* If only the CA coordinates of a model are deposited, nobody will ever be able to validate the model. Actually, in some cases this is possible nowadays.
* One does not need to use non-crystallographic symmetry restraints. The examples to the contrary may make some want to re-do their most recent refinement (7).
* Ramachandran plots are stiflingly boring. On the contrary: they are extremely useful for model validation. We will show some highly entertaining examples from real-life models.

Considering the controversial nature of some aspects of this presentation, the audience is invited to disagree vehemently.

References:
(1) Branden, C.I., & Jones, T.A. (1990). Nature 343, 687-689.
(2) Kleywegt, G.J., & Jones, T.A. (1990). Structure 3, 535-540.
(3) Brunger, A.T. (1992). Nature 355, 472-475.
(4) Jones, T.A., Zou, J.Y., Cowan, S.W., & Kjeldgaard, M. (1991). Acta Cryst. A47, 110-119.
(5) Zou, J.Y., & Mowbray, S.L. (1994). Acta Cryst. D50, 237-249.
(6) Kleywegt, G.J., & Jones, T.A. (1996). Meth. Enzymol., in press.
(7) Kleywegt, G.J. (1996). Acta Cryst. D52, in press.

MS03.04.05 IMPROVED STRUCTURE REFINEMENT THROUGH MAXIMUM LIKELIHOOD. Randy J. Read and Navraj S. Pannu, Departments of Medical Microbiology & Immunology, and Mathematical Sciences, University of Alberta, Edmonton, Alberta T6G 2H7, Canada.

The least-squares target is not theoretically justified for crystal structure refinement, so it is preferable to use a maximum likelihood target instead. With a maximum likelihood treatment, the need for *ad hoc* weighting schemes and resolution cutoffs is eliminated, observational errors are used appropriately and, above all, the refinement is more successful.

When crystal structures of proteins or small molecules are used to address questions of scientific relevance, the accuracy and precision of the atomic coordinates are crucial. Accordingly, the atomic model is generally improved by refining it to improve agreement with the observed diffraction data. The use of least-squares methods would only be justified (by the principle of maximum likelihood) if the probability distribution relating the observed and calculated diffraction measurements were Gaussian. As the relationship is not Gaussian, the least-squares target is inappropriate.

We have implemented two maximum likelihood targets in the program XPLOR: 1) an amplitude-based Gaussian approximation assuming Gaussian errors in the observed amplitudes; and 2) an intensity-based likelihood function assuming Gaussian errors in the observed amplitudes squared. The amplitude-based target can be implemented easily in any least-squares refinement program, while the intensity-based target has a number of advantages including the ability to use negative observed intensities.

Preliminary tests with protein structures give dramatic results. Compared to least-squares refinement, maximum likelihood refinement can achieve more than twice the improvement in average phase error. The resulting electron density maps are correspondingly clearer and suffer less from model bias.

The authors were supported in these investigations by AHFMR, MRC (Canada), NSERC (Canada) and an International Research Scholars award to RJR from HHMI.

MS03.04.06 DESCRIPTION OF PROGRAM USING MAXIMUM LIKELIHOOD RESIDUAL FOR MACROMOLECULAR REFINEMENT, ILLUSTRATED BY SEVERAL EXAMPLES. Eleanor J. Dodson and Garib N. Murshudov, Chemistry Department, University of York, Heslington, York, U.K., and Alexei A. Vagin, UCMB-ULB, Free University of Brussels, Avenue Paul Heger cp160/16 - P2 1050 Brussels, Belgium

We illustrate the advantages of the maximum likelihood refinement method over least-squares for macromolecules. Maximum likelihood refinement has been implemented in the program REFMAC.

At each cycle the program performs two steps. First it estimates the overall parameters of likelihood. This is most successful when the parameters are deduced from the FreeR set of reflections. Secondly it uses these parameters to build the likelihood function and refine the atomic parameters.

At the end of a cycle REFMAC also writes weighted map coefficients to give less biased maps for rebuilding, taking care to restore missing data. Absent reflections cause unpredictable noise in map calculations which may lead to errors in interpretation.

Several examples are described. In each case the refinement was carried to convergence from an existing model. Results were compared to maps and phases generated from the final coordinates.

Different parts of structure may be assigned different expected errors and methods for doing this have been explored and implemented. Two important applications for this are being analysed. In the first case the structure contains several U atoms as well as protein atoms. In the second part of the structure has been interpreted from a poor MIR map but the other part is being modelled from the uninterpretable electron density. There is also an option to include available phase information, for example from MIR or MAD calculations.

PS03.04.07 PROTEIN PRECISION RE-EXAMINED: LUZZATI PLOTS DO NOT ESTIMATE FINAL ERRORS. D W J Cruickshank, Chemistry Department, UMIST, Manchester, M60 1QD, UK

The misuse of Luzzati plots of the residual R versus $\sin\theta/\lambda$ to estimate final coordinate errors has stimulated a re-examination of protein precision. Luzzati (1952, Acta Cryst.) gave a theory for **uncompleted** refinements which estimated the r.m.s. shifts still needed to reach R = 0. His theory assumed no errors in F_{obs} and that **the F_{calc} model was perfect apart from coordinate errors.** The Gaussian error distribution **was the same for all atoms.** These assumptions are invalid for proteins. Quite apart from the dependence on atomic number, it is well established that errors depend very strongly on atomic B values. Nor do Luzzati plots provide an upper limit for $\langle\Delta r\rangle$.

Restrained refinement will be examined theoretically. As applied to the simplest protein model of 2 like atoms in one dimension, restrained refinement determines a length which is the weighted mean of the diffraction-only length and the geometric-dictionary length.

By extending the order-of-magnitude error formula for small molecules given by Cruickshank (1960, Acta Cryst.), the e.s.d. for protein atom i with $B = B_i$ is, very roughly,

$$\sigma(x_i) = k(N_i/p)^{1/2}\,[g(B_i)/g(B_W)]\;C^{-1/3}\;d_{min}\;R,$$

where k is about 1.0, $N_i = \sum Z_j^2/Z_i^2$, $p = N_{obs} - N_{params}$, [provisionally] $g(B) \approx (1 + 0.04B + 0.003B^2)$, B_w is the Wilson B for the structure, and C is the fractional completeness of the data to d_{min}. For example if $N_i = 1000$, $p = 15000 - 4000$, $B_i = B_w$, $C = 0.9$, $d_{min} = 1.4$Å, and $R = 0.15$, then $\sigma(x_i) = 0.07$Å. This approach reveals the basic statistical flaws in the use of Luzzati plots.

Some authors have been able to invert the full LS matrix, and so obtain proper estimates of e.s.d.'s. Even when this is not possible, determined efforts should be made to use the information in a partial LS matrix.