Multiple Heavy-Atom Sites in Protein Crystals having Centrosymmetric Projections: Interpretation of Vector Maps. II. Correlation of Sites in Different Adducts

By C. W. Bunn

The Royal Institution of Great Britain, Davy-Faraday Laboratory, London, England

For different heavy-atom adducts of the same protein crystal, the relative signs of the structure factors $F_{H(A)}$, $F_{H(B)}$ etc. for a particular X-ray reflexion in a centric zone are known experimentally, though the absolute signs are not. The Fourier synthesis with coefficients $F_{H(A)}F_{H(B)}$ which makes the best use of this information gives a vector map ('correlation map') showing no vectors between A sites and no vectors between B sites, but only AB vectors. By itself, the correlation map for two multi-site adducts is of little use, but superposition of the ordinary Patterson maps of A and B on the correlation map AB, with appropriate displacements of origins, can reveal the sites in each adduct as sets of peak coincidences, which for *mm* projections form concentric rectangles. Another superposition method uses three correlation maps AB, AC and BC; any two of these (say AB and AC), superposed with origins displaced by a vector shown in the third (BC), reveal the A sites as a set of peak coincidences, which for *mm* projections form concentric rectangles; the interpretation is checked by adding the A Patterson map to the AB and AC correlation maps, when triple peak coincidences forming concentric rectangles should appear. When several adducts are available, additional checks are possible. The methods have indicated the probable positions of principal sites in adducts of calf rennin.

The difficulties of interpreting multi-site orthorhombic heavy-atom vector maps for centrosymmetric projections (Bunn, Camerman, Liang Tung-T'sai, Moews & Baumber, 1970) arise from the large number of vector peaks, which overlap in ridges and broad plateaux, and from lack of knowledge of the effective scattering powers of the several sites owing to the occurrence of fractional occupancies. It is also worth noting that there is an intrinsic ambiguity in plane groups *cmm* and *pmm*: for any two sites of similar occupancy having coordinates x_1y_1 and x_2y_2, the coordinates can be interchanged to x_1y_2 and x_2y_1 without affecting the cross vectors or the axial self-vectors between symmetry-related atoms (Fig. 1); the only difference is in the position of the non-axial self-vectors, but as these are weak they may not show up. In principle, cross vectors with a third site should distinguish between the two, but in practice the 'many peaked' confusion may lead to doubts.

When interpretation for each individual heavy-atom adduct is ambiguous, the correlation of information from different adducts should be more discriminating. The extra information consists of the relative signs of the heavy-atom structure factors $F_{H(A)}$, $F_{H(B)}$, etc. (where $F_{H(A)} = |F_{PH(A)}| - |F_P|$, etc.) for each reflexion. These relative signs are known experimentally,
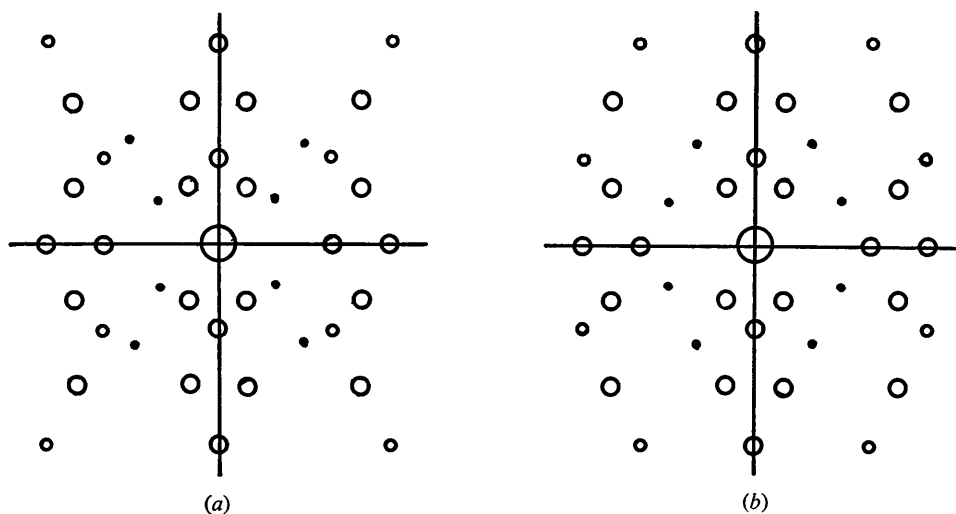

Fig. 1. (*a*) Vectors (○) for a structure of two sites (●), at x_1y_1 and x_2y_2. Plane group with *mm* symmetry. Peak heights indicated by sizes of circles. (*b*) Vectors for structure with sites at x_1y_2 and x_2y_1. All strong peaks in the same positions; only two weak self-vectors are in different places.

though the absolute signs are not; for if heavy atoms A and B increase the intensity of a given protein crystal reflexion, or both decrease it, the signs of $F_{H(A)}$ and $F_{H(B)}$ must be the same, while if A and B have opposite effects, the signs must be opposite. To use this information, the A and B data must be combined in some manner that preserves the sign relations. Of the three combination possibilities: addition, subtraction and multiplication, only the last is generally useful for multi-site adducts. The addition synthesis [coefficients $(F_{H(A)} + F_{H(B)})^2$] would give a map showing all the heavy-atom vectors in a crystal containing both A and B atoms, and would be much more complex than the individual Patterson maps of A and B. The subtraction synthesis [coefficients $(F_{H(A)} - F_{H(B)})^2$] also gives a very complex map, with the additional disadvantage that in some places negative AB peaks may overlap positive AA or BB peaks (or both), cancelling them wholly or partly, so that evidence is lost. It has been used by Rossmann (1960), even in three dimensions where fractional phases complicate the situation, but only for very simple adducts, to settle the relative y coordinates of A and B atoms in a monoclinic structure; the negative AB peaks were quite separate from the positive AA and BB peaks. For multi-site adducts both these syntheses are of little use. Only the multiplication synthesis (coefficients $F_{H(A)}F_{H(B)}$), which gives only positive AB peaks and nothing else (Steinrauf, 1963; Kartha, Bello, Harker & De Jarnette, 1963; Phillips, 1966) is potentially useful.

By itself, however, a correlation map given by the $F_{H(A)}F_{H(B)}$ synthesis for two multi-site adducts is too complex for unambiguous interpretation – even more so than the individual Patterson maps. (For two adducts with 4 sites each, in general positions in mm plane groups, the correlation map would contain 64 vector peaks per lattice point, inevitably overlapping seriously. Each Patterson map would have 36 peaks.) The purpose of this paper is to suggest that the way out of these difficulties is to combine correlation maps with Patterson maps, or different correlation maps AB, AC, BC, etc. with each other, by superposition procedures analogous to those used by Buerger (1959) and others for Patterson maps. Such combinations are more discriminating than operations with individual maps, for interpretation rests on peak coincidences in very different maps.

Characteristics of correlation maps and Patterson maps with mm symmetry

A correlation map for the simplest example – two one-site structures A and B – is shown in Fig. 2, which may be compared with Fig. 1(a), the Patterson map of a structure containing both sites in the same crystal. The same cross vectors are present in both maps, but the correlation map has no peaks on the axes or at the origin. The absence of an origin peak, in terms of structure factors means that the sums of positive and

negative $F_{H(A)}F_{H(B)}$ terms are equal. In terms of vectors the absence of an origin peak means that there are no zero distances between A and B atoms. An important feature of Fig. 2 is that a B atom lies at the centre of the rectangle 1234 representing the four A atoms as seen from B_2, and an A atom lies at the centre of the rectangle 4567 representing the four B atoms as seen from A_2.

For multi-site structures, correlation maps show sets of concentric rectangles: each B atom is the common centre of a set representing the whole A structure, and each A atom is the common centre of a set representing the whole B structure. Fig. 3 illustrates this for the simplest example.

Multi-site adducts of two different heavy-atom-containing substances sometimes have one or more sites in common but the rest in quite different places. The type of correlation map given by such a pair is illustrated in Fig. 4 for the simplest case of a pair of two-site structures with one site in common. There is
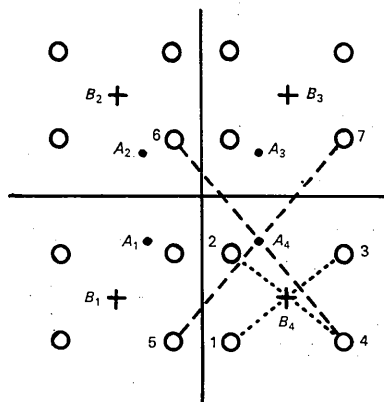

Fig. 2. Correlation vector map for two different structures with one site each. An A site is at the centre of a rectangle of vectors representing the B structure, and a B site is at the centre of a rectangle representing the A structure. There are no vectors at the origin or on the axes.
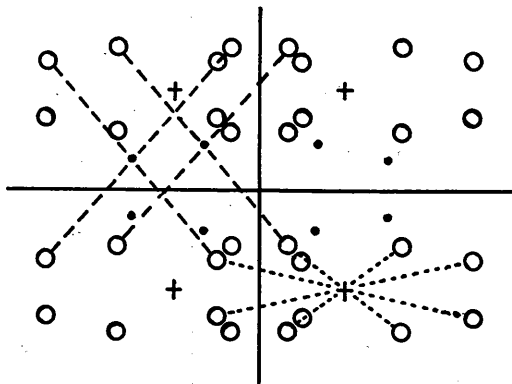

Fig. 3. Correlation map for a pair of structures, A with two sites and B with one. Two concentric rectangles, representing the A structure, have a B atom at the common centre. Each A atom is at the centre of a rectangle representing the one-site B structure.

an origin peak, together with peaks along the axes. The origin peak, in terms of structure amplitudes means that the sum of the positive $F_{H(A)}F_{H(B)}$ terms is greater than the sum of negative ones. In terms of vectors it means that there exist zero distances between A and B atoms in this projection. The correlation map, has, for the common site only, the characteristics of an ordinary Patterson map, including peaks on the axes at twice the coordinates of the common site. The correlation map thus isolates and identifies the common site. A common-site atom is the common centre of rectangles representing all the sites in both adducts.

In practice one has to deal not with a set of points but with contour maps of limited resolution and
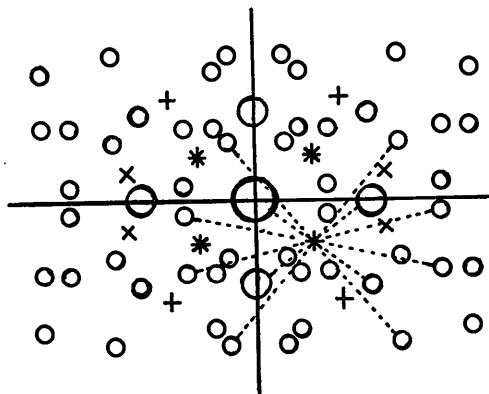

Fig. 4. Correlation map for a pair of two-site structures having one site in common. Peaks at the origin and along the axes identify the common site, which is the common centre of rectangle representing both structures.
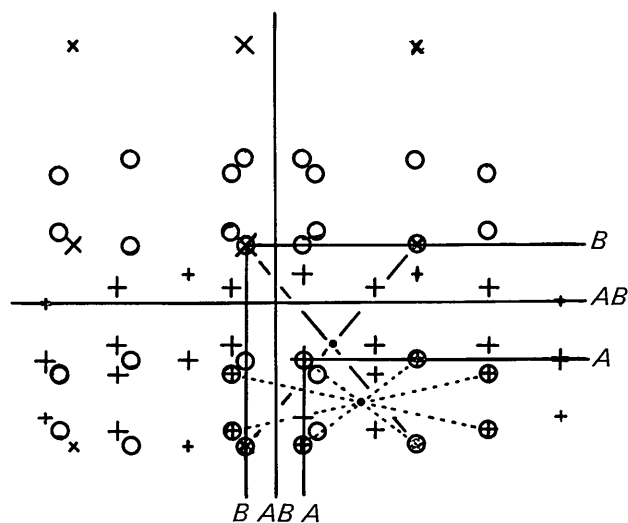

Fig. 5. Patterson maps of A and B superposed on the correlation map AB, with origins on diametrically opposite AB vector peaks. $+$ A vectors, $\times$ B vectors, $\bigcirc$ AB vectors. Sizes of symbols indicate peak heights. Coincidences of A and AB peaks form rectangles representing the two A sites; coincidences of B and AB vectors form a rectangle representing the B site. Coincidences on the Patterson axes lie at equal distances from the AB axes.

limited accuracy (since many $|F_H|$ values are based on small differences between large $|F_{PH}|$ and $|F_P|$ figures); moreover, the 'many peaked' confusion is increased by the intrusion of vectors originating in neighbouring lattice points. Consequently, the sets of concentric rectangles referred to in the foregoing discussion may not be recognized, and false interpretations are only too likely. The difficulties are similar to those encountered in attempting the interpretation of a Patterson map, in which every site is the centre of a set of concentric rectangles representing the whole structure; alternative interpretations of apparently equal merit are possible. The ambiguities of interpretation may, however, be reduced or in favourable cases eliminated, by combining correlation maps with Patterson maps.

Superposition of correlation and Patterson maps

The procedure is simplest when a correlation map shows an origin peak, proving the existence of at least one common site. The origin peak should be accompanied by peaks along the axes, which should correspond with axial peaks on both the individual Patterson maps. In these circumstances superposition of the correlation map AB, first on Patterson map A with origins coincident, and then on Patterson map B, is informative. If there is only one common axial peak with coordinates, say $2x,0$ the common site is at $x,0$. If there is one peak along each axis, at $2x,0$ and $0,2y$, there may be one common site at x,y, or alternatively two common sites on the axes, one at $x,0$, and the other at $0,y$. If there is more than one common peak along each axis, there are additional possibilities, including the ambiguity illustrated in Fig. 1.

To decide which interpretation is correct, and to locate the other sites in each adduct, peak coincidences forming a set of concentric rectangles are sought. In the Patterson map of A, each site is the centre of a set of concentric rectangles representing all the A sites. In the correlation map AB, the common site is the centre of two sets of concentric rectangles, one representing the A sites and the other the B sites. In both maps there is much confusion due to overlapping of many other peaks, but the overlapping is likely to be very different in the two maps, which are quite different except for common-site peaks; consequently, the places where they coincide should form rectangles representing the A structure. Similarly, superposition of the B Patterson map on the AB correlation map, with origins coincident, should reveal peak coincidences forming rectangles representing the B structure. The word 'peak' should not be taken literally; the highest vector densities on both types of maps are due to the summation of several closely placed vector peaks, and the more moderate vector densities are more likely to be due to single vectors. Consequently, when looking for significant rectangles by observing vector densities at points at x,y; $-x,y$; $x,-y$; and $-x,-y$ from the chosen common-site position, the lowest of the four values is the signi-

ficant quantity; the principle here is Buerger's (1959) 'minimum function' method for interpreting Patterson maps. A convenient procedure is to have a computer program choose such minima for all rectangles round the chosen common-site position, at convenient intervals of x and y, leading to a map of possible sites for each vector map. The places where these possible sites coincide in the Patterson and the correlation maps should be the correct site positions. If there is little similarity, it probably means that the common-site evidence has been incorrectly interpreted, and the other possibilities mentioned in the previous paragraph should be tried. An alternative procedure would be to add the maps together first, and then apply the 'minimum at rectangle corners' criterion; if the first common-site interpretation leads to a featureless result, other possibilities are tried.

When there is no common site, Patterson and correlation maps cannot be superposed with origins coincident. Suitable procedures are indicated by considering Fig. 2, the correlation map for two one-site structures. The rectangle 1234 represents the A structure, and the Patterson map of A, if placed with its origin on 2, would show axial peaks coinciding with 1 and 3 and a non-axial peak coinciding with 4. Similarly the Patterson map of B, placed with its origin on 6 (note that 6 is diametrically opposite 2) would show axial peaks coinciding with 5 and 7, and a non-axial peak coinciding with 4. The coincidences 4, 1, and 5 lie

on a line parallel to the horizontal axis, while 4, 3, and 7 lie on a line parallel to the vertical axis. The procedure therefore is to place the two Patterson maps with origins on diametrically opposite peaks of the correlation map and look for coincidences disposed in the way indicated in Fig. 2. An alternative procedure is to place both Patterson maps with their origins on 4; coincidences would be found along the Patterson axes, at 3 and 1 for the A Patterson and 7 and 5 for the B Patterson.

The application of the first mentioned procedure to the correlation map shown in Fig. 3 (for structure A with two sites and structure B with one site) is illustrated in Fig. 5. Coincidences along the axes of the Patterson maps give twice the coordinates of one site in each structure, and show which sites are involved in the correlation peak chosen for placing the Patterson map origins. The centre of the A rectangle is a B atom (referred to the AB axes), which is also the centre of a rectangle of coincidences representing the second A site.

In practice, a correlation map for two multi-site structures contains many peaks, and the superposition procedure may be repeated, choosing different correlation peaks for the Patterson origins, which lead to different initial identifications of the one site in each (by coincidences with axial Patterson peaks); but the other sites are indicated by a set of concentric rectangles, and the same set should appear for all the super-
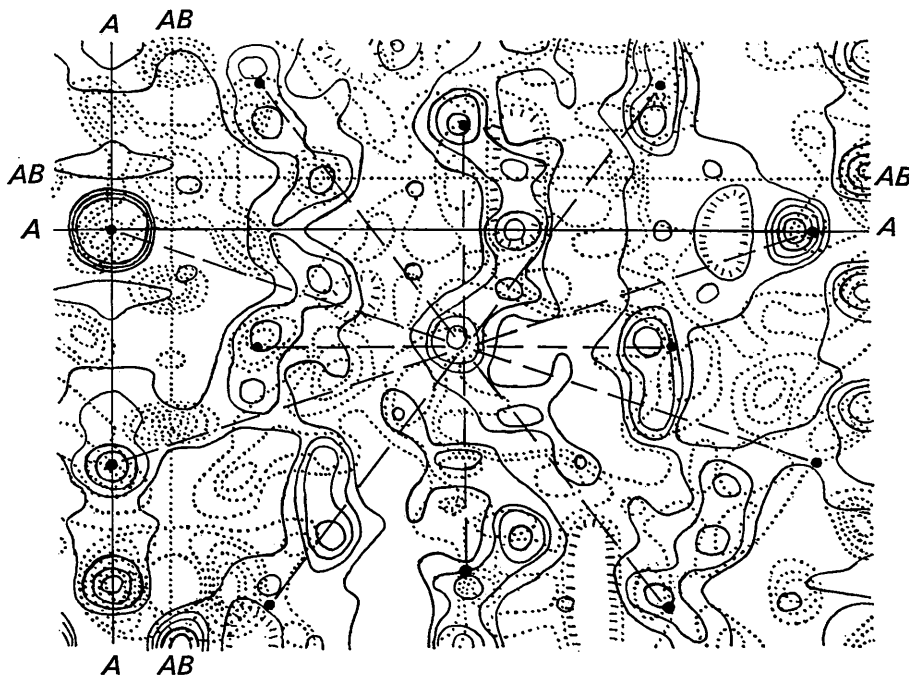

Fig. 6. Superposition of Patterson map A (full lines) for thallous malonate (in calf rennin) on the correlation map AB (dotted lines) of thallous malonate with p-chloromercuribenzenesulphonate, the origin of A being placed on a peak of AB. The initial identification of an A site was made with the origins of A and B Patterson maps on diametrically opposite peaks of AB; B was then withdrawn. The lowest vector density regarded as significant for a coincidence is the lowest contour in each map. Concentric rectangles of coincidences represent the principal sites in the A structure. (For sites on the axes, the rectangles contract to lines.) This set of sites is confirmed by superpositions with the origin of A on other AB peaks.

positions. In any one superposition, spurious rectangles may occur; only those sites indicated consistently should be accepted. To minimize confusion, it is advisable to draw the maps in different colours, and to withdraw one Patterson map as soon as the initial identification of single sites has been made.

An example is shown in Fig. 6, which is the superposition of the Patterson map A for thallous malonate (in calf rennin) on the correlation map AB of thallous malonate with p-chloromercuribenzenesulphonate, the origin of A being placed on a peak of AB. The initial identification of an A site was made with the origins of A and B Patterson maps on diametrically opposite peaks of AB; B was then withdrawn. Concentric rectangles of coincidences represent the principal sites of the A structure. (For sites on the axes, the rectangles contract to lines.) The centre of this system of concentric rectangles is an A site with respect to the A axes and a B site with respect to the AB axes.

In spite of the checks imposed by superpositions at several different places, there is still a possibility of false indications of sites when all the maps are very complex. It is therefore advisable to make correlation maps of A and B with a third adduct C; if the same set of A sites is indicated by an AC map and an AB map, confidence of its correctness is increased; and of course the same applies to B and C.

Superposition of two or more correlation maps

An AB correlation map consists of several repetitions of the A structure, each a group of vector peaks forming a set of concentric rectangles with a B atom at the centre; repetition at each B atom builds the whole assembly. Similarly, an AC correlation map consists of repetitions of the same A structure, but repeated at
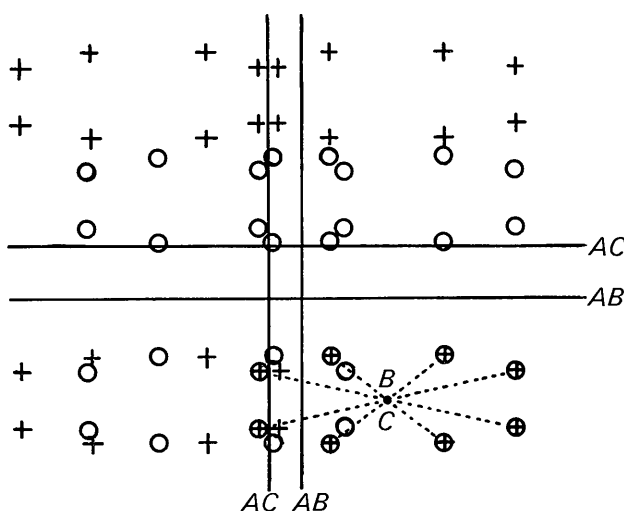

Fig. 7. AB and AC correlation maps superposed with origins displaced by a BC vector. $\bigcirc$ AB vectors, $+$ AC vectors. Coincidences form rectangles representing A sites. The common centre is a B site referred to AB axes, and a C site referred to AC axes.

each C atom, and since the C atoms are in general in quite different positions from B atoms, the AB and AC maps are quite different. It should be possible to reveal the A structure as a group of peak coincidences by superposing the AC map on the AB map. Initially, the positions of B and C atoms are unknown but BC vectors are all contained in the BC correlation map, and therefore peaks in this map indicate the displacements of the origins of AB and AC maps which should reveal the A structure as a set of concentric rectangles of peak coincidences. The procedure is therefore to place the AC map on the AB map with its origin displaced by the vector indicated by any peak on the BC map; concentric rectangles of peak coincidences give the A sites. This is illustrated for the simplest situation (an A structure with two sites, B and C structures of one site each, in different places) in Fig. 7. In practice, there are several sites in B and C, and we do not know initially where they are, though the vectors between them are given by the BC map; therefore we do not know where the concentric rectangles representing A sites will appear. When they have been located, the common centre is a B atom with respect to the AB axes and a C atom with respect to the AC axes. Several peaks on the BC map may be used as displacement vectors for the origins of the AB and AC maps; the same set of concentric rectangles of peak coincidences should be revealed. Some spurious rectangles must be expected when the maps are complex, and therefore only those consistently revealed in several displacement superpositions should be accepted. The most suitable BC peaks for displacement vectors are obviously isolated ones of moderate height.

An example is shown in Fig. 8, which is the superposition of AB, the correlation map of thallous malonate with p-chloromercuribenzenesulphonate, on AC, the correlation map of thallous malonate with potassium chloroplatinate, the origins being displaced by a vector in the BC map. The set of concentric rectangles representing the A structure is the same as in Fig. 6. The common centre is a B site with respect to the AB axes and a C site with respect to the AC axes. Note that since the corner of one rectangle lies within the origin peak of AB, the A site involved is very near a site in B; this conclusion is confirmed by the superposition of A, B and AB maps with origins nearly coincident; in fact, this near-common site could have been the starting point for deriving the other sites in A and B.

The same procedure may be used to find B sites by superposing AB and BC maps with origins displaced by an AC vector, and to find C sites by superposing AC and BC maps with origins displaced by an AB vector.

Strategy

Experience so far with several adducts of calf rennin shows that at the outset, when nothing is known about the sites, the first method, in which a correlation map

is combined with the two Patterson maps, is the most useful. But as soon as at least one site in each adduct is considered provisionally to be known, the second method, in which different correlation maps are combined, becomes more practicable. For if one *B* site and one *C* site are postulated, the *AB* and *AC* maps can be superposed with a *B* site in the *AB* map coinciding with a *C* site in the *AC* map; the *A* structure should then appear as a set of concentric rectangles of peak coincidences round this point. Furthermore, the *A* Patterson map can be added with its origin on one corner of one of the rectangles; if the coincidences already observed become triple coincidences, the *A* structure is confirmed.

These procedures can be used for several adducts, each providing a cross check on all the others. Perhaps the most convenient method, when at least one site in each is considered to be known, is to apply the 'minimum at rectangle corners' criterion to produce possible site maps. For instance, if an *A* site is known, it can be used in an *AB* map to give a possible site map of *B*, and in other correlation maps *AC*, *AD*, *AE*, etc., to give possible site maps of *C*, *D*, *E*, etc. If a *B* site is known, it can be used to give site maps for the other adducts, and so on. All the site maps for *A* can be compared, and where they agree, the sites can be regarded as established. The individual Patterson maps can be brought into this scheme also: at any site in *A*, the 'minimum at rectangle corners' map for all rectangles round the chosen site should show up all the sites in *A*; it may show too many, and only those that agree with the site maps obtained from the correlation maps can be accepted.

When several sites in any one adduct are considered to be known, a different procedure can be used. The atomic positions in *A*, for instance, are plotted on transparent paper as a set of points at the corners of concentric rectangles, and this whole set is moved about on an *AB* correlation map to find places where all the points fall on places of at least medium vector density; the centre of the set of rectangles (the origin of the *A* structure) gives, for each match position, a *B* site. The set of *A* points is an example of Buerger's (1959) 'image-seeking function' applied to correlation maps.

By using these methods on several multi-site adducts, it should be possible to find sets of sites which are compatible in the sense that the vectors between sites in different adducts are correct. The larger the number of cross checks between different adducts, the better. The test of correctness is that when the sites have been used in structure-factor calculations with refinement procedures to adjust site coordinates and occupancies and to detect minor sites, the same set of protein signs should be obtained for all the adducts. Correlation maps of course imply a uniform set of protein signs – this is implicit in the relative heavy-atom signs used in the product synthesis; and if interpretation of the maps is correct, this uniform set of protein signs should be correct. If a uniform set of protein signs is not ob-
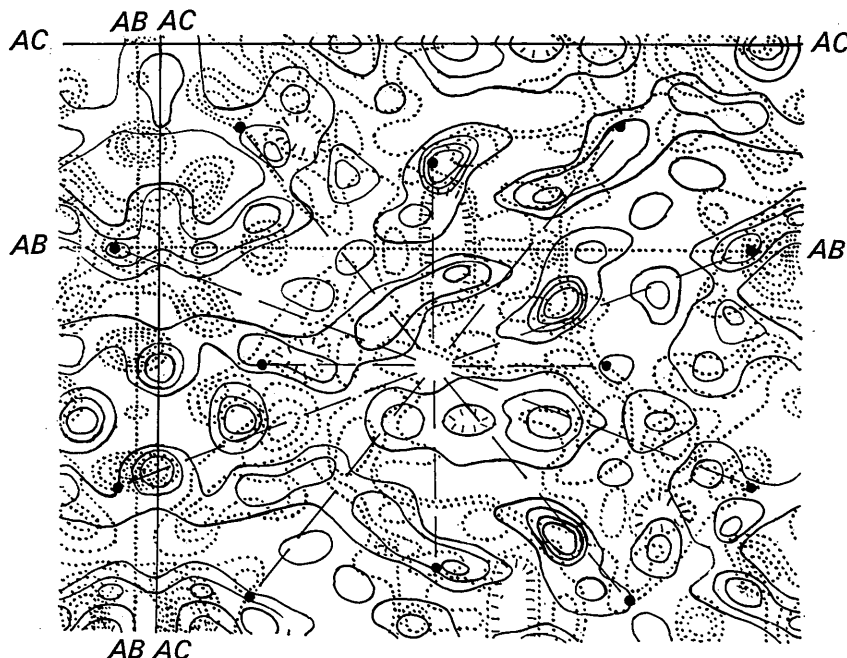

Fig. 8. Superposition of *AB* (dotted lines), the correlation map of thallous malonate with *p*-chloromercuribenzenesulphonate, on *AC* (full lines), the correlation map of thallous malonate with chloroplatinate, the origins being displaced by a vector in the *BC* map. The lowest vector density regarded as significant for a coincidence is the lowest contour in each map. The set of concentric rectangles representing the *A* structure is the same as in Fig. 6. The common centre is a *B* site with respect to the *AB* axes and a *C* site with respect to the *AC* axes. The same sites are indicated by superpositions with origins displaced by other *BC* vectors.

tained, interpretation has been incorrect. A further test is that the same sites with the same occupancies should be found in different centrosymmetric projections of any one adduct.

Similar procedures could be used for other centrosymmetric plane groups. The statement that in an AB correlation map an A site is at the centre of a group of vector peaks representing the B sites, and vice versa, is valid wherever there is a centre of symmetry; but the multiplicity and orientation of the group of peaks for any one site depend on the plane group symmetry. For instance, for the one centrosymmetric projection of monoclinic protein crystals (symmetry $p2$), there would be, around each A site, pairs of diametrically opposite peaks representing B sites, and the diameters would be in different radial directions, not parallel to any axis; and in plane group $p4$ (for the c projection of some tetragonal crystals), the groups are squares with their sides inclined to the cell axes. A peak at the origin is always an indication of a common site, and the superposition procedures, for Patterson maps on a correlation map or for different correlation maps on each other, are the same as those given for the orthorhombic example.

There are obviously limitations to the complexity of the structures that might be solved by these methods. The number of sites, multiplicity of equivalent positions, range of occupancies and the effective resolution of the data employed are all involved:

(i) For the twofold multiplicity of a monoclinic centric projection, it should be possible to deal with a larger number of sites than for the fourfold orthorhombic multiplicity.

(ii) It will probably be possible to locate only the sites of relatively high occupancy; sites of low occupancy would have to be detected by established methods based on electron density maps phased by the sites of high occupancy.

References

BUERGER, M. J. (1959). Vector Space, pp. 218, 239. New York: John Wiley.
BUNN, C. W., CAMERMAN, N., LIANG TUNG-T'SAI, MOEWS, P. C. & BAUMBER, M. E. (1970). Phil. Trans. B257, 153.
KARTHA, G., BELLO, J., HARKER, D. & DE JARNETTE, F. E. (1963). In Aspects of Protein Structure. Ed. G. N. RAMACHANDRAN, p. 13. London: Acadamic Press.
PHILLIPS, D. C. (1966). In Advances in Structure Research by Diffraction Methods. Edited by R. BRILL & R. MASON, Vol. 2, p. 75. New York: Interscience Publishers.
ROSSMANN, M. G. (1960). Acta Cryst. 13, 221.
STEINRAUF, L. K. (1963). Acta Cryst. 16, 317.

The Crystal Structure of UZn₁₂*

BY J. T. MASON AND P. CHIOTTI

Institute for Atomic Research and Department of Metallurgy, Iowa State University, Ames, Iowa 50010, U.S.A.

The crystal structure of the most zinc-rich compound in the uranium–zinc system is hexagonal with space group symmetry $P6/mmm$. Crystals formed by cooling a 95 wt.% zinc alloy have a composition range of $UZn(10\cdot4 - 11\cdot0) \pm 0\cdot1$ as determined by microprobe analysis; lattice parameters are $a = 8\cdot950 \pm 0\cdot001$ and $c = 8\cdot902 \pm 0\cdot002$ Å. The structure is based on an ideal UZn_{12} stoichiometry, but substitution between pairs of zinc atoms and uranium atoms leads to lower concentrations of zinc, $UZn_{10\cdot4}$ being found for the crystal studied. The reported equivalence of the $SmZn_{12}$ and UZn_{12} structures based on powder work is confirmed. They differ only in the extent of substitution at different lattice sites.

Introduction

Two intermetallic compounds are known to exist in the uranium–zinc system. U_2Zn_{17} is reported to have both a rhombohedral and a hexagonal form. The status of the structural work has been summarized by Johnson, Smith & Wood (1968).

The existence of a second compound richer in zinc than U_2Zn_{17} has been shown by a series of investigations by researchers at the Argonne National Laboratory (Martin & Wach, 1960, 1962a & 1962b; Velekis & Goetzinger, 1960; Argonne National Laboratory, 1965). They report a hexagonal compound with a composition range of $UZn_{9\cdot36}$ to $UZn_{11\cdot47}$. Veleckis, Schablaske & Tani (1966), using power methods, found that this compound, designated UZn_{12}, is isostructural with the high-temperature form of $SmZn_{12}$. A detailed

---