the chances of success in "bootstrapping" from an unpromising molecular replacement starting point to a complete structure.

We will also discuss the two main concerns at the moment in the fields of structure refinement and validation where Bayesian methods have much to offer, namely (1) getting better reliability indicators for the final results of structure refinement, and (2) ensuring that these indicators are effectively optimised during refinement.

References

[1] G. Bricogne (1984). *Acta Cryst.* A40, 410-445.
[2] G. Bricogne (1988). *Acta Cryst.* A44, 517-545.
[3] G. Bricogne (1992). In *The Molecular Replacement Method*, edited by W. Wolf, E.J. Dodson and S. Gover, pp. 62-75. Warrington: Daresbury Laboratory.
[4] G. Bricogne, *Acta Cryst.* D49, 37-60 (1993).
[5] D.E. Tronrud (1987). L.F. Ten Eyck, and B.W. Matthews, *Acta Cryst.* A43, 489-501.

PS03.04.12 MACROMOLECULAR ANISOTROPIC TEMPERATURE-FACTOR REFINEMENT WITH STRICT CONSTRAINTS. Jennifer A. Kelly and Todd O. Yeates. The Department of Chemistry and Biochemistry and The Molecular Biology Institute, University of California at Los Angeles, USA

Macromolecular models typically have high crystallographic residuals; the discrepancy between observed and calculated structure factor amplitudes is usually from 15% to 25%. A significant component of the residual is due to inadequate representations of atomic motion in protein models. Data collected from most protein crystals are insufficient to refine individual atomic anisotropic temperature-factors. Anisotropic b-factor refinement introduces six times the number of parameters as isotropic b-factor refinement, and most often results in over-fitting macromolecular models to the data. In order to reliably model protein motion as anisotropic, strict constraints are required. Our method introduces a new form for constraints on temperature-factors described by Fourier series, and uses an FFT-based algorithm for refining the relevant values. Anisotropic temperature-factors are defined by a position-dependent B matrix whose elements are constrained to vary smoothly over the unit cell in a crystal. Each of the six matrix elements is represented by a Fourier series with few terms. Individual anisotropic temperature-factors are determined by refining the coefficients of the six Fourier series which comprise the B matrix. Gradients are computed using FFT's by a modification of the method of Agarwal*. Since each Fourier series has only a few terms, the number of refinement parameters is kept low and over-fitting is avoided.

*Ramesh C. Agarwal, *Acta Cryst.* (1978). A34, 791-809

PS03.04.13 PREDICTING AND ANALYZING DETERMINANTS OF WATER-MEDIATED LIGAND RECOGNITION. Leslie A. Kuhn[1], Michael L. Raymer[1,2], William F. Punch[2], Paul C. Sanschagrin[1], and Erik D. Goodman[3], Depts. of [1]Biochemistry, [2]Computer Science, and [3]Case Center for Computer-Aided Engineering and Manufacturing, Michigan State University, East Lansing, MI 48824, USA

Protein recognition of ligands, from nucleic acids to small molecule inhibitors, is usually mediated by bound water molecules bridging the protein-ligand interface. These water molecules influence both the shape and chemistry of interaction. A barrier to appropriately incorporating active-site bound water to improve molecular simulations and ligand design has been the absence of a method for determining which water sites are likely to be conserved upon ligand binding. Our *Consolv* technique, using a hybrid k-nearest-neighbor classifier/genetic algorithm, predicts which water molecules will mediate ligand binding by examining the structural and chemical environment of each water molecule in the free protein structure, without knowledge of the ligand. After training on 13 non-homologous proteins, *Consolv* correctly predicted conservation or displacement of 74.6% of the active-site water molecules in 7 new proteins. Moreover, water sites mispredicted to be conserved typically were displaced by a polar (oxygen or nitrogen) atom from the ligand. Overall accuracy for predicting conserved water or polar ligand atom binding was 89.6%. The ability to predict water-mediated interactions from the free protein structure implies that the majority of conserved active-site water binding is independent of the ligand, and that the protein micro-environment of each water molecule is the dominant influence. We are now using genetic algorithms with linear weighting and genetic programs allowing non-linear scaling to evaluate the relative importance of several features of the site - temperature factor, hydrogen-bonding capacity, local atomic packing, and atomic hydrophilicity - for determining conserved water binding.

PS03.04.14 AN INTENSITY-BASED LIKELIHOOD FUNCTION FOR STRUCTURE REFINEMENT. Navraj S. Pannu and Randy J. Read, Department of Mathematical Sciences and Medical Microbiology & Immunology, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

In order to improve the quality of a model, structural refinement attempts to optimize the agreement between the observed and the calculated diffraction measurements. The process of structural refinement is commonly based on a least-squares analysis. Since the probability distribution of the observed structure factor amplitude given the calculated structure factor amplitude is not a Gaussian centered at k|Fc|, where k is a scale factor, least-squares is not theoretically justified. Accordingly, least-squares refinement does not make optimal use of the discrepancies between observed and calculated diffraction measurements. Therefore, in order to improve the process of structural refinement, a more general maximum likelihood analysis is considered.

A maximum likelihood target function has been derived and implemented in XPLOR. This function takes into account errors both in the model and in the measurements. Furthermore, this function is intensity-based allowing the use of negative intensities derived from the values observed in the diffraction experiment.

Preliminary tests show that the intensity-based likelihood function can achieve more than twice the improvement in average phase error compared to a conventional least-squares refinement. As a result, the electron density maps are clearer and suffer less from model bias.

PS03.04.15 CRYSTAL STRUCTURE AT 1.0Å RESOLUTION OF HUMAN p56*lck* SH2 DOMAIN IN COMPLEX WITH A SHORT PHOSPHOTYROSYL PEPTIDE. L. Tong, T. C. Warren, J. King, R. Betageri, J. Rose and S. Jakes, Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road, P. O. Box 368, Ridgefield, CT 06877, U.S.A

SH2 domains are modules of about 100 amino acid residues and bind to phosphotyrosine-containing motifs in a sequence-specific manner. They play important roles in intracellular signal transduction and represent potential targets for pharmacological intervention. The protein tyrosine kinase p56*lck* is a member of the *src* family and is involved in T-cell activation. The crystal structure of its SH2 domain with an 11-residue high-affinity peptide [1] showed that the phosphotyrosine (pY) and the Ile residue at the pY+3 position are recognized by the SH2 domain. We have determined the crystal structure of this SH2 domain in complex with the short phosphotyrosyl peptide Ac-pTyr-Glu-Glu-Ile (pYEEI peptide) at 1.0Å resolution [2]. The structural analysis at atomic resolution reveals that residue Arg 134 (αA2), which interacts with the phosphotyrosine side chain, is present in two conformations in the complex. This structure will be compared with those of other complexes.

[1] Eck, M.J., Shoelson, S.E., Harrison, S.C. (1993). *Nature*, **362**, 87.
[2] Tong, L., Warren, T.C., King, J., Betageri, R., Rose, J., Jakes, S. (1996). *J. Mol. Biol.* **256**, 601.