

PACS: 548.737

Statistical modelling and likelihood-based choice in macromolecular crystallography

Lunin V.Y. *, Afonine P.V.^{1**}, Urzhumtsev A.G.^{2***}

Institute of Mathematical problems of Biology, Russian Academy of Sciences, 142290 Pushchino, Russia

¹*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, BLDG 64R0121, Berkeley, CA 94720 USA*

²*Département de Physique, Université H.Poincaré Nancy 1, 54506 Vandoeuvre-lès-Nancy, France*

Abstract. Statistical modeling allows considering the structure under study as an element of an ensemble of structures and assigning common properties of the ensemble structures to a particular structure. Such common properties may be derived sometimes more easily for the whole ensemble than for the particular structure separately. The utility of the properties derived depends on how adequately the ensemble is introduced. The statistical likelihood may be used as the basis to choose the most reasonable statistical model. Several examples of the use of this approach in studies of structures of biological macromolecules by X-ray diffraction methods are discussed.

Key words: *structure of biological macromolecules, X-ray structure analysis, statistical modeling*

1 Introduction

1.1 Crystallographic background

The scattering of X-rays is defined by the electrons in the sample. The electron composition of the sample is described by *electron density distribution function* $\rho(\mathbf{r})$ so that $\rho(\mathbf{r})dV_{\mathbf{r}}$ is the average over the time of experiment electron charge in a primary volume $V_{\mathbf{r}}$. The determination of this distribution is the final goal of high precision X-ray study. If one studies a crystal sample this distribution is considered as periodic one in three dimensions with the periods $\mathbf{a}, \mathbf{b}, \mathbf{c}$:

$$\rho(\mathbf{r}) = \rho(\mathbf{r} + \mathbf{a}) = \rho(\mathbf{r} + \mathbf{b}) = \rho(\mathbf{r} + \mathbf{c}) . \quad (1)$$

The parallelepiped V built on these three vectors is called '*unit cell*' of the crystal. Being periodic the function $\rho(\mathbf{r})$ may be presented by a 3-D Fourier series

$$\rho(\mathbf{r}) = \frac{1}{|V|} \sum_{\mathbf{s} \in S^*} F(\mathbf{s}) \exp[i\varphi(\mathbf{s})] \exp[-2\pi i(\mathbf{s}, \mathbf{r})] , \quad (2)$$

$$F(\mathbf{s}) \exp[i\varphi(\mathbf{s})] = \int_V \rho(\mathbf{r}) \exp[2\pi i(\mathbf{s}, \mathbf{r})] dV_{\mathbf{r}} , \quad (3)$$

where the sum (2) is expanded on all vectors $\mathbf{s} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$, with integer h, k, l and $\{\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*\}$ is the basis conjugate to the $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ one. The complex Fourier coefficient $F(\mathbf{s}) \exp[i\varphi(\mathbf{s})]$ is called '*structure factor*' and its magnitude $F(\mathbf{s})$ and argument $\varphi(\mathbf{s})$ are called '*structure factor magnitude*' and '*structure factor phase*', respectively. For short we call them simply as the magnitude and phase below. We call the vector \mathbf{s} as '*reflection*', and the

* lunin@impb.psn.ru, <http://www.impb.ru/lmc>

** PAfonine@lbl.gov

*** Alexander.Ourjountsev@stmp.uhp-nancy.fr

integers h,k,l as '**reflection indexes**'. Conventional X-ray experiment allows obtaining structure factors magnitudes for a set of reflections S , while the phase values are lost in the experiment. The restoring the phase values composes the central problem of X-ray structure analysis, namely the **phase problem**. It is worthy of noting that beside the phases lost, a number of structure factor magnitudes (mostly for high frequency Fourier harmonics) are also lost in the experiment. So that even with the phases $\{\varphi(\mathbf{s}), \mathbf{s} \in S\}$ restored one cannot calculate the exact electron density distribution by (2), but an image of the electron density distribution $\rho_S(\mathbf{r})$ only. This image is called '**Fourier synthesis of the electron density distribution**'. It contains distortions caused both by the series truncation effects and by errors in the restored phase values.

The 'atomicity' hypothesis suggests that the electron density distribution is not an arbitrary function, but a sum of 'localized' atomic contributions

$$\rho(\mathbf{r}) = \sum_{j=1}^N \rho^{atom}(\mathbf{r} - \mathbf{r}_j), \quad (4)$$

where N is the number of atoms composing the studied object and $\{\mathbf{r}_j\}$ are their coordinates. In this paper for simplicity of presentation we suppose that all atoms are similar. The distribution of electron density in the atom $\rho^{atom}(\mathbf{r})$ is supposed to be known and be spherically symmetric, so that $\rho^{atom}(\mathbf{r}) = \rho^0(r)$. A gaussian three-dimensional function gives an approximate image of an electron density distribution in a single atom. If the atomic coordinates are known the summary electron density distribution may be calculated by (4). The inverse problem, *i.e.* the decomposition of electron density distribution into a sum of atomic contributions, may be much more difficult, especially when an erroneous image $\rho_S(\mathbf{r})$ of electron density is used as the input. This decomposition is called the '**interpretation**' of the Fourier synthesis or '**model building**'.

The atomicity allows expressing the structure factors in the form

$$F^{mod}(\mathbf{s}) \exp[i\varphi^{mod}(\mathbf{s})] = f^0(s) \sum_{j=1}^N \exp[2\pi i(\mathbf{s}, \mathbf{r}_j)], \quad (5)$$

where $f^0(s)$ is sinus-Fourier transform of the single atom density distribution $\rho^0(r)$, *i.e.* it is supposed to be known. The problem of defining atomic coordinates $\{\mathbf{r}_j\}$ may now be formulated as a minimization problem in $3N$ -dimensional space:

$$\sum_{\mathbf{s} \in S} (F^{mod}(\mathbf{s}; \{\mathbf{r}_j\}) - F^{obs}(\mathbf{s}))^2 \Rightarrow \min, \quad (6)$$

where $F^{obs}(\mathbf{s})$ stands for experimentally derived magnitude values. Nevertheless high dimensionality (N is about 10^4 for macromolecular structures) and highly oscillating nature of the right-hand expression in (5) do not leave a room for such straightforward approach. At the same time a local minimization procedure may be performed with contemporary algorithms and computer facilities, provided some reasonable preliminary atomic coordinates were defined.

A conventional X-ray determination of macromolecules structure passes through three main stages:

- estimating of the values of structure factors phases and calculating of corresponding Fourier synthesis $\rho_S(\mathbf{r})$;
- model building, *i.e.* the decomposition of the Fourier synthesis into the sum of atomic contributions (4) with a reasonably accuracy;
- refining of atomic coordinates (and other parameters in general case) by means of minimisation (6).

1.2 Statistical modelling

The goal of X-ray macromolecular study is to define a quite deterministic object, *i.e.* the atomic coordinates for a particular molecule. At the same time some statistical approaches

were found to be very useful to solve this problem. In this paper we do not discuss the 'traditional' use of statistical methods in the processing of experimental errors, where they are natural. The scope of this paper is to discuss some approaches to the use of the theory of probabilities in situations 'non-probabilistic' at the first glance.

The main idea of the approach may be formulated as follows:

- the studied particular structure is considered as a member of some ensemble of possible structures;
- the study of common properties of this ensemble is sometimes more simple problem than the study of the particular structure and it allows some 'typical' properties of the structures to be established;
- it is supposed that the particular studied structure obeys these common properties, so that these may be used to reduce an uncertainty in structure parameters.

The following example illustrates this general idea [1,4-6,21,23,24].

Let N be the number of atoms in the studied structure and let us consider all the structures as equally possible. To be more precise, let us consider the atomic coordinates as independent random variables uniformly distributed in the unit cell. In such the case the structure factor calculated as (5) becomes a random variable and a mathematical task to derive the joint probability distribution $P(F, \varphi)$ of its magnitude and phase may be posed and solved. The found probability distribution may be used to derive marginal distributions $P(F)$ and $P(\varphi)$ that reflect how frequently different values of the magnitude and phase may occur. The found distribution $P(F)$ is used in practice to scale the experimental data. For example, the proper scale might produce the magnitudes values as close as possible to the expected ones. At the same time the marginal distribution $P(\varphi)$ is the uniform one and produces no information. The situation changes if we modify the mathematical task and include some knowledge on the structure studied.

Let us consider three reflections $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ such that $\mathbf{s}_1 + \mathbf{s}_2 + \mathbf{s}_3 = \mathbf{0}$, and the corresponding magnitudes are known from an X-ray experiment. Let us introduce so called 'phase invariant' $T = \varphi_1 + \varphi_2 + \varphi_3$. The mathematical task is now formulated as to derive the conditional probability distribution $P(T|F(\mathbf{s}_1), F(\mathbf{s}_2), F(\mathbf{s}_3))$ for the value of the invariant provided that the three magnitudes corresponding to the random models are fixed by the results of the experiment. In this case the conditional distribution (in some approximation) has the form of the Von Mises (normal circular, *etc.*) distribution

$$P(T|F_1, F_2, F_3) \propto \exp[\kappa \cos T] \quad , \quad \kappa = \frac{2F_1 F_2 F_3}{f^0(s_1) f^0(s_2) f^0(s_3) N^2} \quad . \quad (7)$$

This distribution tends to the uniform distribution when κ tends to zero, but produces a strong restriction for phase values

$$\varphi(s_1) + \varphi(s_2) + \varphi(s_3) \approx 0 \quad (8)$$

when κ is large, *i.e.* when the corresponding experimental structure factors magnitudes are large. This suggests a practical approach to the solution of the phase problem:

- triplets of reflections satisfying $\mathbf{s}_1 + \mathbf{s}_2 + \mathbf{s}_3 = \mathbf{0}$ and possessing of strong experimental magnitudes $F(\mathbf{s}_1), F(\mathbf{s}_2), F(\mathbf{s}_3)$ are selected;
- approximate restrictions (8) are postulated for corresponding structure factor phases; these are used to estimate individual phase values.

It must be emphasised that this procedure contains a principal methodological break. The condition (8) is satisfied 'as usual' for randomly generated structures and there is no guaranty that it is satisfied for a particular structure. Nevertheless, we impose this restriction to the unique structure under study. Obviously, the mathematics is not 'responsible' for consequences of such a decision.

In spite of this methodological break such approaches (in their numerous variations) are widely used in crystallographic practice and produce quite reasonable results when the ensemble of structures is chosen adequately. The choice of the ensemble may be considered as

a way to transform our additional information on the structure studied into mathematical form, and the results obviously depend on the information involved. In what follows we consider the simplest way to define the ensemble. It is supposed that atomic coordinates of different atoms are independent and probability distribution $p_j(\mathbf{r})$ is defined for each atom (in general, these distributions are different for different atom).

1.3 Likelihood ranking

In this paper we discuss likelihood-based approaches to the choice of the ensemble of structures [10]. Statistical likelihood is a widely used tool of mathematical statistics and theory of probabilities and discussion of its mathematical aspects is beyond the scope of this paper. We remind only the main ideas how the likelihood is used to choose a hypothesis. The standard situation when we may apply the likelihood may be described in short as follows:

- there exist a set of 'experimental' measurements $x_1^{obs}, x_2^{obs}, \dots, x_M^{obs}$;
- there exist a hypothesis H that these values were obtained as a result of an independent generation of random numbers x_1, x_2, \dots, x_M with probability distributions $p_1(x), p_2(x), \dots, p_M(x)$, respectively.

A possible way to estimate a consistency of this hypothesis is to calculate how large is the probability to reproduce the result $x_1^{obs}, x_2^{obs}, \dots, x_M^{obs}$ when generating independent random values with probabilities $p_1(x), p_2(x), \dots, p_M(x)$, *i.e.* to calculate

$$L = p_1(x_1^{obs}) p_2(x_2^{obs}) \cdots p_M(x_M^{obs}). \quad (9)$$

The value $L=L(H)$ is called the likelihood of the hypothesis H . If several alternative hypotheses H_1, H_2, \dots, H_K are suggested to explain the results $x_1^{obs}, x_2^{obs}, \dots, x_M^{obs}$ then the likelihood performs a ranking of the hypothesis; the one possessing of the maximal likelihood, *i.e.* the one producing maximal chance to reproduce the experimental results may be considering as the most reasonable explanation of the experimental results. Obviously, likelihood-based ranking is one of many methods used in mathematical statistics. If the set of alternative hypotheses H_t is infinite, *e.g.* it is parameterised by continuous parameter t , the likelihood becomes a function $L(t)$ of a continuous variable (or variables).

If there exists some preliminary information that gives different preference to different hypotheses, and if this information is presented in the form of 'prior probability distribution for the hypotheses' (*e.g.* in the form of a prior probability distribution $P_{prior}(t)$ for the parameter t), then likelihood function may be used to calculate a 'posterior' distribution by means of the Bayesian formula

$$P_{post}(t) \propto L(t) P_{prior}(t). \quad (10)$$

This posterior distribution produces a ranking of hypothesis taking into account both the preliminary information and experimental results.

The definition of the likelihood as the probability to reproduce the experimental results hints at a Monte-Carlo type computer procedure to estimate the likelihood value [9,17,18]. This probability may be calculated through a computer simulation when a large number of generations of random variables are performed to estimate the chance to reproduce with a reasonable accuracy the experimental results. On the other hand, such a procedure is time-consuming and an analytical expression of the likelihood function may be very useful when it exists.

2 Likelihood based choice of an envelope and *ab initio* phasing

At first stages of X-ray study of macromolecular structures the knowledge of a molecular envelope may be very important for a further progress. Molecular envelope is a region of the unit cell that contains the most of molecular atoms. Mathematically the envelope may be defined by a binary (characteristic) function $\chi(\mathbf{r})$. The knowledge of envelope allows defining the position of the molecule in the unit cell and its shape. Sometimes several

alternative envelopes may be suggested as a solution. Likelihood ranking present a possible way to make a choice [9,11,17].

Let us suppose that observed structure factor magnitudes $\{F^{obs}(\mathbf{s})\}, \mathbf{s} \in S$ are known and two alternative envelopes Ω_1 and Ω_2 are suggested. The envelope Ω_i may be linked to the statistical hypothesis H_i : the values $\{F^{obs}(\mathbf{s})\}, \mathbf{s} \in S$ were calculated by formula (5) where the atomic coordinates $\{\mathbf{r}_j\}$ were chosen randomly (independently and uniformly) in the region Ω_i . The likelihood-based choice of the envelope in this case means the choice of the envelope that has the maximal probability to reproduce the observed magnitudes when placing atoms randomly inside it.

The likelihood ranking of envelopes may be applied to the solution of the phase problem [2,18]. Let the full set of reflections with known magnitudes be divided into two subsets: S_1 (work reflections) and S_2 (test reflections) and the problem considered is to define phases for reflections from the work set. For any trial phase set we can define the trial envelope as the region of highest values in the Fourier synthesis (2) calculated with the observed magnitudes $\{F^{obs}(\mathbf{s})\}, \mathbf{s} \in S_1$ and trial phase values

$$\Omega = \Omega(\{\varphi(\mathbf{s})\}) = \{\mathbf{r} : \rho_s(\mathbf{r}) \geq \rho_{crit}\}. \quad (11)$$

The probability to reproduce the magnitudes from S_2 through (5) when placing atoms randomly into this envelope becomes in this way an estimate of reliability of the trial phases and may be used as a criterion to choose the best phase set.

3 Phase errors estimates

3.1 Statistical modelling of phase errors

Coordinates of an atomic model built at intermediate stages of X-ray structure investigation usually contain some errors. Furthermore this model is often incomplete, *i.e.* it may miss a part of atoms of the studied object. The phases calculated by (5) from such a model contain errors. The statistical modelling may be used to estimate the quality of the calculated phases [7,19]. Let the ensemble of the structures be introduced as follows:

- let $\{\mathbf{r}_j^{mod}\}, j = 1, \dots, M$ be the coordinates of atoms of the preliminary model and N be the full number of atoms in the molecule studied;
- we consider all structures composed from N atoms;
- for $j=1, \dots, M$ the probability to find j -th atom in the position \mathbf{r} is defined by $p^0(\|\mathbf{r} - \mathbf{r}_j^{mod}\|)$ where the distribution $p^0(r)$ is supposed to be known; for $j=M+1, \dots, N$ the coordinates \mathbf{r}_j are supposed to be distributed uniformly in the unit cell.

The distribution of errors in the model coordinates $p^0(r)$ (that is supposed here to be isotropic at similar for all atoms) brings some additional knowledge of the quality of the model. Statistical model allows transforming this information on coordinate errors into the estimates of phases and errors of these estimates. To be more precise, the ensemble introduced allows deriving for every reflection the probability distribution for the corresponding structure factor phase:

$$P_s(\varphi) \propto \exp\left[2 \frac{\alpha_s}{\beta_s} F^{mod}(\mathbf{s}) F^{obs}(\mathbf{s}) \cos(\varphi - \varphi^{mod}(\mathbf{s}))\right]. \quad (12)$$

This suggests the phase value $\varphi^{mod}(\mathbf{s})$ as the more probable (that is not surprising) and allows the expected deviation of the phase values from the mean value to be estimated.

3.2 The choice of the statistical model

The estimates of the reliability of phases are derived from the values of parameters α_s and β_s that in turn are derived from the primary information $p^0(r)$ on the model quality and model completeness. In the considered case they are calculated as

$$\alpha_s = \int \cos 2\pi(\mathbf{s}, \mathbf{r}) p^0(r) dV_r, \quad \beta_s = (1 - \alpha_s^2) \sum_{k=1}^M f_k^2(s) + \sum_{k=M+1}^N f_k^2(s). \quad (13)$$

The determination of these parameters is the key step when estimating the phase quality. Being formally different for different reflection these parameters may be considered as constant inside thin spherical shells $s \approx const$, so that it would be necessary to define two parameters α and β for every shell. This determination may be based on the likelihood ranking.

The same statistical model as before allows obtaining probability distributions for magnitudes $F(s)$ as well:

$$P_s(F) \propto \exp \left[-\frac{F^2 + \alpha_s^2 (F^{mod}(\mathbf{s}))^2}{\beta_s} \right]. \quad (14)$$

Together with the experimentally observed magnitudes these distributions allow one to calculate the likelihood

$$L = \prod_{\mathbf{s} \in S} P_s(F^{obs}(\mathbf{s})) \quad (15)$$

that reflects the probability to reproduce the experimental magnitude values after the random corrections were introduced into model coordinates and the necessary number of lost atoms was added randomly. The maximization of the likelihood (15) provides reasonable estimates of model parameters.

4 Likelihood-based model refinement

In the conventional least-squares (LS) refinement every set of the model parameters (for example, atomic coordinates) is associated with a set of calculated structure factors. This refinement is aimed to choose the set of atomic parameters such that the corresponding calculated magnitudes are the most consistent with the experimental data. Traditionally, this consistency is expressed through the residual

$$\begin{aligned} Q_{LSQ} &= \sum_{\mathbf{s} \in S} w_s (\kappa F^{mod}(\mathbf{s}) - F^{obs}(\mathbf{s}))^2 \\ &= const + \sum_{\mathbf{s} \in S} \{w_s \kappa_s^2 (F^{mod}(\mathbf{s}))^2 - 2w_s \kappa_s F^{mod}(\mathbf{s}) F^{obs}(\mathbf{s})\} \end{aligned} \quad (16)$$

while other measures may be introduced, for example, those based on the comparison of intensities. It is supposed that the criterion (16) reaches its minimum, in the idealized situation equal to zero, for the exact parameter values.

A need in a statistical modelling appears when the atomic model contains irremovable errors so that the structure factor magnitudes calculated from the conventional model are different from the experimentally obtained magnitudes even for the exact values of parameters [13]. The simplest example is refinement of an incomplete atomic model. In this case the structure factors magnitudes calculated from the exact coordinates of the partial model are still different from the 'true' values since the addition of the missed atoms is necessary to make these quantities equal to each other.

Statistical refinement uses simultaneously two quite different objects: a conventional atomic model and a statistical model for corrections of the residual irremovable errors. For example, one may expect that the calculated magnitudes would be equal to the observed ones if the partial model with the exact atomic coordinates is completed by the necessary number of the lost atoms placed also exactly. This equality cannot be obtained without an explicit determination of these lost atoms. On the other hand, it is possible to estimate the probability to obtain this equality, precisely or at least approximately, after the lost atoms have been added randomly to the current partial model. One can expect that this probability is the highest one if we try to complete the exact partial model and has lower values if the partial model contains positional errors. Therefore, this probability, *i.e.* the likelihood in statistics, may be

used as a score function to estimate the quality of a partial model, making the fundamentals of the statistical refinement.

To be more formal, in this statistical consideration every current set of atomic parameters is associated with a joint probability distribution of the magnitudes of corrected structure factors rather than with a single set of calculated magnitudes. The corrected structure factors are random variables as they correspond to the current atomic model transformed by random corrections. These distributions are different for different sets of atomic parameters and obviously depend on the given law (probabilistic model) for the necessary corrections. The goal of statistical refinement of the atomic parameters may be formulated as the choice of the set of parameters, for which the corresponding joint probability distribution of corrected magnitudes is the most consistent with the experimental data. The statistical model for the correction of the residual errors is supposed to be defined in advance and fixed in the course of the refinement. The likelihood value (*i.e.* the probability to reproduce the set of observed magnitude values in the framework of this probability distribution) is an example of this measure of consistency.

In the procedure referred as *ML*-refinement [3,12,13,15,16] the criterion to be minimized is the negative logarithm of a simple type of the likelihood function [7,19], the model-dependent part of which may be formally presented as

$$Q_{ML} = \sum_{s \in S} \left\{ \frac{\alpha_s^2 (F^{mod}(s))^2}{\beta_s} - \ln \left(I_0 \left(\frac{2\alpha_s F^{mod}(s) F^{obs}(s)}{\beta_s} \right) \right) \right\}. \quad (17)$$

Here $F^{obs}(s)$ is the observed calculated structure factor magnitude for the reflection s and $F^{mod}(s)$ is the corresponding value calculated from the available atomic model. This type of a likelihood function was derived first for the error correction statistical model, which supposes that the lost atoms are added independently and uniformly in the unit cell, and the atomic position corrections are independent and have the same radial distribution for all model atoms. Some more complicated statistical models may be considered [8,14,20,22] that result in the same type of the likelihood function.

Parameters α_s and β_s are the same as in (12, 14). They reflect the scale and nature of the residual irremovable errors, as they are expected to be in the *refined* model. It must be emphasized that different suggestions about the nature of the residual irremovable errors in the refined atomic model result in different likelihood functions, while presented in the same form (17). The quality of the atomic model obtained by minimization of (17) may strongly depend on the values of α_s and β_s parameters used, and an inappropriate choice of the hypothesis about the residual model error may essentially spoil the refinement. The likelihood-based procedure described above may be used to derive these parameters.

Acknowledgments

This work was supported by RFBR grants. A part of this work was supported by the CNRS – RAS collaboration. AU thanks Pole “Intelligence Logicielle” and CRVHP, LORIA, Nancy for financial support. AU is a member of GdR 2417 CNRS. The authors thank N.Lunina, T.Petrova, T.Skovoroda, E.Vernoslova and A.Podjarny for their contribution to different parts of the work presented in this article.

REFERENCES

1. Bricogne G. 1984. Maximum Entropy and the Foundations of Direct Methods. *Acta Cryst.* **A40**. 410-455.
2. Bricogne G., Gilmore C. J. 1990. A multisolution method of phase determination by combined maximization of entropy and likelihood. I. Theory, algorithms and strategy. *Acta Cryst.* **A46**. 1990. 284-297.

3. Bricogne G., Irwin J. 1996. Maximum-Likelihood Refinement of incomplete models with BUSTER + TNT. Proceedings of the CCP4 Study Weekend. Daresbury Laboratory, Warrington. England. 85-92.
4. Cochran W. 1955. Relations between the phases of structure factors. *Acta Cryst.* **5**. 473–478.
5. Giacovazzo C. 1999. *Direct Phasing in Crystallography*. Oxford University Press. Oxford.
6. Hauptman H., Karle J. 1953. Solution of the phase problem. I. The centrosymmetric crystal. American Crystallographic Association Monograph, 3. Wilmington: The Letter Shop.
7. Lunin V.Yu., Urzhumtsev A.G. 1984. Improvement of Protein Phases by Coarse Model Modification. *Acta Cryst.* **A40**. 269-277.
8. Lunin V.Yu., Skovoroda T.P. 1995. R-free Likelihood-Based Estimates of Errors for Phases Calculated from Atomic Models. *Acta Cryst.* **A51**. 880-887.
9. Lunin V.Yu., Lunina N.L., Petrova T.E., Urzhumtsev A.G., Podjarny A.D. 1998. On the Ab initio solution of the Phase Problem for Macromolecules at Very Low Resolution. II. Generalized Likelihood Based Approach to Cluster Discrimination. *Acta Cryst.* **D54**. 726-734.
10. Lunin V.Y. 1997. The likelihood based choice of priors in statistical approaches to the phase problem. In: S.Fortier (ed.): Direct Methods for Solving Macromolecular Structures, NATO ASI Series C, Vol.507. Kluwer Academic Publishers, the Netherlands. 451-454.
11. Lunin V.Y., Lunina N.L., Petrova T.E., Skovoroda T.P., Urzhumtsev A.G., Podjarny A.D. 2000. Low-resolution ab initio phasing: problems and advances. *Acta Cryst.* **D56**. 1223-1232.
12. Lunin V.Y., Urzhumtsev A.G. 1999. Maximal Likelihood Refinement. It works, but why? CCP4 Newsletter on Protein Crystallography. **37**. 14-28.
13. Lunin V.Y., Afonine P.V., Urzhumtsev A.G. 2002. Likelihood-based refinement. I. Irremovable model errors. *Acta Cryst.* **A58**. 270-282.
14. Luzzati V. 1952. Traitement Statistique des Erreurs dans la Determination des Structures Cristallines. *Acta Cryst.* **5**. 802-810.
15. Murshudov G. N., Vagin A. A., Dodson E. J. 1997. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst.* **D53**. 240-255.
16. Pannu N. S. & Read R. J. 1996. Improved Structure Refinement Through Maximum Likelihood. *Acta Cryst.* **A52**. 659-668.
17. Petrova T.E., Lunin V.Y., Podjarny A.D. 1999. A likelihood-based search for the macromolecular position in the crystalline unit cell. *Acta Cryst.* **A55**. 739-745.
18. Petrova T.E., Lunin V.Y., Podjarny A.D. 2000. Ab initio low-resolution phasing in crystallography of macromolecules by maximization of likelihood. *Acta Cryst.* **D56**. 1245-1252.
19. Read R. J. 1986. Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. *Acta Cryst.* **A42**. 140-149.
20. Read., R.J. 1990. Structure-Factor Probabilities for Related Structures. *Acta Cryst.* **A46**. 900-912.
21. Sheldrick G.M., Hauptman H.A., Weeks C.M., Miller R., Usón I. 2001. Direct methods. In: Rossmann, M., Arnold, E. (eds.): International Tables for Crystallography, Vol.F. Kluwer Academic Publishers, Dordrecht Boston London. 333-345.
22. Srinivasan R., Parthasarathy S. 1976. *Some Statistical Applications in X-ray Crystallography*. Pergamon Press. Oxford.
23. Wilson A.J.C. 1949. The Probability Distribution of X-ray Intensities. *Acta Cryst.* **2**. 318-321.
24. Woolfson M.M. 1954. The statistical theory of sign relationship. *Acta Cryst.* **7**. 61-64.

Received: April 21, 2006. Published: May 5, 2006.