PS03.04.08 A NEW APPROACH TO MACROMOLECULAR REFINEMENT VIA BAYES' LAW AND BOLTZMANN'S DISTRIBUTION. Rob Grothe, Washington University, Dept. of Electrical Engineering, St. Louis MO, USA

The refinement of a macromolecular structure from crystal diffraction data can be formulated as follows: find the *most likely* mini-ensemble of structures given molecular energetics and observed data. The (posterior) conditional probability to be maximized can be expressed via Bayes' law as the *multiplicative product* of two distributions, prior and data likelihood. The prior assigns probability to each mini-ensemble, viewed as a single state, via Boltzmann's distribution of states for a canonical ensemble at ambient temperature; the mini-ensemble energy is the mean of the energy values computed for individual structures under an energetic model. For a given mini-ensemble, a virtual asymmetric unit is constructed by averaging the structures and a virtual crystal by symmetric replication. The data likelihood is the probability that the measurement of x-rays (modeled by a random process) diffracted by this crystal results in the observed data.

The Boltzmann relationship converts energy into probability, the common currency through which two disparate information sources, molecular energetics and diffraction data, can be unified via Baye's law. Viewed in reverse, the relationship yields the molecular dynamics interpretation: the most likely ensemble *minimizes* the model energy function derived from the model probability. As energy depends upon the log of probability, the posterior energy is the sum of two terms corresponding to the factors in the posterior probability. *X-PLOR*, a widely used refinement package, minimizes the sum of a model molecular energy and a term penalizing disagreement between structure and data. The user chooses the form of the term along with weighting factors. In this new approach, a model for diffraction data is chosen, and the data-dependent term is derived from it.

A refinement algorithm, using jump-diffusion random sampling, has been implemented on a 16k-processor parallel machine. Preliminary results have been obtained for refining BPTI, using diffraction data from the Protein Data Bank and the published structure as the initial state.

PS03.04.09 HYDROPHILICITY OF CAVITIES IN PROTEINS. Jan Hermans, Li Zhang, Christopher VanDeusen, and Xinfu Xia, Department of Biochemistry and Biophysics University of North Carolina Chapel Hill, NC 27599-7260

Water molecules inside cavities in proteins constitute integral parts of the structure. We have sought a quantitative measure of the hydrophilicity of the cavities by calculating energies and free energies of introducing a water molecule into these cavities. The computations required to survey the atomic coordinates of a protein molecule in terms of low-energy water positions are rapid. A proper assessment of hydration should be based on free energy, not energy; however, much lengthier dynamics simulations are required to obtain free energies of transfer of water molecules into interior sites. These methods are most direct when applied to cavities able to hold a single water molecule. A simple consistent picture of the energetics of isolated buried water molecules has emerged from this study. A threshold value of the water-protein interaction energy at -12 kcal/mol was found to be able to distinguish hydrated from empty cavities. This is nearly the same value as the energy of ice, and, since the threshold must correspond to a free energy of zero, it follows that buried waters in proteins have entropy comparable to that of ice. The results of this study have enabled us to address the reliability of buried waters assigned in experiments.

We have extended this work to two instances of cavities large enough to contain several water molecules. In one case (uteroglobin; 1UTG), the computed energetics support the presence of 8 water molecules, where the x-ray structure reports 12 sites, some of them rather weak. In the other case, interleukin-1beta, the computed energies and free energies of transferring one or two water molecules into the cavity are insufficiently low, and this suggests that the cavity is not hydrated, as reported in crystallographic studies, and at odds with a report based on nmr experiments that the cavity is hydrated.

The program and instructions for rapidly locating possible water interior water positions and discriminating between these on the basis of the energy of transfer are available from the authors (request DOWSER program from xia@femto.med.unc.edu).

PS03.04.10 MULTICOPY MODELING OF THE SOLVENT DISTRIBUTION IN MACROMOLECULAR CRYSTALS. Eduardo I. Howard and J. Raul Grigera, Instituto de Fisica de Liquidos y Sistemas Biologicos (IFLYSIB), Universidad de La Plata, c.c. 565, 1900, La Plata, Argentina; Alberto D. Podjarny and Alexander Urzhumtsev, IGBMC, UPR de Biologie Structurale, B.P. 163, 67404, Illkirch, Cedex, France

Hydration models of biomacromolecular crystals, as obtained by crystallographic diffraction studies, usually position water molecules on precisely defined sites. Other experimental results, such as NMR, indicate that a large part of the water content has high mobility and is delocalized. The objective of this work is to find a hydration model that describes these mobile water molecules, while keeping the agreement with the observed diffraction amplitudes. A multicopy water model is proposed to describe the mobility. A set of water molecules, positioned by conventional methods, is used to generate several non-interacting copies. The system is set to an initial temperature (typically 400° K) through assigning different initial velocities to each water molecule of the different copies. Then the system is very slowly cooled (5° steps) until it reaches the desired temperature (300° K). This dynamics simulation, implemented in XPLOR, includes the usual modeling forces and an X-ray term. The free R-factor is used to monitor the validity of the process. The method was applied to X-ray diffraction data from BPTI and RNA crystals. The results show that while some water molecules are highly localized (the different copies remain clustered in specific hydration sites), the rest are more widely distributed, sometimes forming water channels. The shape of the multicopy distribution agrees precisely with the Fo-Fc difference maps; in the BPTI case, simulations and difference maps using neutron data were used to cross-check the results. The obtained models agree with the crystallographic data and are more compatible with other experimental observations than the ones with single fixed sites.

This work is supported by the CNRS through the UPR 9004, by the CONICET through the IFLYSIB and by the EU through the collaborative project CII-CT 93 0014 (DG 12 HSMU); JRG is an invited fellow financed by the MENESRIP (France).

PS03.04.11 A TEST OF MAXIMUM-LIKELIHOOD REFINEMENT OF MACROMOLECULAR STRUCTURES WITH BUSTER & TNT. John Irwin and Gérard Bricogne, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

The Bayesian viewpoint [1,2,3,4] has long suggested that structure refinement should be carried out by maximising the log-likelihood gain (LLG) rather than by minimising the conventional least-squares residual, as only the maximum-likelihood (ML) method can take into account the uncertainty of the phases associated to model incompleteness and model imperfection by suitably downweighting the corresponding amplitude constraints. It was predicted [3] that ML refinement would allow the refinement of an incomplete model by using the structure factor statistics of randomly distributed scatterers to represent the effects of the missing atoms, in such a way that the latter would not be wiped out; and that the final would then provide indications about the location of these missing atoms.

These predictions have now been confirmed by actual tests carried out by combined use of BUSTER [4] and TNT [5] on an incomplete (60% of molecule) and imperfect (1Å rms positional error) model. The maximum-likelihood result is more accurate than that from least-squares, and the final LLG gradient map is much more informative than the usual difference map, thus greatly increasing