

RESEARCH PAPERS

Acta Cryst. (1994). **A50**, 287–292

Automated Structure Analysis in Electron Crystallography: Phase Determination with the Tangent Formula and Least-Squares Refinement

BY DOUGLAS L. DORSET AND MARY P. MCCOURT

Electron Diffraction Department, Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203-1196, USA

(Received 19 January 1993; accepted 31 August 1993)

Abstract

The tangent formula (computer program *QTAN*) is used as an automated phasing technique to solve the crystal structure of 2,5-piperazinedione from published electron-diffraction intensity data. Unlike the case of the thiourea polymorphs studied previously, the correct phase set does not correspond to the lowest value of NQUEST, so that four potential maps must be calculated before a chemically recognizable structure is seen. Even though dynamical scattering presents some difficulty in identifying the correct structure by changing some low-magnitude $|E_h|$ 'cross terms', hence affecting the NQUEST figure of merit, it is still possible to refine the atomic coordinates found on the potential map by full-matrix least squares, leading to a structure similar to the earlier X-ray determination.

Introduction

In recent years, it has been shown that experimental electron-diffraction intensity data from organic microcrystals can be used for *ab initio* quantitative crystal structure determinations, despite the long-standing skepticism of some members of the crystallographic community (*e.g.* Lipson & Cochran, 1966). If conditions are favorable for the collection of diffraction intensities that are adequately close to the Fourier transform of the kinematic Patterson function, predicated, for example, by crystal growth and orientation and by selection of an adequately high electron-accelerating voltage, then there is no reason why these experimental data cannot be used for a direct analysis. This has been demonstrated for a number of materials, including small molecules (Dorset, 1991*a,b*), linear polymers (Dorset, 1991*c*) and polymethylene chain compounds such as *n*-paraffins (Dorset & Zemlin, 1990) and phospholipids (Dorset, Beckmann & Zemlin, 1990). In some of these studies, based on direct phase determination,

earlier data sets taken at rather low voltage (*e.g.* 40–60 kV) were also found to be useful, thus independently justifying analyses that had previously relied on contemporary X-ray studies to provide a phasing model. More recently (Dorset, 1992*a*), intensity data from some layer silicates have been found to lead to the same atomic arrangements predicted earlier by model building or interpretation of Patterson maps (Zvyagin, 1967).

Most of the direct phase determinations in electron crystallography reported so far have been based on the evaluation of individual structure invariants (Hauptman, 1972). Hence, if

$$\psi = \varphi_{h_1} + \varphi_{h_2} + \varphi_{h_3} + \dots,$$

where the Miller indices

$$\sum_i h_i \equiv 0$$

represent an invariant phase sum, then one could predict the probability of $\psi = 0, \pi$ depending upon the value of some magnitude *A* or *B* calculated from the normalized values of the structure-factor magnitudes:

$$|E_{h_1}|, |E_{h_2}|, |E_{h_3}|, \dots$$

Therefore, the ψ_i represent simultaneous equations that can be ranked from the most probable to the least probable. If one can also define the origin with a small number of reflections φ_k , having appropriate index parity, then the ψ_i are solved for enough new phase values to permit an interpretable potential map to be computed.

Although its efficacy has been demonstrated for a large number of structures so far, the evaluation of individual structure invariants, as described, might be criticized for possibly including an unconscious biasing of the outcome, particularly if the crystal structure is already known. Of course, one seeks to avoid this by using only triple or quartet invariants above a suitable probability threshold (*A* value for

Σ_2 triples; B value for quartets) and by looking for signs of internally discrepant phase assignments that may affect the outcome of the analysis. One way to avoid the occurrence of an unconscious bias is to employ an automated phasing procedure – based, for instance, on the well known tangent formula (Karle & Hauptman, 1956) – where likely solutions are identified by an appropriate figure of merit. This option has already been discussed for two thermotropic polymorphs of thiourea (Dorset, 1992*b*) in which the structure analysis *via QTAN* (Langs & DeTitta, 1975) progressed as it would in X-ray crystallography. On the other hand, there may be instances where the figures of merit used to identify the correct structure in the multiple solutions might break down, *e.g.* if dynamical scattering perturbs the observed intensities. If this is the case, can a meaningful structure be identified?

There is another aspect of electron crystallography that also needs to be explored. Suppose that, as copiously demonstrated in earlier work, a structure model can be found by direct phasing. Can the next step of structure refinement, obvious to X-ray crystallographers, be realized? Aside from constrained linked-atom techniques, where the valence parameters (*i.e.* bond distances and angles) are constrained *a priori* to idealized values (*e.g.* Brisse, 1989), there is almost no experience with least-squares refinement in electron crystallography. Most often, Fourier refinement techniques have been used in early (see the review by Vainshtein, Zvyagin & Avilov, 1992) and recent (Dorset, Tivol & Turner, 1991, 1992) work. Thus, if enough data are measured for each unique atom to be refined, can least-squares refinement, in which atomic positions can be incremented, be employed to reach a chemically reasonable molecular geometry?

In this paper, the direct phase determination for diketopiperazine with the tangent formula is reported. The data set is chosen deliberately, since the documented perturbation of diffraction intensities by dynamical scattering will indicate where the procedure is most likely to encounter difficulties. Related to this, it is found that, under appropriate conditions, the initial structure can be refined by full-matrix least-squares techniques to a molecular geometry quite like the original X-ray crystal structure, although modifications of the usual procedure are necessary to constrain the refinement to a local, rather than a global, minimum.

Data and methods

Three-dimensional electron-diffraction intensity data (289 observed reflections for 317 hkl terms given in the original paper) from diketopiperazine (2,5-piperazinedione) $C_4H_6N_2O_2$ were collected from

oblique texture patterns by Vainshtein (1955), who then corrected the observed structure-factor magnitudes $|F_o|$ for a phenomenological Lorentz factor. The monoclinic unit cell has space-group symmetry $P2_1/a$ with unit-cell constants $a = 5.20$, $b = 11.45$, $c = 3.97$ Å, $\beta = 81.9^\circ$. Normalized structure factors $|E_h|$ were calculated from $|F_h^{obs}|$ by

$$|E_h| = |F_h^{obs}|^2 / \varepsilon \sum_i f_i'^2,$$

where the f_i' are atomic form factors (Doyle & Turner, 1968) corrected for thermal motion and ε is a multiplicity factor to correct zones with systematic absences. As reported earlier (Dorset, 1991*b*), a Wilson (1942) plot of the experimental intensities indicates that the overall isotropic temperature factor $B_{iso} = 0.0$ Å². [This is an often-observed indicator of dynamical scattering (see *e.g.* Vainshtein & Lobachev, 1956), even though it says nothing about specific changes to individual reflections.] When $|E_h|$ values were calculated, their distributions were shown (Dorset, 1991*b*) to correspond to those expected for a centrosymmetric unit cell (Karle, Dragonette & Brenner, 1965). A control experiment was also carried out with calculated kinematical structure factors based on Vainshtein's (1955) atomic positions for the molecule, including H atoms. The Miller indices of 317 unique experimental data were used for calculation of $|E_h|$.

Instead of evaluating phase invariants singly, one considers all vector contributions to a phase such that

$$\varphi_h = \langle \varphi_k + \varphi_{h-k} \rangle_k,$$

over all \mathbf{k} , in the data set. After origin definition and the location of other phases for a basis set (*e.g.* from Σ_1 -triple estimates), the tangent formula of Karle & Hauptman (1956) is the usual multisolution procedure used to find new phase values. The reliability of φ_h depends upon its variance $V(\varphi_h)$. This is directly related to the magnitude of $\alpha_{h(est)}$ used in the program *QTAN*, as defined by Langs & DeTitta (1975). The figure of merit most often used with *QTAN* is NQUEST, as stated by DeTitta, Edmonds, Langs & Hauptman (1975). This relies on the estimate of negative-quartet invariants. Hence, the most negative values of NQUEST are usually associated with a correct structure determination.

In practice, one defines the unit-cell origin and accepts a few Σ_1 triples predicted with high probability to form the known basis phase set. A number of other unknown phases p that interact with numerous other reflections (determined by generating Σ_2 triples) are assigned algebraic values a_i that can be permuted 0, π (for a centrosymmetric structure). A hierarchy of reflections that can be assigned phase values from phases preceding them is then estab-

lished. With the acceptance of known phase values with a higher weight than trial unknown phase values from one permuted set of a_i , the tangent formula is then used to generate values for unknown reflections, accepting values above a threshold value $\alpha_h > q$, with several cycles used to find all the linked reflections. The unknowns are permuted again and the process is repeated, calculating the figure of merit for each of 2^p trial phase sets.

Phase determination

In the original direct phase analysis of diketopiperazine, based on the evaluation of the triple and quartet structure invariants, 196 reflections were used with $|E_h| \geq 0.5$. For this analysis with the tangent formula, we accept 287 reflections where $|E_h| \geq 0.09$. Three reflections are used to define the origin [consistent with Vainshtein's (1955) choice]:

$$\varphi_{\bar{3}12} = 0, \quad \varphi_{111} = \pi, \quad \varphi_{3,10,1} = 0$$

and $\varphi_{220} = 0$ is accepted from a Σ_1 triple. Interactions between reflections are found from 3214 Σ_2 triples calculated with $A \geq 0.5$ and NQUEST is based on 136 negative quartets generated with $|B| \geq 0.7$. Algebraic values are assigned:

$$\varphi_{\bar{2}81} = a, \quad \varphi_{370} = b, \quad \varphi_{\bar{1}72} = c, \quad \varphi_{\bar{3}22} = d,$$

which are allowed to be permuted. For the control study, there are 303 $|E_h| \geq 0.06$ generated from the kinematical structure factors (again, based on the same 317 hkl values used before). We used the same reflections as before to define the unit-cell origin.

For the experimental data, if one lists the a most negative NQUEST values (Table 1), a correct solution to the ambiguities can be found, for example, at the value -0.377 , where there are 48 phase errors for 239 reflections. However, the most negative NQUEST $= -0.420$ does not correspond to a correct solution, since there are 110 errors in this set. Using only the $0kl$ and $hk0$ data (Table 1), it is seen that four of the nine trials yield correct solutions whereas the other five do not. From the zonal reflections, it is found that these correspond to either of two structures, as depicted in Figs. 1 and 2. With kinematical data, 211 reflections are assigned phase values after the basis set is defined. In this case, the correct solution does correspond to the lowest value for NQUEST (-0.872), where there are 37 phase errors found for the total three-dimensional data set. For the zonal data sets, there is one error for 30 $hk0$ reflections and also one for 23 $0kl$ reflections. Both sets will lead to correctly interpretable potential maps.

Least-squares refinement

The electrostatic potential map obtained from the $QTAN$ phases (Fig. 1) is essentially the result

Table 1. Phase determination for diketopiperazine with $QTAN$

NQUEST	Ambiguities				Number of errors	
	$\varphi_{\bar{2}81}$	φ_{370}	$\varphi_{\bar{1}72}$	$\varphi_{\bar{3}22}$	$0kl$ (27 data)	$hk0$ (42 data)
-0.377	π	0	0	π	2	0
-0.386	0	π	π	0	13	18
-0.389	π	π	0	0	2	2
-0.390	π	0	0	0	2	0
-0.393	0	π	0	π	12	18
-0.395	0	0	0	0	1	0
-0.401	0	0	π	π	13	17
-0.420	0	π	π	π	12	18
-0.423	π	π	π	π	13	18

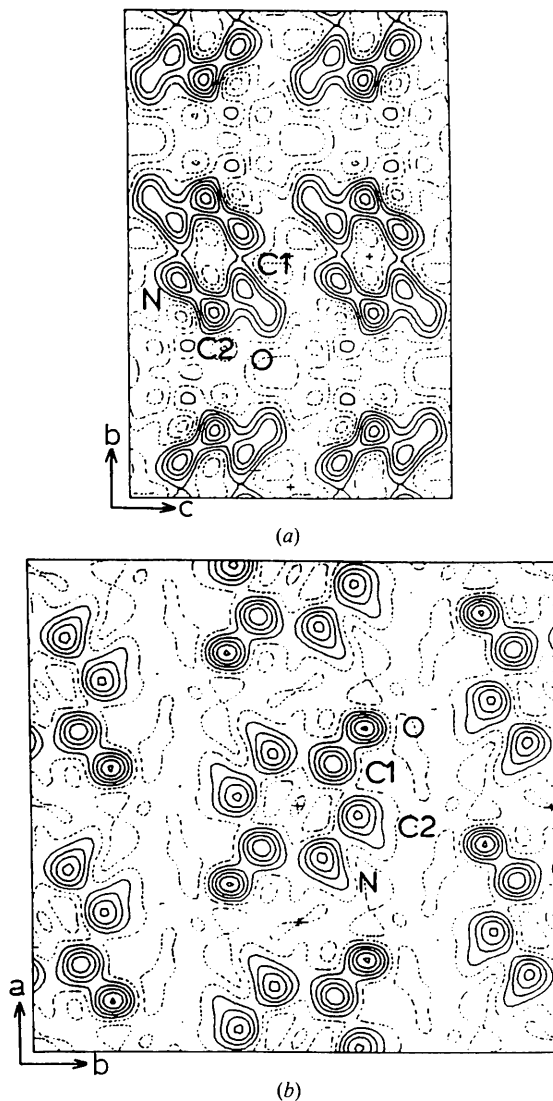


Fig. 1. Correct phase solution for diketopiperazine with $QTAN$, corresponding to NQUEST = -0.377 , -0.389 , -0.390 , -0.395 (Table 1). (a) (100) projection, (b) (001) projection. Depicted are electrostatic potential maps representing the Fourier transforms of phased observed structure factors. [Atomic positions are identified on the basis of reasonable hydrogen-bonding pairs $N-H \cdots O=C$; see Vainshtein (1955).]

obtained before (Dorset, 1991*b*) when individual structure-invariant sums were evaluated to find new phases. Atomic coordinates obtained from measurement of zonal maps are listed in Table 2.

A full-matrix least-squares refinement (Enraf-Nonius software) was carried out using all 317 reflections to minimize $\sum(F_o - F_c)^2$. Atomic form factors were adjusted (by alteration of a computer program in this software) to reproduce values for electron scattering from neutral atoms given by Doyle & Turner (1968). The variables consisted of

Table 2. Fractional coordinates for diketopiperazine

	From initial potential map			After least-squares			X-ray data		
	x	y	z	x	y	z	x	y	z
C1	-0.175	0.072	0.708	0.181	0.073	0.708	-0.182	0.070	0.717
C2	0.051	0.120	0.502	0.046	0.121	0.514	0.052	0.123	0.515
N	0.212	0.049	0.308	0.223	0.047	0.306	0.220	0.043	0.310
O	-0.345	0.130	-0.138	-0.343	0.132	-0.106	-0.331	0.133	-0.097

the 12 positional parameters for the four heavier atoms and an overall scale factor. Isotropic mean-square displacement parameters were fixed at $B = 0 \text{ \AA}^2$, since efforts to refine them gave unrealistic values. Contributions to the structure factors from the H atoms were neglected. After five cycles of refinement, employing atomic shifts dampened by a factor of 0.2, convergence was obtained with $R(F) = 0.27$ for all reflections ($R = 0.25$ for observed reflections). Final atomic parameters are given in Table 2. Final bond distances and angles are in closer agreement with the X-ray structure (Fig. 3) and the essential planarity of the molecule is established by comparing torsion angles (Table 3).*

Discussion

Because very little difficulty was experienced with the original phase determination of electron-diffraction intensity data from diketopiperazine (Dorset, 1991*b*) with individual structure invariants, it is somewhat surprising that the correct solution from the tangent formula does not correspond to the lowest NQUEST value. The problem does not seem to result from incorrectly predicted Σ_2 triples since, for $A \geq 2.0$, there are only two erroneous phase sums out of 146.

* A list of measured structure factors compared to final calculated structure factors obtained after least-squares refinement has been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 71532 (4 pp.). Copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

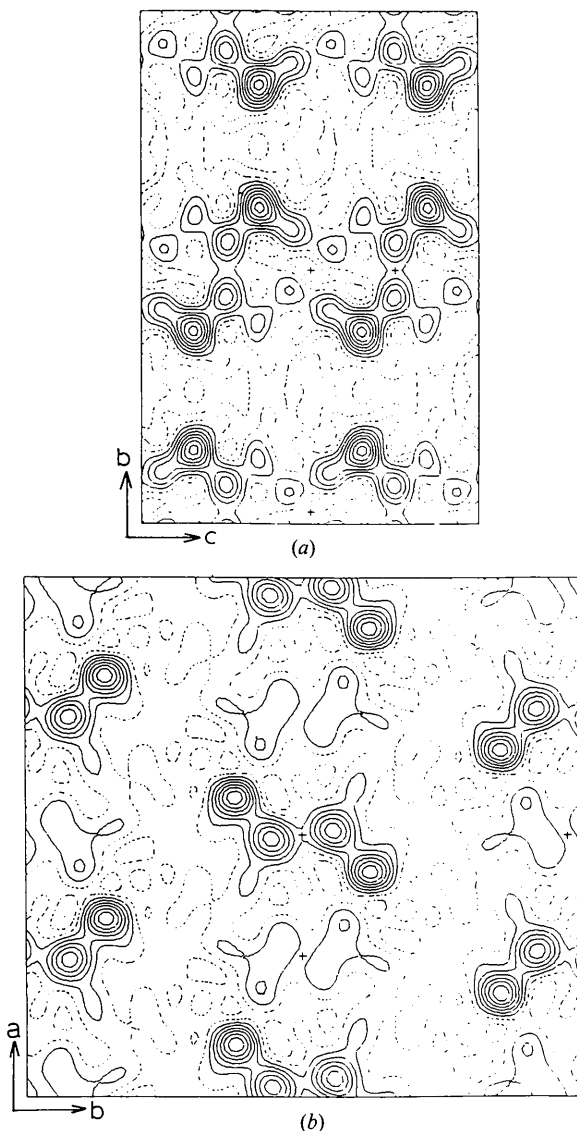


Fig. 2. False phase solutions for diketopiperazine, with $QTAN$, corresponding to NQUEST = -0.386 , -0.393 , -0.401 , -0.402 , -0.423 (Table 1). (a) (100) projection, (b) (001) projection. Electrostatic potential maps are computed from observed structure factors and corresponding phase values from those solutions.

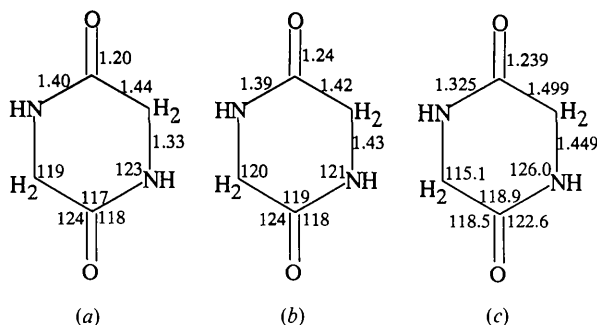


Fig. 3. Bond distances and angles for diketopiperazine (a) from potential map coordinates, (b) after least-squares refinement, (c) from X-ray diffraction by Degeilh & Marsh (1959).

Table 3. *Derived torsion angles for diketopiperazine* (°)

	From electron diffraction after least squares	From X-ray determination (Degeilh & Marsh, 1959)
O-C1-N-C2'	-179.9	-178.4
C1-N-C2'-C1'	-1.1	-1.2
N'-C2-C1-N	1.1	1.1
C2-C1-N-C2'	-1.1	-1.3

Table 4. *Distribution of low $|E_h|$ values for diketopiperazine*

	Experimental	Kinematical
% $ E_h \leq 0.10$	9.7	5.3
% $ E_h \leq 0.20$	11.9	10.1
% $ E_h \leq 0.30$	19.2	18.2
% $ E_h \leq 0.40$	31.4	26.1
% $ E_h \leq 0.50$	37.1	35.5

Even positive quartets have a low number of errors since only two out of 93 sums are incorrectly determined for $B > 2.0$. On the other hand, negative-quartet estimates are a likely cause of this problem since the NQUEST figure of merit depends on accurate estimates of low $|E_h|$ values. Even though the computation of the NQUEST is an average, only 55 out of 85 negative quartets sampled with $|B| \geq 0.70$, *i.e.* 65%, were correct for this structure. When kinematical data were used, this figure was somewhat better, with 80 out of 112 correctly predicted quartets (71%). However, the distribution of $|E_h|$ values does not differ much for experimental and theoretical (kinematical) data (Table 4). More significant, perhaps, is that, within the range $|E_h| \leq 0.5$, there is only a 79% overlap of reflections from the experimental and kinematical sets. Thus, the inaccuracies caused by incorrectly identifying low $|E_h|$ are enough to cause the NQUEST prediction of a correct structure to be somewhat ambiguous.

Despite this difficulty, the correct structure would be determined from the experimental data by inspection if the four most negative solutions were used to calculate potential maps. As seen from Fig. 2, the false solution corresponds neither to the molecular stoichiometry nor to the true structure. Actually, a fragment resembling one half of a molecule appears to be shifted to another unit-cell origin.

It is perhaps surprising that, given these problems of identifying a structure solution, a least-squares refinement, where only B values are constrained, should lead to a structural model that, geometrically, is a better match to the X-ray crystal structure than the model constructed from the original potential map. To our knowledge, this is the first report of such a refinement for electron diffraction data. Recently, we have experienced similar success with the refinement of polyethylene and this will be reported elsewhere. In this case, one carbon and two hydrogen positions are refined against 51 observed data (Hu & Dorset, 1989), starting with a C-atom coordinate obtained after a direct phase determination (Dorset, 1991*d*). This resulted in physically reasonable bond distances and angles as well as thermal parameters. On the other hand, it is fruitless to attempt such a refinement for poly(ϵ -caprolactone) since there are only 47 measured hkl data for eight unique atoms (Dorset, 1991*c*). Thus,

even though there may be some n -beam dynamical perturbation to the measured intensity data so that the overall thermal motion of the molecule cannot be refined, it may still be useful to attempt to improve the crystal structure by least-squares refinement as long as there are sufficient measured data for each refinable parameter. Nevertheless, it is clear from this analysis and earlier Fourier refinements that, because of perturbed intensity data, one must attempt to locate a local minimum rather than a true 'global' minimum, which will correspond to a geometrically distorted molecular geometry.

In conclusion, this analysis of the diketopiperazine structure demonstrated that automated methods for phase determination *via* the tangent formula can successfully retrieve a structure from observed electron diffraction intensities, showing that there was no bias imposed in the original analysis (Dorset, 1991*b*). As we have seen, errors in large $|E_h|$ values are not important in terms of their absolute hierarchy but, if too many errors appear in low $|E_h|$ magnitudes, it may be more difficult to identify the correct structure in the multisolution list if NQUEST is employed as the figure of merit. Perhaps more appropriate figures can be devised for electron diffraction applications based on the Σ_2 triples. Finally, even with deviations from kinematical scattering, refinement of the structure can be carried out, provided that appropriate constraints are applied.

Research was supported by a grant from the National Science Foundation (CHE-9113899), which is gratefully acknowledged. (This grant is supported by both the Theoretical and Computational Chemistry Program and the Polymers Program of the NSF.)

References

- BRISSE, F. (1989). *J. Electron Microsc. Tech.* **11**, 272-279.
 DEGEILH, R. & MARSH, R. E. (1959). *Acta Cryst.* **12**, 1007-1014.
 DETITTA, G. T., EDMONDS, J. W., LANGS, D. A. & HAUPTMAN, H. (1975). *Acta Cryst.* **A31**, 472-479.
 DORSET, D. L. (1991*a*). *Ultramicroscopy*, **38**, 23-40.
 DORSET, D. L. (1991*b*). *Acta Cryst.* **A47**, 510-515.
 DORSET, D. L. (1991*c*). *Proc. Natl Acad. Sci. USA*, **88**, 5499-5502.
 DORSET, D. L. (1991*d*). *Macromolecules*, **24**, 1175-1178.
 DORSET, D. L. (1992*a*). *Ultramicroscopy*, **45**, 5-14.
 DORSET, D. L. (1992*b*). *Ultramicroscopy*, **45**, 357-364.
 DORSET, D. L., BECKMANN, F. & ZEMLIN, F. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 7570-7573.

- DORSET, D. L., TIVOL, W. F. & TURNER, J. N. (1991). *Ultramicroscopy*, **38**, 41–45.
- DORSET, D. L., TIVOL, W. F. & TURNER, J. N. (1992). *Acta Cryst.* **A48**, 562–568.
- DORSET, D. L. & ZEMLIN, F. (1990). *Ultramicroscopy*, **33**, 227–236.
- DOYLE, P. A. & TURNER, P. S. (1968). *Acta Cryst.* **A24**, 390–397.
- HAUPTMAN, H. A. (1972). *Crystal Structure Determination. The Role of the Cosine Seminvariants*. New York: Plenum.
- HU, H. & DORSET, D. L. (1989). *Acta Cryst.* **B45**, 283–290.
- KARLE, I. L., DRAGONETTE, K. S. & BRENNER, S. A. (1965). *Acta Cryst.* **19**, 713–716.
- KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.
- LANGS, D. A. & DETITTA, G. T. (1975). *Acta Cryst.* **A31**, S16.
- LIPSON, H. & COCHRAN, W. (1966). *The Determination of Crystal Structures*, p. 381. London: Bell & Sons.
- VAINSHTEIN, B. K. (1955). *Zh. Fiz. Khim.* **29**, 327–344.
- VAINSHTEIN, B. K. & LOBACHEV, A. N. (1956). *Sov. Phys. Crystallogr.* **1**, 370–371.
- VAINSHTEIN, B. K., ZVYAGIN, B. B. & AVILOV, A. S. (1992). *Electron Diffraction Techniques*, edited by J. M. COWLEY, Vol. 1, pp. 216–312. IUCr/Oxford Univ. Press.
- WILSON, A. J. C. (1942). *Nature (London)*, **150**, 151–152.
- ZVYAGIN, B. B. (1967). *Electron-Diffraction Analysis of Clay Mineral Structures*. New York: Plenum.

Acta Cryst. (1994). **A50**, 292–295

Kinematical Structure Factors from Dynamical Electron Diffraction?

BY J. A. EADES

Center for Microanalysis of Materials, Frederick Seitz Materials Research Laboratory,
University of Illinois, 104 South Goodwin Avenue, Urbana, Illinois 61801, USA

(Received 30 July 1993; accepted 19 October 1993)

Abstract

Electron diffraction intensities are not simply related to the corresponding structure factors except in the case of ‘dynamical extinctions’. These extinctions, explained by Gjønnes & Moodie [*Acta Cryst.* (1965), **19**, 65–67], occur for reflections that are kinematically forbidden – thus the resulting intensity is zero. If it was possible to find conditions when ‘dynamical extinction’ occurs for reflections that are not forbidden, it would be possible to use electron diffraction intensities for the determination of structure factors in a simple way. Unfortunately, it can be shown that this is not possible.

Introduction

A student asked me the following question: ‘If G - M lines can be understood as the result of pair-wise cancellation of multiple-diffraction routes in the case of reflections that are kinematically forbidden, can this same cancellation occur for a reflection that is not kinematically forbidden?’

This is a good question because, if such cancellation did occur, the intensity in such a reflection would depend only on the structure factor for that reflection. Then, we could use those intensities in a very direct way to contribute to structure determination.

The answer, unfortunately, is no.

Background

In electron diffraction, reflections that have structure factor zero (kinematically forbidden reflections), because of the presence of screw axes or glide planes in the

crystal, often appear with intensities that are as high as those of allowed reflections. This is because of the importance of dynamical diffraction or multiple-diffraction routes. However, there are certain conditions under which the intensity of diffraction into these kinematically forbidden reflections is identically zero. These are known as dynamical extinctions.

The conditions under which dynamical extinctions occur refer to both particular symmetry elements and particular orientations. The development of these ideas occurred in three papers (Cowley & Moodie, 1959; Miyake, Takagi & Fujimoto, 1960; Cowley, Moodie, Miyake, Takagi & Fujimoto, 1961); Gjønnes & Moodie (1965) then provided the definitive description in a paper that is clear, concise and correct. In order to discuss the case where reflections that are not kinematically forbidden are involved, we repeat, in the next section, the argument of Gjønnes & Moodie for forbidden reflections (generated by a single symmetry element).

Dynamical extinctions appear in convergent-beam patterns as lines of extinction along the locus of the appropriate conditions. As a result, the following terms have all been used as synonyms for ‘dynamical extinction’: Gjønnes–Moodie line, G - M line, dark bar and black cross. These features have come to play an important role in symmetry determination in electron microscopy. See, for example, the work of Eades (1988) and Tanaka & Terauchi (1985).

The Gjønnes–Moodie theory

It has been shown by Cowley & Moodie (1962) that the amplitude of a particular diffracted beam can be written