48 02-Methods for Structure Determination and Analysis, Computing and Graphics

02.07 – Phasing and Refinement of Macromolecular Structures

DS–02.07.01 ENTROPY MAXIMIZATION CONSTRAINED BY SOLVENT FLATNESS: MACROMOLECULAR PHASE EXTENSION AND REFINEMENT *C. W. Carter, Jr., S. Xiang, S. Doublié, .G. Bricogne‡, and C. J. Gilmore+ Dept. of Biochemistry and Biophysics, UNC, Chapel Hill, North Carolina 27599-7260; ‡Lure, Orsay 91405 France and Dept. of Molecular Biology, BMC, Uppsala; SE +Department of Chemistry, Glasgow University Glasgow, UK

Entropy and likelihood maximization constrained by solvent flatness outside a well-defined molecular envelope (Xiang, et al., 1993 Acta *Cryst.* **D49:**193-212) is an implementation of Bayesian phasing methods for macromolecular crystal structure determination. Tests with both simulated and experimental data confirm that this model-independent phase refinement path remains faithful to the information provided by a basis set of reliable experimental phases, and provides optimal phases for extrapolated reflections outside the basis set. The algorithm can be iterated advantageously in the same fashion as with conventional solvent flattening, by recombining maximum entropy and initial experimental phases and redefining the molecular envelope.

The resulting improvement of the electron density lies almost directly along the path between initial and target maps, avoiding substantial errors that can be introduced by conventional solvent flattening. It is thus a conservative density modification algorithm, making minimally committal departures from the constraints, as expected from the maximum entropy criterion.

Maximum-entropy extrapolation also provides a statistic, the Log-likelihood gain, evaluated over reflections outside the basis set, that is in many cases an accurate figure of merit for the phases in the basis set, analogous to the "Free R-value" used conventionally to cross-validate refinement of atomic models. The extrapolation pattern can reveal the presence of strongly observed, but weakly extrapolated reflections outside the basis set, whose contributions to the electron density are disproportionately large. The LLG has been used successfully to score phase permutation experiments aimed at phasing these critical, but poorly phased reflections directly from the interaction of their amplitudes with those inside the basis set. When coupled with ANOVA significance testing, this approach has considerable direct phasing power. The incomplete factorial design algorithm (Carter and Carter, 1979 *J. Biol. Chem.* **254**:12219-12223) provides a way to carry out such permutation very efficiently, with many fewer nodes than would be required in a full-factorial design. Similar efficiency has been obtained for experiments permuting alternative hypotheses regarding an incompletely known molecular envelope, and from which the correct features were identified.

These methods have been employed in solving the cytidine deaminase and tryptophanyl-tRNA synthetase structures, thereby documenting their effectiveness in real, practical situations. The latter structure solution was critically dependent on this approach, because the initial phase information available from heavy-atom derivatives was rather poor, and therefore represents the first successful application of maximum entropy methods to an unknown structure.

This work illustrates the power of entropy and likelihood maximization (Bricogne, 1988, *Acta Cryst.* A**44**:517-545) and the potential efficiency of supplemental, sampled phase permutation (Bricogne, 1993, *Acta Cryst.* D**49**:37-60) methods. Together, these capabilities provide necessary and sufficient elements for model-independent phase determination and refinement of rather poor starting phases for large macromolecular structures to yield maps comparable to 2Fo - Fc maps obtained from refined atomic models.

DS–02.07.02 SQUASH - COMBINING BOTH REAL AND RECIPROCAL SPACE CONSTRAINTS FOR MACROMOLECULAR PHASE REFINEMENT AND EXTENSION. By Kam Y. J. Zhang, Molecular Biology Institute & Department of Chemistry and Biochemistry, University of California, Los Angeles, CA90024-1570, USA.

An integrated system for macromolecular phase refinement and extension, SQUASH, will be presented. It includes solvent flattening (Wang, *Methods in Enzymology*, 1985, **115**, 90-112), histogram matching(Zhang & Main, *Acta Cryst.*, 1990, **A46**, 41-46), Sayre's equation(Zhang & Main, *Acta Cryst.*, 1990,**A46**, 377-381) and non-crystallographic symmetry refinement and averaging (Zhang, *Acta Cryst.*, 1993, **D49**, 213-222). This method combines the constraints of correct electron density distribution, solvent flatness, correct local shape of electron density and equal molecules for the phase refinement and extension of macromolecules. These constraints on electron densities are satisfied simultaneously by solving a system of non-linear equations by conjugate gradient method using FFT's (Main, *Acta Cryst.*, 1990, A**46**, 372-377). The formulation of the system of constraint equations is general, which enables any known constraints on the electron densities to be incorporated easily.

The electron density solution is further filtered by a phase combination procedure. Since the structure factor amplitudes are measured with much higher accuracy than the phases, the difference between the calculated structure amplitudes and the observed ones serves as a measure of reliability of the calculated phases after the solution to the system of constraint equations. The calculated phases are combined with the observed MIR phases to produce combined phases for the next iteration of phase improvement. This restraint in reciprocal space widens the radius of convergence of the system and makes it less noisy.

The non-crystallographic symmetry operations can be refined by a rotation and translation space search and subsequently by a least squares minimization method (Zhang, *Acta Cryst.*, 1993, **D49**, 213-222), thereby reducing the chance of introducing systematic phase errors during averaging. The implementation of non-crystallographic symmetry averaging could handle both proper and improper symmetry operations.

The effect of each constraint on phase refinement and extension will be examined. The constraints are found to work synergistically in phase improvement. Each constraint when applied alone, could leave the system trapped in a local minimum or could even prevent the system from converging. If the constraints were applied sequentially, the solution might oscillate and not converge. The simultaneous application of the constraints makes the system more stable, having a wider range of convergence.

The method was tested on the known structure of 2Zn pig Insulin. It successfully refined the initial MIR phases of 1677 reflections at 3.0Å from a mean phase error of 46° to 38°. The phases were further extended from 3.0Å to 1.5Å with a mean phase error of 62° for 10729 new reflections. Examples will also be presented of the improvement of MIR maps for several unknown structures which facilitated their interpretation.

DS–02.07.03

CONNECTIVITY AND THE PHASE PROBLEM IN MACROMOLECULAR CRYSTALLOGRAPHY D. Baker, C. Bystroff, A. Krukowski, C. Wilson and D. Agard. Dept of Biochemistry and Biophysics, UCSF.

The crystallographic phase problem is indeterminate in the absence of additional chemical information. The commonly employed chemical constraints--positivity, atomicity, and a solvent boundary--leave the phase problem greatly underdetermined for Fourier data sets of moderate (2.5-3.0Å) resolution. A successful *ab initio* approach must make use of high resolution Fourier data and/or

stronger chemical constraints. One such constraint is the connnectivity of the macromolecule. We have developed a rapid algorithm for measuring the connectivity of a map which shows promise in reducing the multiplicity of solutions to the phase problem. We have also developed a refinement method (PRISM) which exploits the connectivity constraint to iteratively improve phases. An initial electron density map is generated with inaccurate phases derived from a partial structure or from isomorphous replacement. A linear connected skeleton is then constructed from the map using a modified version of Greer's algorithm and a new map is created from the skeleton. This "skeletonized" map is Fourier transformed to obtain new phases, which are combined with any starting phase information and the experimental structure factor amplitudes to produce a new map. The procedure is iterated until convergence is reached. The method has been applied to problems with starting phase information from either molecular replacement or isomorphous replacement and appears to be a significant improvement over solvent flattening in both cases.

DS-02.07.04 DIRECT METHODS AND MACROMOLECULAR CRYSTALLOGRAPHY: LIGHTS AND LIMITS. By C. Giacovazzo*, A. Guagliardi, Dipartimento Geomineralogico, Universita' di Bari, 70124 Bari, Italy; D. Siliqi, Department of Inorganic Chemistry , University of Tirana, Tirana, Albania.

Several papers can be found in literature which describe the application of Direct Methods to macromolecules. Their efficiency is tested both for *ab initio* phasing and for phase refinement and extension. In this paper the role of direct methods in the field of macromolecular crystallography is analyzed. A criterion is formulated which suggests the necessary conditions for the success or the failure of the *ab initio* direct procedure. Most of the experimental protein data do not satisfy such a criterion, therefore their ab initio solution is a quite improbable event.

DS-02.07.05 JOINT X-RAY AND NMR REFINEMENT. By B Shaanan, Department of Biological Chemistry, The Institute of Life Sciences, The Hebrew University of Jerusalem, Israel

PS-02.07.06

AMoRe, An Integrated Molecular Replacement Program in Protein Crystallography : Some Applications to Multibody Systems. By J.Navaza[1], Y.Mauguen[1], P.Saludjian[2], T. Prangé[2], P.Alzari[3] and G.A.Bentley[3]. [1]Laboratoire de Physique, Faculté de Pharmacie, 92290 CHATENAY-MALABRY, France. [2]Chimie Biomoléculaire, URA 1430 CNRS, UFR Biomédicale, 93012-BOBIGNY CEDEX, France. [3]Unité d'Immunologie Structurale, URA 359 CNRS, Institut Pasteur. 25 rue du Docteur Roux, 75724 PARIS CEDEX, France.

A new strategy for Molecular Replacement calculations in protein crystallography has been implemented in the AMoRe package of programs. The algorithms have now been extensively tested in several crystal structures containing multiple copies of the proteins in the asymmetric unit. The examples discussed in the present communication include :

-The complex between Fab F9.13.7 and Guinea-fowl lysozyme with two molecules in the asymmetric unit, using three different search probes (lysozyme, variable and constant regions of the Fab), five out of the six subunits could be sequentially positioned in a single run of AMoRe.

-The trigonal form of Tumor Necrosis Factor, with six copies in the asymmetric unit (a dimer of trimers), using a trimer as the search model.

-A new orthorhombic form of Erabutoxin-b cristallized in presence of KSCN (two copies in the asymmetric unit).

-The complex of a bacterial ribonuclease, barnase, with its specific proteic inhibitor, bastar. There are three copies of the complex in the asymmetric unit. The barnase structure, representing approximately 1/6th of the a.u., was used as the search model.

-An hexagonal form of bastar with four molecules in the a.u., using as the search model the inhibitor subunit, taken out from the above refined complex.

PS-02.07.07 **PHASE PERTURBATION AS A MEANS OF REDUCING MODEL BIAS IN MACROMOLECULAR CRYSTALLOGRAPHY.** By M.V.Hosur and K.K.Kannan Solid State Physics Division, Bhabha Atomic Research Centre,Trombay, Bombay-400085,INDIA.

The **Molecular Replacement Method** is being increasingly used to solve protein structures by X-ray crystallography. It has also been recognized that the search model used in the above method, introduces a phase bias that complicates interpretation of the calculated electron density maps. A number of attempts have been made to reduce this model bias either by calculating OMIT maps or by using modified amplitudes in Fourier calculations. However, the features of a Fourier map are determined more by the phase rather than the amplitude of the coefficient. We have therefore explored the possibility of altering the phases of the coefficients as a means of reducing model bias in Fourier calculations. A variety of schemes of phase perturbation have been tried, with interesting results. These results and their implications to solving protein structures by the Molecular Replacement Method will be discussed.

PS-02.07.08

ASSESSMENT OF BULK SOLVENT MODELS BY CROSS-VALIDATION. By A.T. Brünger and J.-S. Jiang*, Department of Molecular Biophysics and Biochemistry, Yale University, U.S.A.

Bulk solvent models can play an important part in the modeling of macromolecular diffraction data. Bulk solvent models are aimed at reducing the residual for the low-resolution reflections. A low R value is not necessarily an indicator for the quality of the model. For example, we have shown earlier (Brünger, A.T., *Nature* **355**,472-474, 1992) that a bulk solvent model consisting of a disordered liquid of point atoms actually worsens the information content of the crystal structure. Cross-validation showed considerable promise to avoid this type of overfitting.