New Algebraic and Probabilistic Relations
between Structure Factors and Direct Space Information

By M. Knossow

Centre National d'Etudes des Télécommunications, 92220 *Bagneux, France*

and C. de Rango, Y. Mauguen, M. Sarrazin and G. Tsoucaris

E.R. n° 180, Laboratoire de Physique, Centre Pharmaceutique, 92290 *Chatenay Malabry, France*

Further investigation of the Karle–Hauptman determinants leads to several new relations: the D_N determinant is simply related to interatomic vectors; also obtained is a probabilistic restriction on the atomic positions inside the algebraically allowed regions, as defined by von Eller. On the other hand new Gram determinants are defined and shown to be useful in relating moduli, phases and interatomic vectors. A practical application is given of fitting a stereochemically known fragment in the unit cell.

The aim of this paper is the study of useful properties of new Gram determinants (1) in connexion with probabilistic relations and with the notion of 'forbidden regions' introduced a few years ago by von Eller (1962, 1964) and implicitly by Goedkoop (1952). These authors have shown that the use of the positivity of a Karle–Hauptman (K.H.) determinant associated with a partial structure can yield information about the atomic positions from the structure factors involved in the determinant. In this paper we report further results on probabilistic properties of K.H. determinants, as well as on new forms of determinants dealing with the following related topics: forbidden regions; use of stereochemical information. Indeed, (2) leads to a K.H. determinant which enables new phase restrictions to be deduced from the observed moduli. However, if one wants to introduce additional *a priori* information, the $\mathbf{V}_p$ set can be adapted accordingly to this information. This is the main idea underlying this paper.

I. Recall on Gram determinants and probabilistic theory

(a) *Gram determinants*

Let $\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_m$ be m vectors in an N-dimensional vector space, the associated Gram determinant is defined by the scalar product matrix U:

$$D_m = D(\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_m) = \text{Det } \mathsf{U} = \begin{vmatrix} (\mathbf{V}_1 . \mathbf{V}_1)...(\mathbf{V}_1 . \mathbf{V}_m) \\ \vdots \qquad \vdots \\ (\mathbf{V}_m . \mathbf{V}_1)...(\mathbf{V}_m . \mathbf{V}_m) \end{vmatrix} \tag{1}$$

These determinants were first introduced in crystallography by Kitaigorodsky (1961) and von Eller (1955). The $\mathbf{V}_p$ vector is defined by

$$\mathbf{V}_p = \sum_{j=1}^{N} g_j \exp(2\pi i \mathbf{H}_p . \mathbf{r}_j)\mathbf{e}_j \quad p = 1, ..., m \tag{2}$$

where N is the number of atoms in the unit cell, $\{\mathbf{e}_j\}$ is an orthonormal frame of the N-dimensional

vector space, $\mathbf{r}_j$ is an atomic position, $g_j = f_j / \sqrt{(\Sigma f_j^2)}$ where f_j is the atomic scattering factor of the jth atom, $\mathbf{H}_p$ is a reciprocal-space lattice vector.

The scalar products $(\mathbf{V}_p . \mathbf{V}_q)$ are then the unitary structure factors $U_{\mathbf{H}_p - \mathbf{H}_q}$ (denoted $U_{p,q}$) and the vector set Gram determinant is a K.H. determinant. Different choices of vector sets or vector spaces lead to new forms of determinants which will be discussed in § III. On the other hand, the algebraic and probabilistic extension of the K.H. determinant so as to contain explicitly atomic positions will be discussed in § II.

(b) *Multivariate probability theory*

The well known Gram determinants have been used primarily to obtain information in reciprocal space by means of inequality theory. But D_m, (1), may also be considered as the determinant of the covariance matrix of m random variables which are normalized structure factors: $E_q = E_{\mathbf{L} - \mathbf{H}_q}$, $q = 1, ..., m$, where $\mathbf{L}$ is a variable vector and $\mathbf{H}_q$ a fixed vector.

It is possible to deduce from this remark the joint probability density of these m structure factors, as well as other direct-methods formulae (Tsoucaris, 1970). One of these formulae is the maximum determinant rule which will be used hereafter. These formulae use the information contained in a determinant explicitly involving reciprocal space. Here we will derive some new mathematical relations where direct and reciprocal space information is 'mixed'.

II. Algebraic and probabilistic properties of K.H. determinants

In (a) we recall von Eller's and Goedkoop's arguments and then show their equivalence.* In (b) and (c) we develop respectively new algebraic and probabilistic properties.

---

* This equivalence has been derived independently by Dr G. von Eller (private communication).

(a) *General form of determinants involving atomic vectors. Forbidden regions*

$\{U_\mathbf{H}\}$ being the set of structure factors of a given crystal and $\mathbf{r}$ being an atomic position,

$$\left\{ U_\mathbf{H} - \frac{1}{N} \exp(2\pi i \mathbf{H} . \mathbf{r}) \right\} \tag{3}$$

is the set of structure factors of an imaginary structure, obtained by subtracting an atom from the first one. Any K.H. determinant associated with this new structure must be non-negative. Let us choose m vectors in reciprocal space, denoted by $\mathbf{H}_p$ ($p = 1,...,m$) and consider the differences $\mathbf{H}_p - \mathbf{H}_q$ ($p,q = 1,...,m$). The corresponding structure factors of the imaginary structure form a square matrix of order m; we call $^s\mathcal{D}_m$ its determinant (s to account for one atom substracted; Table 1). The regions in direct space which are 'forbidden' (for algebraic reasons) are those in which $^s\mathcal{D}_m$ is negative. Once the phases involved in the determinant are known, the forbidden region may be computed by mere evaluation of $^s\mathcal{D}_m$ for any value of $\mathbf{r}$.

Table 1. $^s\mathcal{D}_m$, *m-order Gram determinant, the forbidden regions K.H. determinant*

Notation: $e_{p,q} = \frac{1}{N} \exp 2i\pi(\mathbf{H}_p - \mathbf{H}_q)\mathbf{r}$.

$$\begin{vmatrix} \left(1 - \frac{1}{N}\right) & \cdots & (U_{1,p} - e_{1,p}) & \cdots & (U_{1,m} - e_{1,m}) \\ \vdots & & \vdots & & \vdots \\ (U_{p,1} - e_{p,1}) & \cdots & \left(1 - \frac{1}{N}\right) & \cdots & (U_{p,m} - e_{p,m}) \\ \vdots & & \vdots & & \vdots \\ (U_{m,1} - e_{m,1}) & \cdots & (U_{m,p} - e_{m,p}) & \cdots & \left(1 - \frac{1}{N}\right) \end{vmatrix}$$

Table 2. $\delta^{(r)}_{m+1}$, *m + 1-order Gram determinant, proportional to $^s\mathcal{D}_m$*

The dashed lines are drawn only to emphasize the special role of the last column and row which contain the contribution of the subtracted atom. Notation: $e_p = \frac{1}{N} \exp 2i\pi(\mathbf{H}_{m+1} - \mathbf{H}_p).\mathbf{r}$.

$$\delta^{(r)}_{m+1} = \begin{vmatrix} 1 & \cdots & U_{1,p} & \cdots & U_{1,m} & e_1^* \\ \vdots & & \vdots & & \vdots & \vdots \\ U_{p,1} & \cdots & 1 & \cdots & U_{p,m} & e_p^* \\ \vdots & & \vdots & & \vdots & \vdots \\ U_{m,1} & \cdots & U_{m,p} & \cdots & 1 & e_m^* \\ \hline e_1 & \cdots & e_2 & \cdots & e_m & \frac{1}{N} \end{vmatrix}$$

On the other hand, Goedkoop (1952) introduced a determinant (Table 2) defined as follows: m vectors $\mathbf{V}_p$ (2) define an ordinary K.H. matrix and the elements of the last row are:

$$e_p = \frac{1}{N} \exp[2\pi i(\mathbf{H}_{m+1} - \mathbf{H}_p).\mathbf{r}]. \tag{3a}$$

$\mathbf{r}$ is now a 'test atomic position', the same as that of the subtracted atom in von Eller's treatment. It is easy to show by linear combinations of columns and rows, which do not change the value of the determinant, that the following form may be given to $\delta^{(r)}_{m+1}$:

$$\delta^{(r)}_{m+1} = \begin{vmatrix} ^s\mathcal{D}_m & \vdots & 1 \\ \hline 0 & \vdots & \frac{1}{N} \end{vmatrix}$$

and we have

$$\delta^{(r)}_{m+1} = \frac{1}{N} {}^s\mathcal{D}_m. \tag{3b}$$

Once U is known, $\delta^{(r)}_{m+1}$ can be evaluated by straightforward algebra:

$$\delta^{(r)}_{m+1} = \frac{1}{N} D_m - \frac{D_m}{N^2} \sum_{p,q=1}^{m} D_{pq} \exp[2\pi i(\mathbf{H}_p - \mathbf{H}_q).\mathbf{r}]$$

where D_{pq} is the element of U^{-1} at intersection of row p and column q. Generally the inequality criterion defined above is not sufficiently restrictive to allow a straightforward determination of the structure, even for moderately complex structures. Nevertheless, for a special case discussed below in (b) the allowed regions are strictly the atomic positions.

(b) *Special case $m = N + 1$*

Let us consider the determinant of (3) with $m = N + 1$; following the same argument as Lajzerowicz & Lajzerowicz (1966), it can be shown that $^s\mathcal{D}_{N+1}$ is the product of two Slater determinants given in Table 3 (A_{N+1} and $\tilde{B}_{N+1}$). Factorizing the term i in the last row of each of them and calculating the product of the resulting matrix (A'_{N+1}, $\tilde{B}'_{N+1}$), it is clear that $^s\mathcal{D}_{N+1}$ may be expressed as:

$$^s\mathcal{D}_{N+1} = (i)^2 D'_{N+1} = -D'_{N+1} \leq 0$$

where D'_{N+1} is the K.H. determinant of basis vectors $\mathbf{H}_1,...,\mathbf{H}_{N+1}$, product of A'_{N+1} and B'_{N+1}, associated with the $N + 1$ ('positive') structure atoms at positions $\mathbf{r}_1,...,\mathbf{r}_N,\mathbf{r}$. $^s\mathcal{D}_{N+1}$ is non-positive, and it is null only if $\mathbf{r}$ coincides with $\mathbf{r}_1$, $\mathbf{r}_2$, or $...\mathbf{r}_N$.

Case of determinant depending solely on interatomic vectors. Moreover, with the following special choice of basic vectors: $\mathbf{H}_1 = \mathbf{H}$, $\mathbf{H}_m = m\mathbf{H}$, $\mathbf{H}_{N+1} = (N+1)\mathbf{H}$, it has been shown (Knossow, 1975) that the K.H. determinant D_N is given by

$$D_N \propto \prod_{i < j \leq N} \sin^2 \pi \mathbf{H}.(\mathbf{r}_i - \mathbf{r}_j). \tag{4}$$

In the present case, Det A_{N+1} and Det B_{N+1}, which are well known as van der Monde determinants may be easily calculated:

$$\text{Det } A_{N+1} = i \prod_{0 < i < j \leq N} [\exp(2\pi i \mathbf{H}.\mathbf{r}_i) - \exp(2\pi i \mathbf{H}.\mathbf{r}_j)]$$

$$\text{Det } B_{N+1} = i \prod_{0 < i < j \leq N} [\exp(-2\pi i \mathbf{H}.\mathbf{r}_i) - \exp(-2\pi i \mathbf{H}.\mathbf{r}_j)]$$

and

$$^s\mathscr{D}_{N+1} = \text{Det } A_{N+1} \text{ Det } B_{N+1}$$

$$= -\prod_{0 < i < j \le N} [2 - 2\cos^2 \pi \mathbf{H}.(\mathbf{r}_i - \mathbf{r}_j)]$$

$$\propto -\prod_{0 < i < j \le N} \sin^2 [\pi \mathbf{H}.(\mathbf{r}_i - \mathbf{r}_j)].$$

We retrieve, of course, the preceding results on the sign of $^s\mathscr{D}_{N+1}$ and obtain the form of the maximum reached when $\mathbf{r}$ is in the neighbourhood of an atomic position:

$$^s\mathscr{D}_{N+1} = -k\|\mathbf{r} - \mathbf{r}_i\|^2, \quad k \text{ positive.} \tag{5}$$

(c) Probabilistic restrictions inside the allowed regions

We have seen that, generally, the constraint imposed by the positivity of the 'forbidden regions determinant' is not restrictive enough to determine unambiguously the atomic positions. The probability theory may, however, give a better criterion, the maximal character of $\delta^{(r)}_{m+1}$ (or $^s\mathscr{D}_m$).

We suppose here that the structure consists of N atoms, and that all the elements of D_m are known, in modulus and phase. A determinant $\delta^{(r)}_{m+1}$ may then be built, by edging D_m with a column and row labelled by the reciprocal space vector $\mathbf{L}$, the general element of which is:

$$\mathbf{e}_p = \frac{1}{N} \exp[2\pi i(\mathbf{L} - \mathbf{H}_p).\mathbf{r}].$$

First we write $\delta^{(r)}_{m+1}$ as:

$$\delta^{(r)}_{m+1} = D_m \left\{ \frac{1}{N} - \sum_{p,q} \frac{D_{pq}}{N^2} \exp[2\pi i(\mathbf{H}_p - \mathbf{H}_q).\mathbf{r}] \right\}.$$

Then we make use of the 'maximum determinant rule' (Tsoucaris, 1970) to deduce the maximal character of $\delta^{(r)}_{m+1}$ at a true atomic position. Under the usual assumption of the equipartition of atomic positions and of their statistical independence, a well known result is the maximal character of the following expression for the true value of phases of $E_{\mathbf{L}-\mathbf{H}_p}$ ($p = 1, m$), once the elements of D_m are known:

$$\frac{\Delta^{\mathbf{L}}_{m+1}}{D_m} = \tau_{\mathbf{L}} = 1 - \frac{1}{N} \sum_{p,q} E_{\mathbf{L}-\mathbf{H}_p} E_{\mathbf{H}_q - \mathbf{L}} D_{pq}. \tag{6}$$

Here we denote by $\Delta^{\mathbf{L}}_{m+1}$ the determinant defined by Tsoucaris [1970, formula (7)]; this determinant does not depend explicitly on atomic coordinates, as opposed to $\delta^{(r)}_{m+1}$.

Next, $\tau_{\mathbf{L}}$ is expanded:

$$\tau_{\mathbf{L}} = 1 - \frac{1}{N^2} \left\{ \sum_{j=1}^{N} \exp[2\pi i(\mathbf{H}_p - \mathbf{H}_q).\mathbf{r}_j] \right.$$
$$\left. + \sum_{j \ne i} \exp 2\pi i[\mathbf{H}_q.\mathbf{r}_j - \mathbf{H}_p.\mathbf{r}_i + \mathbf{L}.(\mathbf{r}_i - \mathbf{r}_j)] \right\} D_{pq}.$$

Averaging over $\mathbf{L}$ in reciprocal space, and assuming the independence of the atomic positions, cross-terms vanish and we obtain:

$$\langle \tau_{\mathbf{L}} \rangle_{\mathbf{L}} = 1 - \frac{1}{N^2} \sum_{pq} [\sum_j \exp 2\pi i(\mathbf{H}_q - \mathbf{H}_p).\mathbf{r}_j] D_{pq}$$
$$= 1 - \text{Tr } U^{-1}C, \tag{7}$$

C being the matrix of atomic contributions. After inverting the double sum, the atomic contributions appear:

$$\langle \tau_{\mathbf{L}} \rangle_{\mathbf{L}} = \sum_j \frac{\delta^{(\mathbf{r}_j)}_{m+1}}{D_m} = \sum_j \tau(\mathbf{r}_j). \tag{7a}$$

The terms of the second member having statistical independence and being positive, their sum is maximal only if any one of them is maximal. We so reach the required result: *at any coincidence of* $\mathbf{r}$ *with an atomic position*,

$$\tau(\mathbf{r}) = \frac{\delta^{(\mathbf{r})}_{m+1}}{D_m} \tag{7b}$$

reaches a (relative) maximum. This result provides us with a new way of determining atomic position using $\delta^{(r)}_{m+1}$ as a 'most probable region determinant'; however it needs the knowledge of some 100–200 phases for a current structure; Gram determinant theory, developed in the next paragraph, may give a way of overcoming this drawback.

III. New Gram determinants

Instead of the Hilbert space defined by (2), we will set up determinants in the tensorial square of it. We introduce now V_p vectors associated with each reciprocal lattice vector $\mathbf{H}_p$:

$$V_p = \sum_{i=1}^{N} \sum_{j=1}^{N} [g_i g_j \exp 2\pi i \mathbf{H}_p.(\mathbf{r}_i - \mathbf{r}_j)] \mathbf{e}_{ij}. \tag{8}$$

It is clear that this vector set may be obtained by the same derivation as in § I(a), starting from the Patterson function instead of the electron density function. The scalar products of these vectors are the squared moduli of unitary structure factors:

Table 3. A_{N+1} and B_{N+1}, Slater determinants of order $N + 1$, N being the number of atoms in a unit cell. Their product is a K.H. forbidden region determinant

Notation: $e_{i,j} = N^{-1/2} \exp(2i\pi \mathbf{h}_i.\mathbf{r}_j)$. a corresponds to the subtracted atom, at position $\mathbf{r}$.

$$A_{N+1} = \begin{vmatrix} e_{1,1} & e_{1,2} & \cdots\cdots & e_{1,N} & ie_{1,a} \\ e_{2,1} & e_{1,2} & \cdots\cdots & e_{2,N} & ie_{2,a} \\ \vdots & \vdots & & \vdots & \vdots \\ e_{N,1} & e_{N,2} & \cdots\cdots & e_{N,N} & ie_{N,a} \\ e_{N+1,1} & e_{N+1,2} & \cdots & e_{N+1,N} & ie_{N+1,a} \end{vmatrix}$$

$$B_{N+1} = \begin{vmatrix} e^*_{1,1} & e^*_{1,2} & \cdots\cdots & e^*_{1,N} & ie^*_{1,a} \\ e^*_{2,1} & e^*_{2,2} & \cdots\cdots & e^*_{2,N} & ie^*_{2,a} \\ \vdots & \vdots & & \vdots & \vdots \\ e^*_{N,1} & e^*_{N,2} & \cdots\cdots & e^*_{N,N} & ie^*_{N,a} \\ e^*_{N+1,1} & e^*_{N+1,2} & \cdots & e^*_{N+1,N} & ie^*_{N+1,a} \end{vmatrix}$$

$$(\mathbf{V}_p \cdot \mathbf{V}_q) = \sum_{i=1}^{N} \sum_{j=1}^{N} [g_i^2 g_j^2 \exp 2\pi i(\mathbf{H}_p - \mathbf{H}_q) \cdot (\mathbf{r}_i - \mathbf{r}_j)]$$
$$= |U_{p,q}|^2.$$

Next, we will build up new Gram determinants δ_{m+1} starting from the above vector set, by addition of one or more vectors belonging to the same space. Doing so we will exploit the restriction set up on the elements of the last column and row, because of the known 'inner part' D_m of the new determinant (which is the Gram determinant of the vectors $\mathbf{V}_1, ..., \mathbf{V}_m$). In a different context the same step has been used in the regression equation and related formulae (de Rango, 1969; de Rango, Tsoucaris & Zelwer, 1974). We shall provide evidence that the true values of the parameters governing variations in the last-column elements reach their most probable value when the determinant reaches its maximum.

(a) A determinant involving phases, moduli and the atomic model

To the m vectors of (8) we add the following:

$$\mathbf{V}_{m+1} = \sum_{i=1}^{n} \sum_{j=1}^{n'} \{g_i g_j \exp [2\pi i \mathbf{H}_{m+1} \cdot (\mathbf{r}_i - \mathbf{r}_j)]\} \mathbf{e}_{ij}$$
$$n, n' < N$$

where i and j describe two independent subsets of atoms of the N structure. Let us write the last column's elements of δ_{m+1} for $n' = N$, as:

$$(\mathbf{V}_{m+1} \cdot \mathbf{V}_q) = \sum_{i=1}^{n} \sum_{j=1}^{N} g_i^2 g_j^2 \exp [2\pi i(\mathbf{H}_{m+1} - \mathbf{H}_q)$$
$$\cdot (\mathbf{r}_i - \mathbf{r}_j)]$$

which is written:

$$(\mathbf{V}_{m+1} \cdot \mathbf{V}_q) = \left\{ \sum_{i=1}^{n} g_i^2 \exp [2\pi i(\mathbf{H}_{m+1} - \mathbf{H}_q) \cdot \mathbf{r}_i] \right\}$$

$$\times \left\{ \sum_{j=1}^{N} g_j^2 \exp [-2\pi i(\mathbf{H}_{m+1} - \mathbf{H}_q) \cdot \mathbf{r}_j] \right\}$$

$$(\mathbf{V}_{m+1} \cdot \mathbf{V}_q) = \left\{ \sum_{i=1}^{n} g_i^2 \exp [2\pi i(\mathbf{H}_{m+1} - \mathbf{H}_q) \cdot \mathbf{r}_i] \right\} U_{m+1, q}^*$$

$$(\mathbf{V}_{m+1} \cdot \mathbf{V}_q) = C_{m+1, q} U_{m+1, q}^*, \qquad (9)$$

denoting $C_{m+1, q}$ the contribution of the n atoms subset to the structure factor $U_{m+1, q}$.

This formula shows that, whereas the principal minor of order m of the determinant involves only moduli, last column and row elements involve both phases (in $U_{m+1, q}$) and an atomic model (in $C_{m+1, q}$). This formalism allows us to select from among the sets of phases and atomic models provided by multi-solution methods the most probable ones, i.e. those which maximize the determinant value.

(b) A determinant involving moduli and stereochemistry

We take now a different choice of the last row vector

as follows:

$$\mathbf{V}_{m+1} = \sum_{i=1}^{n} \sum_{j=1}^{n} [g_i g_j \exp(2\pi i \mathbf{H}_{m+1}) \cdot (\mathbf{r}_i - \mathbf{r}_j)] \mathbf{e}_{ij}$$
$$n < N;$$

i and j describe here the same subset of the structure's atoms (a partial model or a fragment), all other vectors $\mathbf{V}_p$ are given by (8).

Let us write the element of the qth column and last row in the obtained Gram determinant:

$$(\mathbf{V}_{m+1} \cdot \mathbf{V}_q) = \sum_{i=1}^{n} \sum_{j=1}^{n} g_i^2 g_j^2 \exp [2\pi i(\mathbf{H}_{m+1} - \mathbf{H}_q)$$
$$\cdot (\mathbf{r}_i - \mathbf{r}_j)] = |C_{m+1, q}|^2. \quad (10)$$

These elements depend only on the interatomic vectors of the subset, they are the square of the moduli of the fragment contribution to the unitary structure factors. For a given stereochemistry of the fragment, they depend only on its orientation and not on its position in the unit cell. The three orientational parameters of this fragment are the only parameters governing the δ_{m+1} $(\theta_1, \theta_2, \theta_3)$ value. We develop in § IV(b) a practical application of this determinant. For the translation case a similar procedure may be developed.

IV. Practical application

Among the various theoretical aspects developed above, two formulae have been explored through numerical calculations: equation (7b) (of § IIc), involving phase information, has been tested on a real structure; equation (10) (of § IIIb), involving only moduli, has been used to develop a structure determination method, ab initio.

(a) Determination of atomic positions via the use of the 'most probable regions determinant'

In this application, most probable regions are determined by subtracting one atom from the unit cell. First calculations have been performed on the centro-symmetric structure of jamine (space group $P\bar{1}$) having 48 atoms in the unit cell (Karle & Karle, 1964) with
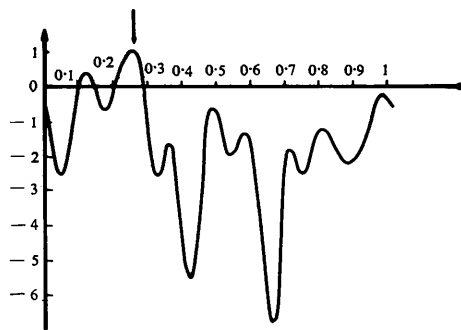

Fig. 1. Variations of $^s\mathcal{D}_m$ (on an arbitrary scale) as a function of the position of the subtracted atom on an axis defined by $x = 0.940$, $z = 0.500$. The arrow indicates the true atomic position.

calculated structure factors; we used the form of the most probable regions determinant denoted $^s\mathcal{D}_m$ in § II(a). We have shown in § II(c) that $\delta^{(r)}_{m+1}$ reaches a maximum at an atomic position, as does $^s\mathcal{D}_m$ which is proportional to it. A one-dimensional plot of $^s\mathcal{D}_m$ clearly shows that the atomic positions are unambiguously determined from the maximal character of $^s\mathcal{D}_m$ (Fig. 1).

Next, another attempt has been performed with the experimental data of a protein structure, that of insulin (space group $R3$), by using a 400-order determinant with observed moduli down to 1·5 Å resolution, and refined phases obtained from the maximal determinant method (de Rango, Mauguen & Tsoucaris, 1975). Fig. 2 shows clearly that the maxima of the $\tau(\mathbf{r})$ function coincide with those of the electron density $\varrho(\mathbf{r})$. Hence this function can be used in a new fast refinement process which is under investigation.

(b) Introduction of the stereochemistry in determining the most probable regions

Performing most probable region calculations with the help of von Eller's formalism [or using the equivalent determinant $\delta^{(r)}_{m+1}$] already needs the knowledge of the phase of 100–200 structure factors. However, valuable information may be obtained from the stereochemistry of a fragment of the structure and from the moduli of the structure factors. Indeed most probable region calculation, using the determinant $\delta_{m+1}(\theta_1,\theta_2,\theta_3)$ obtained in § III(b) (see Table 4), is equivalent to assigning the most probable position of the pseudo-atoms set associated with a stereochemically known fragment, among the whole set of pseudo-atoms in the vector space. The same problem has been studied by using rotation functions (Tollin, 1966; Rossmann, 1972); it is shown in the Appendix that the second-order term of the determinant $\delta_{m+1}(\theta_1,\theta_2,\theta_3)$ expansion is proportional to a rotation function.

Before giving a practical application, we set forward symmetry considerations which permit a reduction of the computing time and make the determination of most probable regions possible for any structure of current importance (40 atoms in the asymmetric unit).
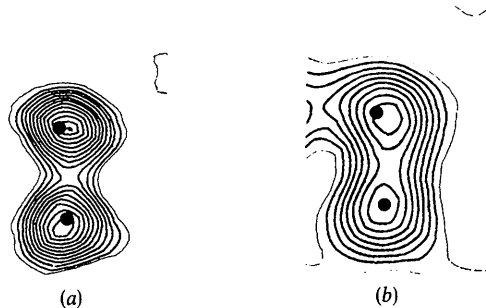
Table 4. $\delta_{m+1}(\theta_1,\theta_2,\theta_3)$, $(m+1)$-order determinant

The last column and row contain the squared contribution to the structure factors of an oriented fragment which is stereochemically known (they depend on $\theta_1,\theta_2,\theta_3$). The inner part contains only observed moduli.

$$\begin{vmatrix} 1 & |U_{1,2}|^2 & \cdots & |U_{1,m}|^2 & |C_1|^2 \\ |U_{2,1}|^2 & 1 & \cdots\cdots & |U_{2,m}|^2 & |C_2|^2 \\ \vdots & \vdots & & \vdots & \vdots \\ |\dot{U}_{m,1}|^2 & |\dot{U}_{m,2}|^2 & \cdots & 1 & |\dot{C}_m|^2 \\ |C_1|^2 & |C_2|^2 & \cdots\cdots & |C_m|^2 & \sum_{i=1}^{n}\sum_{j=1}^{n} g_i^2 g_j^2 \end{vmatrix}$$

We will deal successively with the building of the determinant and the scanning range of orientational parameters.

Recent experience in determinant construction shows that the information they contain becomes practically useful only for an order of $(N/4 - N/10)$, N being the number of atoms in the unit cell. Since the number of atoms in a pseudo-atom structure is roughly the square of that of the associated structure, some mathematical device must be used to reduce the order. Goedkoop's determinants (Goedkoop, 1952; Mauguen, de Rango & Tsoucaris, 1973) provide an appropriate way for doing so: a K.H. determinant may be factorized in s Goedkoop determinants (s being the space-group order of the structure) and this last determinant becomes efficient for an order $(N/4s - N/10s)$. A Goedkoop determinant may be built up from any representation of the structure space group; but we limit ourselves to the totally symmetric one: it can be shown that this representation is the only one that allows us to take into account all special reflexions as basic elements (Knossow, 1975). We report on Fig. 3 the variation of the ratio D_{m+1}/D_m as a function of its order m; it can be noted that the observed ratio's value is always lower than the theoretical mean value $(1 - m/N)$.

Fig. 2. (a) A section of the insulin electron density $\varrho(\mathbf{r})$. (b) The same section of the $\tau(\mathbf{r})$ function. The shaded points indicate the position of the sulphur atoms according to the model structure.
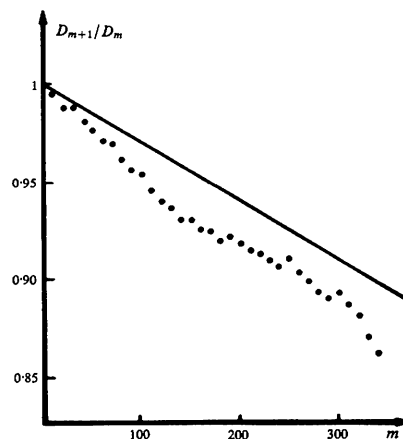

Fig. 3. Variations of the actual value of the ratio (D_{m+1}/D_m) as a function of the determinant's order m. The straight line indicates the theoretical mean value $(1 - m/N)$.

Scanning range of orientational parameters. Cheshire groups

The computing time of most probable regions depends on the order of the determinant involved as well as on the range to be scanned by the parameters. The problem of range determination has been studied by Hirshfeld (1968). The main result obtained by this author is the following: to avoid redundancies and to be exhaustive, the parameters must scan the asymmetric unit of the associated Cheshire group, taking into account also the symmetry group of the oriented fragment (we recall that the Cheshire group of a structure,
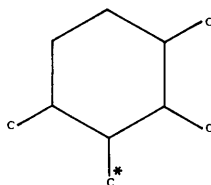

Fig. 4. The stereochemically known fragment of TOT. The three identical fragments which build the molecule are bound through the oxygen and starred carbon atoms.

$\theta_1 = 30°$

θ_3 θ_2	0	15	30	45	60	75
0	700	640	648	404	412	617
15	764	776	635	369	285	273
30	655	292	-31	188	465	538
45	564	91	391	663	700	658
60	545	607	723	646	565	662
75	405	635	620	673	620	686
90	303	749	692	697	716	599
105	762	781	723	766	734	435
120	730	690	738	751	600	535
135	684	655	699	447	576	600
150	719	597	139	264	693	673
165	763	782	489	680	651	597

$\theta_1 = 75°$

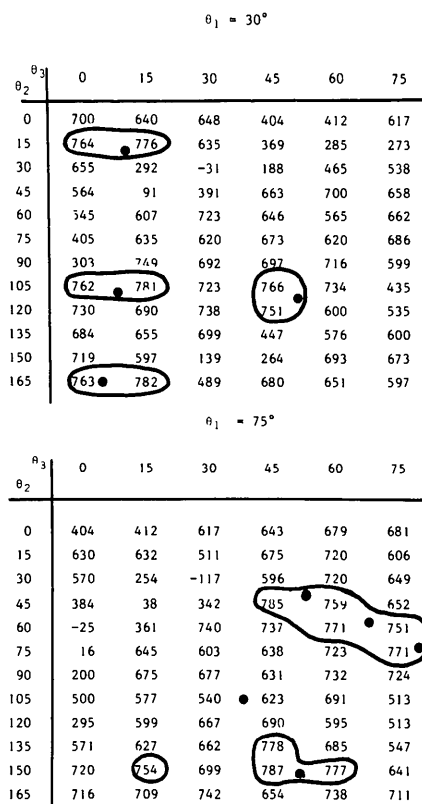| θ_3 θ_2	0	15	30	45	60	75
0	404	412	617	643	679	681
15	630	632	511	675	720	606
30	570	254	-117	596	720	649
45	384	38	342	785	759	652
60	-25	361	740	737	771	751
75	16	645	603	638	723	771
90	200	675	677	631	732	724
105	500	577	540	623	691	513
120	295	599	667	690	595	513
135	571	627	662	778	685	547
150	720	754	699	787	777	641
165	716	709	742	654	738	711

Fig. 5. Rotation search for the orientation of the stereochemically known fragment of TOT. The variations of the determinant are scaled to the interval (−1000, 1000). Dots indicate the true positions of the fragment.

as defined by Hirshfeld, leaves invariant as a whole the equivalent positions and the crystallographic axes).

Application to real structures

Here we limit ourselves to the use of the determinant $\delta_{m+1}(\theta_1,\theta_2,\theta_3)$ (see Table 4) obtained in § III(b). As this determinant involves only the moduli of the structure factors and the stereochemistry of a fragment of the structure, we will compare our results with those of a Patterson search method.

The calculations have been carried out for the structure of tri-o-thymotide, TOT, containing 40 atoms in the asymmetric unit (Brunie & Tsoucaris, 1974). The TOT molecule consists of three identical fragments which are stereochemically known (see Fig. 4). We determined most probable regions in the space of Euler angles of the oriented fragment, using a 340-order Goedkoop determinant. According to the Cheshire group symmetry, the scanning range of the parameters $(\theta_1,\theta_2,\theta_3)$ is for $P2_12_12_1$. $0 \le \theta_1 < \pi/2$, $0 \le \theta_2 < \pi/2$, $0 \le \theta_3 \le \pi$.

Scanning this range, strictly we should find three maxima of the most probable region determinant. But as can be seen on Fig. 4, the pseudo-atom set of a fragment is left nearly invariant in a $k\pi/3$ ($k=1,6$) rotation about the axis normal to the benzene ring. So we must expect 18 peaks in the whole range.

Remembering that D_m is constant, the equi-determinant contour drawn in Fig. 5 indicates the level of 750 when 1000 corresponds to the absolute maximum which can be reached by $G_{m+1}^{(\theta_1,\theta_2,\theta_3)}/G_m$ [G_{m+1} and G_m are the symmetry-adapted determinants associated, respectively, with $\delta_{m+1}^{(r)}$ and D_m defined by (10)]. Three of the twenty highest zones do not correspond to a fragment orientation; however, their associated peaks are among the lowest. On the other hand, one of the true positions is 15° from a maximum. But for these spurious determinations the fragment orientation is fairly well represented by the maxima of the 'most probable region determinant' variations.

In response to a pertinent comment of the referee we compare in Table 5 our results with those of a Patterson search method (Braun, Hornstra & Leenhouts, 1969) (program kindly supplied by J. J. Hornstra). The results of the determinant method seem encouraging; moreover, in difficult cases the choice of a better determinant may improve the contrast. Further improvements under investigation may lead to easier and fairly fast exploitation of this method.

Table 5. Comparison of the results of the determinant method and of the Patterson search method

Notation: P.V. value of the function on a correct position, M.V. mean value of the calculated function. P1: |P.V.−M.V.|/M.V. P2: percentage of points associated with a height above (P.V.−15% M.V.).

	P1	P2
Patterson search method	22%	10%
Determinant method	42%	4%

APPENDIX

First terms of δ_{m+1} expanded in powers of $|U|^2$

We show here that the first term of the δ_{m+1} expansion may be expressed as a rotation function. The term in the δ_{m+1} development depending on the orientation of the unknown fragment can be written:

$$\delta'_{m+1} = -\sum_{p=1}^{m} \sum_{q=1}^{m} (-1)^{p+q} \delta_{pq} |C_{L-H_p}|^2 |C_{L-H_q}|^2, \quad \text{(A1)}$$

δ_{pq} being the cofactor of the D_m element in column q and row p.

The first term in the δ_{pq} expansion in powers of U^2 is written:

$$\delta_{pq}^{(2)} = (-1)^{q+p-1} |U_{H_q - H_p}|^2 \quad \text{(A2)}$$

and we have the first term of the δ'_{m+1} expansion in powers of $|U|^2$:

$$\delta_{m+1}^{(2)} = \sum_{p=1}^{m} \sum_{q=1}^{m} |U_{H_p - H_q}|^2 |C_{L-H_p}|^2 |C_{L-H_q}|^2. \quad \text{(A3)}$$

Let us denote by $P(\mathbf{r})$ the Patterson function of the crystal and by $P_{\text{rot}}(\mathbf{r})$ the Patterson function of the rotated unknown fragment, then

$$P(\mathbf{r}) = \sum_{p=1} |U_{H_p}|^2 \exp 2i\pi H_p \cdot \mathbf{r}$$

$$P_{\text{rot}}^2(\mathbf{r}) \propto \sum_{p'q'} |C_{H_{p'}}|^2 |C_{H_{q'}}|^2 \exp 2i\pi (H_{p'} - H_{q'}) \cdot \mathbf{r}.$$

The integral of $P(\mathbf{r})P_{\text{rot}}^2(\mathbf{r})$ in the unit cell is written

$$I = \int_{UC} P(\mathbf{r})P_{\text{rot}}^2(\mathbf{r})d\tau \propto \int_{UC} (\sum_p |U_{H_p}|^2 \exp 2i\pi H_p \cdot \mathbf{r})$$

$$\times [\sum_{p'q'} |C_{H_{p'}}|^2 |C_{H_{q'}}|^2 \exp 2i\pi (H_{p'} - H_{q'})\mathbf{r}]d\tau$$

$$I \propto \sum_{p,p',q'} |U_{H_p}|^2 |C_{L-H_{p'}}|^2 |C_{L-H_{q'}}|^2$$

$$\times \int_{UC} \exp 2i\pi (H_p + H_{p'} - H_{q'}) \cdot \mathbf{r}d\tau.$$

The integral differs from zero only if $H_p = -(H_{p'} - H_{q'})$. Using the index transformation: $H_{p'} = L - H_{p''}$, $H_{q'} = L - H_{q''}$, the integral becomes

$$\int_{UC} P(\mathbf{r})P_{\text{rot}}^2(\mathbf{r})d\tau$$

$$= \sum_{p'',q''} |U_{H_{p''} - H_{q''}}|^2 |C_{L-H_{p''}}|^2 |C_{L-H_{q''}}|^2. \quad \text{(A4)}$$

Comparison with (A3) shows clearly that the rotation function involving $P(\mathbf{r})$ and $P_{\text{rot}}^2(\mathbf{r})$ is the first term in the expansion of δ_{m+1}.

References

BRAUN, P. B., HORNSTRA, J. J. & LEENHOUTS, J. I. (1969). *Philips Res. Rep.*, **24**, 85.

BRUNIE, S. & TSOUCARIS, G. (1974). *Cryst. Struct. Commun.* **3**, 481–484.

ELLER, G. VON (1955). *Acta Cryst.* **8**, 641–645.

ELLER, G. VON (1962). *Acta Cryst.* **15**, 590–595.

ELLER, G. VON (1964). *C.R. Acad. Sci. Paris*, **258**, 5418–5419.

GOEDKOOP, J. A. (1952). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. PEPINSKY. The Pennsylvania State College.

HIRSHFELD, F. L. (1968). *Acta Cryst.* A **24**, 301–311.

KARLE, I. L. & KARLE, J. (1964). *Acta Cryst.* **17**, 1356–1360.

KITAIGORODSKY, A. I. (1961). *Theory of Crystal Structure Determination.* New York: Consultants Bureau.

KNOSSOW, M. (1975). Thèse de 3ème Cycle, Paris.

LAJZÉROWICZ, J. & LAJZÉROWICZ, J. (1966). *Acta Cryst.* **21**, 8–12.

MAUGUEN, Y., DE RANGO, C. & TSOUCARIS, G. (1973). *Acta Cryst.* A **29**, 574–575.

RANGO C. DE (1969). Thèse, Paris.

RANGO C. DE, MAUGUEN, Y. & TSOUCARIS, G. (1975). *Acta Cryst.* A **31**, 227–233.

RANGO C. DE, TSOUCARIS, G. & ZELWER, C. (1974). *Acta Cryst.* A **30**, 342–353.

ROSSMANN, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon and Breach.

TOLLIN, P. (1966). *Acta Cryst.* **21**, 872–876.

TSOUCARIS, G. (1970). *Acta Cryst.* A **26**, 492–499.