Computing II

General Advances & Application

MS03.02.01 PHASE IMPROVEMENT AND EXTENSION IN PROTEIN CRYSTALLOGraphY. Giuseppe Zanotti, Dritan Siliqi, and Carmelo Giacovazzo. Dipartimento di Chimica and Biologia, Centro Studi Biopolimeri, Via Marzolo 1, 35131 Padova, Italy. *Dip. Geomineralogico, U. di Bari, Via Orobana, Bari, Italy.

In protein structure solution with the MIR technique, very often the isomorphism of heavy-atom derivatives does not extend beyond 3 Å resolution or so. Usually, a set of phases obtained with the isomorphous replacement method must be improved further, in order to obtain an interpretable electron density map. In the single isomorphous replacement (SIR) approach, this is even a conditio sine qua non. Two methods aimed to overcome these problems will be discussed.

In the first, single isomorphous replacement techniques are integrated with direct methods in order to solve the ambiguity of the phases.

The second consists of a procedure of phase extension and improvement, based on discrete Hilbert transforms. This procedure is based on a completely different principle from those previously described and has the advantage of being absolutely model-independent. The method has been tested using simulated diffraction data of a small molecule and of a protein crystal. Starting from a randomly incomplete set of correct phases, it allows calculation of the unknown phases. Moreover, a set of phases affected by a mean phase error of ± 20° can be improved to a mean error of ± 25°, if the correct figures of merit for the reflections are known. The performance and the limitations of the technique, as well as the perspectives for further developments, will be discussed.

References

MS03.02.02 SOFTWARE FOR LATTICE IDENTIFICATION IN G-6. Herbert J. Bernstein, Bernstein+Sons, 5 Brewster Lane, Bellport, NY 11713-2803, USA and Lawrence C. Andrews, Thuridion S.E., 269 Mt. Hermon Rd., Scotts Valley, CA 95066 USA

We present the current state of software for Bravais lattice identification in G-6, based on a Fortran program which will accept experimental cell information on-line or from mmCIF or PDB format data files. The coordinate system of G-6, defined by (a,a), (b,b), (c,c), 2(a,c), 2(b,c), 2(c,a), 2(c,b), 2(b,a) is one in which a meaningful measure for quality of lattice identification is provided by Euclidean distances from appropriate linear subspaces. A modest combinatorial complexity is introduced by the need to evaluate multiple alternative distances not only from a given cell, but from nearby nearly-reduced cells. If the original reduced cell is not highly skewed the identification of a minimal distance requires only one to two shells of operations from the group of reduction transformations.

MS03.02.03 A LOCAL MACROMOLECULAR STRUCTURE DATABASE FOR CRYSTALLOGRAPHIC LABORATORIES. Philip E. Bourne and Ilya N. Shindyalov, San Diego Supercomputer Center, Box 85608, San Diego CA 92186-9784

As the number of macromolecular structures continues to grow exponentially, the need for a compact, easy to load and easy to query laboratory based database system would seem important. Such a database should be capable of loading all or a subset of the structures found in the Protein Data Bank (PDB) as well as maintaining local data which is in PDB format. The ideal system should contain native and derived data, should run on a variety of Unix platforms and should have a Web-based graphical interface to query the database. This paper reports on the design, capabilities and availability of such a database system. The San Diego Supercomputer Center (SDSC) version of the database is available via the World Wide Web on multiple servers (http://www.sdsc.edu/moose) and is within 24 hours of being current with the PDB native distribution as found in the PDB ftp archives. Using similar compression algorithms as found in WPDB [1] reduces data storage requirements 10 fold over native data without any loss of precision and also includes additional derived data. Apart from the obvious types of queries based on authors, protein names and other basic information, queries can be made with respect to characteristics of the polyamide chain, for example, sequence patterns (with gaps), patterns of secondary structure elements,