On behaviur of statistical reliability indicators during crystal structure refinement

Garib Murshudov

MRC-LMB

It is well known that quality of the experimental data as well as the fit of atomic model parameters into the data are monitored by such indicators as R-factors, correlation coefficients. These indicators work sufficiently well for controlling changes in the model analysed using one set of data set. However often they are also used to compare the quality of different data as well as models derived using different crystals, sometimes resulting in misleading conclusions. Statistical properties of crystals affect the probability distribution of the intensities/amplitudes. And therefore behaviour of the reliability indicators calculated using such data sets could be dramatically different. For example the probability distribution of the intensities without noise collected from single crystal is proportional to the chi-squared random variable with degrees of freedom two. This distribution becomes chi-squared distribution of degrees of freedom four for the intensities from hemihedrally perfectly twinned crystals. And underlying distributions do affect the behaviour of widely used indicators, e.g. R-factors. Statistical analysis of these indicators depending on the properties of crystals, noise in the data and on the accuracy of the errors in the atomic model is long due.

In this talk, behaviour of these indicators depending on such properties of crystals as twinning, modulations, and overall isotropic/anisotropic temperature factors will be analysed and alternative indicators will be discussed. Effect of maximum likelihood atomic parameter refinement based on the amplitudes and intensities of structure factors on these indicators will be analysed. The effects of such atomic model refinement tools as local and global TLS restraints, jelly body refinement, local/global non-crystallographic restraints on these indicators will also be demonstrated.

The gap between currently acieved and maximum achievable information from the data with various statistical properties as well as the effect of various crystal/data properties on the calculated maps also will be discussed.