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

