

**On the *ab-initio* solution of the phase problem for macromolecules at very low  
resolution:  
the Few Atoms Model method**

By V.Yu.Lunin, N.L.Lunina, T.E.Petrova, E.A.Vernoslova,

*Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow  
Region, 142292, Russia*

A.G.Urzhumtsev\* and A.D.Podjarny

*UPR de Biologie Structurale, 15, rue Descartes, 67084 Strasbourg, France*

*(\*permanent address : Institute of Mathematical Problems of Biology, Russian Academy of Sciences,  
Pushchino, Moscow Region, 142292, Russia)*

## **Introduction**

The phase problem arises from the experimental difficulty of measuring phases in a diffraction experiment from molecular crystals. If no information other than the diffraction amplitudes from the native crystal is available, the phase problem is not solvable, since any set of phases will provide a density distribution. Therefore, additional information is necessary. In general, this information can be of two forms:

- 1) amplitudes from other crystals, related to the native amplitudes through a phase-dependent function; this is the case for the isomorphous replacement method, anomalous scattering, multiwave anomalous scattering, solvent contrast and related methods where the phase is measured indirectly through the amplitudes;
- 2) information about the nature of the electron density distribution, which limit the set of possible phases; this is the case for small molecule crystallography, where the strong constraint of atomicity has led to algorithms capable of providing the correct set of phases.

In protein crystallography, phases are currently obtained through methods of type (1), e.g., isomorphous replacement. Methods of type (2), using general constraints, are currently being applied to improve an existing phase set, e.g., density modification.

Several attempts are currently being done to develop methods of type (2) applicable to solve the phase problem *ab-initio*. One type of methods is based on statistical approaches, which transform the known density restraint (e.g., atomicity, positivity, connectivity, flat solvent envelope, etc...) to a statistical relationship between structure factors. These methods do not generate explicitly all possible densities which fit the used restraints. Some examples are the direct methods (Roth, 1989) and the maximum entropy methods (Bricogne, 1984; Bricogne, 1993; Navaza, 1985). Another type of methods attempt a detailed exploration of the phase space. These methods generate a very large quantity of individual solutions, and try to identify the correct one using both the information on diffraction amplitudes and on electron density. Some examples are the condensation of a large number of spheres (Subbiah, 1991) and the histogram methods (Lunin et al., 1990). A major problem in this approach is how to assure that phase space is explored exhaustively, so the right solution is not missed.

If a model is known, it provides strong constraints of type (2) and leads to the molecular replacement methods. Several efforts are based on the idea of extending these methods to the case where a very approximate model is known (in the limit, a single gaussian sphere). This has led naturally to work in a very low resolution range, where such a model can be applicable (Podjarny et al., 1987).

Working at very low resolution range the number of variables to be explored is much smaller than in the medium and high resolution ranges. This number might be small enough to envisage a quasi-exhaustive

investigation of all the possible density distributions, even in the absence of any model. Such an investigation has been done (Lunin et al., 1990) using the phases as variables and the histogram of the resulting density as the constraint. It showed the feasibility of a Monte Carlo approach to find the right solution by sampling the space of possible solutions, where every point represents a full set of structure factors. However, four problems arose:

- 1) a number of incorrect solutions which agree with the given criteria was found;
- 2) the number of amplitudes that can be phased is very small, limiting strongly the resolution of the resulting image;
- 3) the calculation of the exact histogram is not always possible;
- 4) since the histograms are calculated in real space, completeness of reflections in the inner core is necessary.

To address (1), a procedure was developed to identify points in phase space which are representative of groups of solutions. These points (clusters) can then be individually investigated. However, problems (2)-(4) remain unsolved.

## 1. The FAM method

To find an alternative method, a description of the electron density in terms of a Few Atoms Model (FAM) was developed. In this approach, the variables are the positions of a small number of huge gaussian scatterers, and the search criteria is the agreement between structure factor amplitudes calculated from these scatterers and the observed ones. A low resolution density can indeed be approximated (see below) with a small number of such pseudo-atoms. Compared to the histogram method, the number of variables is significantly reduced. Moreover, as the check criteria is in reciprocal space, it is less sensitive to missing reflections. The knowledge of the exact histogram is replaced by the knowledge of only two variables, the number and size of the spheres. A large number of possible FAM models are generated and those which agree better with the search criteria are analysed by the cluster method.

Even when this method has points in common with the molecular replacement approach (it can be assimilated to a multi-body molecular replacement) it is of more general character, to the extent that the spheres can overlap and they are meant to simulate any general density and not a fixed model. Note that the imposed model is not a series of separate atoms, but a compact molecular region surrounded by a flat solvent region. The output of FAM is not the position of the few atoms (as in molecular replacement) but the phases associated with them. In particular, FAM models which are very different in atomic positions but lead to similar phase sets will be considered equal. The different aspects of the method are described in detail below.

## 2. Approximation of density with a few atom model

To show the feasibility of the Few Atom Model, it is necessary to test whether at very low resolution the structure factors can be approximated with those calculated from such a model. To do so, the low resolution neutron data ( $\infty$ -16 Å) from the cubic form of the tRNA<sup>Asp</sup>-Synthetase complex were used. Three different H<sub>2</sub>O/D<sub>2</sub>O contrast neutron data sets (Moras et al., 1983) were used, corresponding to the full complex, the synthetase moiety and the tRNA moiety. This crystal form has been solved independently (see accompanying paper by Urzhumtsev et al.), using X-ray data and a model from another crystal form solved at high resolution (Ruff et al., 1991). Neutron diffraction data can be fitted correctly with this model. A 50 Å resolution map calculated with model phases shows a clear ovoidal peak corresponding to the position of the synthetase dimer. Symmetry-related peaks are joined by "arms" corresponding to the tRNA's.

There were two objectives of this test:

- 1) to find the range of resolution at which the neutron data of the cubic complex can be fitted correctly with a model consisting of a small number of spherical scatterers;

2) to check the extent to which the position of these scatterers can be obtained by a combination of single sphere searches and difference maps.

To fulfil the first objective, a 4 sphere model was built by putting 2 atoms in the position of the synthetase and 2 atoms near the tRNA "arms". This model could be refined to an R-factor of 16% at 50 Å

Y  
X Z

