SHORT COMMUNICATIONS

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 1000 *words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.*

Acta Cryst. (1977). B33, 3618–3620

Application of a restrained least-squares refinement procedure to sickling deer hemoglobin.

By W. C. SCHMIDT JR, R. L. GIRLING and E. L. AMMA, *Department of Chemistry, University of South Carolina, Columbia, SC 29208, USA*

(*Received* 13 *May* 1977; *accepted* 12 *July* 1977)

The application of a restrained least-squares refinement procedure to deer III hemoglobin is reported. Initial atomic coordinates were obtained directly from the solution of the rotation and translation functions. After 16 cycles of least-squares calculation, the R factor dropped from 0·42 (4–9 Å data) to 0·28 (1·98–5·00 Å data). With a graphics display system, it was possible to superpose least-squares-refined atomic positions onto regions in F_o maps (2·60–5·00 Å data) where specific atoms had been omitted from the phase calculation. Examination of these regions showed that most such atoms fit quite well the observed electron density, thus implying a high reliability of the phases calculated using the least-squares-refined atomic coordinates. From an F_o map (2·60–5·00 Å data) phased using the coordinates of the 4110 refined atoms, it was possible to place 392 of the 458 missing atoms.

Whilst least-squares refinement of small molecules has been standard practice for some time now, the technique has only rarely been applied to biological macromolecules with any degree of success (Watenpaugh, Sieker, Herriott & Jensen, 1973). Recently, however, Konnert (1976) has proposed an extension of the conditional structure-factor least-squares approach of Waser (1963), where subsidiary conditions based on known structural parameters are treated as observational equations. Subsidiary conditions that Konnert employs are bond distances, bond angles (nonbonded interatomic distances) and planarity of the various groups within the macromolecule. Use of these restraints, therefore, permits refinement of individual atomic positions while the structure is constrained to retain chemically reasonable geometry. This is in contrast to rigid-group refinement (Scheringer, 1963), where the position and orientation of a defined group of atoms are refined with respect to the origin and unit-cell axes, but where the atoms within the group are not permitted to move with respect to one another. The normal equations are then solved by an iterative approach termed the method of conjugate gradients, in which the matrix of the equations remains unchanged as the solution proceeds, thus requiring that only non-zero elements be stored for the calculations.

To test this approach, Konnert chose carp myogen (Nockolds, Kretsinger, Coffee & Bradshaw, 1972), a protein with a molecular weight of ~12 000 containing 108 amino acid residues (812 non-hydrogen atoms). The initial atomic coordinates were obtained from a Kendrew model constructed with the aid of an optical comparator (Richards, 1968). Initial results were quite encouraging; with 1370 reflections having d spacings of between 3–5 Å, two cycles of restrained least-squares refinement reduced the R factor from 0·42 to 0·20, with average deviations from ideal values of 0·04 Å for the bonded distances and 0·06 Å for the nonbonded distances.

This communication presents preliminary results from the application of Konnert's restrained refinement procedure to deer type III cyanomethemoglobin [β-chain variant type III; Hb(DIII)], the structure solution of which has been reported elsewhere (Schmidt, Girling, Houston, Sproul, Amma & Huisman, 1977). This application differs from that of Konnert's to the carp myogen in two important respects: first, with respect to the size of the problem, in that the myogen has 812 non-hydrogen atoms, whereas the Hb(DIII) has ~4500 such atoms (see below); second, with respect to the reliability of the initial coordinates of the atoms in each molecule. As noted above, the myogen atomic coordinates were obtained directly from an electron density map, whereas the Hb(DIII) atomic coordinates were obtained by placing the horse oxy-Hb molecule in the Hb(DIII) cell using the rotation and translation functions. Although there was no systematic attempt to make certain that all atoms so obtained were located in electron density, vanishingly low electron density in the neighborhood of some of the side chains in an F_o Fourier map (d spacings of 4 to 11 Å) phased with the 4372 atoms common to both hemoglobins, indicated that a number of the side chains would require repositioning. No attempt was made to reposition any of the poorly placed atoms prior to least-squares refinement. Thus, refinement began and proceeded without manual intervention of any sort.

To identify very badly placed atoms, a single least-squares cycle was calculated including all 4372 atoms and the 5589 reflections included in the 4 to 9 Å shell. The (restrained) interatomic distances were recalculated using the single-cycle-refined coordinates, and those atoms involved in distances deviating by more than 1 Å from the ideal were deleted. All atoms deleted in this manner were at either the N- or C-termini of the peptide chains or were at the end of the long, flexible side chains of amino acids such as lysine or arginine, *etc.* 262 atoms were deleted using this distance criterion, leaving 4110 atoms.

The progress of the refinement is shown in Table 1. We ran several least-squares cycles using low-resolution data (4–9 Å) so that the calculated derivatives would be valid over a wider range of atomic coordinates than would be the case for higher-resolution data. Use of this data gives poorly placed atoms a better chance to properly reposition themselves than if higher-resolution data were used initially. Refinement using the 4–9 Å data was halted after four cycles when the atomic

shifts became negligibly small. Higher-resolution data were added by extending the data to 3 Å, while the lower-resolution portion between 5 and 9 Å was excluded. While these latter data do have the advantage that their derivatives are valid over a wider coordinate range than higher-resolution data, they have the disadvantage that they contain significant contributions from solvent molecules and disordered side chains. As can be seen from Table 1, four least-squares cycles were performed with the 3–5 Å data, two with the 2·80–5·00 Å data, four with the 2·60–5·00 Å data, and two with the 1·98–5·00 Å data. The criterion for adding additional data at each stage was a combination of a leveling out of the R factor and a decrease in the magnitude of the atomic shifts. At the end of each cycle, the interatomic distance and planar restraints were examined to ensure that there were no portions of the molecule deviating significantly from chemical ideality.

After the final least-squares cycle using the 2·60–5·00 Å data, Fourier methods were used to assess the quality of the refined atomic positions. In all cases to be discussed, the electron density maps were phased from atomic coordinates refined through this final cycle. Three F_o maps were calculated with the 2·60–5·00 Å data, the first with all 4110 atoms whose positions were refined included in the phase calculation, the second with residues 1–19 of the α_1 chain omitted (A helix, AB turn; 119 atoms) and the third with residues 89–117 of the α_1 chain omitted (FG turn, G helix, GH turn; 218 atoms). All three maps were subjected to a program written by one of us (WCS) which calculates the electron density at an atomic position by interpolation of the electron density at the eight nearest-neighbor grid points surrounding the atomic position. The results of these calculations indicate that virtually all the atoms remained in regions of positive electron density whether or not some of them were omitted from the phase calculation, thus implying a high reliability of the phases calculated using the least-squares-refined atomic coordinates. In further support of this point, the F_o map based on phases calculated with all 4110 refined atomic positions was clear enough to permit placement of 392 previously missing atoms.* One disappointing aspect of the refinement is that, of several unrealistically close intermolecular contacts observed before refinement, a number still remained after refinement. This implies that atoms which are very badly misplaced prior to refinement cannot be corrected by the restrained least-squares refinement procedure, but rather require correction by manual intervention at some stage.

* The map was examined on a PDP-11/40 linked to a Digital 17" CRT using a locally modified version of a program (Love & Wishner, 1976) which displays a stereopair of designated sections of a pre-contoured electron density map on a CRT.

Table 1. *Progress of the refinement*

The least-squares calculations were performed on an IBM 370/168. The standard deviations used for calculating the weights on the restraints were as follows: 0·062 Å and 0·125 Å, respectively, for the bonded and non-bonded distances; 0·10 Å for the planes. The value of $\langle|\Delta F|\rangle$ was taken as the structure amplitude standard deviation in weighting the reflection data.

Resolution (Å)		R (%) In	R (%) Out*	Average shift (Å) All†	Main‡	Side‡	All§	Number of reflections	Time (h per cycle)
4·00–9·00	(1)	41·9	38·7	0·239				5589	2·72
	(2)	38·7	35·7	0·185					
	(3)	35·7	33·7	0·177					
	(4)	33·7		0·083	0·306	0·681	0·459		
3·00–5·00								9524	4·50
	(1)	40·0	35·8	0·206					
	(2)	35·8	33·6	0·148					
	(3)	33·6	32·1	0·125					
	(4)	32·1		0·119					
2·80–5·00								11647	5·35
	(1)	32·1	30·6	0·128					
	(2)	30·6		0·123					
2·60–5·00								13934	6·33
	(1)	30·6	29·0	0·096					
	(2)	29·0	27·5	0·095					
	(3)	27·5	26·5	0·085					
	(4)	26·5		0·072					
1·98–5·00								19930	8·93
	(1)	28·9	28·4	0·107					
	(2)	28·4		0·085	0·591	0·790	0·652		
					0·698	1·035	0·831		

* Blanks in this column represent R factors which were not calculated.
† Average shift in an atomic position for the least-squares cycle indicated.
‡ Average shift in a main or side-chain atomic position at the conclusion of the least-squares cycles indicated.
§ Average shift in atomic position at the conclusion of the least-squares cycles indicated.

We wish to express our thanks to J. H. Konnert and W. A. Hendrickson for a copy of the restrained least-squares refinement program, and for valuable discussions regarding its implementation. We also wish to thank the University of South Carolina Computer Services Division; B. McFadden, J. Berger and W. Horger for their aid in streamlining the least-squares program and G. Walters and L. Zarpentine for their efforts in the special scheduling required for the program's execution.

This work was supported by NIH grant HL-15158.

References

KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.

LOVE, W. E. & WISHNER, B. C. (1976). Private communication.

NOCKOLDS, C. E., KRETSINGER, R. H., COFFEE, C. J. & BRADSHAW, R. A. (1972). *Proc. Natl Acad. Sci. USA*, **69**, 581–584.

RICHARDS, F. M. (1968). *J. Mol. Biol.* **37**, 225–230.

SCHERINGER, C. (1963). *Acta Cryst.* **16**, 546–550.

SCHMIDT, W. C. JR., GIRLING, R. L., HOUSTON, T. E., SPROUL, G. D., AMMA, E. L. & HUISMAN, T. H. J. (1977). *Acta Cryst.* B**33**, 335–343.

WASER, J. (1963). *Acta Cryst.* **16**, 1091–1094.

WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943–956.

Acta Cryst. (1977). B**33**, 3620

P-Methyl-*P*-phenylpropylphosphine selenide: correction of a printer's error.

By Z. GAŁDECKI and M. L. GŁÓWKA, *Institute of General Chemistry, Technical University*, 36 *Żwirki*, 90-924 *Łódź*, *Poland* and J. MICHALSKI, A. OKRUSZEK and W. J. STEC, *Polish Academy of Science, Centre of Molecular and Macromolecular Studies*, 90-362 *Łódź*, *Poland*

(*Received* 20 *September* 1977)

Improved reproductions are given of Figs. 1 and 2 of the paper by Gałdecki, Główka, Michalski, Okruszek & Stec [*Acta Cryst.* (1977), B**33**, 2322–2324].

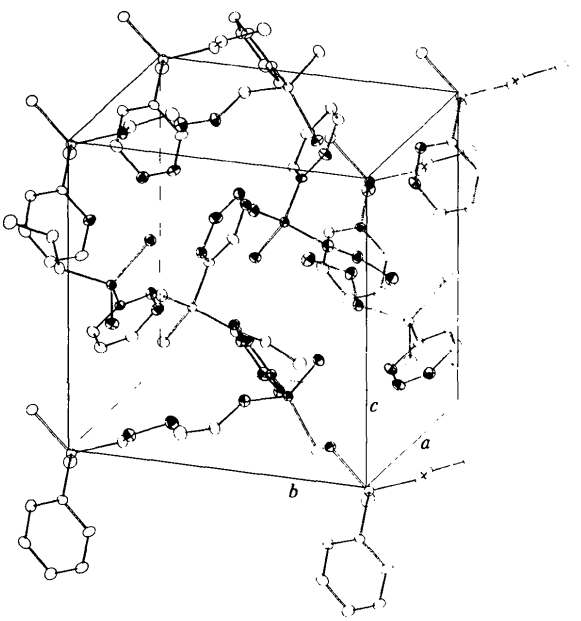

Fig. 1. Computer drawing of *P*-methyl-*P*-phenylpropylphosphine selenide. The thermal ellipsoids have been scaled to include 40% probability. Hydrogen atoms are omitted for clarity.

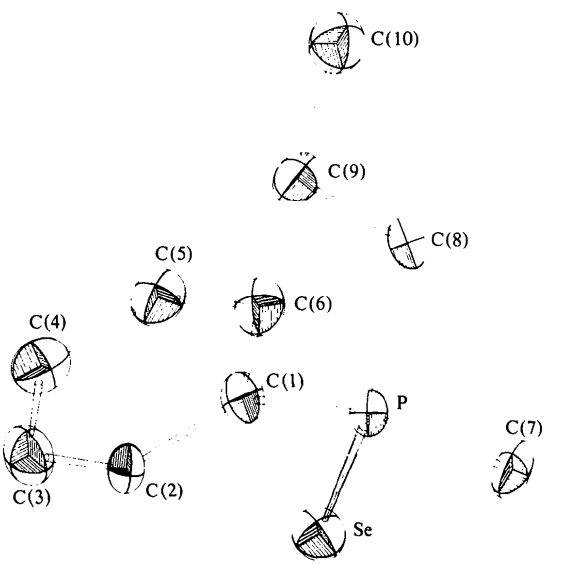

Fig. 2. Projection of the unit-cell contents.