Table 7. *Distances less than* 3·5 Å *between methyl carbon atoms and neighbouring atoms*

C(1) (I) – O(2) (VII)	3·22 Å	C(2) (I) – O(2) (II)	3·20 Å
– O(2) (VIII)	3·24	– O(4) (V)	3·42
– O(3) (VI)	3·27	– O(4) (I)	3·45
– O(4) (VI)	3·43	– O(7) (III)	3·20
– O(4) (VIII)	3·44	– O(7) (I)	3·48
– O(5) (IX)	3·44	– O(8) (I)	3·38
– O(7) (VI)	3·18		
– O(8) (IX)	3·29		

References

ALBANO, V., BELLON, P. L., POMPA, F. & SCATTURIN, V. (1963). *Ric. Sci.* 3A, 1067.

COULTER, C. L., GANTZEL, P. K. & MCCULLOUGH, J. D. (1963). *Acta Cryst.* **16**, 676.

CRUICKSHANK, D. W. J. (1961). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, p.32. Oxford: Pergamon Press.

FLETCHER, R. O. W. & STEEPLE, H. (1964). *Acta Cryst.* **17**, 290.

HUGHES, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.

HUGHES, E. W. & LIPSCOMB, W. N. (1946). *J. Amer. Chem. Soc.* **68**, 1970.

IITAKA, Y. & HUSE, Y. (1965). *Acta Cryst.* **18**, 110.

International Tables for X-ray Crystallography (1962). Vol. III, p.202. Birmingham: Kynoch Press.

MCCULLOUGH, J. D. (1964). *Acta Cryst.* **17**, 1067.

OKAYA, Y., AHMED, M. S., PEPINSKY, R. & VAND, V. (1957). (1957). *Z. Kristallogr.* 109, p.367.

PHILLIPS, D. C. (1954). *Acta Cryst.* **7**, 746.

STEWART, J. M. (1964). Technical Report TR-64-6, Univ. of Maryland Computer Science Center.

TRUTER, M. R. (1961). *Acta Cryst.* **14**, 318.

TRUTER, M. R., CRUICKSHANK, D. W. J. & JEFFREY, G. A. (1960). *Acta Cryst.* **13**, 855.

Acta Cryst. (1968). B24, 504

A New Method of Locating Heavy Atoms Bound to Protein Crystals

BY THOMAS A. STEITZ*

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, *U.S.A.*

(*Received* 5 *May* 1967 *and in revised form* 28 *July* 1967)

It is shown that heavy atoms bound isomorphously to protein crystals can be located by using a direct method (Sayre's equation) to phase X-ray diffraction reflections in centrosymmetric projections. This method has been tested successfully with three derivatives of carboxypeptidase A, one of which contains four heavy atoms per protein molecule. Quite possibly several heavy atoms can be located in this manner with greater ease and assurance than by difference Patterson methods alone and thus the method may be especially useful in the X-ray study of proteins larger than 125,000 molecular weight, where binding of more than one or two heavy atoms will probably be both common and necessary.

Multiple isomorphous replacement phasing of X-ray diffraction data is the method which has proven most useful for determining protein structures. Its application requires the preparation of several isomorphous heavy atom derivatives of the native protein and the determination of the positions of these heavy atoms in the derivative crystals. The difference Patterson synthesis (Green, Ingram & Perutz, 1954), which is used to locate heavy atoms when no estimate of the protein phases is available, is quite adequate if only one or two heavy atoms are bound per asymmetric unit. However, by this method the location of larger numbers of heavy atoms becomes increasingly difficult.

It is shown here that a direct phase determination using Sayre's equation may be applied to reflections in centrosymmetric zones to locate heavy atoms bound isomorphously to protein crystals. The method has been successfully tested with three heavy atom derivatives of carboxypeptidase A_α (CPA_α), one of which contains eight mercury atoms per unit cell (four per molecule), and hence may prove to be a useful and powerful alternative to the difference Patterson syntheses for locating large numbers of heavy atoms.

* Present address: Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, England.

The method

Sayre's (1952) equation expresses a relationship between the signs (phases) of certain structure factors for centrosymmetric structures of non-overlapping atoms:

$$s(F_{\mathbf{h}}) = s \sum_{\mathbf{k}} (F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}}),$$

where $s(\ )$ means 'sign of' and $F_{\mathbf{h}}$, $F_{\mathbf{k}}$ and $F_{\mathbf{h}-\mathbf{k}}$ are the structure factors of reflections $\mathbf{h}(=hkl)$, $\mathbf{k}$, and $\mathbf{h}-\mathbf{k}$. Using this equation, and assuming the signs of a few structure factors, the signs of other structure factors can be predicted and these in turn can be used to pre-

dict still others. With this method the structures of many small molecules in centrosymmetric crystals have been solved directly from the intensity data, *e.g.* $B_{20}H_{16}(NCCH_3)_2.CH_3CN$ (Enemark, Friedman & Lipscomb, 1966).

Since proteins can crystallize only in non-centrosymmetric space groups, Sayre's method is not applicable to determining the phases of general reflections *hkl*. However, it can be used to determine phases for centrosymmetric projections. From measurements of the native and derivative intensities for these projections the scattering due to the heavy atoms alone, $|\Delta F|$, can be calculated, assuming the native and derivative F's to have the same sign:

$$|\Delta F|_{\mathbf{h}} = ||F_D|_{\mathbf{h}} - |F_P|_{\mathbf{h}}| \, .$$

Here $|F_D|_{\mathbf{h}}, |F_P|_{\mathbf{h}}$ are the derivative and native protein structure factor amplitudes of reflection $\mathbf{h}(=hkl)$. The $|\Delta F|$'s, then, can be phased by Sayre's equation and the positions of the heavy atoms located directly in a difference Fourier map.

In the application of this method to CPA_α the values of $|\Delta F|_{\mathbf{h}}$ for the centric *h0l* data were converted to normalized structure factors, $\Delta E_{\mathbf{h}}$ (Hauptman & Karle, 1953),

$$|\Delta E|^2_{\mathbf{h}} = |\Delta F|^2_{\mathbf{h}} / \left(\sum_{j=1}^{N} f^2_{j,\mathbf{h}} \right),$$

where N is the number of heavy atoms in the unit cell and the f_j are their atomic scattering factors, evaluated at the scattering angle of reflection $\mathbf{h}$ and corrected for the effects of thermal motion. Since accurate determination of a protein absolute scale is difficult and since one initially has no knowledge of the number of heavy atom binding sites or their occupancy, one mercury atom ($N=1$) was assumed in all cases described below and $\langle \Delta E^2 \rangle$ was normalized to 1·00 by adjustment of the scale factor.

Sayre's equation was applied to these ΔE's using the multiple-solution computer program of Long (1965) on an IBM 7094 computer. On the basis of large values of $|\Delta E_{\mathbf{h}}| \sum_{\mathbf{k}} |\Delta E_{\mathbf{h}}| \, |\Delta E_{\mathbf{h}-\mathbf{k}}|$ the signs of two reflections were chosen by the program to fix the origin in space group $P2_1$ (Woolfson, 1961), and four additional signs were assumed to initiate phasing. Each of the sixteen combinations of assumed signs was iterated to self-consistency; *i.e.* the phasing was continued until no new signs were determined and no signs changed after a pass through the data.

Long defines a consistency index:

$$C = \frac{\langle |E_{\mathbf{h}} \sum_{\mathbf{k}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \rangle}{\langle |E_{\mathbf{h}}| \sum_{\mathbf{k}} |E_{\mathbf{k}}| \, |E_{\mathbf{h}-\mathbf{k}}| \rangle} \, ,$$

where the average is taken over all $\mathbf{h}(=hkl)$. In the case of small structure determination the correct solution usually has the largest value of C and requires the smallest number of cycles to reach self-consistency. However, this has not always been true in the phasing of the normalized $|\Delta F|$'s for CPA_α.

Results and discussion

This method was tested with three derivatives of CPA_α, a zinc-containing enzyme, which has a molecular weight of 34,600 and crystallizes in space group $P2_1$ with two molecules per unit cell of dimensions $a=51{\cdot}4$, $b=59{\cdot}9$, $c=47{\cdot}2$ Å, $\beta=97°35'$ (Lipscomb, Coppola, Hartsuck, Ludwig, Muirhead, Searl & Steitz, 1966). The three derivatives examined were a single-site mercury derivative, Hg(1), in which the zinc is replaced by mercury (Hartsuck, Ludwig, Muirhead, Steitz & Lipscomb, 1965), a double-site lead derivative, Pb(2) (Ludwig, Paul, Pawley & Lipscomb, 1963), and a four-site mercury derivative, Hg(4) (Steitz, 1966). The *h0l* data used from the first two derivatives extended to a minimum Bragg spacing of 2 Å resolution, while those used from the third extended to 2·8 Å. Derivative and native protein reflections of small intensity (about one-fourth of the 2 Å projection data) were not included, since erroneously large $|\Delta F|$'s might arise from errors in their measurement.

The distributions of ΔE's obtained for the three derivatives were as predicted theoretically (Hauptman & Karle, 1953) and are given in Table 1. Although in the case of small structures solutions are usually calculated for reflections of $|E| \geq 1{\cdot}5$, it was found that better phasing of the normalized $|\Delta F'|$s was obtained if all ΔE's $\geq 1{\cdot}0$ were included.

Table 1. *The distributions of* ΔE's*

	Observed			Theory†
	Pb	Hg(1)	Hg(4)	
$\langle \lvert\Delta E\rvert \rangle$	0·793	0·802	0·799	0·798
$\langle \lvert\Delta E^2 - 1\rvert \rangle$	0·965	0·953	0·968	0·968
$\lvert\Delta E\rvert \geq 3{\cdot}0$	0·22%	0·5%	0·0%	0·3%
$\lvert\Delta E\rvert \geq 2{\cdot}0$	3·7%	4·0%	4·6%	5·0%
$\lvert\Delta E\rvert \geq 1{\cdot}0$	33·4%	32·8%	32·4%	32·0%

* $\langle \Delta E^2 \rangle$ normalized to 1·00 by adjustment of the scale factor
† Hauptman & Karle (1953).

For each of the three derivatives the correct solution of the sixteen was found by comparing the ΔE maps with the difference Fourier maps (ΔF) calculated with use of the protein phases determined by the multiple isomorphous replacement method (Ludwig, Hartsuck, Steitz, Muirhead, Coppola, Reeke & Lipscomb, 1967): (*a*) The correct solution for Hg(1) had the highest consistency index ($C=0{\cdot}89$) and required the fewest cycles (8). The difference E map calculated with the 132 phased ΔE's contained one peak whose height was more than four times the next highest background peak and whose position was consistent with the difference Patterson map and with the ΔF map. (*b*) The solution of the Pb(2) derivative which had the highest consistency index and required the least number of cycles

was not correct. A difference map calculated with 154 signed ΔE's from the correct solution contained two peaks having the proper positions and relative heights. (*c*) Again, the correct solution of the Hg(4) derivative was not obvious from the number of cycles or the consistency index. Table 2 lists the five solutions with the highest consistency index; number four is correct. The ΔE map [Fig. 1(*a*)] contains four peaks clearly above background. For comparison, a ΔF map calculated using protein phases deduced from the lead derivative is shown in Fig. 1(*b*). The peak-to-background ratio of the former map is somewhat greater than that of the latter. The peaks in both the ΔE map and the ΔF map are consistent with the difference Patterson map [Fig. 1(*c*)] which contains noise peaks as high as some of the double-weight cross-interaction peaks and which we had been unable to solve earlier.

Table 2. *Solutions of* Hg(4) *phasing with highest consistency indices*

No.	Consistency index	Number of cycles
1	1·0	5
2	0·942	7
3	0·913	5
4	0·890	7
5	0·878	7

Although neither the consistency index nor the number of cycles indicates unambiguoulsy which of the sixteen solutions is correct, both comparison of the ΔE maps with the difference Patterson map and heavy-atom refinement can distinguish the correct solution. Probably ΔE maps need to be calculated only for several of the solutions with high consistency indices. The

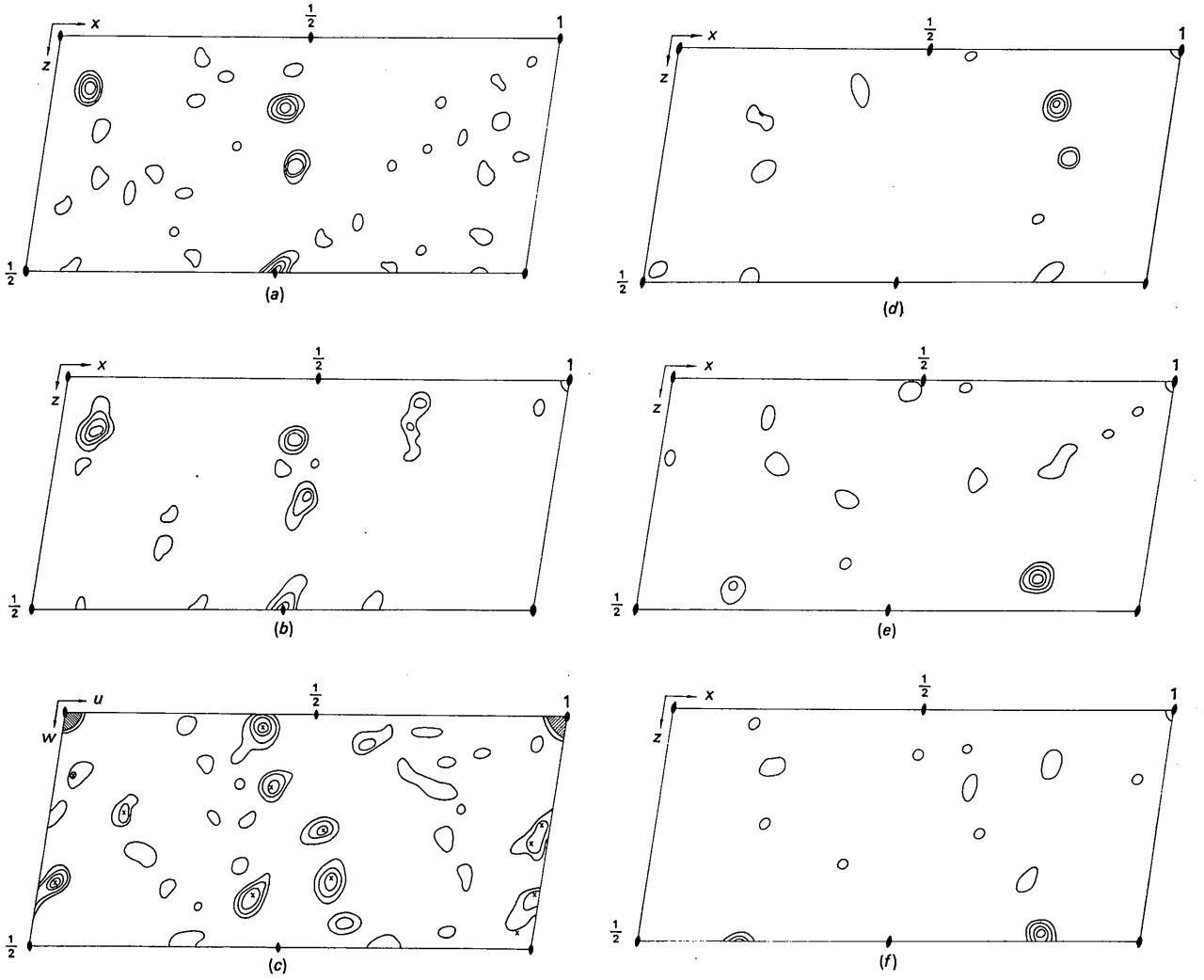

Fig. 1. Four-site mercury difference maps at 2·8 Å resolution, [010] projection. Contours are at equal but arbitrary intervals above the zero contour. (*a*) Difference *E* map calculated using 85 ΔE's phased with Sayre's equation. (*b*) Difference Fourier map calculated applying protein phases deduced from the lead derivative. (*c*) Difference Patterson map. The expected interactions are marked by crosses. (*d*), (*e*), and (*f*) Incorrect ΔE maps.

incorrect ΔE maps [Fig. 1(*d*), (*e*), (*f*)] also contain peaks above background (usually only one or two), but these peaks do not adequately account for the significant difference Patterson map peaks and do not refine. Table 3 shows the results of six cycles of refinement of the coordinates derived from the ΔE maps of the correct and of two incorrect solutions for the Hg(4) derivative; these refinements were carried out on an IBM 7094 computer using a program written by H. Muirhead (Lipscomb *et al.*, 1966). The correct solution can be easily distinguished since it yields a higher figure of merit, lower root-mean-square lack-of-closure error and a lower R index.

Table 3. *Refinement of correct and two incorrect* Hg(4) *solutions*

	Correct solution	Incorrect solution	
Solution number	4	3	2
*Average figure of merit ($\bar{m}$)	0·33	0·19	0·21
†Lack of closure error in electrons	47	65	63
‡R Index	0·44	0·62	0·58

$$*m_{hkl} = \frac{\int_0^{2\pi} p(\alpha) \cos(\alpha - \alpha_0) d\alpha}{\int_0^{2\pi} p(\alpha) d\alpha} = \langle \cos(\alpha - \alpha_0) \rangle;$$

$$\bar{m} = \frac{\sum_{hkl} m_{hkl}}{\text{number of reflections}},$$

where α is the phase angle of reflection (hkl) and $p(\alpha)$ is the probability that the phase angle has the value α.

$$\dagger E_j = [\sum_{hkl} (|F_D|_{obs} - |F_D|_{calc})^2 / n_j]^{1/2},$$

where n_j is the number of observed amplitudes of scattering $|F_D|_{obs}$ for derivative j.

$$\ddagger R = \sum_{hkl} |F_D - F_P|_{calc} / \sum_{hkl} |F_D| - |F_P|_{obs}.$$

The results presented here show that by application of Sayre's equation in centrosymmetric projections at least four heavy atoms per asymmetric unit can be located as unambiguously as with a conventional ΔF map. Although the limit of this method is not known, it appears quite possible that the positions of large numbers of bound heavy atoms can be found by this method with greater ease and assurance than by difference Patterson methods alone and therefore the method will be of greatest use in locating the heavy atoms in the first protein derivative, when no protein phase information is available. Furthermore, the ability to locate many heavy atoms may facilitate the structure determination of proteins whose molecular weights exceed 125,000. In these cases binding of more than one or two heavy atoms per protein molecule will probably be rather common and indeed may be required in order to obtain measurable intensity changes.

I thank Professor W. N. Lipscomb, whose support and encouragement made this study possible. I also want to thank Drs J. Enemark, D. Voet, and J. Hartsuck for able advice and Drs R. E. Long and K. N. Trueblood for the computer program for the Sayre method. I acknowledge support by the Office of Naval Research, the National Institutes of Health, and the Advanced Research Projects Agency.

References

ENEMARK, J. H., FRIEDMAN, L. B. & LIPSCOMB, W. N. (1966). *Inorg. Chem.* **5**, 2165.

GREEN, D. W., INGRAM, V. M. & PERUTZ, M. F. (1954). *Proc. Roy. Soc.* A**225**, 287.

HARTSUCK, J. A., LUDWIG, M. L., MUIRHEAD, H., STEITZ, T. A. & LIPSCOMB, W. N. (1965). *Proc. Nat. Acad. Sci. Wash.* **53**, 396.

HAUPTMAN, H. & KARLE, J. (1953). *The Solution of the Phase Problem. I. The Centrosymmetric Crystal.* ACA Monograph No. 3. Ann Arbor, Michigan: Edwards Brothers.

LIPSCOMB, W. N., COPPOLA, J. C., HARTSUCK, J. A., LUDWIG, M. L., MUIRHEAD, H., SEARL, J. & STEITZ, T. A. (1966). *J. Mol. Biol.* **19**, 423.

LONG, R. E. (1965). Ph. D. Thesis, Part III, U.C.L.A.

LUDWIG, M. L., HARTSUCK, J. A., STEITZ, T. A., MUIRHEAD, H., COPPOLA, J. C., REEKE, G. N. & LIPSCOMB, W. N. (1967). *Proc. Nat. Acad. Sci. Wash.* **57**, 511.

LUDWIG, M. L., PAUL, I. C., PAWLEY, G. S. & LIPSCOMB, W. N. (1963). *Proc. Nat. Acad. Sci. Wash.* **50**, 282.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60.

STEITZ, T. A. (1966). Ph. D. Thesis, Harvard Univ.

WOOLFSON, M. M. (1961). *Direct Methods in Crystallography.* London: Oxford Univ. Press.