Phase III

Les octaèdres ont subi, dans cette phase une seconde rotation de 6,8 (2)° autour de **c**. A ce mouvement est associé un raccourcissement des paramètres de la maille [0,04 (2), 0,03 (2) et 0,07 (4) Å pour *a*, *b'* et *c* respectivement] qui correspond à un nouveau tassement de la structure. Ce second mouvement n'affecte pas fondamentalement l'organisation générale du composé; on remarque néamoins qu'il existe maintenant deux types différents de dépressions entre octaèdres, correspondant aux atomes non équivalents K(1) et K(2), ce qui entraîne des environnements différents pour les deux potassiums: Dans l'une de ces dépressions, correspondant par exemple à $K(1^{xvi})$ (Fig. 3*a*), les rotations autour de **c** ont rapproché $F(1^{ii})$ et $F(1^{xvi})$ et éloigné F(21) de $F(22^{xii})$. La conclusion est inversée pour l'autre type de dépression [correspondant par exemple à $K(2^{xiv})$]. Ces mouvements ont peu d'influence sur les positions des fluors terminaux: Les potassiums se placent dans des positions relatives très proches de celles trouvées dans la phase II; seules, les distances des ions K^+ avec leurs seconds voisins (Tableau 2) ont été modifiées. La rotation caractérisant la phase III laisse pratiquement inchangés les contacts opposés entre les potassiums et les deux fluors terminaux du feuillet suivant (Fig. 3).

Géométrie des octaèdres

Dans les deux phases étudiées, les octaèdres FeF_6 paraissent réguliers et peu déformés (Tableau 2). Il existe cependant des différences, que nous n'avons pu interpréter, suivant le type de phase dans lequel on se place: Dans la phase II on observe une distance Fe-fluor terminal F(3), nettement plus courte: 1,878 (2) Å que les distances Fe-F liés, soit 1,963 Å en moyenne, comme dans NH_4FeF_4 (Leblanc, Ferey, De Pape & Teillet, 1985) où les distances correspondantes moyennes sont de 1,956 et 1,867 Å. Ce résultat est inversé pour les octaèdres de la phase III: les distances Fe-F liés, qui s'écartent peu de leur valeur moyenne 1,972 Å, proche de celle trouvée pour la phase II, sont plus courtes que les distances Fe-F terminal: 2,022 Å en moyenne.

Références

BILLIET, Y., SAYARI, A. & ZARROUK, H. (1978). *Acta Cryst.* A**34**, 414-421.

BUSING, W. R., MARTIN, K. O., LEVY, H. A., ELLISON, R. D., HAMILTON, W. C., IBERS, J. A., JOHNSON, C. K. & THIESSEN, W. E. (1971). *ORXFLS3*. Oak Ridge National Laboratory, Tennessee.

COPPENS, P. & HAMILTON W. C. (1970). *Acta Cryst.* A**26**, 71-83.

ESCANDE, A. (1971). Thèse de doctorat de 3e cycle, Montpellier.

GLAZER, A. M. (1972). *Acta Cryst.* B**28**, 3384-3392.

GLAZER, A. M. (1975). *Acta Cryst.* A**31**, 756-762.

HEGER, G., GELLER, R. & BABEL, D. (1971). *Solid State Commun.* **9**, 335-340.

HIDAKA, M., GARRARD, B. J. & WANKLYN, B. M. R. (1979). *J. Phys. C*, **12**, 2737-2743.

International Tables for Crystallography (1983). Tome *A*. Dordrecht: D. Reidel.

International Tables for X-ray Crystallography (1974). Tome IV. Birmingham: Kynoch Press. (Distributeur actuel D. Reidel, Dordrecht.)

LEBLANC, M., FEREY, G., DE PAPE, R. & TEILLET, J. (1985). *Acta Cryst.* C**41**, 657-660.

SAINT-GRÉGOIRE, P., PÉREZ, A., ALMAIRAC, R. & LOPEZ, M. (1985). *Phys. Status Solidi A*, **87**, K1-K6.

Acta Cryst. (1986). B**42**, 262-272

Analysis of the Diffuse Scattering from Disordered Molecular Crystals: Application to 1,4-Dibromo-2,5-diethyl-3,6-dimethylbenzene at 295 K

BY T. R. WELBERRY AND J. SIRIPITAYANANON

Research School of Chemistry, Australian National University, PO Box 4, *Canberra City, ACT* 2601, *Australia*

(*Received* 27 *August* 1985; *accepted* 23 *December* 1985)

Abstract

The disorder diffuse X-ray scattering (DDS) in 1,4-dibromo-2,5-diethyl-3,6-dimethylbenzene, $C_{12}H_{16}Br_2$ (BEMB1), has been analysed using the least-squares procedure of Epstein & Welberry [*Acta Cryst.* (1983), A**39**, 882-892]. Data from five reciprocal-lattice sections were used to obtain values for all intermolecular correlation coefficients within a 20 Å neighbourhood. Of the 43 correlations within this neighbourhood, 18 were found to be significantly different from zero, and of these three were significant at the 3σ level. The largest correlation of $-0{\cdot}208(25)$ was between a central molecule and its nearest neighbour in the [201] crystal direction. The correlation coefficients were generally smaller for this non-polar molecule

0108-7681/86/030262-11$01.50

than in previously studied polar molecules where values in excess of 0·4 have been common. A comparison of results for two different methods of analysis has indicated that the e.s.d.'s reported by Epstein & Welberry (1983) are probably underestimated by about 25%.

1. Introduction

In recent publications (Welberry & Jones, 1980; Welberry, Jones & Epstein, 1982; Epstein, Welberry & Jones, 1982; Welberry, 1983; Epstein & Welberry, 1983) we have described the development of methods for recording and analysing diffuse X-ray scattering data from disordered molecular crystals, based on long-exposure Weissenberg photographs. The methods have developed from initial semi-quantitative procedures involving visual comparison of the observed X-ray patterns with optical diffraction patterns of computer-generated models of the disorder (Welberry & Jones, 1980), to recent more quantitative procedures in which the observed scattering is fitted using least squares to a calculated model of the disorder (Epstein & Welberry, 1983). In all cases the aim of the experiment is to derive values for a set of intermolecular correlation coefficients, $C_{n,m}$, which define the local orientational ordering in compounds in which disorder occurs because of the similarity in size of some common substituent groups. In particular we have applied the methods to a number of systems in which molecules containing CH_3 and Br or Cl substituents form static orientationally disordered crystals.

These analyses indicated that short-range ordering occurred when the disordered substituent sites in neighbouring molecules were in close proximity and this ordering could be explained qualitatively by considering the interactions of the partial charges on atoms resulting from the presence of the electronegative halogen substituents and electropositive methyl substituents. To explore fully such relationships between the magnitude of the short-range order and the mutual disposition of the electrostatic fields of neighbouring molecules, data from a large number of examples with a wide range of different mutual interactions are required. As part of our long-term aim to compile systematically such information we sought a molecular crystal system which would provide a range of different crystal structures and states of disorder.

The series of compounds which was chosen for this study is the set of isomers of dibromodiethyldimethylbenzene (see Fig. 1). Of the eleven isomers which are depicted, simple consideration of the van der Waals shape of the molecule would suggest that seven would be likely to form disordered structures in which each molecular site can be occupied by either of two different molecular orientations. In each case the two different orientations result in altering the disposition of the substituent methyl and bromo groups about the diethylbenzene nucleus. The alternative molecular orientation may be obtained from that depicted by a rotation about a twofold axis, as indicated in Fig. 1. Note that in some cases the conformation of the ethyl group is required to change also. For isomers 1, 2, 3, 10 and 11 the disorder is expected to involve both methyls and both bromines, while for isomers 5 and 6 it would involve only one bromine and one methyl. The molecular dipole moment for these molecules would be expected to vary considerably from zero for isomer 1, which is centrosymmetric, to a large value for isomer 3 in which both the bromo substituents are maximally separated from the methyl substituents.

The three *p*-ethyl isomers (1, 2, 3) form a particularly interesting set since they are isostructural with closely similar unit-cell parameters, although they exhibit markedly different diffuse-scattering patterns. The crystal-structure determination of isomers 1 and 2 (hereafter referred to as BEMB1 and BEMB2) has been reported elsewhere (Wood, Welberry & Puza, 1984). In this paper we describe our analysis of the diffuse scattering of BEMB1 (1,4-dibromo-2,5-diethyl-3,6-dimethylbenzene). Analysis of the diffuse scattering for BEMB2 and subsequent isomers is in progress and will be reported in due course.

As our methods of analysis have become more quantitative, the number of significant correlation

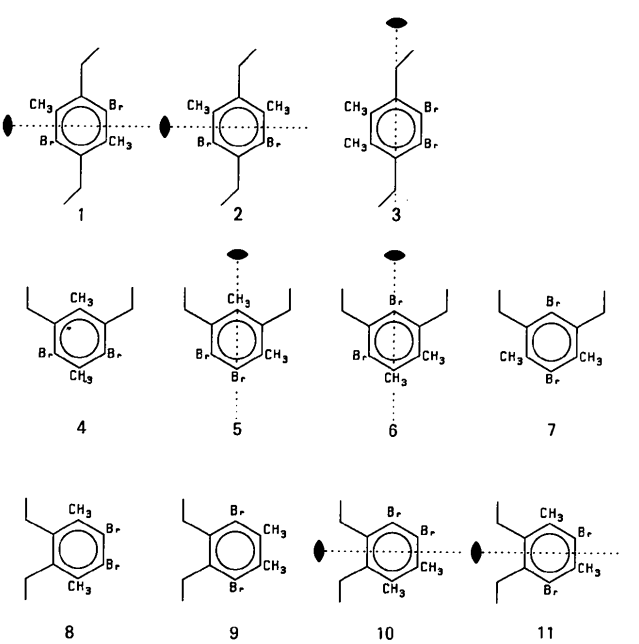

Fig. 1. The 11 isomers of dibromodiethyldimethylbenzene. Disorder might be expected to occur in the structures of isomers 1, 2, 3, 5, 6, 10 and 11 because of the similarity in shape of the molecule in the orientation depicted to that obtained by a twofold rotation as indicated. In some cases this requires a change in conformation of the ethyl substituents.

coefficients ($C_{n,m}$) that it is claimed may be determined has increased from the one or two strongest correlations which could be assigned with any confidence, to the most recent work using the least-squares procedure where thirteen $C_{n,m}$ values were claimed to be significant at the 3σ level. However, since this latter work was the first attempt at such quantitative analysis it was considered that further efforts should be made to assess the validity of the claimed level of significance.

2. Theoretical background

Following Epstein *et al.* (1982) and Epstein & Welberry (1983), if thermal diffuse scattering is neglected the diffuse scattering (DDS) of X-rays from a two-component substitutionally disordered molecular crystal may be represented (with some change of nomenclature) by

$$I_{\rm diff}(\mathbf{S}) = Nm_Am_B \sum_n^{\rm cell} \sum_m C_{n,m}\, \Delta F_n(\mathbf{S})\, \Delta F_m^*(\mathbf{S}), \quad (1)$$

where the magnitude of the scattering vector **S** is $4\pi \sin\theta/\lambda$; N is the number of unit cells; m_A is the concentration of molecules of type A; $C_{n,m}$ is a correlation coefficient defined by $C_{n,m} = (P_{AA} - m_A^2)/m_Am_B$, where P_{AA} is the joint probability that both sites n and m are occupied by molecules of type A. $\Delta F_n(\mathbf{S})$ is the difference between the vibrationally averaged form factors of a type A molecule and a type B molecule at site n. The summation over m is over all molecules within a given neighbourhood of a particular molecule and the summation over n is over molecules within one unit cell.

In this experiment we seek to determine $C_{n,m}$ values for all intermolecular vectors within a given neighbourhood of a particular molecule. For each pair of sites n, m (1) gives a distribution in reciprocal space which is characteristic of the intermolecular vector $R_{n,m}$. Since it is expected that symmetry-related vectors will have the same values of $C_{n,m}$ such terms are grouped together to make up a contribution to the total intensity which we call a *correlation distribution.* Of particular importance is the term involving only self-vectors, $C_{n,n}$ ($=1$, by definition), which corresponds to the intensity occurring when no intermolecular correlations are present. We refer to this as the *random distribution.* Both the random distribution (which is positive everywhere) and the correlation distributions (which are not) may be computed from a knowledge of the average structure. These distributions are used as basis functions in the least-squares analysis which establishes the values of $C_{n,m}$ which give the best fit to the observed data.

Collection of a full 3D set of diffuse scattering data is not a viable proposition since diffuse intensities are typically $\sim 10^{-3}$ of Bragg intensities. Consequently we make use of a number of 2D reciprocal-lattice sections obtained from Weissenberg photographs taken with the crystal mounted about several different axes. The data from these different sections may be scaled together and analysed as described above, but in practice such scaling is difficult because of the intrinsically low accuracy of the data and the small number of common data points. Consequently Epstein & Welberry (1983) discarded this option in favour of a method in which each section was first analysed in terms of so-called 2D reduced distribution coefficients, x_l', and subsequent solution of a set of simultaneous equations (their equation 11),

$$x_l' = C_{n,m} + \alpha C_{n,m'} + \beta C_{n,m''} + \ldots - x_l' C_{n,p} - x_l' C_{n,p'} \ldots \quad (2)$$

to obtain values for the 3D $C_{n,m}$ values. In this paper we present for comparison, for BEMB1, results using both methods.

3. Data collection

Long-exposure (~3 d) Weissenberg photographs were taken of five reciprocal-lattice sections ($h0l$, $hk0$, $0kl$, $hk\bar{h}$ and $hk2\bar{h}$) of BEMB1 using a conventional Weissenberg camera with a layer-screen gap width of 1 mm and graphite-monochromated Cu $K\alpha$ radiation at room temperature (295 K). Although it was necessary to use different crystals for each mounting, spherical melt-grown crystals of ~0·5 mm diameter were used in each case. To aid in preliminary investigations the photographs were digitized using an Optronics P1700 photomation system, corrected for background, and processed to remove the Weissenberg distortion using the procedure described by Welberry & Jones (1980). This yielded the enhanced undistorted photographic images of the reciprocal-lattice sections which are shown in Figs. 2(*a*), 3(*a*), 4(*a*), 5(*a*) and 6(*a*), for later comparison. Such a digitized image is composed of ~650 000 picture elements (pixels). For quantitative intensity measurements, individual densitometer readings were made directly from the Weissenberg films at points on a much coarser grid in reciprocal space. The grid chosen had increments of $a^*/10$, $b^*/20$ and $c^*/5$ (see Table 1 for the cell data for BEMB1). These values were chosen to provide approximately equal spatial resolution in all directions, and to provide adequate sampling of reciprocal space in a situation where the spatial frequencies involved in the calculated distribution functions might correspond to vectors extending over ~30 Å.

The data were measured at these points by means of the procedure described by Welberry & Glazer (1985). The intensity measured at each point was taken as the mean optical density of a 3×3 array of pixels centred on the calculated position, and the

variation of these nine measurements was taken as an estimate of the experimental standard deviation, σ_i. Two symmetry-related sets of data were obtained from the two sides of the Weissenberg film for the $h0l$ section and four sets for each of the other sections. Data were collected in the range $2{\cdot}5 < \theta < 25{\cdot}6°$. An empirical background correction was applied to the data using the procedure also described by Welberry & Glazer (1985), and the symmetry-related sets of data were then averaged to yield a single set of data for each section.

Table 1. *BEMB*1 *cell data and average structural information assuming space group* $P2_1/c$

$a = 9{\cdot}082$, $b = 4{\cdot}455$, $c = 17{\cdot}929$ Å, $\beta = 122{\cdot}79°$.

	Coordinates for Br and $C(CH_3)$ for molecule in orientation 1 (Site occupancy $m_A = 0{\cdot}56$)			orientation 2 (Site occupancy $1 - m_A = 0{\cdot}44$)		
	x	y	z	x	y	z
Br(1)	0·2240	0·2479	0·0467	0·9150	0·3821	0·1326
Br(2)	0·7760	0·7521	−0·0467	0·0850	0·6179	−0·1326
C(1)	0·9150	0·3821	0·1326	0·2240	0·2479	0·0467
C(2)	0·0850	0·6179	−0·1326	0·7760	0·7521	−0·0467

To calculate the distribution functions we specify the molecular site 55501 as (x, y, z) and the site 55502 as $(x, \frac{1}{2} - y, \frac{1}{2} + z)$.

4. Data selection

At this stage in the development of our methods for extracting the correlation coefficients, $C_{n,m}$, we have sought to side-step the problem of thermal diffuse scattering (TDS) by eliminating from the analysis any data that may be supposed to be significantly affected by TDS. Epstein & Welberry (1983) attempted to avoid TDS by choosing to measure the diffuse scattering data only in selected regions of reciprocal space which appeared devoid of TDS problems. However, in the least-squares analysis that ensued, it was found that in some regions where no measurements were made the calculated intensity took on unrealistic negative values. It seems likely that this effect was directly attributable to having data from only limited

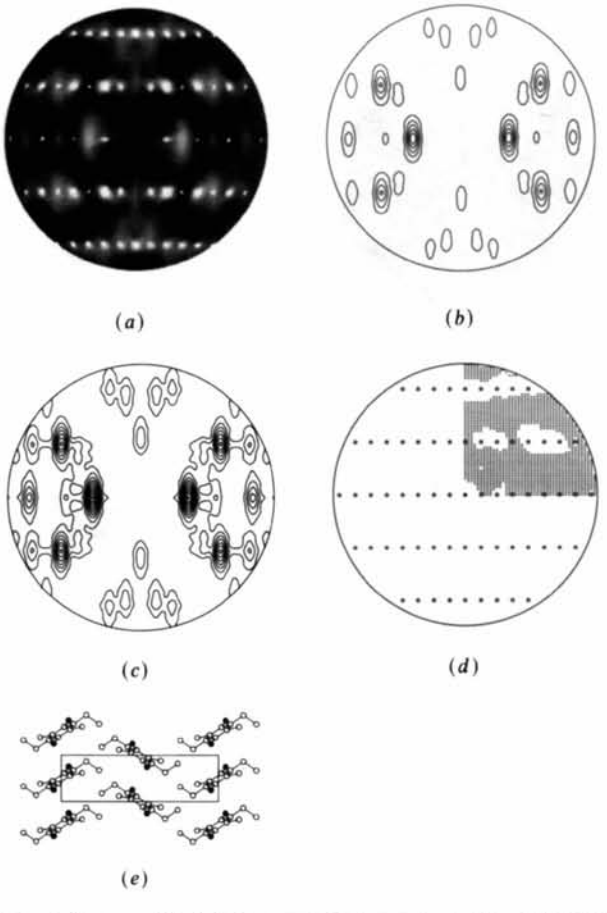

Fig. 2. Section $0kl$. (*a*) Observed X-ray pattern. (*b*) Calculated intensity from individual section analysis. (*c*) Calculated intensity using $C_{n,m}$ values obtained from final 2D solution. (*d*) The data points used in the analysis. (*e*) The projection of the structure corresponding to this section.

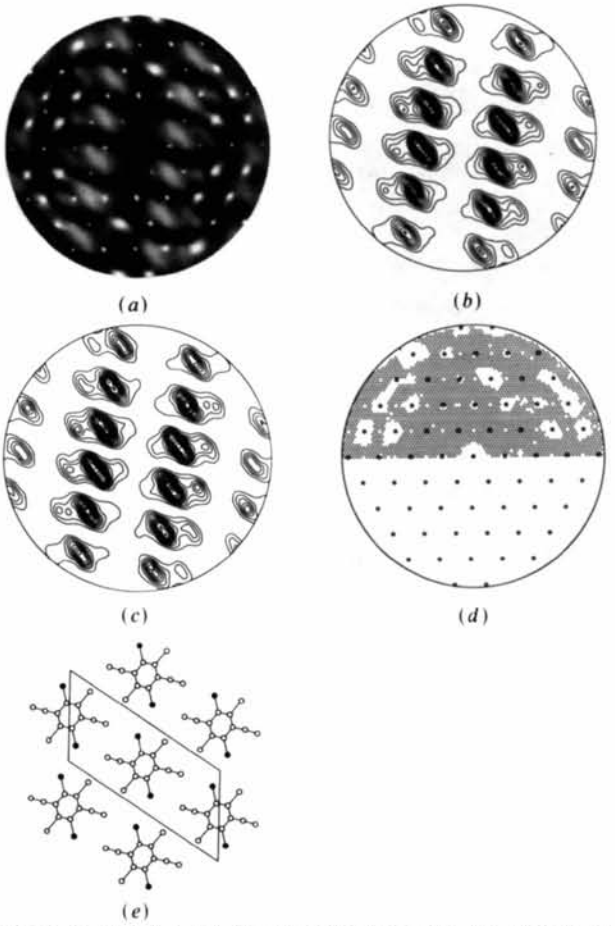

Fig. 3. Section $h0l$. (*a*) Observed X-ray pattern. (*b*) Calculated intensity from individual section analysis. (*c*) Calculated intensity using $C_{n,m}$ values obtained from final 2D solution. (*d*) The data points used in the analysis. (*e*) The projection of the structure corresponding to this section.

regions of reciprocal space. Although the problem was overcome by applying positivity constraints in the unrepresented regions, in the present work we have adopted the point of view that as complete sections of reciprocal space as possible should be employed, and data only discarded when TDS contributions were considered significant. As a result of this we have found it unnecessary to resort to the use of such constraints.

In this work, as in the previous work, we limited the extent of the data to values of $\theta < \sim 25°$ for Cu $K\alpha$ radiation. This has the double purpose of minimizing any residual effects of TDS (which greatly increase at higher θ) and minimizing problems of spatial resolution resulting from the use of Weissenberg geometry (see Welberry, 1983).

The removal of unwanted data affected by TDS was carried out by making use of the fact that the strongest TDS occurs close to the strong Bragg peaks. A simple computer algorithm was devised which enabled points within a certain distance of a Bragg peak to be rejected, where the distance depended on the magnitude of the Bragg peak. For weak Bragg peaks only the Bragg peak itself was rejected while, for the very strongest, the area of rejection extended well into the centre of the reciprocal cell.

The final sets of data that were obtained for use in the analysis consisted of the points shown in Figs. 2(*d*), 3(*d*), 4(*d*), 5(*d*) and 6(*d*). Comparison of these figures with the corresponding photographic images of the intensity indicates the effective removal of those and only those data affected substantially by TDS.

5. Experimental considerations

Although the empirical background correction employed in the initial stages of data reduction is effective in removing most of the unwanted intensity coming from air scattering, fluorescence *etc.*, it was found necessary by Epstein & Welberry (1983) to include in the analysis a further correction in the form of a constant (over the whole reciprocal-space section) which could be refined in the least-squares fitting to the observed intensity distribution. In the present

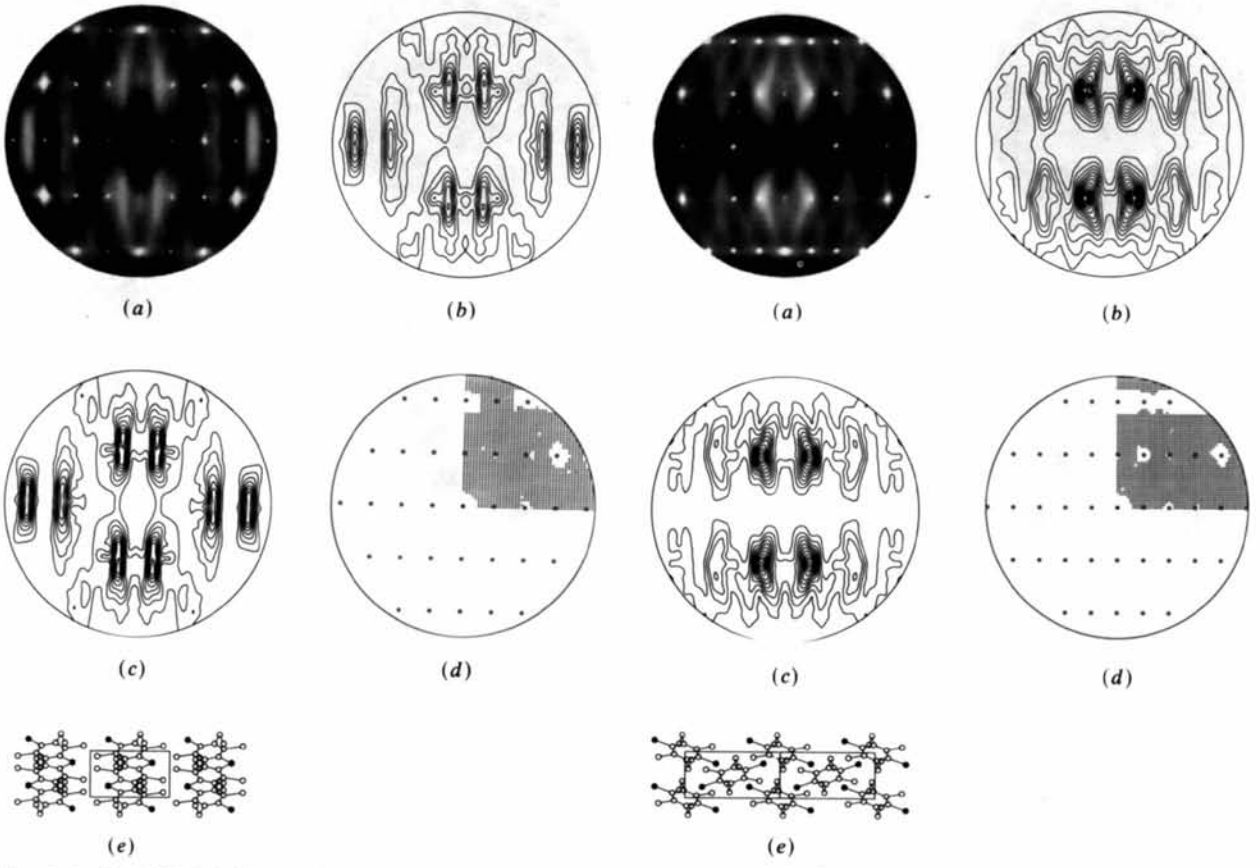

Fig. 4. Section *hk*0. (*a*) Observed X-ray pattern. (*b*) Calculated intensity from individual section analysis. (*c*) Calculated intensity using $C_{n,m}$ values obtained from final 2D solution. (*d*) The data points used in the analysis. (*e*) The projection of the structure corresponding to this section.

Fig. 5. Section $hk\bar{h}$. (*a*) Observed X-ray pattern. (*b*) Calculated intensity from individual section analysis. (*c*) Calculated intensity using $C_{n,m}$ values obtained from final 2D solution. (*d*) The data points used in the analysis. (*e*) The projection of the structure corresponding to this section.

experiment we have confirmed the need for such a constant by performing tests on a film exposed to the background scattering when the crystal had been removed from its mounting. (Note that this is mainly air scattering and does not include the fluorescence component which is present when the crystal is illuminated.) For this film we measured the background directly and also by our empirical estimation method. It was found in this way that the empirical method consistently underestimates the average background and that to a good approximation the difference is constant with θ and of a comparable magnitude to the values of the constant obtained in the least-squares analysis.

A second consideration in relation to the background determination is the effect of the assumption that at any θ the intensity will be zero for some ω. There is no *a priori* way of testing the validity of this assumption for the observed intensity, but we can do so for the calculated random distribution - *i.e.* the intensity that would be present if there were no short-range order. Although the presence of correlation will affect the distribution of the intensity, the regions of zero or very low intensity will be unaffected. Plots depicting the maximum and minimum of the random intensity occurring in each 1° range of 2θ for each of the five reciprocal sections used showed that for all sections over much of the range of θ the minimum intensity that occurs is a small fraction ($\ll 10\%$) of the maximum intensity. This gives reasonable credence to the validity of the assumption. However, there were small ranges of θ in some sections (around $\theta = 12$ and $\theta = 20°$ in $hk\bar{h}$, around $\theta = 14°$ in $hk0$ and around $\theta = 7°$ in $hk2\bar{h}$) where the assumption appears to be less valid. In the subsequent analysis an awareness of the possibility of problems in these regions was maintained, but no attempt was made to correct for them.

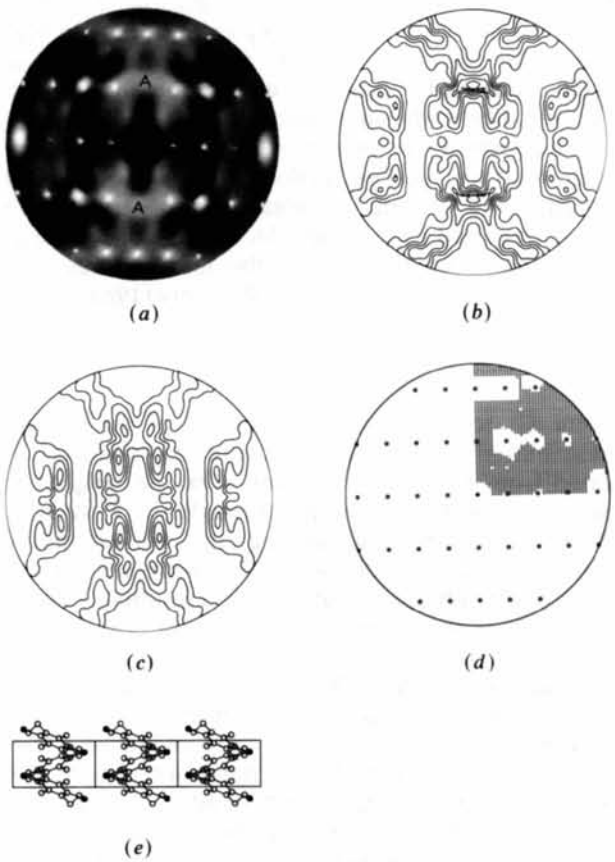

Fig. 6. Section $hk2\bar{h}$. (*a*) Observed X-ray pattern. (*b*) Calculated intensity from individual section analysis. (*c*) Calculated intensity using $C_{n,m}$ values obtained from final 2D solution. (*d*) The data points used in the analysis. (*e*) The projection of the structure corresponding to this section.

In the previous analysis of Epstein & Welberry (1983) the values of σ_i obtained as described in § 3 were used to provide weights, σ_i^{-2}, for use in the least-squares analysis. However, following further tests on the validity of these σ_i and the results of a study (Welberry & Glazer, 1985) in which a set of film data was compared to a set of counter data, we consider it more appropriate to use unit weights. Although the experimental σ_i provide some measure of that component of the error which is random from point to point, this is a fairly small proportion of the total error in a given measurement. The largest part of the error is undoubtedly due to systematic effects. The constant term arising from the way in which the background is estimated is one systematic error, account of which has been taken; but other unaccountable sources of systematic error are certainly present (see Welberry & Glazer, 1985).

6. 2D Analysis

The cell dimensions, space group and the average crystal structure information for BEMB1 obtained from the Bragg scattering experiment were reported by Wood, Welberry & Puza (1984). Although the space group of the average structure is $P2_1$ the structure is very close to $P2_1/c$ with only a small number of fairly weak ($h0l$, $l = 2n+1$) reflections dictating the adoption of $P2_1$. However, following the calculation of the diffuse intensity distributions required for the analysis, we found those distributions which were due to molecular sites that would have been exactly related by symmetry in $P2_1/c$ were, in the low-θ range in which we work, still practically identical in $P2_1$. This meant that in the least-squares analysis of each section using $P2_1$ the least-squares matrix was so ill conditioned that no meaningful distinction could be made between these pairs of distributions. Consequently it became necessary to carry out the analysis in the space group $P2_1/c$.

A structure refinement in this space group using the data of Wood *et al.* (1984) was available (Jones,

Welberry & McLaughlin, unpublished). In this refinement the diethylbenzene fragment of the molecule was assumed to be ordered and the disordered $-CH_3$ and $-Br$ substituents were represented by single composite atoms in a manner similar to that reported for 9-bromo-10-methylanthracene (Jones & Welberry, 1980). The calculations of the diffuse scattering distributions were then made using this average structural information. The atomic coordinates for the Br and C(methyl) in each of the two molecular orientations that can occur in a given site are given in Table 1, together with the site occupancy m_A, the cell data and the equivalent-position specifiers used. An isotropic temperature factor corresponding to a mean-square atomic displacement of 0·06 Å^2 was used throughout. We adopted this isotropic value since the anisotropic values that are obtained from the structure refinement relate to a composite disordered atomic site and do not truly reflect the thermal motion of an individual atom. It should be noted that the asymmetric unit is only half a molecule but since we need to use the whole molecule in the calculation we require only two equivalent-position specifiers to define the molecular sites. For convenience we use the usual *ORTEP* (Johnson, 1976) method of referring to sites related by cell translations. In the present case 55501 refers to the molecule at (x, y, z) and 55502 to that at $(x, \frac{1}{2}-y, \frac{1}{2}+z)$ which is related by the c glide. Because the molecular site contains a centre of symmetry the 2_1 symmetry operation is redundant.

All molecular sites within a neighbourhood of 20 Å of a particular site were included in the calculations. The number of molecular sites within this range was 114 but with the assumption that the same correlation applies to symmetry-related sites the number of distinct $C_{n,m}$ needed to be determined was only 43. The designation codes for these sites may be seen in the table of final results, Table 3.

For the analysis of each 2D section the least-squares procedure determines the value of the relative distribution coefficients x'_l. The numbers of such coefficients within the chosen neighbourhood were found to be 13, 9, 28, 22 and 27 for the $0kl$, $h0l$, $hk0$, $hk\bar{h}$ and $hk2\bar{h}$ sections respectively. Some details of the 2D analyses are given in Table 2. The R factor quoted is that obtained directly from the least-squares analysis in which a constant is included in the calculated intensity. R_{real} is the R factor that is obtained when the same constant is subtracted from the data instead, and this more properly expresses the agreement between observed and calculated data. In view of the fact that most reports of diffuse-scattering studies, even the most quantitative ones, do not usually include a statement of an agreement factor, it is somewhat difficult to assess what magnitude of R should be considered satisfactory. In one study where R factors were quoted, values as high as ~0·48 were obtained (Cenedese, Bley & Lefebvre, 1984). It has been estimated (Welberry & Glazer, 1985) that the accuracy of our intensity measurements is only ~20% at best. In the present case the R factor obtained for the $h0l$, $hk0$ and $hk\bar{h}$ sections we consider quite satisfactory, that for $0kl$ and $hk2\bar{h}$ less so. In the case of $0kl$ the higher R is undoubtedly due to the fact that the film for this section was appreciably less intense than those for the other sections (see later). Much of the poor agreement for the $hk2\bar{h}$ section is associated with the region marked A in Fig. 6(a). This region of intensity appears not to be of DDS origin and is probably attributable to the poor performance of the empirical background procedure for this section (see later also).

Table 2. *Details of the least-squares analyses of individual 2D sections for BEMB*1

Values of the refined background constant, c, and the R and R_{real} values obtained in the 3D analysis are included in parentheses for comparison.

	Reciprocal-space section				
	$0kl$	$h0l$	$hk0$	$hk\bar{h}$	$hk2\bar{h}$
Number of accepted data	1235	2621	1528	1655	1372
Number of excluded data	397	649	170	330	374
Number of distributions, x_l	13	9	28	22	27
Background constant, c	2·2 (2·5)	4·7 (4·7)	1·6 (2·9)	0·4 (0·6)	1·4 (4·0)
$R=[\sum(I_o-I_c)^2/\sum(I_o)^2]^{1/2}$	0·38 (0·39)	0·29 (0·29)	0·29 (0·31)	0·25 (0·27)	0·39 (0·48)
$R_{\text{real}}=[\sum(I_o-I_c)^2/\sum(I_o-c)^2]^{1/2}$	0·50 (0·48)	0·35 (0·35)	0·32 (0·37)	0·25 (0·28)	0·43 (0·59)
Scale factor for 3D analysis	1·60	1·00	1·14	0·67	1·23

In total for the five sections the analyses yielded 99 equations of type (2) relating the observed x'_l to the 43 required $C_{n,m}$ values. An attempt was made to solve these equations using the least-squares procedure described by Epstein & Welberry (1983) which minimizes the quantity.

$$\varepsilon=\sum_l w_l(\hat{x}'_l-x'_l)^2.$$

Here $\hat{x}'_l$ are the measured estimates of the relative distribution coefficients and x'_l are the values calculated from (2). In the present work the weights w_l were assumed to be unity throughout. Epstein & Welberrry (1983) in solving the corresponding equations for 2,3-dichloro-6,7-dimethylanthracene were able to obtain a solution with a residual R factor of 0·103. In the present case the result obtained for the 43 $C_{n,m}$ from the full set of 99 equations gave a residual R of 0·304. This was considered unacceptably large and indicated that the procedure had failed to yield a satisfactory solution to the equations. Furthermore, the intensity distributions for the individual sections calculated from the derived $C_{n,m}$ visibly did not agree with the observed distributions.

After some investigation it was found that this failure of the solution procedure was attributable to the fact that there existed strong correlations, $\rho_{l,l'}$, between some of the distributions used in the 2D-section analyses, as indicated by the least-squares

Table 3. *Final results for the correlation coefficients obtained from the 2D and the 3D methods of analysis*

Only the 43 unique designation codes are listed. For each 01-type code specified by the three integers (i, j, k) there is another symmetry-equivalent site with a designation code specified by $(10-i, 10-j, 10-k)$. For each 02-type code specified by (i, j, k) there are three additional symmetry-equivalent codes specified by $(i, 11-j, k)$; $(10-i, j, 9-k)$; $(10-i, 11-j, 9-k)$. See text for details.

Site	$C^{2D}_{n,m}$	$C^{3D}_{n,m}$	Δ	$\bar{C}_{n,m}$	$\bar{\sigma}$	$\bar{C}_{n,m}/\bar{\sigma}$
66502	−0·184(25)	−0·231(6)	0·047	−0·208	0·025	8·205
65501	−0·150(24)	−0·131(7)	−0·019	−0·141	0·026	5·485
64501	0·073(14)	0·105(9)	−0·032	0·089	0·022	4·005
63501	0·032(13)	0·049(8)	−0·017	0·041	0·020	2·022
75601	0·060(37)	0·073(6)	−0·013	0·067	0·035	1·903
66601	0·035(15)	0·034(8)	0·001	0·035	0·021	1·636
78502	0·003(9)	0·032(5)	−0·029	0·018	0·013	1·349
73601	−0·013(15)	−0·039(8)	0·026	−0·026	0·021	1·233
67501	−0·016(13)	−0·033(8)	0·017	−0·025	0·020	1·223
56601	−0·021(18)	−0·033(8)	0·012	−0·027	0·023	1·181
56501	0·010(26)	0·047(5)	−0·037	0·029	0·025	1·133
66501	0·038(14)	0·008(8)	0·030	0·023	0·021	1·119
64601	−0·015(14)	−0·030(8)	0·015	−0·022	0·021	1·095
74501	0·023(14)	0·018(7)	0·005	0·021	0·019	1·082
57401	0·008(17)	0·034(6)	−0·026	0·021	0·019	1·081
68402	−0·001(9)	−0·027(5)	0·026	−0·014	0·013	1·079
76601	0·013(18)	0·041(10)	−0·028	0·027	0·026	1·040
68502	0·006(12)	0·024(5)	−0·018	0·015	0·015	1·017
58501	0·018(18)	0·016(5)	0·002	0·017	0·019	0·898
56401	0·014(19)	0·027(8)	−0·013	0·021	0·023	0·873
58502	−0·003(12)	−0·021(5)	0·018	−0·012	0·015	0·813
62501	0·015(12)	0·014(7)	0·001	0·015	0·018	0·812
67502	0·002(13)	0·023(5)	−0·021	0·013	0·015	0·812
67402	0·001(9)	−0·021(5)	0·022	−0·010	0·013	0·771
69502	0·003(9)	0·014(4)	−0·011	0·009	0·011	0·744
59501	0·019(18)	0·008(5)	0·011	0·014	0·019	0·713
67601	0·012(14)	0·017(8)	−0·005	0·015	0·021	0·706
77502	−0·024(9)	0·007(5)	−0·031	−0·009	0·013	0·655
76502	0·015(15)	−0·038(6)	0·053	−0·012	0·018	0·636
57502	0·004(13)	−0·023(5)	0·027	−0·009	0·015	0·617
69501	0·007(12)	0·007(5)	0·000	0·007	0·015	0·475
66402	−0·023(15)	0·037(6)	−0·060	0·007	0·018	0·387
77601	−0·002(16)	−0·013(8)	0·011	−0·008	0·022	0·346
57501	0·006(20)	0·008(5)	−0·002	0·007	0·020	0·343
65601	−0·002(31)	0·020(5)	−0·022	0·009	0·029	0·308
59502	0·007(9)	−0·001(3)	0·008	0·003	0·010	0·298
55601	0·024(36)	−0·005(6)	0·029	0·010	0·034	0·278
75501	−0·023(17)	0·014(4)	−0·037	−0·004	0·017	0·263
76501	0·004(14)	−0·013(7)	0·017	−0·005	0·019	0·237
68501	−0·002(12)	−0·002(7)	0·000	−0·002	0·018	0·112
63601	0·003(14)	0·000(8)	0·003	0·002	0·021	0·073
56502	−0·033(22)	0·032(5)	−0·065	0·000	0·022	0·023
74601	0·004(18)	−0·003(10)	0·007	0·001	0·026	0·019

correlation matrix. This simply means that some distributions were sufficiently similar over the area covered by the data that they could not be reliably distinguished. To overcome the problem it was therefore decided to omit from the solution procedure those equations which involved distribution coefficients with high $\rho_{l,l'}$ values. All equations involving distribution coefficients x'_l, $x'_{l'}$, for which $\rho_{l,l'} > |0{\cdot}6|$ were omitted (24 equations) and the solution procedure repeated using the 75 remaining equations. The R factor for this solution was 0·178. Omission of two further equations involving distributions with $|\rho_{l,l'}|$ in the range 0·5–0·6 led to a solution with $R =$ 0·148. Application of Hamilton's (1965) R-factor ratio test to this series of omissions led us to the conclusion that this solution was the best that could be achieved and no benefit was obtained when further equations were eliminated.

The resulting values for the $C_{n,m}$ are given in Table 3 together with their e.s.d.'s derived from the least-squares matrix used to solve (2). As confirmation of the fact that these $C_{n,m}$ provided a satisfactory solution to the set of equations (2) the values were used to calculate the intensity distribution in each section and plots of these distributions which are shown in Figs. 2(*c*), 3(*c*), 4(*c*), 5(*c*), and 6(*c*) are seen to be in generally good agreement with the observed X-ray patterns, although those for Figs. 6(*c*) and 2(*c*) are somewhat less good. It should be pointed out that while these results indicate that the presence of high correlations between some 2D distributions means that the individual values for the $\hat{x}'_l$ may not be reliable, their combined effect provides a good description of the 2D intensity distribution as shown in Figs. 2(*b*), 3(*b*), 4(*b*), 5(*b*) and 6(*b*). It is interesting to note that the one major area of disagreement is the area A in Fig. 6(*a*). The 2D-section analysis has to some extent been able to model this feature. The final solution, which must take account of the other sections also, is unable to account for this feature which we consequently believe is attributable to the poor performance of the empirical background procedure for this section.

The procedure of omitting those equations which involve high $\rho_{l,l'}$ is one which we would envisage would be useful in future work. It might be anticipated, however, that omission of too many equations may lead to a situation where insufficient equations remain to allow a meaningful solution. In this case, collection of data from further reciprocal sections would appear to be mandatory.

7. 3D analysis

For comparison with the 2D analysis described above we have also carried out an analysis by combining the data from the five reciprocal sections and then performing the least squares using 3D correlation distributions to yield $C_{n,m}$ values directly.

Since the data for each section are not on the same scale it is necessary to determine relative scale factors. This is not a simple matter since for any two sections which intersect along a line in reciprocal space only about 30–35 common reciprocal points were available and the accuracy of these data is intrinsically fairly low. In addition, each section was found in the 2D analysis to have a different residual background. It was not surprising therefore that attempts to use the common data for scaling failed to provide a satisfactory solution.

To obtain suitable scale factors we instead made use of the calculated intensities from the 2D analyses of each section, (*a*) made directly from the distribution coefficients, x_l, and (*b*) made from the $C_{n,m}$

obtained from the solution of (2). Calculations for (*a*) are on the scale of the observed data for each section (as are the plots of Figs. 2*b*–6*b*), while those for (*b*) are all on the same scale (as are the plots in Figs. 2*c*–6*c*). The two together provide the means by which the required scale factors may be obtained. This was done by comparing the intensities of suitable strong regions of diffuse intensity in the two sets of calculations. Note that this method of scaling effectively takes account of the level of the observed intensity over the whole of each reciprocal section rather than over the very limited regions covered by the common data. The data, multiplied by the derived scale factors, were then merged together into one computer file; and a complete set of 3D correlation distribution functions was calculated for use in the least-squares analysis.

As for the 2D refinements it was anticipated that a refinable constant would be required. In fact five constant functions were included in the analysis each of which was unity everywhere on one of the five reciprocal sections and zero on the others. In this way a constant for each different section could be refined while at the same time the fixed set of scale factors was maintained. The constants from the analysis agreed with those obtained from the 2D analysis within experimental error, except for the $(hk2\bar{h})$ section (see later). The scale factors used and the constants derived from the least-squares analysis are given in Table 2.

Although for each section crystals of approximately the same size (~0·5 mm diameter) were used it may be seen that the scale factors for some sections differ substantially from 1·0, particularly $(0kl)$. There are a number of factors which may affect the relative scale. These include the alignment of the X-ray collimation and monochromator system, the condition of the X-ray tube, the condition of the film developer, and the setting of the Weissenberg layer-screen gap. Since the different sections were recorded at different times over a period of several months, differences of scale were not unexpected.

As a result of this direct analysis of the 3D data using unit weights, the $C_{n,m}$ values listed in the third column of Table 3 were obtained together with e.s.d.'s obtained from the least-squares correlation matrix. R and R_{real} factors for each section of data from this refinement are given in Table 2. Plots of the intensity distributions in the five sections using these values for $C_{n,m}$ were very similar to those for the final 2D analysis shown in Figs. 2(*d*), 3(*d*), 4(*d*), 5(*d*) and 6(*d*) and consequently are omitted.

8. Results and discussion

The $C_{n,m}$ values obtained from the 2D and 3D analyses are listed in Table 3, together with their e.s.d.'s in parentheses. Table 3 also contains a column listing the mean of the 2D and 3D values, $\bar{C}_{n,m} = (C^{2D}_{n,m} + C^{3D}_{n,m})/2$, which may be considered our best estimate of the intermolecular correlation coefficients; and another column containing the differences, $\Delta = C^{3D}_{n,m} - C^{2D}_{n,m}$. Some of these differences appear to be quite large and it is necessary to consider whether these discrepancies are consistent with the e.s.d.'s σ_{2D} and σ_{3D} obtained from the analyses. In particular we note that the e.s.d.'s derived from the 3D least-squares matrix are much smaller than those derived from the solution of (2) for the 2D method.

If we assume that $C^{2D}_{n,m}$ and $C^{3D}_{n,m}$ are independent measurements of the same quantity and we expect each to have the same normal probability of error given by the standard deviation σ, then the statistic, $\delta_{n,m} = \Delta/2^{1/2}\sigma$, should be normally distributed with unit variance. This can be tested by use of the normal probability plot method of Abrahams & Keve (1971). Such probability plots were made; first for $\sigma = \sigma_{2D}$ and then for $\sigma = \sigma_{3D}$. In both cases the plots were acceptably linear and passed through the origin, indicating that the errors were normally distributed. The slopes of the plots, however, were 1·25 and 2·9 respectively. This indicates that the σ_{2D} and σ_{3D} both underestimated the true σ. To account satisfactorily for the observed discrepancies between $C^{2D}_{n,m}$ and $C^{3D}_{n,m}$ we should take as the estimate of their standard deviations $1{\cdot}25\sigma_{2D}$ and $2{\cdot}9\sigma_{3D}$ respectively, and consequently as an estimate of the standard deviation of $\bar{C}_{n,m}$ we may use $\bar{\sigma} = 2^{-1/2}(1{\cdot}25^2\sigma^2_{2D} + 2{\cdot}9^2\sigma^2_{3D})^{1/2}$, which is also listed in Table 3.

Underestimation of e.s.d.'s in situations where a significant proportion of the disagreement between observed and calculated intensities is not due to random errors has been discussed at length by a number of authors (Hamilton & Abrahams, 1970; Sakata & Cooper, 1979; Baharie & Pawley, 1983; Hill & Madsen, 1984). In particular, recent attention has been given to the e.s.d.'s which are derived from the Rietveld powder profile refinement method, a problem which has many features in common with the present work. In such cases the e.s.d.'s may often be underestimated by a factor in excess of two. In the present case the problem arises because the data are measured on a grid which is finely spaced relative to the rather slowly varying features of most distribution functions (particularly the most important low-order ones). Use of a large number of data points within one diffuse peak effectively only increases the precision of the measurement and does not increase the accuracy substantially if the major part of the error is caused by other systematic effects not included in the model. In the present experiment, the fact that the 3D method gives e.s.d.'s which are too small by a factor of 2·9 suggests that little benefit is gained by measuring the films on as fine a grid as that actually used. For the 2D method, the e.s.d.'s are an expression of the level

to which the derived values of $C_{n,m}$ agree between the different projections. The solution of (2) does not directly involve the systematic errors and consequently the derived e.s.d.'s are much closer to the correct values.

It is not clear that either the 2D or the 3D method of analysis offers any advantage over the other since both methods have their own problems. Provided that satisfactory scale factors can be obtained the 3D method is much simpler to use, but at the present stage it would seem preferable to use both methods where possible to provide a cross check on the accuracy of the results.

As a result of these considerations, of the 43 values for $\bar{C}_{n,m}$ given in Table 3, 18 were found to be significantly different from zero at the 1σ level and three at the 3σ level. In comparison to the previously published example of 2,3-dichloro-6,7-dimethylanthracene (Epstein & Welberry, 1983), this result may appear somewhat disappointing since in that work 13 correlations were reported to be significant at the 3σ level. It should be noted, however, that the current results suggest that the e.s.d.'s in that experiment were probably underestimated by ~25%, in which case only five $C_{n,m}$ would have been significant at the 3σ level and 15 at the 2σ level. In addition it should be noted that in that compound the low-order $C_{n,m}$ values were larger (there were six coefficients with values greater than 0·2 with the maximum being 0·46), and since the higher-order correlations will generally be dependent on the low-order ones this necessarily implies that a greater number will be significant. At present it appears that in our experiment we are able to derive $C_{n,m}$ values with an e.s.d. of about 0·02 and this value is reasonably independent of the magnitude or the order of the particular $C_{n,m}$.

Plots of the structure viewed in projection down each axis are included in Figs. 2(*e*), 3(*e*), 4(*e*), 5(*e*) and 6(*e*) for reference. Fig. 7 shows a stereo-plot of the molecular environment of the molecule at the site 55501. All neighbouring molecules which have a direct contact with 55501 involving the disordered Br and CH_3 sites are shown (with all molecules in orientation 1 - see Table 1). The remaining nearest neighbours, 55502, 56502, 55402 and 56402, which are only in direct contact with 55501 through their ethyl groups, are omitted for clarity. Table 4 lists all the intermolecular contact distances <4·5 Å which involve the Br or CH_3 groups, and these are also shown as dashed lines in Fig. 7.

The largest correlation involving 66502 is seen to result from only one contact of 4·24 Å between the disordered groups. With both molecules in orientation 1, as in Fig. 7, this is a CH_3–CH_3 contact. The observed negative correlation results in a decrease (compared to the case of random occupancy of the two sites) in the number of CH_3–CH_3 and Br–Br contacts in favour of CH_3–Br and Br–CH_3 as given in Table 5. This is in agreement with the previously observed trend that CH_3–Br contacts are energetically preferred to either CH_3–CH_3 or Br–Br contacts.

Table 4. *Contact distances* (Å) *between the disordered atomic sites in the central* 55501 *molecule with those in various neighbouring molecules*

Atom labels refer to all molecules in orientation 1.

Neighbour	Contact type	Contact distance
56501, 54501	Br–Br, Br–Br	4·45
	Br–CH_3, CH_3–Br	3·92
	CH_3–CH_3, CH_3–CH_3	4·45
65501, 45501	Br–Br	4·12
	Br–CH_3, CH_3–Br	3·91
66501, 44501	Br–Br	4·10
66502, 65502, 46402, 45402	CH_3–CH_3	4·24

The next largest correlation involving 65501 involves three different contacts between disordered substituent sites. With both molecules in orientation 1 these are a Br–CH_3 and a CH_3–Br contact, each of 3·91 Å, and a Br–Br contact of 4·10 Å. Using the simplistic view of the mechanism for the short-range order given above, it might be supposed here that the effects of the two shorter CH_3–Br contacts would outweigh the effect of the one rather longer Br–Br contact and result in a positive correlation. In fact the observed correlation is negative.

It is perhaps not surprising that this simplistic view of molecular ordering is inadequate to explain in detail our observed values for $C_{n,m}$ since it is well known, for example, that lattice-energy sums involving electrostatic interactions are slowly convergent and truncating the summation at too small an interaction distance can give misleading results. (see *e.g.* Kitaigorodskii, 1973, p. 144). Kitaigorodskii (1973, p. 160) has also pointed out that the main problem in obtaining lattice energies is not in computing the energy sums but in obtaining sufficient information regarding the charge distribution. In order to establish a more realistic model for the short-range order in molecular crystals such as BEMB1, it appears necessary to try to establish such a detailed description of the charge distribution over the whole

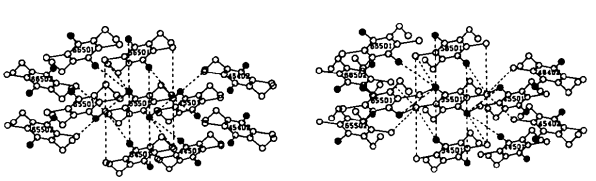

Fig. 7. A stereoscopic view of the molecular environment of the molecule at site 55501. All molecules are shown in orientation 1. Dark atoms represent bromine and light atoms are carbon.

Table 5. *The proportion of different types of molecular pairs occurring for pairs of neighbouring sites in BEMB*1, *compared with the proportion that would be present if each site were independently occupied by* m_A *molecules of type* 1 *and* $(1-m_A)$ *of type* 2. *The pair consists of the site listed together with the central site* 55501

Pair type	Site 66502	Site 65501	Site 56501	Random
1-1	0·262	0·279	0·320	0·314
1-2, 2-1	0·298	0·281	0·240	0·246
2-2	0·142	0·159	0·200	0·194

molecule and to carry out energy summations over a sufficiently large range of interaction distances. When this is done the results of the current experiment can be used together with the results from similar compounds to test such a theory.

We are grateful to K. Owen, M. Puza and G. Lockhart for technical assistance. One of us (JS) gratefully acknowledges the receipt of a Australian National University Scholarship during the tenure of which this work was carried out.

References

ABRAHAMS, S. C. & KEVE, E. T. (1971). *Acta Cryst.* A**27**, 157-165.
BAHARIE, E. & PAWLEY, G. S. (1983). *J. Appl. Cryst.* **16**, 404-406.
CENEDESE, P., BLEY, F. & LEFEBVRE, S. (1984). *Acta Cryst.* A**40**, 228-240.
EPSTEIN, J. & WELBERRY, T. R. (1983). *Acta Cryst.* A**39**, 882-892.
EPSTEIN, J., WELBERRY, T. R. & JONES, R. D. G. (1982). *Acta Cryst.* A**38**, 611-618.
HAMILTON, W. C. (1965). *Acta Cryst.* **18**, 502-510.
HAMILTON, W. C. & ABRAHAMS, S. C. (1970). *Acta Cryst.* A**26**, 18-24.
HILL, R. J. & MADSEN, I. C. (1984). *J. Appl. Cryst.* **17**, 297-306.
JOHNSON, C. K. (1976). *ORTEP* II. Report ORNL-5138. Oak Ridge National Laboratory, Tennessee.
JONES, R. D. G. & WELBERRY, T. R. (1980). *Acta Cryst.* B**36**, 852-857.
KITAIGORODSKII, A. I. (1973). *Molecular Crystals and Molecules.* New York: Academic Press.
SAKATA, M. & COOPER, M. J. (1979). *J. Appl. Cryst.* **12**, 554-563.
WELBERRY, T. R. (1983). *J. Appl. Cryst.* **16**, 192-197.
WELBERRY, T. R. & GLAZER, A. M. (1985). *Acta Cryst.* A**41**, 394-399.
WELBERRY, T. R. & JONES, R. D. G. (1980). *J. Appl. Cryst.* **13**, 244-251.
WELBERRY, T. R., JONES, R. D. G. & EPSTEIN, J. (1982). *Acta Cryst.* B**38**, 1518-1525.
WOOD, R. A., WELBERRY, T. R. & PUZA, M. (1984). *Acta Cryst.* C**40**, 1255-1260.

Acta Cryst. (1986). B**42**, 272-280

Three-Dimensional Structure of Yeast $tRNA^{Asp}$. I. Structure Determination

BY M. B. COMARMOND,* R. GIEGÉ, J. C. THIERRY AND D. MORAS†

Laboratoire de Cristallographie Biologique, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, 67084 Strasbourg CEDEX, France

AND J. FISCHER

Laboratoire de Cristallochimie, ERA 08, Institut Le Bel, 4 rue Blaise Pascal, 67070 Strasbourg CEDEX, France

(*Received* 5 *October* 1984; *accepted* 11 *November* 1985)

Abstract

The three-dimensional structure of yeast $tRNA^{Asp}$, an elongator tRNA with a short variable loop, has been solved to high resolution in two non-isomorphous but closely related orthorhombic crystal forms. A first attempt to solve the structure by molecular replacement using the molecular structure of yeast $tRNA^{Phe}$ as a model was unsuccessful although the tRNA molecule could be correctly oriented within 10° by the rotation function. The phase problem was solved by the MIR approach. Two heavy-atom derivatives (Gd and Au) were used as markers. A 3·5 Å resolution map could thus be interpreted and a Kendrew skeletal model built. The folding of the molecule is similar to that originally found for yeast $tRNA^{Phe}$. Major differences concern the relative positioning of the acceptor and anticodon stems: the more open conformation confers to the $tRNA^{Asp}$ molecule a boomerang-like shape. Crystal packing involves self-complementary GUC anticodon interactions through a crystallographic twofold axis. This is the first visualization of such an interaction at the molecular level.

Introduction

Transfer ribonucleic acids (tRNA's) form a family of small nucleic acids (70-90 nucleotides) crucial for living cells. Their most extensively studied and best known function is their role in messenger RNA mediated protein synthesis where they act as 'adaptor molecules' (see the review in Schimmel, Söll &

* Present address: Immunologie Structurale, Institut Pasteur, 28 rue du Dr Roux, 75015 Paris, France.
† To whom reprint requests should be addressed.

0108-7681/86/030272-09$01.50