30 02-Methods for Structure Determination and Analysis, Computing and Graphics

The further development of the EDH based approach is to consider the set of conditional histograms calculated from the points in the unit cell satisfying some additional restrictions. Such additional constraints may imply position of a point with respect to the molecular region or the values of some other functions connected with the object under investigation.

PS-02.01.08 ELECTRON DENSITY SQUARING METHOD AND NON-CRYSTALLOGRAPHIC SYMMETRY. By A.F.Mishnev, Latvian Institute of Organic Synthesis, Riga, Latvia

Non-crystallographic symmetry imposes restraints on phases of the structure factors. Linear relationships among structure factors due to identical molecules in different crystallographic environment have been obtained by Main & Rossmann (Acta Cryst., 1966, 21, 67-72). For a structure containing like atoms the electron density squaring method (Sayre, Acta Cryst., 1952, 5, 60-65) may be introduced in the analysis of Main & Rossmann, that results in quadratic equations for the structure factors. In the presence of non-crystallographic symmetry the structure factor of the "squared" crystal takes the form

$$G_p = \sum_{n=1}^{N} \int_U \rho_1^2(x_1) \cdot \exp\{2\pi i([C_n]x_1 + d_n)\} dx_1 . \qquad (1)$$

The "squared" structure factor may be expressed by $G_p = g_p/f_p \cdot F_p$. Let $\rho_2(x)$ be the electron density in a second crystal, which contains the same molecule. Since $\rho_2(x_1) = \rho_1(x_1)$ by definition, one can obtain the equations

$$F_p = \frac{f_p}{g_p V_K V_H} \sum_K \sum_H F_K \cdot F_H \cdot S_{KHP}, \qquad (2)$$

where S_{PKH} are functions of molecular envelope, rotation and translation parameters. When the two crystals are identical equation (2) reduces to the Sayre's equation. Numerical test calculations of equations (2) using simulated crystal data will be presented.

PS-02.01.09 DIRECT PHASING FOR MACROMOLECULES BY ENTROPY MAXIMISATION AND LIKELIHOOD RANKING. By G. Bricogne, Department. of Molecular Biology, Biomedical Centre, Box 590, 751 24 Uppsala, Sweden ; and LURE, Bâtiment 209D, 91405 Orsay, France.

A new multisolution phasing method based on entropy maximisation and likelihood ranking, proposed for the specific purpose of extending probabilistic direct methods to the field of macromolecules [Bricogne (1984). *Acta Cryst.* A**40**, 410-445], has been implemented in two different computer programs [Bricogne & Gilmore (1990). *Acta Cryst.* A**46**, 284-297; Bricogne (1993). *Acta Cryst.* D**49**, 37-60] and applied to a wide variety of problems. The latter comprise the determination of small crystal structures from X-ray diffraction data obtained from single crystals [Gilmore, Bricogne & Bannister (1990). *Acta Cryst.* A**46**, 297-308] or from powders [Bricogne (1991). *Acta Cryst.* A**47**, 803-829; Gilmore, K. Henderson & Bricogne (1991). *Acta Cryst.* A**47**, 830-841; Shankland, Gilmore, Bricogne & Hashizume (1993). *Acta Cryst.* A**49**, in the press], and from electron diffraction data partially phased by image processing of electron micrographs [Dong *et al.* (1992). *Nature, Lond.*, **355**, 605-609] or even unphased [Gilmore, Shankland & Bricogne (1993). Submitted to *Proc. R. Soc. London Ser. A.*]; the *ab initio* generation [Bricogne (1993). *Acta Cryst.* D**49**, 37-60] and ranking [Gilmore, A.N. Henderson & Bricogne (1991). *Acta Cryst.* A**47**, 842-846] of phase sets for small proteins; and the improvement of poor quality phases for a larger protein at medium resolution under constraint of solvent flatness [Xiang, Carter, Bricogne & Gilmore (1993). *Acta Cryst.* D**49**, 193-212]. These applications show that the primary goal of this new method – namely increasing the accuracy and sensitivity of probabilistic phase indications compared with conventional direct methods – has been achieved.

The main components of the method as implemented in the computer program BUSTER [Bricogne (1993). *Acta Cryst.* D**49**, 37-60] are (1) a tree-directed search through a space of trial phase sets; (2) the saddlepoint method for calculating joint probabilities of structure factors, using entropy maximisation; (3) likelihood-based scores to rank trial phase sets and prune the search tree; (4) a new method for optimising the choice of reflexions so as to maximise the sensitivity of the likelihood to their phases; (5) efficient schemes, based on error-correcting codes, for sampling trial phase sets; (6) a statistical analysis of the scores for automatically selecting reliable phase indications by multidimensional Fourier techniques coupled with tests of statistical significance. This program has been successfully tested on two small structures and has been applied to data from two small proteins.

The mathematical techniques now available in BUSTER bring closer a number of major enhancements of standard macromolecular phasing methods proposed earlier [Bricogne (1988). *Acta Cryst.* A**44**, 517-545] as an extension of the initial theory. In the molecular replacement method, for instance, the detection and placement of a known fragment described in a reference position and orientation by a density ρ^M with transform F^M can be accomplished by calculating the log-likelihood gain :

$$\mathrm{LLG}(\mathbf{R}, \mathbf{t}) = \log \frac{\mathcal{P}\left(\left|F_{\mathbf{h}}\right| = \left|F_{\mathbf{h}}\right|^{\mathrm{obs}} \text{ for all } \mathbf{h} \mid (\mathcal{H}_1[\mathbf{R}, \mathbf{t}]) \right)}{\mathcal{P}\left(\left|F_{\mathbf{h}}\right| = \left|F_{\mathbf{h}}\right|^{\mathrm{obs}} \text{ for all } \mathbf{h} \mid (\mathcal{H}_0) \right)}$$

where $(\mathcal{H}_0)$ denotes the null hypothesis that all atoms are uniformly distributed in the asymmetric unit while $(\mathcal{H}_1[\mathbf{R}, \mathbf{t}])$ denotes the alternative hypothesis that the known fragment is placed in the asymmetric unit with orientation $\mathbf{R}$ at position $\mathbf{t}$, and the rest of the atoms are distributed at random. A drastic simplification of LLG yields a sum of (1) a Patterson correlation (PC) - based rotation function in which a sum of point-group symmetry-related copies of the self-Patterson of the rotated fragment is correlated with the origin-removed self-Patterson of the whole structure; and (2) a PC-based translation function, expressed as a Fourier series with argument $\mathbf{t}$ itself. This function is already an improvement on the PC functions used in XPLOR [Brünger (1990). *Acta Cryst.* A**46**, 46-57], yet it is in general a poor approximation to LLG. It will be shown how the systematic use of LLG and of its relations to Bayesian statistical methods yields a new procedure for the detection and accurate placement of a known molecular fragment and of its recycling into the phasing process which overcomes every single limitation of the current methodology [Bricogne (1993). In *The Molecular Replacement Method*, edited. by W. Wolf, E.J. Dodson & S. Gover. Warrington: SERC Daresbury Laboratory, in the press].

PS-02.01.10 FOURIER-TRANSFORM-BASED METHODS FOR PHASE EXTENSION AND REFINEMENT AND PERHAPS THE SOLUTION OF MACROMOLECULES. By L Refaat, C Tate and M M Woolfson*, Department of Physics, University of York, UK.

The Sayre equation is known to be effective for phase extension and refinement, either alone (Sayre, D, 1972, Acta Cryst A28, 210-212) or in conjunction with other constraints, as in the SQUASH procedure (Main, P, 1990, Acta Cryst A46, 372-377).

A method is described by which the coefficients of a set of linear equations are derived, solely from FFT operations, leading to phase

modification based on satisfying the Sayre equation. In the event that the structure contains one type of heavy atom then a modified equation can be used which involves both squaring and cubing the current electron density (Woolfson, M M, 1958, Acta Cryst 11, 287-283).

A process has been devised, the ABC method, which enables density to be modified to satisfy the constraints of solvent flattening, histogram matching, Sayre's equation and the magnitudes of structure factors, all in terms of refining only a few (4-6) parameters. It may also be possible to introduce other constraints, for example a model distribution of atomic environments.

Preliminary results will be described and their potentiality discussed - in particular for *ab initio* phasing.

PS-02.01.11 A DENSITY MODIFICATION PROCEDURE FOR SOLVING SMALL & MIDDLE SIZE STRUCTURES AND PHASE REFINEMENT FOR PROTEINS. By M.Shiono, Y.Yada*, Department of Physics, Kyushu University, Higashi-ku, Fukuoka, Japan. L. S. Refaat and M. M. Woolfson, Department of Physics, University of York, Heslington, York, YO1 5DD.

The Low Density Elimination (LDE) procedure (Shiono, M. and Woolfson, M. M., *Acta Cryst.*(1992),A48,451-456) which was developed for phase extension and refinement in order to solve macromolecules has been investigated regarding its power to solve small and middle size structures starting from random phase sets. In fact, the method is competitive against conventional direct methods. The LDE method, however, is time-consuming compared with conventional direct methods (e.g. MULTAN) since the procedure includes two Fourier transforms in one cycle. We have, therefore, combined MULTAN and LDE procedure. The LDE can be run in three different modes as follows.
Mode 1. Run the LDE with phases estimated by anomalous scatterings or isomorphous replacements. Mode 2. Employing multi-solution strategy, run the LDE individually assigning all reflexions random phases. Mode 3. Run MULTAN and then proceed to the LDE using MULTAN phases as initial phase sets in order of figures of merit.
For small and middle size structures, mode 3 is most effective. We might eventually solve the structures with MULTAN trial. Even if MULTAN fails to find any useful structural configurations, MULTAN phases increase the power of the LDE in solving structures.

PS-02.01.12 DIRECT PHASING OF MACROMOLECULAR STRUCTURES BY MULTIPLE BEAM DIFFRACTION
E. Weckert*, W. Schwegle and K. Hümmer
Inst. f. Kristallographie, Universität, D-7500 Karlsruhe 1, Kaiserstr. 12, Germany

The feasibility of experimental phase determination of small protein structures using three-beam diffraction has already been demonstrated (K. Hümmer, W. Schwegle & E. Weckert (1991) Acta Cryst. **A47**. 60-62). It has been shown that triplet-phase invariants $\phi = -\varphi(\mathbf{h}) + \varphi(\mathbf{g}) + \varphi(\mathbf{h}-\mathbf{g})$ can be deduced from three-beam interference profiles (K. Hümmer, E. Weckert & H.Bondza (1989) Acta Cryst. **A45**, 182-187), where the φ's are the phases of individual structure factors of the involved reflections with recipriocal lattice vectors **h**, **g**, and **h-g**.
In experimental phase determination of macromolecular structure significant differences compared to small molecule structures occur. Among others they concern the weaker scattering power of individual reflections, increasing overlap of multiple-beam interference effects due to larger unit cells and in general higher sensitivity to radiation damage.

Because of the large number of overlapping multiple-beam interference patterns in protein crystal structures only three-beam cases of reflections with large structure factors are suitable for phase determination. It was possible to determine about 80 triplet phases of the small protein lysozyme in the low and medium resolution range with a mean phase error of about 17°.
Radiation damage can often be significantly reduced by using higher energy radiation, i.e. $\lambda = 0.7$ Å . Therefore, interference effects of lysozyme were systematically investigated in the range from 0.7 Å to 1.58 Å . As a result also in the short wavelength regime phase determination is possible.
Theoretical calculations by dynamical theory and experimental results confirm the existence of three-beam interference effects even for crystal sizes smaller than the "Pendellösung" lengths. Further investigations show that it is this range where an unique correlation exists between the interference profiles and the triplet phases independent whether the primary reflection is in Bragg- or Laue-diffraction geometry. In general the crystal size of proteins is smaller than the "Pendellösung" length.
Due to the weak reflectivity of protein crystals "Aufhellung" effects are weaker compared to the higher reflectivity of small molecule structures. Therefore, the high number of overlapping three-beam cases does not severely affect the phase exploitation in proteins.
First experiments with catalase crystals (space group $P4_22_12$, a = 106.7Å , c = 106,3 Å) indicate that even for large proteins experimental phase determination may be feasible.
The possibilities to integrate measured triplet-phase invariants into statistical structure solution methods are discussed. First results to extend this phase information by an approach using maximum entropy will be presented.

This work has been funded by the German Federal Minister of Research and Technology (BMFT) under the contract No. 05 5VKIXI and the Deutsche Forschungsgemeinschaft.

PS-02.01.13 **FOURIER SERIES PROBABILITY DISTRIBUTION OF STRUCTURE FACTORS AND ITS RELATION WITH OTHER DISTRIBUTIONS.** By G. B. Mitra*, CSS Department, Indian Association for the Cultivation of Science, Calcutta-700 032, India and Sabita Das, Victoria Institution (College), Calcutta-700 009, India.

In his pioneering work, Wilson (Acta Cryst, 1949, 2, 318), showed that the distribution of structure factor components was Gaussian. Later, introduction of Edgeworth series (Mitra and