Macromolecular *ab initio* Phasing

MS02.06.01 P. Main

MS02.06.02 BAYESIAN *AB INITIO* PHASING: THE ROLE OF STRUCTURE FACTOR STATISTICS WITH BUILT-IN STEREOCHEMISTRY. Gérard. Bricogne, Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England; and LURE, Bâtiment 209D, 91405 Orsay, France.

So far the main efforts in formulating and implementing the Bayesian approach to structure determination have been directed towards the design of (1) a more powerful method (the saddlepoint approximation) for evaluating joint probability distributions of structure factors, capable of handling non-uniform distributions of random atomic positions (using the maximum-entropy method); and of (2) a systematic protocol for forming hypotheses (typically, but not exclusively, trial phase assignments), for sampling them efficiently (e.g. by "magic lattices" based on error-correcting codes) and for testing them against the available data (by examining the log-likelihood gain statistic).

In spite of these elaborations, the initial assumptions on which the Bayesian statistical machinery is set to work remain the same as that of standard direct methods: all atoms are assumed to be statistically independent, so that chemical bonding rules are ignored. Overcoming this embarrassing inadequacy, i.e. finding a way of incorporating *a priori* stereochemical knowledge into structure factor statistics, has proved one of the most elusive questions in theoretical crystallography.

It will be shown here that the key concepts of saddlepoint approximation and maximum-entropy distributions can be applied to this problem to yield joint probability distributions of structure factors with built-in stereochemistry, i.e. *a priori* statistical criteria of stereochemical validity [1]. This procedure can use a hierarchically organised knowledge base incorporating the known clusterings of short hexapeptide building blocks, of secondary and super-secondary motifs, and of domain folds. Sequential Bayesian inference can be conducted on the basis of these more stringent criteria in such a way as to consult all relevant structural information to compensate for the relative paucity of diffraction data which distinguishes the macromolecular from the "small moiety" situation. This procedure provides the natural foundation on which to build a genuine *expert system* for knowledge-based structure determination.

Reference

[1] BRICOGNE, G. (1995). In *ECCC Computational Chemistry* (Bernardi & Rivail, editors). *Amer. Inst. of Phys. Conf. Proceedings*, 330, 449-470.

MS02.06.03 LOW RESOLUTION AB-INITIO PHASING WITH MONTE CARLO AND CLUSTERIZATION TECHNIQUES. V.Y. Lunin, Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow Region, 142292, Russia

A Monte Carlo type approach has been developed for low-resolution ab-initio phasing. It is based on the generation of a large amount of possible phase sets followed by an enrichment procedure which rejects non-admissible sets in accordance with some specified selection criteria. Two approaches to phase sets generation were tried: direct generating of phase values (Lunin, Acta Cryst., D49, 90-99, 1993) and the recently developed Few Atom Models method (Lunin et al., Acta Cryst., D51, 896-903, 1995) in which low resolution phase sets are approximated using structure factors calculated from pseudo-atomic models.

The various selection criteria (suggested by different authors), such as magnitude correlation, electron density map (e.d.m.) histograms, e.d.m. connectivity and local density variation, maximum likelihood estimates of phase errors etc. are not strongly discriminative when applied in the low resolution range. Furthermore, attempts at local refinement without special precautions fits the criteria without necessarily improving the e.d.m.. (Lunin&Skovoroda, Acta Cryst., A51, 880-887, 1995). To overcome this difficulty a cluster analysis procedure was applied to split the enriched collection of phase sets into a small number of clusters, each representing a possible solution. This procedure as well as the averaging of variants inside a cluster require preliminary maps alignment (Lunin & Lunina, Acta Cryst., in print, 1996).Some additional criteria, e.g. high density at dyads, may be used then to reject wrong clusters.

The low resolution phase information obtained may be used to construct non-trivial prior coordinate probability distributions and to modify classical direct methods approaches on the base of explicit representation of saddle-point based approximations for the whole set of structure factors (Lunin et al., to be published).