the polymer-structure analyses cited above, as well as the layer-packing analysis of some phospholipids (Dorset, 1987).] While attempts at *ab initio* modelling of three-dimensional crystal structures, by searching an n-dimensional parameter space and seeking a global internal energy minimum, has remained an active research area, most success so far seems to have been realized with the prediction of two-dimensional layers (Scaringe, 1992). In general, for complicated unit cells, determination of a structure by trial and error is very difficult unless adequate constraints can be placed on the search.

Although Patterson techniques have been very useful in electron crystallography, there are also inherent difficulties in their use, particularly for locating heavy atoms. As will be appreciated from comparison of scattering-factor tables for X-rays [*IT* C (2004), Chapter 6.1] with those for electrons [*IT* C (2004), Chapter 4.3], the relative values of the electron form factors are more compressed with respect to atomic number than are those for X-ray scattering. As discussed in Chapter 2.3, it is desirable that the ratio of summed scattering-factor terms, $r = \sum_{\text{heavy}} Z^2 / \sum_{\text{light}} Z^2$, where Z is the scattering-factor value at $\sin\theta/\lambda = 0$, be near unity. A practical comparison would be the value of r for copper (DL-alaninate) solved from electron-diffraction data by Vainshtein *et al.* (1971). For electron diffraction, $r = 0.47$ compared to the value 2.36 for X-ray diffraction. Orientation of salient structural features, such as chains and rings, would be equally useful for light-atom moieties in electron or X-ray crystallography with Patterson techniques. As structures become more complicated, interpretation of Patterson maps becomes more and more difficult unless an automated search can be carried out against a known structural fragment (Chapter 2.3).

2.5.8.2. Direct phase determination from electron micrographs

The 'direct method' most familiar to the electron microscopist is the high-resolution electron micrograph of a crystalline lattice. Retrieval of an average structure from such a micrograph assumes that the experimental image conforms adequately to the 'weak phase object' approximation, as discussed in Section 2.5.5. If this is so, the use of image-averaging techniques, *e.g.* Fourier filtration or correlational alignment, will allow the unit-cell contents to be visualized after the electron-microscope phase contrast transfer function is deconvoluted from the average image, also discussed in Section 2.5.5. Image analyses can also be extended to three dimensions, as discussed in Section 2.5.6, basically by employing tomographic reconstruction techniques to combine information from the several tilt projections taken from the crystalline object. The potential distribution of the unit cell to the resolution of the imaging experiment can then be used, *via* the Fourier transform, to obtain crystallographic phases for the electron-diffraction amplitudes recorded at the same resolution. This method for phase determination has been the mainstay of protein electron crystallography.

Once a set of phases is obtained from the Fourier transform of the deconvoluted image, they must, however, be referred to an allowed crystallographic origin. For many crystallographic space groups, this choice of origin may coincide with the location of a major symmetry element in the unit cell [see *IT* A (2005)]. Hence, since the Fourier transform of translation is a phase term, if an image shift $[\delta(\mathbf{r} + \mathbf{r}_0)]$ is required to translate the origin of the repeating mass unit $\varphi(\mathbf{r})$ from the arbitrary position in the image to a specific site allowed by the space group,

$$\mathrm{g}(\mathbf{r}) = \varphi(\mathbf{r}) \otimes \delta(\mathbf{r} + \mathbf{r}_0) = \varphi(\mathbf{r} + \mathbf{r}_0),$$

where the operation '$\otimes$' denotes convolution. The Fourier transform of this shifted density function will be

$$G(\mathbf{s}) = F(\mathbf{s}) \exp(2\pi i \mathbf{s} \cdot \mathbf{r}_0) = |F(\mathbf{s})| \exp[i(\phi_s + 2\pi i \mathbf{s} \cdot \mathbf{r}_0)].$$

In addition to the crystallographic phases ϕ_s, it will, therefore, be necessary to find the additional phase-shift term $2\pi i \mathbf{s} \cdot \mathbf{r}_0$ that will access an allowed unit-cell origin. Such origin searches are carried out automatically by some commercial image-averaging computer-software packages.

In addition to applications to thin protein crystals (*e.g.* Henderson *et al.*, 1990; Jap *et al.*, 1991; Kühlbrandt *et al.*, 1994), there are numerous examples of molecular crystals that have been imaged to a resolution of 3–4 Å, many of which have been discussed by Fryer (1993). For π-delocalized compounds, which are the most stable in the electron beam against radiation damage, the best results (2 Å resolution) have been obtained at 500 kV from copper perchlorophthalocyanine epitaxically crystallized onto KCl. As shown by Uyeda *et al.* (1978–1979), the averaged images clearly depict the positions of the heavy Cu and Cl atoms, while the positions of the light atoms in the organic residue are not resolved. (The utility of image-derived phases as a basis set for phase extension will be discussed below.) A number of aromatic polymer crystals have also been imaged to about 3 Å resolution, as reviewed recently (Tsuji, 1989; Dorset, 1994*b*).

Aliphatic molecular crystals are much more difficult to study because of their increased radiation sensitivity. Nevertheless, monolamellar crystals of the paraffin n-$C_{44}H_{90}$ have been imaged to 2.5 Å resolution with a liquid-helium cryomicroscope (Zemlin *et al.*, 1985). Similar images have been obtained at room temperature from polyethylene (Revol & Manley, 1986) and also a number of other aliphatic polymer crystals (Revol, 1991).

As noted by J. M. Cowley and J. C. H. Spence in Section 2.5.1, dynamical scattering can pose a significant barrier to the direct interpretation of high-resolution images from many inorganic materials. Nevertheless, with adequate control of experimental conditions (limiting crystal thickness, use of high-voltage electrons) some progress has been made. Pan & Crozier (1993) have described 2.0 Å images from zeolites in terms of the phase-grating approximation. A three-dimensional structural study has been carried out on an aluminosilicate by Wenk *et al.* (1992) with thin samples that conform to the weak-phase-object approximation at the 800 kV used for the imaging experiment. Heavy and light (*e.g.* oxygen) atoms were located in the micrographs in good agreement with an X-ray crystal structure. Heavy-atom positions from electron microscopic and X-ray structure analyses have also been favourably compared for two heavy-metal oxides (Hovmöller *et al.*, 1984; Li & Hovmöller, 1988).

2.5.8.3. Probabilistic estimate of phase invariant sums

Conventional direct phasing techniques, as commonly employed in X-ray crystallography (*e.g.* see Chapter 2.2), have also been used for *ab initio* electron-crystallographic analyses. As in X-ray crystallography, probabilistic estimates of a linear combination of phases (Hauptman & Karle, 1953; Hauptman, 1972) are made after normalized structure factors are calculated *via* electron form factors, *i.e.*

$$|E_{\mathbf{h}}^2| = I_{\text{obs}} / \varepsilon \sum_i f_i^2, \text{ where } \langle |E|^2 \rangle = 1.000.$$

(Here, an overall temperature factor can be found from a Wilson plot. Because of multiple scattering, the value of B may be found occasionally to lie close to 0.0 Å^2.) The phase invariant sums

$$\psi = \phi_{\mathbf{h}_1} + \phi_{\mathbf{h}_2} + \phi_{\mathbf{h}_3} + \ldots$$

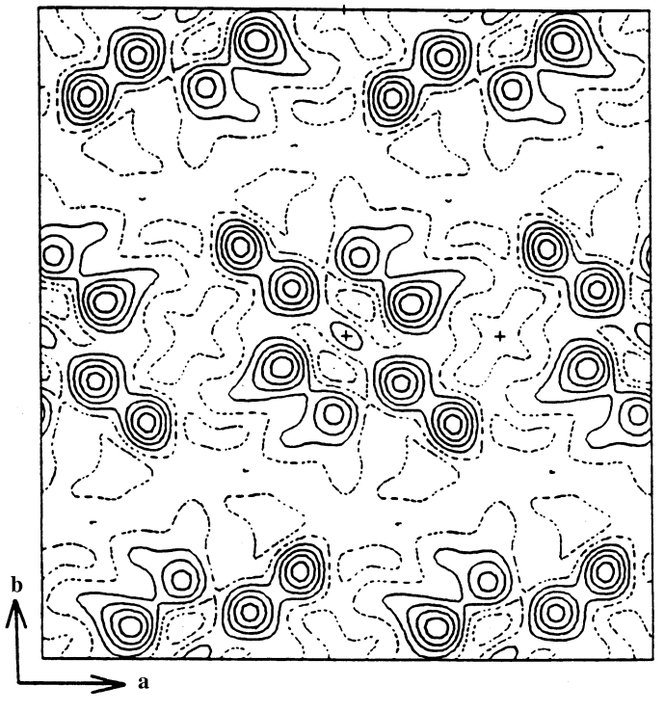

Fig. 2.5.8.1. Potential map for diketopiperazine ([001] projection) after a direct phase determination with texture electron-diffraction intensity data obtained originally by Vainshtein (1955).

can be particularly effective for structure analysis. Of particular importance historically have been the Σ_2-triple invariants where $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$ and $\mathbf{h}_1 \neq \mathbf{h}_2 \neq \mathbf{h}_3$. The probability of predicting $\psi = 0$ is directly related to the value of

$$A = (2\sigma_3/\sigma_2^{3/2})|E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_3}|,$$

where $\sigma_{\mathbf{h}} = \sum_{j=1}^{N} Z_j^n$ and Z is the value of the scattering factor at $\sin\theta/\lambda = 0$. Thus, the values of the phases are related to the measured structure factors, just as they are found to be in X-ray crystallography. The normalization described above imposes the point-atom structure (compensating for the fall-off of an approximately Gaussian form factor) often assumed in deriving the joint probability distributions. Especially for van der Waals structures, the constraint of positivity also holds in electron crystallography. (It is also quite useful for charged atoms so long as the reflections are not measured at very low angles.) Other useful phase invariant sums are the Σ_1 triples, where $\mathbf{h}_1 = \mathbf{h}_2 = -1/2\mathbf{h}_3$, and the quartets, where $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 = 0$ and $\mathbf{h}_1 \neq \mathbf{h}_2 \neq \mathbf{h}_3 \neq \mathbf{h}_4$. The prediction of a correct phase for an invariant is related in each case to the normalized structure-factor magnitudes.

The procedure for phase determination, therefore, is identical to the one used in X-ray crystallography (see Chapter 2.2). Using vectorial combinations of Miller indices, one generates triple and quartet invariants from available measured data and ranks them according to parameters such as A, defined above, which, as shown in Chapter 2.2, are arguments of the Cochran formula. The invariants are thus listed in order of their reliability. This, in fact, generates a set of simultaneous equations in crystallographic phase. In order to begin solving these equations, it is permissible to define arbitrarily the phase values of a limited number of reflections (three for a three-dimensional primitive unit cell) for reflections with Miller-index parity $hkl \neq ggg$ and $\sum_i h_i k_i l_i \neq ggg$, where g is an even number. This defines the origin of a unit cell. For noncentrosymmetric unit cells, the condition for defining the origin, which depends on the space group, is somewhat more complicated and an enantiomorph-defining reflection must be added.

In the evaluation of phase-invariant sums above a certain probability threshold, phase values are determined algebraically after origin (and enantiomorph) definition until a large enough set is obtained to permit calculation of an interpretable potential map (*i.e.* where atomic positions can be seen). There may be a few invariant phase sums above this threshold probability value which are incorrectly predicted, leading either to false phase assignments or at least to phase assignments inconsistent with those found from other invariants. A small number of such errors can generally be tolerated. Another problem arises when an insufficient quantity of new phase values is assigned directly from the phase invariants after the origin-defining phases are defined. This difficulty may occur for small data sets, for example. If this is the case, it is possible that a new reflection of proper index parity can be used to define the origin. Alternatively, $\phi_n = a, b, c \ldots$ algebraic unknowns can be used to establish the phase linkage among certain reflections. If the structure is centrosymmetric, and when enough reflections are given at least symbolic phase assignments, 2^n maps are calculated and the correct structure is identified by inspection of the potential maps. When all goes well in this so-called 'symbolic addition' procedure, the symbols are uniquely determined and there is no need to calculate more than a single map. If algebraic values are retained for certain phases because of limited vectorial connections in the data set, then a few maps may need to be generated so that the correct structure can be identified using the chemical knowledge of the investigator. The atomic positions identified can then be used to calculate phases for all observed data (*via* the structure-factor calculation) and the structure can be refined by Fourier (or, sometimes, least-squares) techniques to minimize the crystallographic R factor.

The first actual application of direct phasing techniques to experimental electron-diffraction data, based on symbolic addition procedures, was to two methylene subcell structures (an *n*-paraffin and a phospholipid; Dorset & Hauptman, 1976). Since then, evaluation of phase invariants has led to numerous other structures. For example, early texture electron-diffraction data sets obtained in Moscow (Vainshtein, 1964) were shown to be suitable for direct analysis. The structure of diketopiperazine (Dorset, 1991*a*) was determined from these electron-diffraction data (Vainshtein, 1955) when directly determined phases allowed computation of potential maps such as the one shown in Fig. 2.5.8.1. Bond distances and angles are in good agreement with the X-ray structure, particularly after least-squares refinement (Dorset & McCourt, 1994*a*). In addition, the structures of urea (Dorset, 1991*b*), using data published by Lobachev & Vainshtein (1961), paraelectric thiourea (Dorset, 1991*b*), using data published by Dvoryankin & Vainshtein (1960), and three mineral structures (Dorset, 1992*a*), from data published by Zvyagin (1967), have been determined, all using the original texture (or mosaic single-crystal) diffraction data. The most recent determination based on such texture diffraction data is that of basic copper chloride (Voronova & Vainshtein, 1958; Dorset, 1994*c*).

Symbolic addition has also been used to assign phases to selected-area diffraction data. The crystal structure of boric acid (Cowley, 1953) has been redetermined, adding an independent low-temperature analysis (Dorset, 1992*b*). Additionally, a direct structure analysis has been reported for graphite, based on high-voltage intensity data (Ogawa *et al.*, 1994). Two-dimensional data from several polymer structures have also been analysed successfully (Dorset, 1992*c*) as have three-dimensional intensity data (Dorset, 1991*c*,*d*; Dorset & McCourt, 1993).

Phase information from electron micrographs has also been used to aid phase determination by symbolic addition. Examples include the epitaxically oriented paraffins *n*-hexatriacontane (Dorset & Zemlin, 1990), *n*-tritriacontane (Dorset & Zhang, 1991) and a 1:1 solid solution of n-$C_{32}H_{66}$/n-$C_{36}H_{74}$ (Dorset,

1990*a*). Similarly, lamellar electron-diffraction data to *ca* 3 Å resolution from epitaxically oriented phospholipids have been phased by analysis of Σ_1 and Σ_2-triplet invariants (Dorset, 1990*b*, 1991*e*,*f*), in one case combined with values from a 6 Å resolution electron-microscope image (Dorset *et al.*, 1990, 1993). Most recently, such data have been used to determine the layer packing of a phospholipid binary solid solution (Dorset, 1994*d*).

An *ab initio* direct phase analysis was carried out with zonal electron-diffraction data from copper perchlorophthalocyanine. Using intensities from a *ca* 100 Å thick sample collected at 1.2 MeV, the best map from a phase set with symbolic unknowns retrieves the positions of all the heavy atoms, equivalent to the results of the best images (Uyeda *et al.*, 1978–1979). Using these positions to calculate an initial phase set, the positions of the remaining light C, N atoms were found by Fourier refinement so that the final bond distances and angles were in good agreement with those from X-ray structures of similar compounds (Dorset *et al.*, 1991). A similar analysis has been carried out for the perbromo analogue (Dorset *et al.*, 1992). Although dynamical scattering and secondary scattering significantly perturb the observed intensity data, the total molecular structure can be visualized after a Fourier refinement. Most recently, a three-dimensional structure determination was reported for C_{60} buckminsterfullerene based on symbolic addition with results most in accord with a rotationally disordered molecular packing (Dorset & McCourt, 1994*b*).

2.5.8.4. The tangent formula

Given a triple phase relationship

$$\phi_{\mathbf{h}} \simeq \phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}},$$

where $\mathbf{h}$, $\mathbf{k}$ and $\mathbf{h}-\mathbf{k}$ form a vector sum, it is often possible to find a more reliable estimate of $\phi_{\mathbf{h}}$ when all the possible vectorial contributions to it within the observed data set $\mathbf{k}_{\mathbf{r}}$ are considered as an average, *viz:*

$$\phi_{\mathbf{h}} \simeq \langle \phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}} \rangle_{\mathbf{k}_{\mathbf{r}}}.$$

For actual phase determination, this can be formalized as follows. After calculating normalized structure-factor magnitudes $|E_{\mathbf{h}}|$ from the observed $|F_{\mathbf{h}}|$ to generate all possible phase triples within a reasonably high $A_{\mathbf{h}}$ threshold, new phase values can be estimated after origin definition by use of the tangent formula (Karle & Hauptman, 1956):

$$\tan \phi_{\mathbf{h}} = \frac{\sum_{\mathbf{k}_{\mathbf{r}}} W_{\mathbf{h}} |E_{\mathbf{k}}||E_{\mathbf{h}-\mathbf{k}}| \sin(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}_{\mathbf{r}}} W_{\mathbf{h}} |E_{\mathbf{k}}||E_{\mathbf{h}-\mathbf{k}}| \cos(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})}.$$

The reliability of the phase estimate depends on the variance $V(\phi_{\mathbf{h}})$, which is directly related to the magnitude of $\alpha_{\mathbf{h}}$, *i.e.*

$$\alpha_{\mathbf{h}}^2 = \left[\sum_{\mathbf{k}_{\mathbf{r}}} A_{\mathbf{h},\mathbf{k}} \cos(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})\right]^2 + \left[\sum_{\mathbf{k}_{\mathbf{r}}} A_{\mathbf{h},\mathbf{k}} \sin(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})\right]^2;$$

$A_{\mathbf{h},\mathbf{k}}$ is identical to the A value defined in the previous section. In the initial stages of phase determination $\alpha_{\mathbf{h}}$ is replaced by an expectation value α_{E} until enough phases are available to permit its calculation.

The phase solutions indicated by the tangent formula can thus be ranked according to the phase variance and the determination of phases can be made symbolically from the most probable triple-product relationships. This procedure is equivalent to the one described above for the evaluation of phase-invariant sums by symbolic addition. This procedure may allow determination of a large enough basis phase set to produce an interpretable map.

An alternative procedure is to use an automatic version of the tangent formula in a multisolution process. This procedure is described in Chapter 2.2. After origin definition, enough algebraic unknowns are defined (two values if centrosymmetric and four values, cycling through phase quadrants, if noncentrosymmetric) to access as many of the unknown phases as possible. These are used to generate a number of trial phase sets and the likelihood of identifying the correct solution is based on the use of some figure of merit.

Multisolution approaches employing the tangent formula include *MULTAN* (Germain *et al.*, 1971), *QTAN* (Langs & DeTitta, 1975) and *RANTAN* (Yao, 1981). *RANTAN* is a version of *MULTAN* that allows for a larger initial random phase set (with suitable control of weights in the tangent formula). *QTAN* utilizes the $\alpha_{\mathbf{h}_{\mathrm{est}}}$ definition, where

$$\alpha_{\mathbf{h}_{\mathrm{est}}} = \left\{\sum_{\mathbf{k}} A_{\mathbf{h},\mathbf{k}}^2 + 2 \sum_{\mathbf{k} \neq} \sum_{\mathbf{k}'} A_{\mathbf{h},\mathbf{k}} A'_{\mathbf{h},\mathbf{k}} \frac{I_1(A_{\mathbf{h},\mathbf{k}}) I_1(A'_{\mathbf{h},\mathbf{k}})}{I_0(A_{\mathbf{h},\mathbf{k}}) I_0(A'_{\mathbf{h},\mathbf{k}})}\right\}^{1/2},$$

for evaluating the phase variance. (Here I_0, I_1 are modified Bessel functions.) After multiple solutions are generated, it is desirable to locate the structurally most relevant phase sets by some figure of merit. There are many that have been suggested (Chapter 2.2). The most useful figure of merit in *QTAN* has been the NQEST (De Titta *et al.*, 1975) estimate of negative quartet invariants (see Chapter 2.2). More recently, this has been superseded by the minimal function (Hauptman, 1993):

$$R(\phi) = \frac{\sum_{\mathbf{h},\mathbf{k}} A_{\mathbf{h},\mathbf{k}} (\cos \phi_{\mathbf{h},\mathbf{k}} - t_{\mathbf{h},\mathbf{k}})^2}{\sum_{\mathbf{h},\mathbf{k}} A_{\mathbf{h},\mathbf{k}}},$$

where $t_{\mathbf{h},\mathbf{k}} = I_1(A_{\mathbf{h},\mathbf{k}})/I_0(A_{\mathbf{h},\mathbf{k}})$ and $\phi_{\mathbf{h},\mathbf{k}} = \phi_{\mathbf{h}} + \phi_{\mathbf{k}} + \phi_{-\mathbf{h}-\mathbf{k}}$.

In the first application (Dorset *et al.*, 1979) of multisolution phasing to electron-diffraction data (using the program *QTAN*), n-beam dynamical structure factors generated for cytosine and disodium 4-oxypyrimidine-2-sulfinate were used to assess the effect of increasing crystal thickness and electron accelerating voltage on the success of the structure determination. At 100 kV samples at least 80 Å thick were usable for data collection and at 1000 kV this sample thickness limit could be pushed to 300 Å – or, perhaps, 610 Å if a partial structure were accepted for later Fourier refinement. NQEST identified the best phase solutions. Later *QTAN* was used to evaluate the effect of elastic crystal bend on the structure analysis of cytosine (Moss & Dorset, 1982).

In actual experimental applications, two forms of thiourea were investigated with *QTAN* (Dorset, 1992*d*), using published three-dimensional electron-diffraction intensities (Dvoryankin & Vainshtein, 1960, 1962). Analysis of the centrosymmetric paraelectric structure yielded results equivalent to those found earlier by symbolic addition (Dorset, 1991*b*). Analysis of the noncentrosymmetric ferroelectric form was also quite successful (Dorset, 1992*d*). In both cases, the correct structure was found at the lowest value of NQEST. Re-analysis of the diketopiperazine structure with *QTAN* also found the correct structure (Dorset & McCourt, 1994*a*) within the four lowest values of NQEST, but not the one at the lowest value. The effectiveness of this figure of merit became more questionable when *QTAN* was used to solve the noncentrosymmetric crystal structure of a polymer (Dorset, McCourt, Kopp *et al.*, 1994). The solution could not be found readily when NQEST was used but was easily identified when the minimal function $R(\phi)$ was employed instead.

MULTAN has been used to phase simulated data from copper perchlorophthalocyanine (Fan *et al.*, 1985). Within the 2 Å