17. COMPUTATIONAL METHODS AND ERROR ANALYSIS


The earliest probability considerations in crystal structure analysis were introduced by Wilson (Nature [1942] 150, 152) who showed that intensity data could be corrected for vibrational effects and placed on an absolute scale by use of a statistical relationship between the intensities and the atomic scattering factors. This relationship is employed for the preparation of experimental data for direct phase determining procedures. Wilson pointed out that the expected values for absolute intensities is given by

$$<I^2_b> = \sum_{j=1}^{N} I_j f_j^2$$

where $f_j$ is the atomic scattering factor for the $j$th atom in a unit cell containing $N$ atoms. He also noted that the observed intensities $I_{obs}$ differ from the $I_j$ by a scale factor, $k$ independent of $I_j$, and a temperature factor. By use of the above relation the observed data can be rescaled and corrected for positional disorder. Wilson (Acta Cryst. [1949] 2, 195) also found a very good approximation, probability distributions for the intensities in non-centrosymmetric and centrosymmetric crystals that could be used to determine the presence or absence of a center of symmetry. A number of studies followed in the attempt to take account of the presence of heavy atoms, special positions, and the regularities that occur in crystal structures. One such study involved the use of the joint probability distribution (Hauptman and Karle, Acta Cryst. [1957] 5, 136).

It had an additional purpose, namely, to develop a facility in deriving joint probability distributions to be associated with phase determining relations.

The basic phase determining relations, used in procedures for phase determination have had their origins in inequality theory (Harker and Kasper, Acta Cryst. [1948] 1, 76; Karle and Hauptman, Acta Cryst. [1950] 3, 171). Inequality theory per se, is only applicable to simple structures. However, it was apparent that the inequality relationships could be extended as probabilistic formulas in order to have formulas applicable to complex structures. For example, it was observed by Cullis (Acta Cryst. [1948] 1, 174) that inequalities of Harker and Kasper still gave correct phase indications even when the appropriate phase factor magnitudes were of too small a magnitude to satisfy the conditions of application of the inequalities. The appropriate mathematical tool for developing the probabilistic aspects of the inequality theory was the joint probability distribution. The first series of investigations concerned centrosymmetric crystals (Hauptman and Karle [1957] Monograph No. 3, American Crystallographic Association, Polytechnic Book Service; Western Springs). Further application of probability theory continued over the years, thus establishing the mathematical foundations of direct methods. The development involves the non-negativity of the electron density distribution, atomicity and the overdeterminacy of the experimental data. The foundation mathematics which is comprised of the phase determining formulas from inequality theory and their probabilistic implications takes proper account of the presence of atoms of unequal atomic number in crystals.

17. X-6 FOURIER REPRESENTATIONS OF PROBABILITY DISTRIBUTIONS IN INTENSITY STATISTICS. By G. H. Weiss U. Shmueli², J. E. Kiefer, and A. J. C. Wilson². 1, National Institutes of Health, Bethesda, MD USA; 2, Tel-Aviv University, Tel-Aviv, Israel; 3, Crystallographic Data Centre, Cambridge, England.

The study of intensity statistics has hitherto been based on a Gaussian distribution for the normalized structure factor, or on expansions in terms of polynomials orthogonal with respect to the Gaussian. These expansions are necessarily approximate and necessitate increasingly complicated calculations as the degree of heterogeneity increases. Barakat has shown that exact solutions of the corresponding random walk problems can be written as a Fourier or Fourier-Bessel series expansion, the coefficients of which are readily computable for many space groups. He showed, in particular that the series for space groups P1 and P1 are expressible in terms of products of Bessel functions, and that the resulting series present no serious numerical problems for even large degrees of heterogeneity. These results are readily generalized to deal with the bicentric distribution for the P1 space group. When the atoms in the unit cell are all equal one can readily derive the known neglect-Wilson approximation, but our results are exact for any degree of heterogeneity. Other space groups can also be analyzed by the Fourier series technique.

17. X-7 CENTRIC, PARTIALLY BICENTRIC AND BICENTRIC INTENSITY STATISTICS IN THE PRESENCE OF HEAVY ATOMS. By Uri Shmueli, Department of Chemistry, Tel Aviv University, 69 978 Tel Aviv, Israel, and George H. Weiss, National Institutes of Health, Bethesda, MD 20205, USA.

Generalized statistical methods, capable of dealing with problems due to heterogeneous atomic composition, effects of space-group symmetry and others, fall into two classes: (1) expansions of the probability density functions (pdf's) for the normalized structure amplitude, \[ \hat{I} \], in terms of orthogonal polynomials and coefficients depending on the problem (e.g., Shmueli & Wilson, Acta Cryst. A 37, 342-353 (1981)); and (2) Fourier series for \[ \hat{I} \], based on the exact solution to the problem of random walk, with steps of unequal size (Shmueli, Weiss, Kiefer & Wilson 1984). The pdf's of class (1) are now available for all the space groups, as four-term expansions, for the case of atoms in general positions and no symmetry. However, these four terms may not suffice for extreme atomic heterogeneities, especially in space groups of the lowest symmetries. This gap of applicability is now very effectively filled by the exact pdf's of class (2), which are available for low-symmetry space groups.

The random-walk approach has recently been extended to the treatment of bicentric distributions and work on partially bicentric distributions is now in progress.

The above two approaches will be compared with respect to their performance. The test distributions comprise (i) computer simulations, (ii) \[ \hat{I} \] values recalculated from published structures and (iii) experimental distributions pertaining to centric and partially bicentric organic macromolecules.

---

²These authors were supported by National Institutes of Health, Bethesda, MD USA.