



Phase problem and Direct Methods

Carmelo Giacovazzo

carmelo.giacovazzo@ic.cnr.it

The phase problem is one of the most difficult problems in X-ray Crystallography. The difficulties directly arise from the diffraction experiment itself: only the diffraction amplitudes are experimentally available, not their phases. This lack of information hinders the straightforward crystal structure solution via the use of the inverse Fourier transform. Special techniques are therefore necessary for recovering phase values, and therefore for solving the crystal structures: they are usually called *phasing methods*.

The phasing methods are divided in two categories: *ab initio* and *non ab initio*. *Ab initio* methods only exploit the experimental data of the target structure, *non ab initio* methods need some supplementary source of information. Using one type or another type of approach only depends on the quality of the experimental data and on the supplementary available information.

In this lecture we discuss the logical bases of any phasing approach: in particular we show that the phase information is not completely lost during the phase experiment, that it survives in the diffraction data, and that it may be recovered under favourable conditions. Thus any phasing procedure may be described as a special method for extracting the phases from the experimental data. The lecture will more treat the logical bases of any phasing process rather than to provide details of the various phasing techniques.

The success of a phasing procedure is usually testified by a low value of the crystallographic residual and by the chemical consistency of the structural model. The crystallographer is then sure that the proposed model is correct and that no other structure may be proposed in the future with the same low value of the crystallographic residual and the same chemical consistency. It will be shown that this belief is not based on a mathematical demonstrated theorem but only on a postulate, which is called ***The basic postulate of the structural crystallography:***

Only one chemically sound crystal structure exists compatible with the experimental diffraction data.

This is a postulate and chemically sound is a necessary attribute.

Any scientist who is attempting to solve a crystal structure should be well aware of the nature of such postulate, because, probably without full conscience, he admits it in a silent way.

There is another important question which is analyzed: is the amount of information stored in the diffraction amplitudes always sufficient to define the structure? It is well known that the Ewald sphere limits data resolution: the vibrational state of the atoms add supplementary limits to the effective resolution, so that data at atomic resolution often cannot be collected.

Since data resolution defines the quantity of information provided by the diffraction experiment, low data resolution may make the crystal structure solution difficult. These

aspects are very severe in macromolecular crystallography where the natural mobility of the molecules and the presence of the solvent facilitate large atomic vibrations.

In other diffraction techniques, e.g. in powder crystallography, it is the diffraction geometry itself which limits the information contained in the experimental data. Peak overlapping is still today one of the largest obstacles for a straightforward crystal structure solution from powder data. As effect, the structure factor amplitudes are ambiguous, and the electron density maps themselves are less informative. The practical consequence is that, as for proteins, the real quantity of structural information collected during the diffraction experiment is often not sufficient to determine the structure, and some supplementary source of information is needed.

In any phasing procedure a special role is played by the structure invariants, even when they are not explicitly mentioned. For example they are involved in any direct space refinement, where electron density modification techniques are used. Phase invariants are defined as combination of phases which do not vary when the origin in the unit cell is moved from one position to another. It is such property which allows both phase determination and phase refinement. The simplest phase invariants are discussed and their possible use is enlightened.

A short overview of the basic probabilistic relations used by Direct Methods will be presented. They solved in practice the phase problem for small molecules and they are, still today, one of the most powerful tools for solving small and macromolecular structures with data at atomic resolution.