Acta Cryst. (1989). A**45**, 349-353

On the Application of Phase Relationships to Complex Structures. XXVIII. *XMY* as a Random Approach to the Phase Problem

By T. Debaerdemaeker

Sektion für Röntgen- und Elektronenbeugung, Universität, Oberer Eselsburg, D-7900 Ulm, Federal Republic of Germany

and M. M. Woolfson

Department of Physics, University of York, York YO1 5DD, England

(*Received* 28 *April* 1988; *accepted* 3 *January* 1989)

Abstract

In modern direct methods a very powerful multisolution approach involves the refinement by some means of initially random phases. In exploring a number of refinement functions Debaerdemaeker & Woolfson [*Acta Cryst.* (1983), A39, 193-196] found by accident a function of phases, the maximization of which was very effective in obtaining substantially correct phase sets. This has led to the *XMY* method which works well although no completely rational explanation can be offered for how it does so. Examples of its use in the *ab initio* solution of unknown structures are given and tests are described indicating its usefulness as a means of carrying out multisolution fragment development.

Introducing some properties of *XMY*

Debaerdemaeker & Woolfson (1983) explored the efficiency of a number of functions whose maximization or minimization might be expected to refine initially random phases. The refinement process which they found to be most effective is the parameter-shift method described by Bhuiya & Stanley (1963). It has the advantage of being able to jump away from a local extremum and find a better one while most other techniques become locked into the first extremum they locate.

In their investigations Debaerdemaeker & Woolfson (1983) found by accident a very effective process, called *XMY* (*X* minus *Y*), involving the function

$$\psi = \sum_{\mathbf{h}} [X(\mathbf{h}) - Y(\mathbf{h})] \qquad (1)$$

where

$$X(\mathbf{h}) = \sum_{\mathbf{h}'} |E(\mathbf{h})E(\mathbf{h}')E(\mathbf{h}-\mathbf{h}')| \times \cos[\varphi(\mathbf{h}) - \varphi(\mathbf{h}') - \varphi(\mathbf{h}-\mathbf{h}')]$$

and

$$Y(\mathbf{h}) = \sum_{\mathbf{h}'} |E(\mathbf{h})E(\mathbf{h}')E(\mathbf{h}-\mathbf{h}')| \times \sin[\varphi(\mathbf{h}) - \varphi(\mathbf{h}') - \varphi(\mathbf{h}-\mathbf{h}')].$$

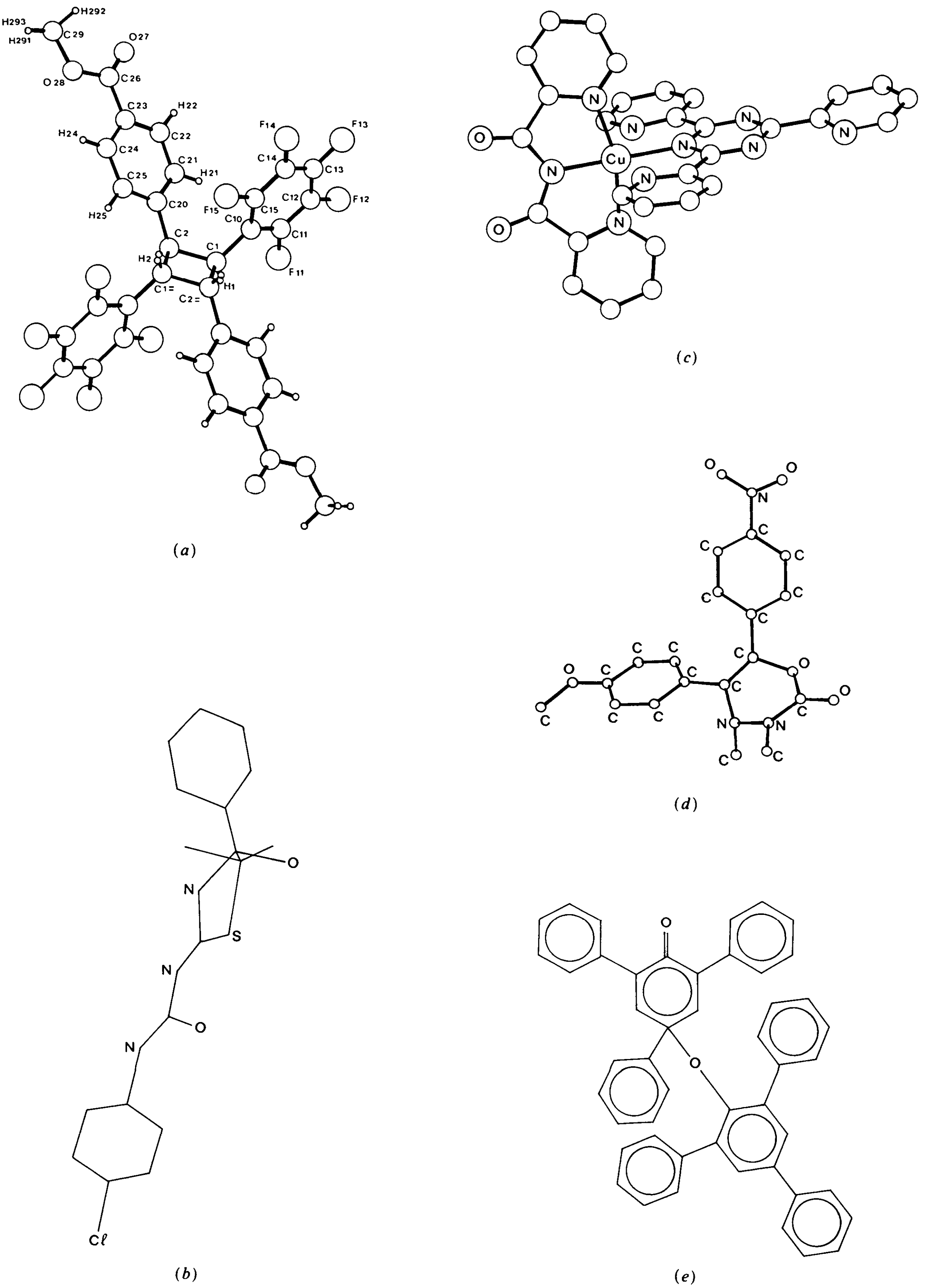
While a correct set of phases would be expected to make $X(\mathbf{h})$ large and positive and $Y(\mathbf{h})$ small in magnitude, logic dictates that $|Y(\mathbf{h})|$ should appear in (1) rather than $Y(\mathbf{h})$. In fact making $Y(\mathbf{h})$ equal to its expectation value, zero, was the way in which Hauptman & Karle (1956) first derived the tangent formula and Debaerdemaeker, Tate & Woolfson (1985) showed that the tangent formula followed also from the maximization of $\sum X(\mathbf{h})$. It seems therefore that the two terms in ψ are both driving towards satisfying the tangent formula. However, there is the nagging doubt that perhaps one could maximize ψ by, for example, making $\sum X(\mathbf{h})$ positive but of moderate magnitude while making $\sum Y(\mathbf{h})$ large and negative - which would be well removed from any realistic values of phase.

Another source of concern is that even for a substantially correct set of phases for which ψ was maximized, then for the enantiomorph structure, which would reverse the values of all the phases, the value of ψ would change. This means that the quantity ψ is not enantiomorph independent and would have two different values for sets of phases of equivalent validity. As far as we know all previous techniques of refinement have the property that the refinement pathway is equivalent, but antisymmetrically related to the original one, if the enantiomorph phases are employed; this seems not to be true for $X - Y$.

We can satisfy ourselves that the condition of having a moderate sized $\sum X(\mathbf{h})$ and large negative $\sum Y(\mathbf{h})$ will be very unlikely. In the maximization of ψ each three-phase invariant appears three times - in $X(\mathbf{h})$ as

$$\varphi(\mathbf{h}) - \varphi(\mathbf{h}') - \varphi(\mathbf{h}-\mathbf{h}')$$

0108-7673/89/060349-05$03.00



(a) (b) (c) (d) (e)

Fig. 1. Structural diagrams for: (a) GUENPH6; (b) G5076; $C_{18}H_{18}ClN_3O_2S$; (c) $[Cu(C_{30}H_{20}N_9O_2)]$, the portion CF_3SO_3 is not shown; (d) $C_{18}H_{17}N_3O_5$; (e) $C_{48}H_{34}O_2$.

Table 1. *A non-centrosymmetric solution of the structure GUENPH6*

The figures of merit ABSFOM, ψ_0, RESID and CFOM are the ones used in the *SAYTAN* (*MULTAN*) system. Solutions marked * show the complete structure while that marked ** shows a considerable fragment.

Set	ψ ($\times10^5$)	$\sum X(\mathbf{h})$ ($\times10^5$)	$\sum Y(\mathbf{h})$ ($\times10^4$)	ABSFOM	ψ_0	RESID	CFOM
1	0·902	0·877	−0·257	0·4804	5·984	45·33	0·823
2	0·586	0·834	−0·226	0·4434	4·123	48·69	1·057
3	0·926	0·903	−0·238	0·4802	4·948	47·21	0·976
4	0·889	0·863	−0·252	0·4027	4·519	51·02	0·891
5	0·965	0·945	−0·207	1·0015	1·223	19·10	2·820**
6	1·01	0·988	−0·224	0·9194	1·203	17·18	2·771*
7	1·05	1·02	−0·272	0·4448	6·432	49·80	0·609
8	0·802	0·779	−0·232	0·3574	5·334	52·53	0·659
9	1·01	0·987	−0·225	0·2435	4·369	60·10	0·555
10	0·899	0·878	−0·213	0·3992	5·545	51·52	0·687
11	0·931	0·908	−0·230	0·4334	5·346	48·54	0·823
12	0·917	0·893	−0·234	0·5736	5·139	41·48	1·162
13	1·01	0·985	−0·219	0·1836	6·128	65·10	0·061
14	0·826	0·807	−0·190	0·3944	6·459	50·40	0·536
15	0·840	0·816	−0·240	0·4222	5·900	49·12	0·696
16	0·958	0·932	−0·254	0·3908	6·408	50·94	0·530
17	0·952	0·931	−0·206	0·3794	4·105	51·73	0·927
18	0·868	0·845	−0·225	0·4050	5·355	49·60	0·768
19	0·980	0·957	−0·228	0·3754	3·822	49·93	1·011
20	1·01	0·985	−0·234	1·0820	1·043	16·29	3·000*

Table 2. *A centrosymmetric solution for the structure GUENPH6*

The set marked * shows the complete structure.

Set	$\sum X(\mathbf{h})=\psi$ ($\times10^5$)	ABSFOM	ψ_0	RESID	CFOM
1	0·673	0·6257	3·869	36·62	0·963
2	0·559	0·4760	5·532	45·50	0·055
3	0·778	0·7642	2·794	30·36	1·632
4	0·609	0·5401	4·404	39·80	0·601
5	0·559	0·4752	2·837	43·17	0·734
6	0·798	0·7922	2·558	27·99	1·810
7	0·723	0·6909	2·469	31·84	1·540
8	0·591	0·5169	4·367	42·79	0·471
9	0·636	0·5768	3·370	41·21	0·841
10	0·534	0·4407	4·257	45·49	0·284
11	0·630	0·5685	4·270	39·93	0·671
12	0·646	0·5898	4·205	37·79	0·792
13	0·751	0·7284	2·564	31·01	1·606
14	1·02	1·0820	1·043	16·29	3·000*
15	0·735	0·7083	2·802	32·16	1·480
16	0·799	0·7923	2·630	27·90	1·797
17	0·643	0·5860	3·579	39·36	0·872
18	0·749	0·7267	2·599	30·81	=Set 13
19	0·670	0·6216	3·578	35·57	1·057
20	0·851	0·8610	4·057	31·39	1·467

and in $X(\mathbf{h}')$ and $X(\mathbf{h}-\mathbf{h}')$ as

$$\varphi(\mathbf{h}')-\varphi(\mathbf{h})+\varphi(\mathbf{h}-\mathbf{h}')$$

and

$$\varphi(\mathbf{h}-\mathbf{h}')-\varphi(\mathbf{h})+\varphi(\mathbf{h}').$$

If the value of the first form of the three-phase invariant is s then the other two both have the value $-s$. Depending on the way in which the program handles $\sum_2$ relationships the three manifestations of the three-phase invariant could have two of one sign and one of the other or they could all be of the same sign. In the first case the component of ψ containing the relationship will be

$$C_1=3\cos(s)\pm\sin(s) \qquad (2)$$

while the second case gives

$$C_2=3\cos(s)-3\sin(s). \qquad (3)$$

It is straightforward to show that $C_1=10^{1/2}\cos(s\pm18°)$ and that $C_2=18^{1/2}\cos(s-45°)$ so that maximizing ψ is tending to drive the three-phase invariant towards the values ±18 or +45° respectively. The normal tangent formula tends to drive towards the value zero but, even so, we can see that it is likely that maximization of ψ will tend to make $\sum X(\mathbf{h})$ large and positive.

What actually happens in the refinement process is that the value of $\sum Y(\mathbf{h})$ is always negative and much smaller in magnitude than $\sum X(\mathbf{h})$. This is illustrated in Table 1 for a recently solved structure code-named GUENPH6, $(C_{16}H_8F_5O_2)_2$, $P2_1/n$, $a=20{\cdot}859$, $b=7{\cdot}127$, $c=9{\cdot}287$ Å, $\beta=94{\cdot}91°$, $Z=4$ (Fig. 1*a*; asymmetric unit is half the molecule). This centrosymmetric structure was treated as though it was non-centrosymmetric and the phases were taken as the nearest acceptable value at the end of the refinement process. In the 20 trials shown in Table 1 there were three substantially correct solutions, being those with the highest *SAYTAN* (successor to *MULTAN*) combined figures of merit. It will also be noted that the values of $\sum Y(\mathbf{h})$ are all negative and about 2% of the values of $\sum X(\mathbf{h})$ in magnitude.

While GUENPH6 is under discussion it is interesting to compare the results in Table 1 with those in Table 2 where the structure is treated as centrosymmetric. The initially random phases are chosen from the alternatives 0 and 180° and then refined with a parameter-shift step of 180°. In this case there is only one correct solution in 20 trials. Our experience with *XMY* seems to show that it is always better to treat centrosymmetric structures as non-centrosymmetric.

The situation, which is abundantly confirmed by experiment, is that maximizing ψ is far more effective in refining phases than is the normal tangent formula or the apparently more logical formula obtained by substituting $\sum |Y(\mathbf{h})|$ for $\sum Y(\mathbf{h})$ in (1). While *XMY* is, in general, not quite as powerful as the Sayre-equation tangent formula (Debaerdemaeker, Tate & Woolfson, 1985) it is well worth having as a back-up method to try in the event that others fail.

Structures recently solved with *XMY*

We now give examples of some unknown, and sometimes difficult, structures which have all recently been solved with *XMY*.

(1) $C_{18}H_{18}ClN_3O_2S$ (code name G5076), $P1$, $Z=6$ (Fig. 1*b*). $a=17\cdot905$, $b=15\cdot940$, $c=10\cdot021$ Å; $\alpha=94\cdot13$, $\beta=104\cdot43$, $\gamma=102\cdot22°$. This structure contains 150 independent non-hydrogen atoms. Twenty sets of initially random phases were refined. That with the highest figure of merit showed almost all of the six molecules in the asymmetric unit. In all the molecules the molecular chain was complete and atoms were only missing from the terminal phenyl groups. Attempts to solve this structure with *MULTAN* and *SAYTAN* were unsuccessful.

(2) $[Cu(C_{30}H_{20}N_9O_2)].CF_3SO_3$, $P\bar{1}$, $Z=2$ (Fig. 1*c*). $a=9\cdot177$, $b=10\cdot735$, $c=18\cdot231$ Å; $\alpha=102\cdot65$, $\beta=106\cdot6$, $\gamma=93\cdot89°$. Again twenty sets of phases were refined. The one with the highest figure of merit showed not only the Cu atom but also eight of the nearest neighbours. This structure could also readily be solved with both *MULTAN* and *SAYTAN*.

(3) $C_{18}H_{17}N_3O_5$, $C2/c$, $Z=8$ (Fig. 1*d*). $a=16\cdot537$, $b=17\cdot109$, $c=12\cdot074$ Å, $\beta=95\cdot48°$. Only five sets of phases were refined. The one with the best figure of merit showed the complete structure. Again this is a structure which could be solved with both *MULTAN* and *SAYTAN*.

(4) $C_{48}H_{34}O_2$, $P2_12_12_1$, $Z=4$ (Fig. 1*e*). This structure turned out to be rather difficult and 100 sets of phases were examined. However, the set with the highest figure of merit showed the complete structure. Unsuccessful attempts to solve this structure have been made in many laboratories using a variety of direct and non-direct methods. Present indications are that *XMY* is the only readily available method capable of solving this structure.

All these structures will be described fully in due course.

Partial structure development with *XMY*

When direct methods generate a structural fragment then objective procedures can be employed to try to derive the whole structure from the fragment - the first such procedure being that suggested by Karle (1968). Yao Jia-xing (1983) introduced the concept of multisolution fragment development based on the *RANTAN* procedure he had developed. We have similarly used *XMY* to develop fragments. The stages in the process are:

(i) Structure factors, F_c, are calculated for the fragment.

(ii) Where $|F_c|$ is greater than some fraction of $|F_o|$ the phase estimate, φ_c, is accepted as a reliable one.

(iii) Non-reliable phases are partially randomized by adding $72x°$ to φ_c, where x is chosen randomly from a uniform distribution between -1 and 1.

(iv) The phases are refined by *XMY* with the 'reliable' phases fixed until the last few cycles of refinement.

(v) The process is repeated from (iii) to give several initial sets of phases and a multisolution approach.

The results of tests with several fairly difficult known structures are now given.

(1) MUNICH1 (Szeimies-Seebach. Harnisch, Szeimies, Van Meerssche, Germain & Declercq, 1978), $C_{20}H_{16}$, $C2$, $Z=8$. A fragment of six randomly chosen correct atomic positions was chosen as the fragment (13% of the structure). Of the twenty phase sets generated five showed the complete structure.

(2) Factor S (Declercq, Germain, Van Meerssche, Hull & Irwin, 1978), $C_{43}H_{49}N_7O_{10}$, $P2_12_12_1$, $Z=4$. A fragment of seven connected atoms was chosen (10% of the structure). One set out of the fifty sets of phases showed the complete structure.

(3) PtKOW1 (Debaerdemaeker, Berhalter, Weisemann & Brune, 1987), $[Pt(C_{11}H_9)_2(C_{18}H_{15}P)_2].CH_2Cl_2$, $P2_1/a$, $Z=4$. Only Pt was taken as the input atom. Out of 20 trial sets four showed a large part of the structure (48 atoms).

Concluding remarks

Although it is not entirely satisfactory that the power of *XMY* cannot be explained in a totally rational way, it cannot be denied that it is an extremely effective way of solving crystal structures - more so than for many methods that *can* be rationalized. A referee of this paper made the comment '... the authors are preventing the method from satisfying all triplets exactly and thus may well be holding the enantiomorph in a better way' - which is as good an explanation as is available at present. *XMY* is now an integral part of *SAYTAN*87, the successor to *MULTAN*, which is the current package we are distributing.

However, the fragment-development component is not available with *SAYTAN*87 as yet. The test results we report have been carried out with a program based on *MULTAN*80 which stores information, including data, in a way different from that in *SAYTAN*87.

We wish to record our gratitude to the European Economic Community, the Science and Engineering Research Council and the Deutsche Forschungsgemeinschaft for their support of this research. We are also grateful to a referee for his helpful comment which we have quoted and other useful suggestions on presentation.

References

BHUIYA, A. K. & STANLEY, E. (1963). *Acta Cryst.* **16**, 981-984.
DEBAERDEMAEKER, T., BERHALTER, K., WEISEMANN, C. & BRUNE, H. A. (1987). *Acta Cryst.* **C43**, 1253-1255.
DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* **A41**, 286-290.
DEBAERDEMAEKER, T. & WOOLFSON, M. M. (1983). *Acta Cryst.* **A39**, 193-196.
DECLERCQ, J.-P., GERMAIN, G., VAN MEERSSCHE, M., HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* **B34**, 3644-3648.
HAUPTMAN, H. & KARLE, J. (1956). *Acta Cryst.* **9**, 45-55.
KARLE, J. (1968). *Acta Cryst.* **B24**, 182-186.
SZEIMIES-SEEBACH, U., HARNISCH, J., SZEIMIES, G., VAN MEERSSCHE, M., GERMAIN, G. & DERCLERCQ, J.-P. (1978). *Angew. Chem. Int. Ed. Engl.* **17**, 848-850.
YAO JIA-XING (1983). *Acta Cryst.* **A39**, 35-37.

Acta Cryst. (1989). **A45**, 353-357

Equations for Diffuse Scattering from Disordered Alloys with H.C.P. Structure

BY RITA KHANNA

Materials Science Division, Indira Gandhi Centre for Atomic Research, Kalpakkam 603 102, *Tamil Nadu, India*

(*Received* 5 *December* 1984; *accepted* 3 *January* 1989)

Abstract

General equations are presented for the diffuse scattering due to local atomic arrangements and displacements in disordered alloys having h.c.p. structure. The scattering due to static and dynamic displacements is treated separately. The calculations show that the second-order terms in displacements are sufficient for observing the direct effect of temperature factors, common to all contributions of diffuse scattering. A new data analysis scheme, using asymmetry of diffuse scattering around superlattice reflections, is presented for a complete separation of various components.

Introduction

Since the early experiments of Wilchinsky (1944) and Cowley (1950), most quantitative studies of diffuse X-ray or thermal neutron scattering to reveal local atomic arrangements in disordered alloys have been carried out on systems having f.c.c. structure. The techniques for interpreting the diffraction pattern, data and error analysis are still evolving; see Borie & Sparks (1971), Gragg, Hayakawa & Cohen (1973), Hayakawa & Cohen (1975), Tibballs (1975), Khanna (1984) for the latest procedures. Although there are a large number of alloy phases which possess h.c.p. structure at high temperatures and which undergo ordering on cooling, general equations for diffuse scattering from such alloys have not yet been presented. Development of these equations and the separation procedures will be the principal subject of this paper.

After a brief review of the diffraction theory for disordered alloys, general equations are derived for the diffuse scattering due to short-range order and displacements (both static and dynamic) from h.c.p. alloys. Dynamic displacements due to thermal vibrations are treated separately to account properly for their effect on other components of diffuse scattering. Procedures for data analysis and separation of various diffuse scattering components are also discussed.

1. Diffraction theory

From kinematic theory, the total scattered intensity from a disordered binary alloy can be written as

$$I_{\rm tot}=\sum_{\substack{l,l'\\k,k'}} f_{lk}f_{l'k'}\exp\left[i\mathbf{Q}\,.\,(\mathbf{r}_{lk}-\mathbf{r}_{l'k'})\right]. \tag{1}$$

f_{lk} represents the atomic form factor of an atom located at position $\mathbf{r}_{lk}$. l and k are lattice and sublattice indices respectively. $\mathbf{Q}$ is the diffraction vector. Let x^i_k be the sublattice fraction of atom i on sublattice k and let $p^{ij}_{kk'}$ represent the conditional probability of finding a j-type atom on the k'th sublattice of the l'th lattice point if there is an i-type atom on the kth sublattice of the lth lattice point. Equation (1) may then be written as (Hayakawa & Cohen, 1975)

$$I_{\rm tot}=\sum_{\substack{l,l'\\k,k'\\i,j}} x^i_k f_{lk}f_{l'k'}P^{ij}_{kk'}\exp\left[i\mathbf{Q}\,.\,(\mathbf{r}_{lk}-\mathbf{r}_{l'k'})\right]. \tag{2}$$

0108-7673/89/060353-05$03.00