | | — | | — |
| $\beta_2$ | 103–114 | | — | | — | | — |
| $\beta_3$ | 129–137 | | — | | — | | — |
| $\beta_4$ | 140–148 | | — | 42 | 1·3 | | — |
| $\alpha_3$ | 159–168 | 119 | 1·6 | | — | 149 | 7·3 |
| $\alpha_4$ | 177–187 | 159 | 1·9 | | — | 36 | 0·7 |
| $\alpha_5$ | 188–200 | 16 | ∞ | | — | 12 | 0·1 |
| $\beta_5$ | 213–223 | | — | | — | | — |
| $\beta_6$ | 226–238 | | — | | — | | — |
| $\alpha_6$ | 246–256 | | — | | — | 68 | 2·5 |
| $\alpha_7$ | 258–271 | | — | | — | 119 | 3·2 |
| $\beta_7$ | 274–283 | | — | | — | | — |
| $\alpha_7'$ | 284–291 | | — | | — | | — |
| $\beta_8$ | 310–319 | | — | | — | | — |
| $\alpha_8$ | 323–331 | | — | 141 | 1·8 | 71 | 3·2 |
| $\alpha_9$ | 347–366 | 20 | 5·7 | 128 | 1·1 | 276 | 2·0 |
| $w$ | 369–419 | 397 | 1·5 | 1292 | 1·8 | | — |
| $\beta_9$ | 422–429 | 287 | 0·7 | 15 | 0·7 | 155 | 2·3 |
| $\alpha_{10}$ | 439–449 | | — | | — | | — |
| $\alpha_{11}$ | 452–469 | | — | | — | 13 | 1·0 |
| $\alpha_{12}$ | 470–484 | 44 | 3·8 | | — | 43 | 1·3 |
| $\alpha_{13}$ | 485–500 | | — | | — | | — |
| Non-secondary structural elements | | 295 | 3·0 | 1062 | 1·7 | 1733 | 1·5 |

Note: C/≠C is the ratio of carbon atom area to non-carbon atom area that is in contact with the neighboring subunit. This ratio is, therefore, larger for hydrophobic contacts.

graphic symmetry. The contact area III (Table 17) is not possible in PVC as this area is covered by the additional 'flavodoxin-like' domain (Vainshtein *et al.*, 1986).

We are grateful to Sharon Wilder, Kathy Shuster and Bill Boyle for assistance in the preparation of this manuscript, as well as to Cele Abad-Zapatero and Jiliang Chiu for helpful discussions. The work was supported by National Institutes of Health grant No. GM 10704 and National Science Foundation grant No. PCM82-07747.

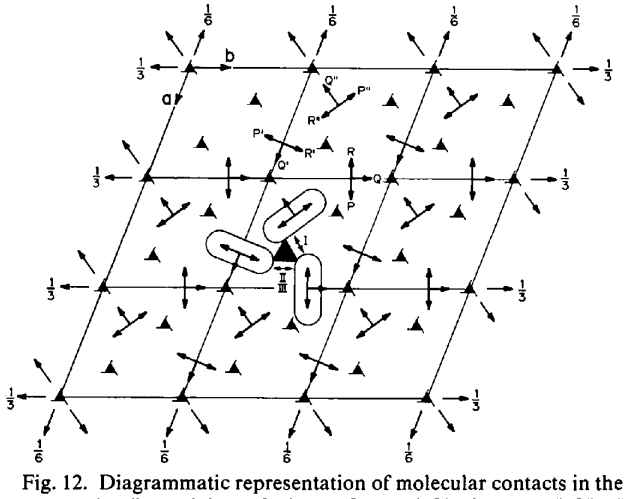



Fig. 12. Diagrammatic representation of molecular contacts in the crystal. The origins of the *P,Q,R*, *P',Q',R'* and *P'',Q'',R''* molecular axial systems are located, respectively, at $(0, y_0, \frac{5}{6})$, $(y_0-1, 0, \frac{7}{6})$ and $(-y_0, 1-y_0, \frac{1}{2})$.

Table 16. *Total area of contact* (Å$^2$)

| Between dimers containing the: | Total area | C/≠C |
|---|---|---|
| *P* axis | 1300 | 1·6 |
| *Q* axis | 3100 | 1·8 |
| *R* axis | 3200 | 1·8 |

Note: C/≠C has the same significance as in Table 15.

Table 17. *Crystal contacts between different molecules*

The position of the molecular origin in the standard molecule is at $(0, y_0, \frac{5}{6})$.

| Standard molecule: Subunit | Residue | Atom | Neighbor molecule: Subunit | Residue | Atom |
|---|---|---|---|---|---|
| Interface (I): Position of molecular origin of neighbor molecule is $(y_0, 1-y_0, \frac{1}{2})$ | | | | | |
| 2 | His 423 | ND1 | 2 | Asp 238 | OD1 |
| 2 | His 423 | NE2 | 2 | Ser 275 | OG |
| 2 | His 423 | ND1 | 2 | Lys 314 | NZ |
| 2 | Arg 421 | NH1 | 2 | Thr 277 | OG1 |
| 2 | Arg 421 | NH1 | 2 | Tyr 279 | OH |
| Interface (II): Position of molecular origin of neighbor molecule is $(y_0-1, 0, 1+\frac{1}{6})$ | | | | | |
| 2 | Glu 85 | OE2 | 2 | Pro 45 | CD |
| 2 | Val 86 | O | 2 | Pro 45 | CG |
| 2 | Thr 87 | CG2 | 2 | Pro 45 | CG |
| 2 | Ile 102 | CG2 | 2 | Pro 45 | CG |
| 2 | His 101 | CE1 | 2 | Arg 46 | NH1 |
| 2 | Gly 103 | O | 2 | Arg 46 | NH1 |
| Interface (III): Position of molecular origin of neighbor molecule is the same as for interface (II) | | | | | |
| 2 | Arg 105 | NH1 | 1 | Glu 289 | OE2 |
| 2 | Arg 379 | NH2 | 1 | Ile 290 | CD1 |
| 2 | His 107 | NE2 | 1 | Pro 292 | O |
| 2 | Tyr 273 | CD1 | 1 | His 423 | CE1 |
| 2 | Thr 277 | OG1 | 1 | Asp 427 | OD2 |
| 2 | Lys 314 | NZ | 1 | Asp 427 | OD1 |
| 2 | His 88 | NE2 | 1 | Glu 429 | OE1 |
| 2 | Thr 87 | O | 1 | Phe 431 | CD1 |
| 2 | Ile 102 | CD1 | 1 | Asn 432 | O |

## References

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

BAKER, E. N. & HUBBARD, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.

BARRY, C. D., BOSSHARD, H. E., ELLIS, R. A. & MARSHALL, G. R. (1975). *Computers in Life Science Research*, edited by W. SILER & D. A. B. LINDBERG, pp. 137–147. New York: Plenum.

BRANT, D. A. & SCHIMMEL, P. R. (1967). *Proc. Natl Acad. Sci. USA*, **58**, 428–435.

DEISSEROTH, A. & DOUNCE, A. L. (1970). *Physiol. Rev.* **50**, 319–375.

EVENTOFF, W., TANAKA, N. & ROSSMANN, M. G. (1976). *J. Mol. Biol.* **103**, 799–801.

FITA, I. & ROSSMANN, M. G. (1985*a*). *Proc. Natl Acad. Sci. USA*, **82**, 1604–1608.

FITA, I. & ROSSMANN, M. G. (1985*b*). *J. Mol. Biol.* **185**, 21–37.

FORSYTH, J. B. & WELLS, M. (1959). *Acta Cryst.* **12**, 412–415.

HAMILTON, W. C. (1964). *Statistics in Physical Sciences*, pp. 157–162. New York: Ronald Press.

HAMILTON, W. C. (1965). *Acta Cryst.* **18**, 502–510.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.26. Bangalore: Indian Academy of Sciences.

JAMES, M. N. G., HSU, I. N., HOFMANN, T. & SIELECKI, A. R. (1981). *Structural Studies on Molecules of Biological Interest*, edited by G. DODSON, J. P. GLUSKER & D. SAYRE, pp. 350–389. Oxford: Clarendon Press.

JAMES, M. N. G. & SIELECKI, A. R. (1983). *J. Mol. Biol.* **163**, 299–361.

JOHNSON, J. E. (1978). *Acta Cryst.* B**34**, 576–577.

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

KAMPHUIS, I. G., KALK, K. H., SWARTE, M. B. A. & DRENTH, J. (1984). *J. Mol. Biol.* **179**, 233-256.
KIRKMAN, H. N. & GAETANI, G. F. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 4343-4348.
KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344-350.
LEE, B. & RICHARDS, F. M. (1971). *J. Mol. Biol.* **55**, 379-400.
MURTHY, M. R. N., REID, T. J. III, SICIGNANO, A., TANAKA, N. & ROSSMANN, M. G. (1981). *J. Mol. Biol.* **152**, 465-499.
NEWCOMER, M. E., JONES, T. A., ÅQVIST, J., SUNDELIN, J., ERIKSSON, U., RASK, L. & PETERSON, P. A. (1984). *EMBO J.* **3**, 1451-1454.
RAO, S. T. & ROSSMANN, M. G. (1973). *J. Mol. Biol.* **76**, 241-256.
REID, T. J. III, MURTHY, M. R. N., SICIGNANO, A., TANAKA, N., MUSICK, W. D. L. & ROSSMANN, M. G. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4767-4771.
RICHARDSON, J. S., GETZOFF, E. D. & RICHARDSON, D. C. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 2574-2578.
ROSSMANN, M. G. (1976). *Acta Cryst.* A**32**, 774-777.
ROSSMANN, M. G., ABAD-ZAPATERO, C., HERMODSON, M. A. & ERICKSON, J. W. (1983). *J. Mol. Biol.* **166**, 37-83.
ROSSMANN, M. G. & ARGOS, P. (1975). *J. Biol. Chem.* **250**, 7525-7532.
SCHONBAUM, G. R. & CHANCE, B. (1976). *The Enzymes*, Vol. XIII, edited by P. D. BOYER, 3rd ed., pp. 363-408. New York: Academic Press.
SCHROEDER, W. A., SHELTON, J. R., SHELTON, J. B., APELL, G., EVANS, L., BONAVENTURA, J. & FANG, R. S. (1982). *Arch. Biochem. Biophys.* **214**, 422-424.
SCHROEDER, W. A., SHELTON, J. R., SHELTON, J. B., ROBBERSON, B., APELL, G., FANG, R. S. & BONAVENTURA, J. (1982). *Arch. Biochem. Biophys.* **214**, 397-421.
SHERIFF, S., HENDRICKSON, W. A., STENKAMP, R. E., SIEKER, L. C. & JENSEN, L. H. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 1104-1107.
SHRAKE, A. & RUPLEY, J. A. (1973). *J. Mol. Biol.* **79**, 351-371.
TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183-191.
TEN EYCK, L. F. (1977). *Acta Cryst.* A**33**, 486-492.
VAINSHTEIN, B. K., MELIK-ADAMYAN, W. R., BARYNIN, V. V., VAGIN, A. A. & GREBENKO, A. I. (1981). *Nature (London)*, **293**, 411-412.
VAINSHTEIN, B. K., MELIK-ADAMYAN, W. R., BARYNIN, V. V., VAGIN, A. A., GREBENKO, A. I., BORISOV, V. V., BARTELS, K. S., FITA, I. & ROSSMANN, M. G. (1986). *J. Mol. Biol.* **188**, 49-61.

*Acta Cryst.* (1986). B**42**, 515-522

# A Systematic Pairwise Comparison of Geometric Parameters Obtained by X-ray and Neutron Diffraction

BY FRANK H. ALLEN

*Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England*

(*Received* 11 *November* 1985; *accepted* 24 *March* 1986)

## Abstract

The effects of X-ray atomic asphericity on derived molecular geometry have been examined for ten common substructures involving C, N, O. The Cambridge Structural Database has been used to locate pairs of X-ray ($X$) and neutron ($N$) structure determinations of the same compound. Rigid criteria based on experimental temperature, refinement procedures, crystalline form and precision have been applied in selecting suitable $X$, $N$ structure pairs. Corresponding $X$ and $N$ values for derived parameters have been paired for 46 chemically unique subsets (39 bond-length sets and 7 valence-angle sets). The statistical significance of the distribution of $X-N$ differences ($D_i$) has been assessed in each case *via* a Wilcoxon matched-pairs signed-ranks test. The signed magnitude of the mean difference ($\bar{D}$) is large, as expected, for bonds to H, ranging from −0·096 (7) Å for C–H to −0·155 (10) Å for O–H. Some highly significant differences are also observed for bonds involving C, N, O alone. Representative $\bar{D}$ values (Å) are −0·0096 (9) for C=C (ethylenic), −0·0052 (9) for C=C (benzenoid), −0·0078 (16) for C≡N, +0·0035 (12) for C–OH (carboxyl), +0·0046 (10) for N=O (nitro), +0·0054 (8) for C–O (carbohydrate) and +0·0082 (21) for C–C (C–C≡N); the significance of the $D_i$ distribution for C=O (keto) was marginal. No significant differences were observed for C=O (COOH and $COO^-$), nor for any of the valence angles tested. Some implications for the combination or comparison of structural results obtained from different techniques are noted.

## Introduction

It is well known that the spherically symmetric atomic charge density distribution around H is subject to a relatively large distortion on formation of a bond $A$–H (Stewart, Davidson & Simpson, 1965). The H position obtained from X-ray data ($X$) with normal spherical scattering factors is displaced towards $A$ along the bond vector when compared with H positions obtained from neutron data ($N$). Thus the mean $X-N$ shifts for C–H and O–H in sucrose are −0·13 (1) and −0·18 (1) Å, respectively (Hanson, Sieker & Jensen, 1973).

Measurable asphericity shifts are not, however, restricted to H (see for example Dawson, 1964). A significant $X-N$ positional shift of O(H) in oxalic

0108-7681/86/050515-08$01.50 