SHARP : A Maximum-Likelihood Heavy-Atom Parameter Refinement Program for the MIR and MAD Methods.

E. de La Fortelle, J. Irwin
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH England
eric@mrc-lmb.cam.ac.uk, ji10@mrc-lmb.cam.ac.uk
http://www.mrc-lmb.cam.ac.uk

and G. Bricogne
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH England and
LURE, bât 209d, F-91405 Orsay Cedex, France
gb10@mrc-lmb.cam.ac.uk
http://www.mrc-lmb.cam.ac.uk

Abstract

The problem of estimating heavy-atom parameters (esp. occupancies) from acentric reflexions in the MIR method has a long history of difficulties, and a conceptually satisfactory solution allowing bias-free refinement of all parameters (including the lack of isomorphism) has only recently been obtained by a recourse to the method of maximum-likelihood estimation. The situation is essentially identical in the case of MAD phasing. The maximum-likelihood method needs to be invoked to exploit incomplete phase information in a heavy-atom parameter refinement while preventing that information from biasing the results.

We have designed and written from scratch a computer program –SHARP(Statistical Heavy-Atom Refinement and Phasing) – that fully implements the maximum-likelihood approach. It can refine all parameters specifying the internal rescaling of data as well as the models for lack-of-isomorphism and for all relevant heavy atoms, against data from MIR and/or MAD data in any combination. The program has been systematically tested, both on synthetic and on measured data, and compared to the standard program MLPHARE. The results show the superiority of our approach, especially in cases of low signal-to-noise ratio. The likelihood function has also been used as a detection tool to compute residual Fourier maps which can be inspected to probe for minor sites, and to calculate phase probability distributions encoded in Hendrickson-Lattman coefficients.

1 Introduction

Bias-free refinement of heavy-atom parameters in the MIR and MAD methods, which is an essential step towards obtaining the best possible electron-density maps given the available data, has remained for a long time a troublesome issue in macromolecular crystallography. The conventional approach to this problem was originally conceived [1,2] as a straightforward adaptation of the least-squares method previously used on centric data by Hart [3]: the "most probable" or the "best" estimates of the phases, as defined by Blow & Crick [4], were simply made to play a rôle analogous to that of the signs of centric reflexions. Dickerson, Weinzierl & Palmer [5] pointed out that more than two derivatives were needed for this type of refinement, and Blow & Matthews [6] found this method to have poor convergence properties unless steps were taken to ensure that the acentric phase estimates used in the refinement were independent of the parameters that were being refined. With hindsight, these difficulties are easily rationalised : this 'phased' least-squares refinement was, in effect, violating the first cardinal rule of the least-squares method, namely that any quantity involved in the observational equations should be either a *model parameter* or an *observation*. Treating the native phase as a known constant within each cycle, but recalculating it after each refinement step, introduces bias on the parameters, especially in the case of mostly bimodal phase distributions.

At the same time as the first attempts were being made to use phase estimates, an alternative refinement scheme was devised by Rossmann [7], based on a difference-Patterson correlation criterion, and evolved towards the "F_{HLE} method" [8,9], and finally the "origin-removed Patterson-correlation function" [10]. Here the use of acentric phase estimates is avoided altogether, but at the price of impoverishing the available information in the sense that multiple derivatives are not allowed to assist each other's refinement through the generation of phase information.

Sygusch [11] recognized that a middle-ground could perhaps be found if the acentric phases were no longer deemed to be "estimates", but were instead treated as extra parameters and refined along with the others. Unfortunately, the enormous increase in the number of variables dictated the use of a diagonal approximation, which rather defeated the original purpose of accommodating the correlations between phases and parameters. Bricogne [12,13] proposed a solution that partially overcame these difficulties. The main idea was that structure-factor estimates for acentric reflexions are *implicit functions* of the parameters that are being refined. This dependence was shown to result (*via* the chain rule) in a correction to the partial derivatives from which the normal equations of the least-squares method are to be formed. Many previously observed pathologies, such as the rapid divergence of the site occupancies of good derivatives, were cured by this analysis, but slower-moving instabilities were observed that resulted in divergent behaviour of the estimates for the *lack of isomorphism* of the various derivatives. Moreover, the problem of bimodality remained.

At this point, compliance with the first cardinal rule of the least-squares method had been essentially restored, but attention was drawn to the violation of a second cardinal rule : the inverse-variance 'weights' in the expression for the least-squares residual should be kept *fixed* as if they were part of the observed data. Since the method of least-squares is a special case of the maximum-likelihood method when errors are normally distributed with fixed (co)variances, it is clear that the problem of properly estimating the lack-of-isomorphism parameters demanded a fully-fledged maximum-likelihood rather than a least-squares treatment.

Perusal of the literature shows that two-dimensional statistical 'phasing' (probability distribution on the phase and on the modulus of the native structure factor) had been considered as early as 1970 [14], leading to the first mention of likelihood in this context by Einsein [15]. The first mention of parameter estimation by maximum-likelihood, in a very limited context, is found in Green [16]. Maximum-likelihood (ML) refinement for heavy-atom parameters was then advocated by Bricogne [17,18,19], Read [20], and an approximation to it was implemented by Otwinowski [21] in the program MLPHARE. This program is only a partial implementation of ML refinement – best described as 'phase-integrated least-squares' – in the sense that (i) it integrates the exponential of the least-squares residual and its partial derivatives only over the phase of the native structure factor (not over its modulus) ; and that (ii) the lack of isomorphism is still *re-estimated* at the end of each refinement cycle rather than being *refined*, and may often converge to non-optimal values. Nevertheless, this approach has been shown in numerous cases to yield better results than earlier refinements, drawing attention to the potential of maximum-likelihood methods.

The maximum-likelihood formalism outlined in Bricogne [19] for the MIR and SIR cases forms the basis of the present work. We will describe here its extension to probability distributions incorporating anomalous diffraction effects as well as measurement error and non-isomorphism. Integrating these distributions in the whole complex plane leads to likelihood functions that can be used for heavy-atom detection and refinement, and for producing phase probability distributions. We will also describe the current implementation of this formalism in a computer program, named SHARP (for Statistical Heavy-Atom Refinement and Phasing) [22].

2 Likelihood functions for parameter refinement

2.1 Outline

Generally speaking, bias is introduced in a model incorporating some degree of randomness whenever a *distribution* for a random quantity is replaced by a *value* for that quantity. The likelihood formalism avoids this pitfall by consistently emphasizing that *distributions* are involved.

More specifically, a least-squares (LS) model is always formulated as a prescription for turning given values of model parameters into 'calculated' (error-free) values to be compared with the observables. Error estimates are obtained *a posteriori*, by examining the residual discrepancy between the 'calculated' and the 'observed' quantities.By contrast, a likelihood-based model casts its predictions directly in the form of probability distributions for the observables, the quantities called 'calculated' in the LS formalism usually appearing as parameters in these distributions.

2.2 The native structure factor

The most important thing to bear in mind when building up the likelihood function for heavy-atom refinement is that the complex value of the native structure factor $\mathbf{F}^P(\mathbf{h})$ is not known. The measurement of a native amplitude for an acentric reflexion **h**, if present, gives rise to a two-dimensional probability distribution $p\left(\mathbf{F}^P(\mathbf{h})\right)$. A measurement for the structure factor of a derivative crystal will also give rise to a two-dimensional probability distribution $p\left(\mathbf{F}^P(\mathbf{h})|\{g\}\right)$ for the *native* structure factor, conditional to the values {g} of the set of global parameters for the heavy-atom model, for the scaling model and for the lack-of-isomorphism model.

For a centric reflexion, the probablity distribution becomes one-dimensional, but the theory is essentially similar.

2.3 The likelihood function

For a given reflexion **h**, the probability distribution of the native complex-valued, conditional to all the information available, is obtained by multiplying the probability distributions of $\mathbf{F}^{P}(\mathbf{h})$ for independent measurements.

This probability distribution is then transformed into a likelihood distribution for that reflexion, via the simple rule (in the absence of prior phase information) :

$$\Lambda\left(\{g\},\mathbf{F}^{P}{}_{*}(\mathbf{h})\right) = p\left(\mathbf{F}^{P}_{*}(\mathbf{h})|\{g\}\right)$$

Note that this equation is valid at each *trial point* $\mathbf{F}^{P}_{*}(\mathbf{h})$ in the Harker plane. In order to have a likelihood function that is independent of assumptions on the native complex structure factor, we must now integrate the likelihood function over all possible values of $\mathbf{F}^{P}_{*}(\mathbf{h})$:

$$\Lambda(\{g\}) = \iint \Lambda\left(\{g\},\mathbf{F}^{P}{}_{*}(\mathbf{h})\right)\, d^{2}\mathbf{F}^{P}_{*}$$

In the case of a centric reflexion, the integration is one-dimensional only, along the axis defined by the centric phase.

3 Parametrisation

3.1 Heavy-atom structure factors

This parametrisation amounts to a physical description of diffraction properties , involving heavy-atom coordinates, occupancies, isotropic and (if need be) anisotropic temperature factors, as well as normal and anomalous scattering factors. This was prefered to 'isomorphous' and 'anomalous' occupancies, because the physical parameters f' and f" are either known precisely (MIR experiment off an absorption adge) or can be measured (Cromer curve in a MAD experiment). Our implementation uses a hierarchical organisation for these parameters, that enables common attributes to be shared appropriately (Fig. 2). A list of site coordinates is determined that contains all known sites in all derivatives, and for each level of the hierarchy, these sites are 'qualified' (by a chemical identity, by an occupancy etc.). In this way, the long-standing problem of the same site being refined independently at each wavelength of a MAD experiment cannot occur, And common sites in a MIR experiment are parametrised correctly.

Future developments will incorporate a parametrisation of the anisotropy of anomalous scattering [23,24] and will allow a refinement of the corresponding parameters from unmerged data carrying suitable goniometric information for each measurement.

3.2 Scale factors

Currently, scale factors are parametrised by a constant scale K_{j}^{sc}, an isotropic relative temperature factor B_{j}^{sc}, and six anisotropic increments $\left(b_{j}^{p,q}\right)$ to B_{j}^{sc} :

$$k_{j}(\mathbf{h}) = K_{j}^{sc} \exp\left[-\frac{1}{4} B_{j}^{sc}(d^{*})^{2}\right] \exp\left[-\left(\sum_{\substack{p=1\ldots3\\ q\leq p}} b_{j}^{p,q}\,\mathbf{h}^{p}\mathbf{h}^{q}\right)\right]$$

3.3 Lack-of-isomorphism variance

Differences between native and derivative structure factors are explained by a heavy-atom model, and by an error model. In the 'null hypothesis' where we know nothing about the heavy-atom structure, all the differences are average attributed to the error, and this error will be refined to smaller values as the heavy-atom model becomes more accurate.

This error can be broken down in three main components :

The measurement error.
It is part of the crystallographic data and not refined.

The physical lack-of-isomorphism error.
In the absence of structural evidence for 'localised' lack-of-isomorphism, our assumption will be that of Luzzati [25] that there is a random isotropic positional perturbation, with spatially uniform mean amplitude and normal distribution, over the whole asymmetric unit. Based on this hypothesis, following the work of Read [26] and Dumas [27], we used a one-parameter model for the physical lack-of-isomorphism variance, increasing with resolution.

The model error.
This error has the same effect on the statistical distribution of the native structure factor as the previous one, but its variance is approximately decreasing with resolution as the mean intensity of remaining heavy atoms. We used a two-parameter model (a constant and a temperature factor) for this error.

A similar parametrisation is used for the error on the anomalous differences. Although there is no physical basis for adopting the same model, it was thought flexible enough as a function of resolution to fit to more diverse functions of resolution.

4 Other uses for the likelihood function

4.1 Residual maps for model updates

The likelihood formalism also provides the opportunity of checking for significant systematic disagreement betweeen the data and the substitution model. For each reflexion **h**, we calculate the gradients of the likelihood function with respect to the real and imaginary parts of the various heavy-atom structure factors $\mathbf{F}^{H}_{j}(\mathbf{h})$. These numbers are then used in Fourier syntheses to produce residual maps, that have the symmetry of the crystal. Similarly, in the case where there is significant anomalous diffraction, the gradients with respect to $\left(\mathbf{F}^{H}_{j+} + \mathbf{F}^{H}_{j-}\right)$ become coefficients for isomorphous residual maps, and those with respect to $\left(\mathbf{F}^{H}_{j+} - \mathbf{F}^{H}_{j-}\right)$ for anomalous residual maps.

These maps enable the detection of minor sites, and perform this task in an optimal fashion because they take into account the full unbiased phase information available from the data at the current stage of refinement. They are essentially Fourier syntheses calculated from inverse-variance weighted difference coefficients between the derivative and native data. Their enhanced sensitivity to any departure from the current heavy-atom model (when the data are accurate enough, and to high enough resolution) makes them the instrument of choice to detect more subtle features, such as anisotropy in the heavy-atom temperature factors or structural disorder at certain sites.

4.2 Final phasing and calculation of Hendrickson-Lattman coefficients

Once the global parameters have been refined to convergence, the likelihood function $\Lambda\left(\mathbf{F}^{P}_{*}, \{\ \ g\}\right)$ considered as a function of the trial native structure factor $\mathbf{F}^{P}_{*}$ only, becomes (after suitable normalisation) the probability distribution function of the modulus *and* phase of the native structure factor (this is a simple application of Bayes's theorem). The two-dimensional centroids $\mathbf{F}^{P}_{\mathrm{best}}(\mathbf{h})$, used as Fourier coefficients of the electron-density map, and the Hendrickson-Lattman 'ABCD' coefficients [28] of the marginal phase distribution can be easily derived from this likelihood function.

4.3 Future developments and perspectives

A natural extension of the quantitative use of residual maps based on log-likelihood gradients is the refinement of heavy-atom clusters of known geometry by real-space techniques of the Agarwal-Lifchitz type (e.g. as implemented in the TNT package). This is currently underway.

In order to offer *ab initio* detection capability, another type of map will be added to the existing program. Its coefficients will initially involve second-order derivatives of the log-likelihood function associated to the null hypothesis defined by "all intensity differences between data sets are caused by lack of isomorphism". This map will have the character of a Buerger sum function over a weighted difference-Patterson function [29]. As major sites are detected and included in the substitution model, the log-likelihood function will develop first-order derivatives giving rise to a difference-Fourier component in the residual map, while the revised second derivatives will keep contributing a component with the character of a sum function over a residual difference-Patterson.

The whole process of detecting sites and of assessing their significance quantitatively can thus be automated, using the log-likelihood gain referred to the null hypothesis as a scoring criterion for the peak-search. The procedure will stop when the highest remaining peak in the residual maps is essentially at the level of the noise.

Once all heavy atoms have been detected and refined, the remaining features in the 'isomorphous' residual maps, if they are significant, can provide the basis for a systematic study of lack of isomorphism. This could improve the rather crude way in which 'global' and 'local' lack of isomorphism have hitherto been described.

5 The Graphical User Interface

In order to facilitate the full use of SHARP's ability to accommodate experimental data from any combination of sources (MIR, with or without anomalous scattering, MAD, or a blend of both), it was necessary to guide the user in the construction of a hierarchical parameter file describing his/her experimental situation. This was achieved by the means of an HTML browser-based graphical user interface. The same system was used to facilitate inspection of the output of the program, and the optional execution of standard peripheral tasks.

5.1 Choice of tools

Our approach is based on a client-server philosophy, in order to make best use of the World Wide Web as a communication tool. As a result, once SHARP is installed on a server (a powerful computer, workstation or other, that will actually do the calculations), any authorised user can run the program from any terminal connected to the Internet. This has proved invaluable to us during the beta-testing stage, and provides high flexibility for all users.

The result is a forms-based interface, written in HTML language and processed by Perl scripts, that exactly mirrors the hierarchy of parameters during the buildup of the parameters file, and that connects automatically to graphical helper applications, to facilitate inspection of the output.

5.2 Input

The input pages consist in a series of embedded forms, that guide the user through our parametrisation of the experiment (list of sites, compounds, crystals, wavelengths, batches). Because the options taken in the higher levels condition the structure of the lower levels, the setting of the parameter tree is unidirectional ; coming back erases what has been set before.

5.3 Output

The output listing file of SHARP is also written (by the FORTRAN program) in the HTML language, and thus contains hyperlinks to both text documents and graphics. This enables the main listing to contain only the minimal amount of information on the refinement, while anciliary detail is spun off into sub-files, which are hyperlinked to the main listing file.

Third-party graphical helper applications such as npo [30] for stacks of 2D plots and O [31] for 3D plots, are triggered from the logfile to visualise residual maps and electron-density maps.

Finally, links to the documentation are scattered through the output, to provide the user with a context-sensitive help facility.

6 Applications

6.1 Synthetic data

The first tests of SHARP were aimed at establishing the numerical validity of the procedure. In statistical terms, the refinement is successful when, starting from a non-optimal point in parameter space, it converges towards a solution at a distance of 1 from the 'perfect' parameters that we introduced in the synthesis. The distance is calculated in the metric of the variance-covariance matrix, and represents units of standard deviations on the parameter estimates, taking into account all correlations between the errors.

The results of these tests are presented in Table 1, in three situations of interest. The starting parameters have been chosen to correspond to a realistic situation. Scales and lack-of-isomorphism parameters are estimated before refinement using robust procedures, and the starting distance is calculated after this estimation.

6.2 Measured data on known structures

Two datasets have been used for testing the accuracy of the refinement and phasing procedure as implemented in SHARP, and comparing it to the widely-used program MLPHARE [21].

The first of these datasets is a SIRAS experiment performed on the 26kD protein elastase, using a crystal un der high pressure of krypton gas as a derivative [35]. Refinement started with one krypton site, and the log-likelihood gradient map showed that the thermal motion of the krypton atom was anisotropic. Once this anisotropy was included in the heavy-atom model the features disappeared from the residual map. The peaks that remained in the map at this stage were a pair of elongated peaks of opposite sign, that correspond to the movement of a side-chain (val 121), and a spherical negative peak that suggests the departure of a water molecule, replaced by the krypton atom in the cavity.

Phase probability distributions were calculated with SHARP, using as a final heavy-atom model one Krypton atom with anisotropic thermal motion. The side-chain displacement and the departing oxygen atom were not included.

Because the phase probability distributions are mostly bimodal at this point (due to scarce anomalous scattering), the quality of the electron-density map is a poor judge of the quality of the phasing, since centroid phases will anyway be poor in information content. We therefore used a density modification procedure [36] to improve the phases , with the rationale that it would yield better results when the input phase probability distributions (encoded as Hendrickson-Lattman coefficients) are statistically more accurate. The density modification procedure was exactly similar for both SHARP and MLPHARE. The results are summarized in Table 2.

The second is a MAD dataset on IF3-C [37,38] (C-terminal part of translational initiation factor 3). The two methionine residues of this 94-residue protein were replaced by selenomethionines and a three-wavelength anomalous diffraction experiment was performed at the Selenium K edge.

The starting heavy-atom model consisted in two selenium atoms with isotropic thermal motion. Refinement of this model showed that, consistently with the results of other refinement procedures, the second selenium atom had a high temperature factor (around 60). Once the refinement was completed, the residual maps showed strong anisotropic features for the first selenium site and weaker anisotropy for the second. We therefore updated the heavy-atom model by allowing an anisotropic temperature factor for both seleniums atoms. The resultsing residual map showed much fewer features above the noise level, except for a 10σ peak at 1.8 Å distance from the first selenium site. The second update of the heavy-atom model allowed for a third selenium atom with an isotropic temperature factor, that refined to a low occupancy (0.2). The remarkable result was that the added occupancies of site 1 and site 3 were equal to the the occupancy of site 2 within the standard deviation of this parameter. This observation, added to the small distance between site 1 and site 3, shows that this Methionine residue has a double conformation.

As in the previous example, we show in Table 3 comparative results of MLPHARE and SHARP, both before and after density modification.

7 Conclusion

The results described here were an important step in the commissioning of SHARP. They showed, on simulated data and on measured data, that its performance was at least as good as we had expected. The sensitivity of the log-likelihood gradient maps enables very precise updating of the heavy-atom model, which in turn results in increased phasing power of a given derivative, especially in the highest resolution range, as witnessed by the consistently lower phase errors and higher correlation coefficients. The advantages of a bias-free refinement of heavy-atom parameters are clearest when a phase modification procedure is subsequently applied to the output of SHARP.

The program is now at the end of the beta-testing stage, and will be released to academic sites very soon. The World Wide Web page of interest is indicated at the top of this article.

This work was supported in part by an International Research Fellowship from the Howard Hughes Medical Institute (to GB) and by an EMBO fellowship (to ELF).

References

[1] R. E. Dickerson, J. C. Kendrew & B. E. Strandberg "The Phase Problem and Isomorphous Replacement Methods in Protein Structures" In *Symposium on computer methods and the phase problem*, p. 84. Glasgow : Pergamon Press, 1960.

[2] R.E. Dickerson, J.C. Kendrew & B.E. Strandberg . In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, J.M. Robertson & J.C. Speakman, pp.236-251. Oxford : Pergamon Press, 1961.

[3] R.G. Hart *Acta Cryst.* **14**, pp.1194-1195, 1961.

[4] D.M. Blow & F.H.C. Crick "The Treatement of Errors in the Isomorphous Replacement Method" *Acta Cryst.* **12**, 794-802, 1959.

[5] R. E. Dickerson, J. E. Weinzierl & R. A. Palmer "A Least-Squares Refinement Method for Isomorphous Replacement" *Acta Cryst.* B**24**, 997-1003, 1968.

[6] D.M. Blow & B.W. Matthews "Parameter Refinement in the Multiple Isomorphous-Replacement Method" *Acta Cryst.* A**29**, 56-62, 1973.

[7] M. G. Rossmann "The Accurate Determination of the Position and Shape of Heavy-Atom Replacement Groups in Proteins" *Acta Cryst*. **13**, 221, 1960.

[8] G. Kartha "Comparison of Multiple Isomorphous Replacament and Anomalous Dispersion Data for Protein Structure Determination.III. Refinement of Heavy Atom Positions by the Least-Squares Method"*Acta Cryst*. **19**, 883-885, 1965.

[9] E. J. Dodson, P. R. Evans & S. French "The Use of Anomalous Scattering in Refining Heavy Atom Parameters in Proteins" *Anomalous Scattering*, pp. 423-436. Edited by S. Ramaseshan and S. C. Abrahams. Copenhagen : Munskgaard, 1975.

[10] T. C. Terwilliger & D. Eisenberg "Unbiased Three-Dimensional Refinement of Heavy-Atom Parameters by Correlation of Origin-Removed Patterson Functions" *Acta Cryst*. A**39**, 813-817, 1983.

[11] J. Sygusch "Minimum-Variance Fourier Coefficients from the Isomorphous Replacement Method by Least-Squares Analysis" *Acta Cryst*. A**33**, 512-518, 1977.

[12] G. Bricogne "Multiple Isomorphous Replacement : The Problem of Parameter Refinement from Acentric Reflexions" In *Computational Crystallography*, edited by D. SAYRE, pp. 223-230. New York: Oxford University Press, 1982.

[13] G. Bricogne "Application of Isomorphous Replacement and Anomalous Dispersion Techiques to Proteins" In *Methods and Applications in Crystallographic Computing*, edited by S.R. HALL & T. ASHIDA, pp. 141-151. Oxford : Clarendon Press, 1984.

[14] V. SH. Raiz & N. S. Andreeva "Determining the Coefficients of the Fourier Series of the Electron-Density Function of Protein Crystals "*Sov. Phys. Crystallogr*. **15**, 206-210. Translated from *Kristallografiya* **15**, 246-251, 1970.

[15] R. J. Einstein "An Improved Method for Combining Isomorphous Replacement and Anomalous Scattering Diffraction Data for Macromolecular Crystals" *Acta Cryst*. **A33**, 75-85, 1977.

[16] E. A. Green "A New Statistical Model for Describing Errors in Isomorphous Replacement Data : The Case of One Derivative"*Acta Cryst*. **A35**, 351-359, 1979.

[17] G. Bricogne Unpublished lecture given at the Bischenberg conference on the Crystallography of Molecular Biology, 1985.

[18] G. Bricogne "A Bayesian Theory of the Phase Problem. I. A Multichannel Maximum-Entropy Formalism for Constructing Generalized Joint Probability Distributions of Structure Factors" *Acta Cryst.* A**44**, 517-545, 1988.

[19] G. Bricogne "A Maximum-Likelihood Theory of Heavy-atom Parameter Refinement in the Isomorphous

Replacement Method" In *Isomorphous Replacement and Anomalous Scattering* Proc. Daresbury Study Weekend, pp. 60-68. SERC Daresbury Laboratory, Warrington, England, 1991.

[20] R. J. Read "Dealing with imperfect isomorphism in multiple isomorphous replacement" In *Isomorphous Replacement and Anomalous Scattering* Proc. Daresbury Study Weekend, pp. 69-79. SERC Daresbury Laboratory, Warrington, England, 1991.

[21] Z. Otwinowski "Maximum Likelihood Refinement of Heavy Atom Parameters" In *Isomorphous Replacement and Anomalous Scattering* Proc. Daresbury Study Weekend, pp. 80-85. SERC Daresbury Laboratory, Warrington, England, 1991.

[22] E. de La Fortelle & G. Bricogne, G. "Maximum-Likelihood Heavy-Atom Parameter Refinement in the MIR and MAD Methods" In *Methods in Enzymology,* (C.W Carter & R.M. Sweet, eds), **276**, Chapter 7, pp472-494, Academic Press, 1996.

[23] D.H. Templeton & L.K. Templeton *Acta Cryst.* A**38**, 62-67, 1982.

[24] L.K. Templeton & D.H. Templeton *Acta Cryst.* A**44**, 1045-1051, 1988.

[25] V. Luzzati *Acta Cryst.* **5**, 802-810, 1952.

[26] R. J. Read "Improved Coefficients for Maps Using Phases from Partial Structures With Errors" *Acta Cryst.* A**42**, 140-149, 1986.

[27] P. Dumas "The Heavy-Atom Problem : a Statistical Analysis. I. *A Priori* Determination of Best Scaling, Level of Substitution, Lack of Isomorphism and Scaling Power" *Acta Cryst.* A**50**, 526-537, 1994.

[28] W. A. Hendrickson & E. E. Lattman "Representation of Phase Probability Distributions for Simplified Combination of Independent Phase Information" *Acta Cryst*. B**26**, 136-143, 1970.

[29] G. Bricogne, In *Methods In Enzymology* **276**, edited by C.W. Carter and R.M. Sweet (in the press), 1996.

[30] Collaborative Computational Project, Number 4 "The CCP4 suite : Programs for Protein Crystallography" *Acta Cryst.* D**50**, 760-763, 1994.

[31] T.A. Jones, J.Y. Zou, S.W. Cowan & M. Kjeldgaard, "Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models" *Acta Cryst.* A**47**, 110-119, 1991.

[32] F. A. Quiocho, N. K. Vyas, "Novel Stereospecificity of the L-arabinose-binding protein" Nature **310**, 381-386, 1984.

[33] S. J. Gamblin *et. al.* "Activity and Specificity of Human Aldolases" *J. Mol. Biol.* **219**, 573-576, 1991.

[34] M.M. Skinner *et. al.* " Structure of the Gene V Protein of Bacteriophage F1 Determined by Multiwavelength X-Ray diffraction on the Selenomethionyl Protein" *PNAS* **91**, No 5, 2071-2075, 1994.

[35] M. Schiltz *et. al.* "High-Pressure Krypton Gas and Statistical Heavy-Atom Refinement : a Successful Combination of Tools for Macromoleular Structure Determination" *Acta Cryst.* D**53**, 78-92, 1996.

[36] J.P. Abrahams & A. G. W. Leslie "Methods used in the structural determination of bovine mitochondrial F_1 ATPase" *Acta Cryst*. D**52**, 30-42, 1996.

[37] V. Biou, F. Shu & V. Ramakrishnan "X-Ray Crystallography Shows that Translational Initiation Factor IF3 Consists of 2 Compact alpha/beta Domains Linked by an alpha-helix" *EMBO. J.* **14**, 4056-4064, 1995.

[38] V. Ramakrishnan, V. Biou In *Methods in Enzymology,* (C.W Carter & R.M. Sweet, eds), **276**, Academic Press, 1997.

type of experiment	protein (pdb ID)	space-group	resolution (Å)	heavy atom (per mol.)	phasing power (iso/ano)		distance at start	distance at convergence
SIR	1abe [32] (306 resid.)	$P2_12_12_1$	20 - 3.	1 Hg	1.42	0.	4.06	1.14
MIRAS	1ald [33] (363 resid.)	$P6_422$	20. - 3.5	4U 2Pb 1Hg 3Ag 3Au	0.66 0.85 0.92 0.38 0.64	1.24 0. 1.11 0. 0.	9.05	0.92
MAD	1bgh [34] (85 resid.)	C2	20. - 2.	1 Se (3 wvl)	0. 0.2 0.68	0.57 1.01 0.6	11.2	0.88

Table 1 : general overview of test refinements on simulated data from PDB files

Resolution (Å)	**ALL**	50.0	5.25	3.73	3.05	2.64	2.36	2.16	2.00	1.87

SHARP refinement and phasing, density modification with SOLOMON (43% solv.)

FOM	**0.92**	0.96	0.97	0.95	0.93	0.92	0.92	0.92	0.87
<Δφ>	**24.3**	19.9	16.6	23.8	24.7	27.4	26.6	28.0	31.7
CORREL	**0.86**	0.81	0.93	0.87	0.86	0.84	0.84	0.84	0.80

MLPHARE (+VECREF) refinement and phasing, density modification with SOLOMON (43% solv.)

FOM	**0.87**	0.88	0.94	0.95	0.93	0.90	0.89	0.86	0.71
<Δφ>	**31.4**	30.5	27.4	29.2	31.7	33.7	33.0	33.5	35.7
CORREL	**0.81**	0.87	0.82	0.84	0.80	0.85	0.77	0.73	0.68

Table 2 : Quality of Kr-Elastase SIRAS phasing, in comparison with the refined model

Resolution (Å)	**ALL**	50.0	5.25	3.73	3.05	2.64	2.36	2.16	2.00	1.87

SHARP refinement and phasing, density modification with SOLOMON

FOM	**0.88**	0.86	0.92	0.92	0.90	0.89	0.87	0.86	0.84
<Δφ>	**29.2**	41.9	24.7	23.6	27.6	28.9	30.5	30.3	34.6
CORREL	**0.82**	0.68	0.88	0.87	0.84	0.83	0.80	0.81	0.74

MLPHARE refinement and phasing, density modification with SOLOMON

FOM	**0.88**	0.82	0.92	0.92	0.90	0.89	0.87	0.86	0.84
<Δφ>	**33.6**	41.7	26.4	30.0	32.6	32.7	35.9	38.9	40.8
CORREL	**0.78**	0.68	0.86	0.82	0.79	0.79	0.75	0.72	0.67

Table 3 : Quality of IF3-C MAD phasing, in comparison with the refined model