Statistical methods in protein crystallography, Randy J. Read*, Airlie J. McCoy and Laurent C. Storoni, *Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Cambridge CB2 2XY U.K.* E-mail: rjr27@cam.ac.uk

Keywords: Maximum likelihood; Experimental phasing; Molecular replacement

Most algorithms for macromolecular crystallography have been based traditionally on either least-squares optimisation or analysis of the Patterson function. In recent years, great progress has been made by reformulating these algorithms in terms of the principle of maximum likelihood. The basic idea behind likelihood is fairly simple: the quality of a model is judged by the probability that model assigns to the set of observations that was measured. If the errors in predicting the observations from the model are Gaussian and arise only from measurement error, likelihood is equivalent to least-squares. Many probability distributions relevant to crystallography are indeed Gaussian, but they describe the relationships among phased structure factors. Elimination of the unknown phase changes the form of the distribution so that it is necessary to apply a proper maximum likelihood treatment, instead of simply using least squares. Some likelihood-based methods are very slow, but reasonable approximations can be computed quickly. Interestingly, some of these approximations turn out to bear a close relationship to Patterson-based methods.

Likelihood has now been applied to a number of areas in protein crystallography: the estimation of model phase probabilities and computation of map coefficients in the program *SIGMAA*; the refinement of atomic models in *CNS*, *Refmac* and Buster/TNT; experimental phasing in *Sharp*; molecular replacement in *Beast*. We have been developing a new program for likelihood-based phasing in macromolecular crystallography. Our program, *Phaser*, implements new methods for solving structures by molecular replacement, and new methods for experimental phasing are under active development.

In my talk, I will describe the current capabilities of *Phaser*: anisotropic normalisation, likelihood-based fast rotation and translation functions, MIR phasing and SAD phasing, and I will show how some difficult structures can now be solved easily. I will discuss some of our future plans to increase the sophistication of *Phaser*, particularly in accounting for correlations in sources of error and in combining the information from molecular replacement and experimental phasing.