

17.X-01 ADVANCES IN DIRECT METHODS FOR SMALL AND MEDIUM-SIZE MOLECULES. By M M Woolfson, Department of Physics, University of York, Heslington, York YO1 5DD, U K.

For small equal-atom structures crystallography is now dominated by automated direct methods. Their main tool is the triple-phase (or sign) relationship supplemented by the tangent formula (t.f.) for phase development and refinement. Other relationships are only marginally useful. Where these methods fail it is often because of uncertainties in the early stages of phase development when few reflexions and relationships are involved. Some recently developed methods use, ab initio, larger sets of reflexions.

(i) MAGIC By the primary-secondary concept some 40-50 phases may be represented in magic integer form. Plausible sets of initial phases are indicated by a ψ map and then extended by the t.f.

(ii) MAGEX Secondary reflexions are given multiple representation in terms of primaries and the method is very insensitive to individual poor relationships. Limited multi-pathway phase extension is available before t.f. phase extension is initiated.

(iii) YZARC Random values are allocated to about 100 phases. These are refined by a least-squares or steepest-descents process, treating relationships as linear equations, before t.f. extension.

(iv) RANTAN Several hundred phases are given random values and then refined by the controlled use of the t.f.

Further progress in solving complex structures may be possible if better estimates are available for the reliability of phase relationships. This is the basis of a great deal of present work.

17.X-02 ADVANCES IN DIRECT PHASING OF LARGE-MOLECULE STRUCTURES. By D. Sayre, IBM Research Center, Yorktown Heights, New York 10598, USA.

A survey of recent developments, mainly in phase extension and refinement in direct and reciprocal space.

17.X-03 PROGRESS AND PROBLEMS IN CRYSTALLOGRAPHIC STATISTICS - GENERAL REVIEW. By Herbert Hauptman, Medical Fndn. of Buffalo, Inc., Buffalo, NY 14203.

With few exceptions the applications of statistics in the physical sciences require first the definition of a suitable random variable, the derivation of its probability distribution, the calculation of some of its parameters, usually the mean and variance, and finally a comparison of theoretical and observed values of these parameters.

For a fixed crystal structure the intensity of a reflection I_H is a function of H , so that, if the latter is assumed to be the primitive random variable uniformly distributed in reciprocal space, the former is also a random variable. Probability distributions for the intensity of a reflection were first obtained in 1949 for the space groups $P1$ and $P1$. By taking into account the effects of the various symmetry elements on these distributions, the latter may serve as the basis for identifying the space group of a crystal via the statistical analysis of the observed intensities. Again, these distributions also serve as the basis of a proper statistical method for estimating the values and standard deviations of the intensities of unobserved reflections. If a large number of reflections are unobserved because they are weak, the procedure may serve an important function in the determination of the crystal structure by direct methods which usually require a great over-redundancy of data for success.

Next, if the reciprocal lattice vector H is fixed and the atomic position vectors r_j are the primitive random variables, assumed to be uniformly and independently distributed in the unit cell, then the normalized structure factors E_H and G_H for an isomorphous pair of structures are functions of the r_j 's and therefore themselves random variables. The joint probability distribution of the pair E_H, G_H has been found, even in the case of imperfect isomorphism. Thus the case that the G-structure is a trial, or model, structure, possibly incomplete, of the "true" E-structure is also included. In the latter case the distribution of the normalized discrepancy index $R = \frac{\langle |E_H| - |G_H| \rangle}{\langle |E_H| \rangle}$ may be worked out and leads to an estimate of the average error $\langle |\Delta r| \rangle$ of the model; thus the distribution of R plays an important role in the process of refinement of crystal structures. In the former case the joint probability distribution of the pair E_H, G_H leads to formulas for the moments of the sum, difference, and product variables, as well as the correlation coefficient of the pair $|E_H|^2, |G_H|^2$. Comparison of these with the observed values, in particular their variation with $\sin\theta/\lambda$, leads to methods for the assessment of the degree of isomorphism between the E- and G-structures, with application to protein crystallography for which the isomorphous replacement technique is the most important tool.

Next, the electron density function ρ is represented by a Fourier series the coefficients of which are the structure factors $F = |F| \exp(i\phi)$. In practice the magnitudes $|F|$ are obtained from experiment and are therefore subject to error; the phases ϕ are derived from the observed magnitudes $|F|$ and are therefore subject to additional errors the distribution of which depends on the nature of the phase determination process. Thus the structure factor F is a random variable, and its probability distribution determines how the Fourier coefficients are to be weighted in order to yield, in some sense, the "best" electron density function ρ . Similar remarks apply, for example, to the use of the Patterson function to locate heavy atom positions employing data from a pair of isomorphous crystals.

Finally, Wilson has recently obtained the probability $p(R_0)$ that a reflection of true intensity R will have an observed value R_0 (possibly negative) for four different counting modes. These distributions have potential application in estimating a more likely positive value for the intensity of a reflection actually measured as negative, improving the estimate of structural parameters, and detecting the presence of structural imperfection or systematic error.

Research supported by NSF Grant No. CHE79-11282 and NIGMS Grant No. GM-26195.