Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 7 March 2002
Accepted 7 May 2002

Confined structures: basic crystallographic aspects

Liberato De Caro,[a] Carmelo Giacovazzo[a,b]* and Dritan Siliqi[b,c]

[a]IRMEC c/o Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, [b]Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and [c]Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana, Albania. Correspondence e-mail: c.giacovazzo@area.ba.cnr.it

The concept of a confined structure is introduced. The statistical properties of the structure factors for such structures are derived and the main features of the Patterson function are described. It is shown that the structure-factor distributions for the confined structures coincide with those derived for the rational index reflections of ordinary structures. Algorithms potentially useful for protein crystal structure solution are identified and checked *via* experimental applications.

1. Symbols and notation

N: number of atoms in the unit cell
h: three-dimensional index with integral components (h_1, h_2, h_3)
p, **q**: three-dimensional indices with rational (integral included) components (p_1, p_2, p_3) and (q_1, q_2, q_3), respectively
$\boldsymbol{\varepsilon}$, **e**: three-dimensional indices with even components $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ and (e_1, e_2, e_3), respectively
o: three-dimensional index with odd components (o_1, o_2, o_3)
$F_{\mathbf{p}} = A_{\mathbf{p}} + iB_{\mathbf{p}} = R_{\mathbf{p}} \exp(i\varphi_{\mathbf{p}})$: structure factor with vectorial index **p**

$$\sum\nolimits_1 = \sum_{j=1}^{N} f_j$$

$$\sum\nolimits_2 = \sum_{j=1}^{N} f_i^2$$

For brevity, the following papers will be denoted as papers I–VI, respectively: Giacovazzo & Siliqi (1998); Giacovazzo, Siliqi, Carrozzini *et al.* (1999); Giacovazzo, Siliqi, Altomare *et al.* (1999); Giacovazzo, Siliqi & Fernández-Castaño (1999); Giacovazzo, Siliqi, Fernández-Castaño & Comunale (1999); Giacovazzo, Siliqi, Fernández-Castaño, Cascarano & Carrozzini (1999).

2. Introduction

Ramachandran (1969) first conjectured about the possible use of the Hilbert transform to solve the phase problem in crystallography. He focused his attention on the following relationships:

$$A(\mathbf{p}) = \pi^{-3} P \int_{S^*} \frac{B(\mathbf{q})}{(q_1 - p_1)(q_2 - p_2)(q_3 - p_3)} \, \mathrm{d}S^* \qquad (1)$$

and

$$B(\mathbf{p}) = -\pi^{-3} P \int_{S^*} \frac{A(\mathbf{q})}{(q_1 - p_1)(q_2 - p_2)(q_3 - p_3)} \, \mathrm{d}S^* \qquad (2)$$

from which

$$F(\mathbf{p}) = (-i/\pi)^{-3} P \int_{S^*} \frac{F(\mathbf{q})}{(q_1 - p_1)(q_2 - p_2)(q_3 - p_3)} \, \mathrm{d}S^* \qquad (3)$$

arises. P denotes the Cauchy principal value of the integral, S^* the reciprocal space. Equations (1) and (2) are known in optics as Kramer–Kronig relations. Ramachandran proposed a set of equations that involve unknown derivatives and therefore were of limited practical usefulness. The problem was revisited by Mishnev (1993) who suggested applying the Shannon sampling theorem (Shannon, 1949; Sayre, 1952) to reconstruct $F(\mathbf{p})$ from the values sampled at the Bragg reciprocal-lattice points **q**. Zanotti *et al.* (1996) applied the Mishnev equations (relating half-integral and integral index reflections) to extend and improve the phase information.

A probabilistic approach to the problem was described in papers I–VI, where the following probability distributions were derived: (*a*) the density $P(F_{\mathbf{p}})$ in $P1$ and in $P\bar{1}$ (papers I and II, respectively); (*b*) the joint probability density $P(F_{\mathbf{p}}, F_{\mathbf{q}})$ and, more generally, the joint densities $P(F_{\mathbf{p}}, \{F_{\mathbf{q}}\})$, where $\{F_{\mathbf{q}}\}$ denotes any infinite or finite set of reflections (papers III and IV); (*c*) the density $P(|F_{\mathbf{p}}||\{|F_{\mathbf{q}}|\})$, to obtain estimates of the moduli of the half-integral reflections $|F_{\mathbf{p}}|$ given the moduli of the integral index reflections (paper V); (*d*) the densities $P(F_{\mathbf{p}}, \{F_{\mathbf{q}}\})$ for reflections of reduced dimensionality (paper VI).

The outcome of the above probabilistic approach may be summarized as: (*a*) unlike the Wilson distribution, the probability $P(F_{\mathbf{p}})$ can provide an estimate of $\varphi_{\mathbf{p}}$ together with the corresponding reliability parameter; (*b*) for reflections of any dimensionality, the phase and modulus of $F_{\mathbf{p}}$ may be estimated given phases and moduli of the set $\{F_{\mathbf{q}}\}$. The method provides also the corresponding reliability factors. When applied to the

canonical case (**p** is then an integral component vector and **q** a half-integral component vector, or *vice versa*), the formulas encompass Mishnev relationships; (*c*) moduli of the half-integral reflections can be evaluated by exploiting the moduli of the standard reflections only.

Following the original ideas of Sayre (1952), a new approach has recently been introduced in crystallography: the reciprocal-space oversampling method. For non-crystalline specimens it has been shown (Miao *et al.*, 1998; Sayre *et al.*, 1998) that sampling the diffraction pattern of a finite specimen at a spacing finer than Nyquist spacing generates a no-density region surrounding the electron density of the specimen, which may be used to retrieve the phase information. The technique has recently been extended (Miao & Sayre, 2000) to crystalline specimens, it does not require atomic resolution, but imposes the measurement of the intensities between the Bragg peaks. Such intensities are generally faint compared with those measurable at Bragg points and weaken with the size of the crystal. As a consequence, the technique is probably practical only for small crystals and, additionally, imposes a very high radiation dose, very unusual even for modern crystallography techniques. In practice, the phase problem is transformed into the radiation-damage problem (*i.e.* recording all the diffraction peaks necessary for the success of the method before the crystal is irreversibly damaged).

In this paper, we show that the study of the probability distributions of the rational index reflections performed in papers I–VI is strictly related to the method described by Miao & Sayre (2000). In order to obtain such a result, we develop the concept of the *confined structure* and we study the related crystallographic properties. Furthermore, we are able to obtain probabilistic estimates of the intensities not available *via* diffraction experiments (*i.e.* the analogy of the non-Bragg intensities for which the Miao & Sayre method imposes the experimental measure). Our experimental tests can be considered as a pilot study for assessing the level of accuracy of the measurements for non-Bragg intensities necessary for the solution of the phase problem.

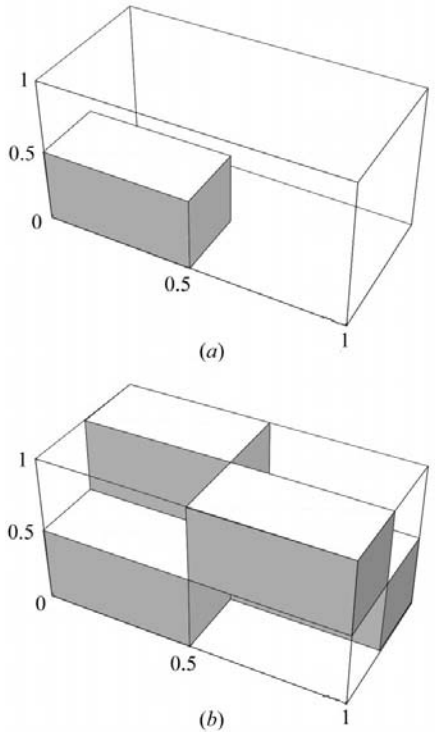

Figure 1
(*a*) Confined structures: the shadowed volume satisfies the conditions (4). (*b*) An *F*-centred confinement.

3. The concept of the confined crystal structure

Let G be the space group of a crystal structure and U_S its standard unit cell. It is usual to consider the atomic positions as the primitive random variables for any probabilistic approach aiming at stating the joint probability distributions of the structure factors. The range over which the atomic coordinates are allowed to vary are usually the following:

$$0 \le x < 1, \quad 0 \le y < 1, \quad 0 \le z < 1.$$

Crystal structures for which the electron density is spread out in the above interval will be denoted as *ordinary structure* (OS) and $\rho(\mathbf{r})$ will be their electron-density function.

Let us now double the length of the U_S axes. In the new cell (say U_C), the atomic coordinates will assume values half of the previous ones and the structure will be confined to the subcell defined by the following limits (see Fig. 1*a*):

$$0 \le x < 1/2, \quad 0 \le y < 1/2, \quad 0 \le z < 1/2. \qquad (4)$$

No atoms will be out of these limits. We will refer to this case as a *confined structure* (CS) and $\rho'(\mathbf{r})$ will be its electron-density function. The volume of U_C is eight times larger than the volume of U_S and the electron density is uniformly zero in seven of the eight subcells. The change (from U_S to U_C) in direct space involves in the reciprocal space the duplication of the indices of the reflections: so the (hkl) indices (relative to U_S) will become ($e_1 = 2h$, $e_2 = 2k$, $e_3 = 2l$) in the new reference. While the (hkl) reflections completely define the electron density in U_S, the ($e_1e_2e_3$) reflections are only 1/8 of the reflections necessary to define the electron density in U_C.

The confinement defined by the conditions (4) is not the only one possible. For example, we can duplicate the lengths of the U_S axes so that the atomic coordinates are confined in the subcell (4), but we may additionally assume that the subcell defined by the conditions (4) repeats identically in the following subcells (see Fig. 1*b*):

$$\begin{aligned} &0 \le x < 1/2, \quad 1/2 \le y < 1, \quad 1/2 \le z < 1, \\ &1/2 \le x < 1, \quad 0 \le y < 1/2, \quad 1/2 \le z < 1, \\ &1/2 \le x < 1, \quad 1/2 \le y < 1, \quad 0 \le z < 1/2. \end{aligned} \qquad (4')$$

Also in this case, the electron density in the remaining subcells is assumed to be identically equal to zero. In this case, U_C becomes an *F*-centred cell and the only non-systematically absent reflections will have parity (e_1, e_2, e_3) or (o_1, o_2, o_3).

The above considerations suggest that one can freely choose U_C as an *A*-, *B*-, *C*- or *I*-centred cell by requiring that the structure defined in the subcell (4) identically repeats in:

$$\begin{aligned} &0 \le x < 1/2, \quad 1/2 \le y < 1, \quad 1/2 \le z < 1, \\ &\qquad\qquad \text{for an } A\text{-centred cell}, \\ &1/2 \le x < 1, \quad 0 \le y < 1/2, \quad 1/2 \le z < 1, \\ &\qquad\qquad \text{for a } B\text{-centred cell}, \\ &1/2 \le x < 1, \quad 1/2 \le y < 1, \quad 0 \le z < 1/2, \\ &\qquad\qquad \text{for a } C\text{-centred cell}, \\ &1/2 \le x < 1, \quad 1/2 \le y < 1, \quad 1/2 \le z < 1, \\ &\qquad\qquad \text{for an } I\text{-centred cell}. \end{aligned} \tag{4''}$$

Other types of confinement can be thought of, involving triplication, quadruplication *etc.*, of the axes of the U_S cell. In fact, the larger the ratio U_C/U_S the larger will be the computer resources necessary to perform the calculations, unless centred cells are used: a primitive confined cell with axes P times those of an ordinary structure will have P^3 more reflections than the standard ones.

It may be worthwhile stressing that the crystal-structure confinement described in this paper is only a mathematical expedient: actually such structures are not crystallochemically consistent and therefore are unrealistic. Additionally, the borders of the subcells where the electron density is allowed to be different from zero involve discontinuity of the electron-density function, incompatible with the presence of atoms on the border. However, the idea of confinement: (*a*) shows features of non-negligible interest for the solution of the phase problem; (*b*) has a real counterpart in nature when part of the unit cell is disordered and therefore does not substantially contribute to X-ray diffraction; (*c*) the problem of the presence of atoms on the borders can be neglected at this stage of the work, since for big crystal structures the percentage of atoms lying on the CS cell border surfaces is very small and it should not hinder the convergence of any phasing algorithm.

4. The role of symmetry in the confined crystal structures

In the following, we will essentially refer to a specific type of confinement: the axes of U_S have been doubled, the structure is confined to the subcell (4), the electron density in the other seven subcells is identically zero.

Let us indicate by $\mathbf{r}_j = (x_j, y_j, z_j)$ and by $\mathbf{r}'_j = (x'_j, y'_j, z'_j) = (x_j/2, y_j/2, z_j/2)$ the positional vectors of the jth atom in U_S and in U_C, respectively. Accordingly, $F_{\mathbf{h}}$ and $F'_{2\mathbf{h}}$ are the corresponding structure factors for the two structures. If $\{C_s\} = \{\mathbf{R}_s, \mathbf{T}_s\}$ is the set of symmetry operators characterizing the space group of the OS, then

$$\begin{aligned} F_{\mathbf{h}} &= \sum_{s=1}^{m} \sum_{j=1}^{t} f_j \exp\{2\pi i \mathbf{h}(\mathbf{R}_s \mathbf{r}_j + \mathbf{T}_s)\} \\ &= \sum_{s=1}^{m} \sum_{j=1}^{t} f_j \exp\{2\pi i [2\mathbf{h}(\mathbf{R}_s \mathbf{r}_j/2 + \mathbf{T}_s/2)]\}, \end{aligned}$$

where t is the number of atoms in the asymmetric unit and m the number of symmetry operators. Then,

$$F'_{2\mathbf{h}} = F_{\mathbf{h}} \tag{5}$$

provided the set of symmetry operators $\{C'_s\} = \{\mathbf{R}'_s, \mathbf{T}'_s\}$ is used, where $R'_s = R_s$ and $T'_s = T_s/2$. Since

$$F_{\mathbf{hR}_s} = \exp(-2\pi i \mathbf{h}\mathbf{T}_s) F_{\mathbf{h}}, \tag{6}$$

using (5) into (6) gives

$$F'_{2\mathbf{hR}_s} = F_{\mathbf{hR}_s} = \exp(-2\pi i \mathbf{h}\mathbf{T}_s) F_{\mathbf{h}} = \exp(-2\pi i \mathbf{h}\mathbf{T}_s) F'_{2\mathbf{h}}. \tag{7}$$

The above results may be summarized as: the set of symmetry operators $\{C_s\} = \{\mathbf{R}_s, \mathbf{T}_s\}$ in U_S transform into the set $\{C'_s\} = \{\mathbf{R}'_s, \mathbf{T}'_s\}$ when the unit cell U_C is chosen. However, the set $\{C'\}$ only relates the reflections F' with index $2\mathbf{hR}_s$ to the reflection F' with index $2\mathbf{h}$; no symmetry relations can be stated among reflections $F'_{\mathbf{hR}_s}$ and $F'_{\mathbf{h}}$ if $\mathbf{h} \neq (e_1, e_2, e_3)$. An analogous behaviour was noted in paper IV for the reflection $F_{\mathbf{p}}$, where $\mathbf{p}$ is a vectorial index with at least one half-integral component. The rationale is the following: the symmetry elements present in the U_S cell are degraded in U_C to local symmetry operators. Accordingly, the space group of a CS is usually $P1$, except for some symmorphic space groups. We show in Fig. 2, as examples of symmorphic groups for which the space-group symmetry is maintained, the U_S and the U_C cells when G is $P2$ and $P4$: the different location in U_C of the symmetry elements adds a translational component to the rotational matrix, *i.e.* the symmetry-equivalent positions for $P2$ in U_C are (x, y, z), $(\bar{x} + \frac{1}{2}, y, \bar{z} + \frac{1}{2})$ so that $F_{\bar{h}k\bar{l}} = F_{hkl} \exp[-i\pi(h + l)]$ instead of the standard relation $F_{\bar{h}k\bar{l}} = F_{hkl}$.

By using the same arguments, the reader will easily verify that, for the case $G = P3$, the symmetry degrades to $P1$ when a CS is considered.

5. The distribution $P(R, \varphi)$ in $P1$

In this section, we will show that the distributions $P(R_{\mathbf{h}}, \varphi_{\mathbf{h}})$ for a CS coincide with the distributions derived in paper II for the rational index reflections of an OS. Let us consider a three-dimensional crystal under the following four assumptions: (*a*) no symmetry element is present; (*b*) the N atoms constituting the entire chemical content of the unit cell U_C are confined in the subcell defined by (4); (*c*) the electron density in the other seven subcells is identically zero; (*d*) the atomic coordinates

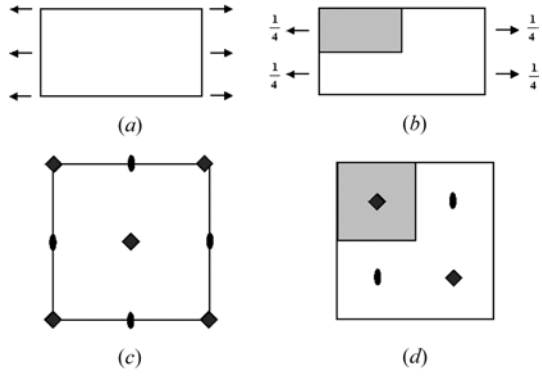

Figure 2
(*a*) The standard $P2$ space-group diagram; (*b*) the $P2$ diagram for the CS; (*c*) the standard $P4$ diagram; (*d*) the $P4$ diagram for the CS.

are randomly distributed in the subcell (4): they constitute the primitive random variables of our probabilistic approach.

Then the characteristic function of the distribution $P(A_{\mathbf{h}}, B_{\mathbf{h}})$, say $C(u, \mathrm{v}) = \langle \exp i(uA_{\mathbf{h}} + \mathrm{v}B_{\mathbf{h}})\rangle$ may be written in terms of the cumulants of the distribution:

$$C(u, \mathrm{v}) = \exp[i(uK_{10} + \mathrm{v}K_{01}) - 0.5(u^2K_{20} + \mathrm{v}^2K_{02} + 2u\mathrm{v}K_{11})], \tag{8}$$

where u and v are carrying variables associated with $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$, respectively. In their turn, the cumulants K_{ij} may be expressed in terms of the moments m_{rs} of

$$\begin{aligned} P(A_{\mathbf{h}}, B_{\mathbf{h}}):& \\ K_{10} &= m_{10} = \langle A_{\mathbf{h}}\rangle, \\ K_{01} &= m_{01} = \langle B_{\mathbf{h}}\rangle, \\ K_{20} &= m_{20} - m_{10}^2 = \langle A_{\mathbf{h}}^2\rangle - \langle A_{\mathbf{h}}\rangle^2, \\ K_{02} &= m_{02} - m_{01}^2 = \langle B_{\mathbf{h}}^2\rangle - \langle B_{\mathbf{h}}\rangle^2, \\ K_{11} &= m_{11} - m_{10}m_{01} = \langle A_{\mathbf{h}}B_{\mathbf{h}}\rangle - \langle A_{\mathbf{h}}\rangle\langle B_{\mathbf{h}}\rangle. \end{aligned}$$

The calculation of the moments may be performed in accordance with the following relationships:

$$\langle \cos(2\pi hx)\rangle = 2\int_0^{1/2} \cos(2\pi hx)\,\mathrm{d}x = c_{h/2}$$

$$\langle \sin(2\pi hx)\rangle = 2\int_0^{1/2} \sin(2\pi hx)\,\mathrm{d}x = s_{h/2},$$

where $c_h = \sin(2\pi h)/(2\pi h)$, $\ldots$, $s_h = [1 - \cos(2\pi h)]/(2\pi h)$. Accordingly,

$$\begin{aligned} \langle \cos[2\pi(hx + ky + lz)]\rangle &= \langle \cos 2\pi hx \cos 2\pi ky \cos 2\pi lz \\ &\quad - \sin 2\pi hx \sin 2\pi ky \cos 2\pi lz \\ &\quad - \sin 2\pi hx \cos 2\pi ky \sin 2\pi lz \\ &\quad - \cos 2\pi hx \sin 2\pi ky \sin 2\pi lz\rangle \\ &= \cos[\pi(h + k + l)/2]c_{h/4}c_{k/4}c_{l/4} = c_{\mathbf{h}}, \\ \langle \sin[2\pi(hx + ky + lz)]\rangle &= \sin[\pi(h + k + l)/2]c_{h/4}c_{k/4}c_{1/4} = s_{\mathbf{h}}, \\ \langle A_{\mathbf{h}}\rangle &= \textstyle\sum_1 c_{\mathbf{h}}, \quad \langle B_{\mathbf{h}}\rangle = \sum_1 s_{\mathbf{h}}, \\ K_{20} &= 0.5\textstyle\sum_2(1-2c_{\mathbf{h}}^2), \quad K_{02} = 0.5\sum_2(1 - 2s_{\mathbf{h}}^2), \\ K_{11} &= \textstyle\sum_2 c_{\mathbf{h}}s_{\mathbf{h}}. \end{aligned}$$

The Fourier transform of (8) gives

$$\begin{aligned} P(A_{\mathbf{h}}, B_{\mathbf{h}}) = (2\pi)^{-1}\Delta^{-1/2} \exp\{-(2\Delta)^{-1}[&K_{02}(A_{\mathbf{h}} - K_{10})^2 \\ &+ K_{20}(B_{\mathbf{h}} - K_{01})^2 - 2(A_{\mathbf{h}} - K_{10})(B_{\mathbf{h}} - K_{01})K_{11}]\}, \end{aligned} \tag{9}$$

where $\Delta = (K_{02}K_{20} - K_{11}^2)$.

Equation (9) formally coincides with the equation (I.1), derived in paper I for describing the statistical properties of the rational index reflections in an OS. Consequently, the joint probability distribution $P(|F_{\mathbf{h}}|, \varphi_{\mathbf{h}})$ and the marginal distributions $P(|F_{\mathbf{h}}|)$ and $P(\varphi_{\mathbf{h}})$ derivable from (9) for a CS will coincide with the distributions (I.4) and (I.6), respectively. For brevity they are not quoted here.

We specialize the probability distribution (9) for the few cases of interest.

(i) At least one of the indices is even (and different from zero). Then,

$$c_{\mathbf{h}} = c_{2\mathbf{h}} = s_{\mathbf{h}} = s_{2\mathbf{h}} = 0, \quad K_{02} = K_{20} = 0.5\textstyle\sum_2$$

and the Wilson statistics hold. In this case fall the reflections (*eee*), (*oee*), (*eoe*), (*eeo*), (*eoo*), (*oeo*), (*ooe*). For them, the structure confinement cannot provide phase distributions other than the uniform distribution in the range $(0, 2\pi)$.

(ii) All the three indices are odd. Then,

$$\begin{aligned} c_{h/4} &= \sin(\pi h/2)/(\pi h/2) = 2(-1)^{(h-1)/2}/(\pi h), \\ c_{\mathbf{h}} &= c_{2\mathbf{h}} = 0, \\ s_{\mathbf{h}} &= -8/(\pi^3 hkl), \quad s_{2\mathbf{h}} = 0, \\ K_{10} &= 0, \quad K_{01} = -8\textstyle\sum_1/(\pi^3 hkl), \quad K_{11} = 0, \\ K_{20} &= 0.5\textstyle\sum_2, \quad K_{20} = 0.5\sum_2[1 - 128/(\pi^6 h^2k^2l^2)]. \end{aligned}$$

The structure confinement destroys, for the (*ooo*) reflections, the uniform distribution of $2\pi\mathbf{hr}_j$ on the trigonometric circle, so that the Wilson distribution does not hold anymore. Since the parameters $c_{\mathbf{h}}$, $c_{2\mathbf{h}}$, $s_{\mathbf{h}}$, $s_{2\mathbf{h}}$ coincide with the corresponding parameters of the distribution $P(A_{\mathbf{p}}, B_{\mathbf{p}})$ derived for an OS when $\mathbf{p}$ is a vectorial index with half-integer components, the marginal distributions $P(|F_{\mathbf{h}}|)$, $P(\varphi_{\mathbf{h}})$ and the conditionals $P(|F_{\mathbf{h}}||\varphi_{\mathbf{h}})$ and $P(\varphi_{\mathbf{h}}||F_{\mathbf{h}}|)$ will coincide with the distributions (I.16), (I.17), (I.18) and (I.19), respectively. Phases can be qualitatively estimated *via*

$$(\varphi_{\mathbf{h}})_{\mathrm{est}} = \tan^{-1}(K_{01}/K_{10}). \tag{10}$$

(iii) One of the indices is zero and the other two are odd numbers [*i.e.* $\mathbf{h}$ = (0*oo*) or (*o*0*o*) or (*oo*0)]. In this case,

$$\begin{aligned} &s_{\mathbf{h}} = s_{2\mathbf{h}} = c_{2\mathbf{h}} = 0, \quad K_{01} = 0, \quad K_{02} = 0.5\textstyle\sum_2, \quad K_{11} = 0, \\ &c_{\mathbf{h}} = 4/(\pi^2 kl), \quad K_{10} = -4\textstyle\sum_1/(\pi^2 kl), \\ &\qquad K_{20} = 0.5\textstyle\sum_2[1 - 32/(\pi^4 k^2 l^2)], \quad \text{if } \mathbf{h} = (0oo), \\ &c_{\mathbf{h}} = -4/(\pi^2 hl), \quad K_{10} = -4\textstyle\sum_1/(\pi^2 hl), \\ &\qquad K_{20} = 0.5\textstyle\sum_2[1 - 32/(\pi^4 k^2 l^2)], \quad \text{if } \mathbf{h} = (o0o), \\ &c_{\mathbf{h}} = -4/(\pi^2 hk), \quad K_{10} = -4\textstyle\sum_1/(\pi^2 hk), \\ &\qquad K_{20} = 0.5\textstyle\sum_2[1 - 32/(\pi^4 k^2 k^2)], \quad \text{if } \mathbf{h} = (oo0). \end{aligned}$$

Again, the Wilson distribution does not hold any more. The parameters $c_{\mathbf{h}}$, $c_{2\mathbf{h}}$, $s_{\mathbf{h}}$, $s_{2\mathbf{h}}$ coincide with the corresponding parameters of the distribution $P(A_{\mathbf{p}}, B_{\mathbf{p}})$ derived for an OS when $\mathbf{p}$ is a vectorial index with two half-integral components and one zero component. This case has been treated in paper VI.

(iv) Two of the indices are zero and the third is odd. Then,

$$\begin{aligned} &c_{\mathbf{h}} = c_{2\mathbf{h}} = s_{2\mathbf{h}} = 0, \quad K_{20} = 0.5\textstyle\sum_2, \quad K_{11} = 0, \\ &s_{\mathbf{h}} = 2/(\pi h), \quad K_{10} = 0, \quad K_{01} = 2\textstyle\sum_1/(\pi h), \\ &\qquad K_{02} = 0.5\textstyle\sum_2[1 - 8/(\pi^2 h^2)], \quad \text{if } \mathbf{h} = (o00), \\ &s_{\mathbf{h}} = 2/(\pi k), \quad K_{10} = 0, \quad K_{01} = 2\textstyle\sum_1/(\pi k), \\ &\qquad K_{02} = 0.5\textstyle\sum_2[1 - 8/(\pi^2 k^2)], \quad \text{if } \mathbf{h} = (0o0), \\ &s_{\mathbf{h}} = 2/(\pi l), \quad K_{10} = 0, \quad K_{01} = 2\textstyle\sum_1/(\pi l), \\ &\qquad K_{02} = 0.5\textstyle\sum_2[1 - 8/(\pi^2 l^2)], \quad \text{if } \mathbf{h} = (00o). \end{aligned}$$

Again the Wilson distribution does not hold any more. The parameters $c_{\mathbf{h}}$, $c_{2\mathbf{h}}$, $s_{\mathbf{h}}$, $s_{2\mathbf{h}}$ coincide with the corresponding parameters of the distribution $P(A_{\mathbf{p}}, B_{\mathbf{p}})$ derived for an OS when $\mathbf{p}$ is a vectorial index with two half-integral components and one zero component.

The above results and, in particular, equation (10) indicate that phase probabilities and, therefore, phase estimates for CS are quite different from those valid for OS: consequently, also their use in the phasing procedure has to be different.

6. The distributions $P(F_{\mathbf{h}} \mid \{F_{\mathbf{k}}\})$

In papers III and IV, distributions of type $P(F_{\mathbf{p}}|\{F_{\mathbf{q}}\})$ were derived that are able to estimate, for the OS's, the modulus and the phase of $F_{\mathbf{p}}$ given the moduli and the phases of the set $\{F_{\mathbf{q}}\}$. Particular emphasis was given to the canonical case, in which $\mathbf{p}$ is a half-integral index reflection and $\mathbf{q}$ is a standard reflection, and *vice versa*. [The conclusive formulas were denoted in paper IV as (CPR1), (CPR2), (CPR3), (CPR4).] Analogous formulas were obtained in paper VI relating the modulus and phase of a *mixed type reflection* $\mathbf{p}$ (*i.e.* one with half-integral and integral indices) to the moduli and to the phases of reflections belonging to specific sections of the three-dimensional reciprocal space [see formulas denoted as (MCPR5), (MCPR6), (MCPR7)]. The case of reflections with reduced dimensionality was also considered.

In paper V, the probabilistic approach was further developed to explore whether the moduli of the half-integral index reflections could be evaluated in the absence of phase information; *i.e.* by exploiting the moduli of the standard reflections only [see formula (V.20)].

The probabilistic results obtained in §5 suggests that each distribution $P(F_{\mathbf{p}}|\{F_{\mathbf{q}}\})$ derived for an OS coincides, in a CS, with a corresponding distribution $P(F_{\mathbf{h}}|\{F_{\mathbf{k}}\})$. For shortness, we will not give further details.

7. The Patterson function of the confined structures

Let us use the following notation:

(*a*) $P(\mathbf{u})$ is the Patterson function for an OS and $P'(\mathbf{u})$ is the corresponding function for the CS;

(*b*) $\rho'_e(\mathbf{r})$ is the electron-density function of an OS calculated by using only the $\mathbf{e}$ reflections. The periodicity of $\rho'_e(\mathbf{r})$ is half the periodicity of $\rho'(\mathbf{r})$;

(*c*) $\Phi'(\mathbf{r})$ is a form function, with periods defined by U_C, equal to unity for $(n_1 \le x < n_1 + 1/2, n_2 \le y < n_2 + 1/2, n_3 \le z < n_3 + 1/2)$ and equal to zero elsewhere. n_1, n_2, n_3 are integral values.

Then,

$$P'(\mathbf{u}) = \rho'(\mathbf{r}) * \rho'(-\mathbf{r}), \tag{11}$$

where

$$\rho'(\mathbf{r}) = \rho'_{\mathbf{e}}(\mathbf{r}) \cdot \Phi'(\mathbf{r}).$$

The Fourier transform of (11) may be written as

$$\begin{aligned} |F'_{\mathbf{h}}|^2 &= T[\rho'(\mathbf{r}) * \rho'(-\mathbf{r})] \\ &= T[\rho'_{\mathbf{e}}(\mathbf{r}) \cdot \Phi'(\mathbf{r})] \cdot T[\rho'_{\mathbf{e}}(-\mathbf{r}) \cdot \Phi'(-\mathbf{r})] \\ &= |F'_{\mathbf{e}} * D'(\mathbf{h})|^2 \\ &= \Big| \sum_{\mathbf{e}} F_{\mathbf{e}} D'(\mathbf{h} - \mathbf{e}) \Big|^2 . \end{aligned} \tag{12}$$

The relation (12) in more explicit terms becomes

$$\begin{aligned} |F'_{\mathbf{h}}|^2 &= \sum_{\mathbf{e}} |F_{\mathbf{e}}|^2 |D'(\mathbf{h} - \mathbf{e})|^2 \\ &\quad + \sum \sum_{\mathbf{e}_1 \neq \mathbf{e}_2} F_{\mathbf{e}_1} D'(\mathbf{h} - \mathbf{e}_1) \overline{F}_{\mathbf{e}_2} \overline{D'}(\mathbf{h} - \mathbf{e}_2) \\ &= \sum_{\mathbf{e}} |F_{\mathbf{e}}|^2 |D'(\mathbf{h} - \mathbf{e})|^2 + \sum \sum_{\mathbf{e}_1 \neq \mathbf{e}_2} |F_{\mathbf{e}_1} F_{\mathbf{e}_2}| \\ &\quad \times \exp[i(\varphi_{\mathbf{e}1} - \varphi_{\mathbf{e}2})] D'(\mathbf{h} - \mathbf{e}_1) \overline{D'}(\mathbf{h} - \mathbf{e}_2), \end{aligned} \tag{13}$$

where $\overline{F}$ and $\overline{D'}$ are the complex conjugates of F and D'.

Equation (13) shows that the Patterson function $P'(\mathbf{u})$ cannot be calculated by using only the coefficients $|F_{\mathbf{e}}|^2$: indeed, the value of any $|F_{\mathbf{h}}|^2$, for $\mathbf{h} \neq \mathbf{e}$, is not unequivocally determined by the prior knowledge of the $|F_{\mathbf{e}}|^2$ but also depends on the phase values $\{\varphi_{\mathbf{e}}\}$. The above result was first stressed by Mishnev (1996), who instead of the function $P'(\mathbf{u})$ (*i.e.* the CS Patterson function) considered the autocorrelation function of a single unit cell, and proposed its estimate by using measured Bragg reflections and non-Bragg intensities obtained *via* the discrete Hilbert transform.

In the absence of any *a priori* available phase information (*e.g.* obtained by direct methods), we can assume that the $\varphi_{\mathbf{e}}$'s are uniformly dispersed. In this case, the following approximation holds:

$$|F'_{\mathbf{h}}|^2 \approx \sum_{\mathbf{e}} |F_{\mathbf{e}}|^2 |D'(\mathbf{h} - \mathbf{e})|^2 . \tag{14}$$

In order to clarify the role of (14) in our probabilistic approach, we note that its use as a simplified form of (13) corresponds to the approximation

$$|F_{\mathbf{h}}|^2 = \sum_{j=1}^{N} f_j^2 + \sum_{j,k=1}^{N} f_j f_k \exp[2\pi i \mathbf{h} \cdot (\mathbf{r}_j - \mathbf{r}_k)] \approx \sum_{j=1}^{N} f_j^2 ,$$

which is usually made when the atomic positions are completely unknown. However, (14) is more efficient than the above approximation owing to the fact that it exploits the prior knowledge that a large part of the CS unit cell is empty. Of course, the prior knowledge of some vectors $\mathbf{r}_j$ can improve the too crude approximation $|F_{\mathbf{h}}|^2 \approx \sum_{j=1}^{N} f_j^2$, as well as the prior knowledge of some phases would enable us to use (13) rather than (14). In this paper, we will explore the case in which no prior information is available besides the experimental moduli.

In order to do this, let us now calculate the expression for D' for a CS according to the conditions (4). We have

$$D'(\mathbf{h}) = \int_0^{1/2} \int_0^{1/2} \int_0^{1/2} \exp 2\pi i(hx + ky + lz)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z = c_{\mathbf{h}} + i s_{\mathbf{h}},$$

from which

$$|D'(\mathbf{h})|^2 = c_{\mathbf{h}}^2 + s_{\mathbf{h}}^2 = c_{h/4}^2 c_{k/4}^2 c_{l/4}^2 = \left(\frac{\sin \pi h/2}{\pi h/2}\right)^2 \left(\frac{\sin \pi k/2}{\pi k/2}\right)^2 \left(\frac{\sin \pi l/2}{\pi l/2}\right)^2.$$

Accordingly, the values of $|D'(\mathbf{h})|^2$ for the different reflection parities are given by:

if $\mathbf{h} \equiv (o_1 o_2 o_3)$ then $|D'(\mathbf{h})|^2 = 64\pi^{-6}(o_1 o_2 o_3)^{-2}$;
if $\mathbf{h} \equiv (0 o_2 o_3)$ then $|D'(\mathbf{h})|^2 = 16\pi^{-4}(o_2 o_3)^{-2}$;
if $\mathbf{h} \equiv (00 o_3)$ then $|D'(\mathbf{h})|^2 = 4\pi^{-2} o_3^{-2}$;
if $\mathbf{h} \equiv (o_1 e_2 e_3)$ or $(e_1 o_2 o_3)$ or $\ldots (e_1 o_2 e_3)$ then $|D'(\mathbf{h})|^2 = 0$.

Then (14) may be rewritten, for the different parities, as follows:

$$|F'_{o_1 o_2 o_3}|^2 \approx \sum_{\mathbf{e}} |F_{e_1 e_2 e_3}|^2 64\pi^{-6}[(o_1 - e_1)(o_2 - e_2)(o_3 - e_3)]^{-2}$$

$$|F'_{\varepsilon_1 o_2 o_3}|^2 \approx \sum_{e_1 e_2 e_3} |F_{e_1 e_2 e_3}|^2 |D'(\varepsilon_1 - e_1, o_2 - e_2, o_3 - e_3)|^2 \equiv \sum_{e_2, e_3} |F_{\varepsilon_1 e_2 e_3}|^2 |D'(0, o_2 - e_2, o_3 - e_3)|^2 \equiv \sum_{e_2, e_3} |F_{\varepsilon_1 e_2 e_3}|^2 16\pi^{-4}[(o_2 - e_2)(o_3 - e_3)]^{-2},$$

valid also when $e_1 = 0$;

$$|F'_{\varepsilon_1 \varepsilon_2 o_3}|^2 \approx \sum_{e_1 e_2 e_3} |F_{e_1 e_2 e_3}|^2 |D'(\varepsilon_1 - e_1, \varepsilon_2 - e_2, o_3 - e_3)|^2 \equiv \sum_{e_3} |F_{\varepsilon_1 \varepsilon_2 e_3}|^2 |D'(0, 0, o_3 - e_3)|^2 \equiv \sum_{e_3} |F_{\varepsilon_1 \varepsilon_2 e_3}|^2 4\pi^{-2}(o_3 - e_3)^{-2},$$

valid also when $e_1 = e_2 = 0$.

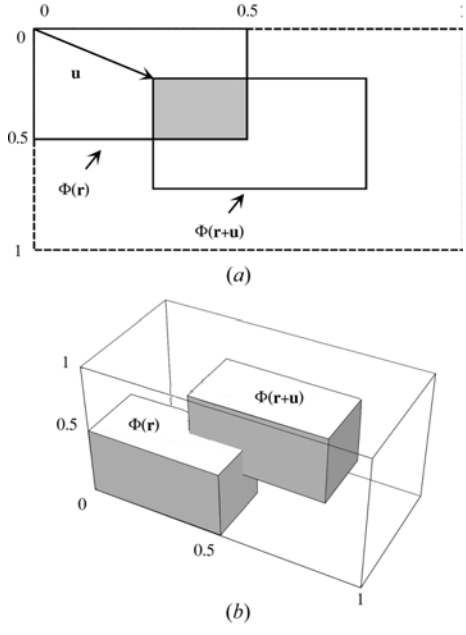

Figure 3
(*a*) Two-dimensional structure confined to the subcell $[0 \le x < \frac{1}{2}, 0 \le y < \frac{1}{2}]$. The shaded area coincides with $P_\Phi(\mathbf{u})$. (*b*) Three-dimensional structure; confined according to the conditions (4). $P_\Phi(\mathbf{u})$ is the volume in common to the two subcells.

The practical estimation of the $|F'_{\mathbf{h}}|^2$ may be quite long in practice: indeed, each F' depends on the contributions arising from all the $|F_{\mathbf{e}}|^2$. An alternative method, much faster and more efficient, may be devised. Let us rewrite (14) as

$$|F_{\mathbf{h}}|^2 = |F'_{\mathbf{e}}| * |D(\mathbf{h})|^2.$$

Its Fourier transform gives

$$P'(\mathbf{u}) = P'_{\mathbf{e}}(\mathbf{u}) \cdot P_\Phi(\mathbf{u}), \tag{15}$$

where

$$P_\Phi(\mathbf{u}) = \int_0^{1/2} \int_0^{1/2} \int_0^{1/2} \Phi(\mathbf{r})\Phi(\mathbf{r}+\mathbf{u})\, d\mathbf{r}. \tag{16}$$

For each $\mathbf{u}$, the function $P_\Phi(\mathbf{u})$ is the volume of the unit cell which belongs both to $\Phi(\mathbf{r})$ and to $\Phi(\mathbf{r}+\mathbf{u})$. To help the reader to evaluate the integral in (16), we return to the CS defined by the conditions (4) and we consider first the two-dimensional case [*i.e.* $\rho'(\mathbf{r})$ is confined to the subcell defined by $0 \le x \le 1/2$, $0 \le y \le 1/2$]. It is evident that $P_\Phi(\mathbf{u})$ is the shaded area in Fig. 3(*a*). Analogously, for the three-dimensional case (see Fig. 3*b*), $P_\Phi(\mathbf{u})$ is the volume common to the subcell defined by (4) and to its shifted image. We note that, unlike $\rho'(\mathbf{r})$, the function $P_\Phi(\mathbf{u})$ is no longer confined to the subcell defined by the conditions (4): it reaches its maximum value at $\mathbf{u} = 0$, where $P_\Phi(\mathbf{u}) = 2^{-3}$, and is equal to zero when u_1 and/or u_2 and/or $u_3 = \frac{1}{2}$. Simple calculations show that

$$P_\Phi(\mathbf{u}) = |(u_1 - \tfrac{1}{2})(u_2 - \tfrac{1}{2})(u_3 - \tfrac{1}{2})|, \tag{17}$$

which for the two-dimensional case degrades to

$$P_\Phi(\mathbf{u}) = |(u_1 - \tfrac{1}{2})(u_2 - \tfrac{1}{2})|. \tag{18}$$

Equation (18) is depicted in Fig. 4. According to (17) and (18), $P'(\mathbf{u})$ is expected to be close to zero in a large central zone. This expectation is legitimated by Figs. 5(*a*) and 5(*b*), where a simple OS of four atoms and its Patterson map $P(\mathbf{u})$ are shown; the corresponding functions $\rho'(\mathbf{r})$ and $P'(\mathbf{u})$ are shown in Figs. 5(*c*) and 5(*d*), respectively. To complete our analysis, we state the expression for $P_\Phi(\mathbf{u})$ when

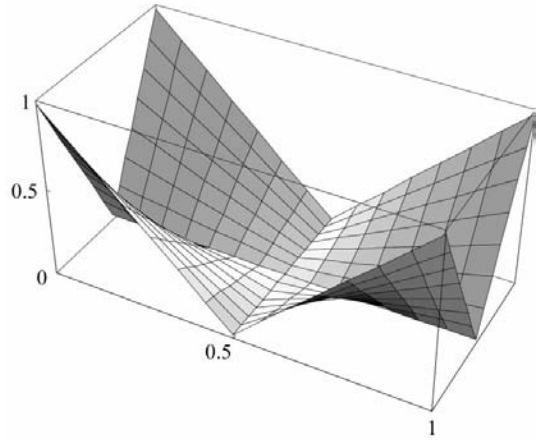

Figure 4
$P_\Phi(\mathbf{u})$ for a two-dimensional confinement.

(a) the confinement is described by an F-centred cell:

$$P_\Phi(\mathbf{u}) = |u_1u_2(\tfrac{1}{2}-u_3) + u_2u_3(\tfrac{1}{2}-u_1) + u_1u_3(\tfrac{1}{2}-u_2) - (\tfrac{1}{2}-u_1)(\tfrac{1}{2}-u_2)(\tfrac{1}{2}-u_3)|;$$

(b) the confinement is described by a C-cell:

$$P_\Phi(\mathbf{u}) = |u_1u_2(\tfrac{1}{2}-u_3) - (\tfrac{1}{2}-u_1)(\tfrac{1}{2}-u_2)(\tfrac{1}{2}-u_3)|.$$

Expressions for the A-, B- and I-centred cells can be derived by analogy. The relation (15) suggests a quite fast tool for estimating the values of $|F'_h|^2$ on the assumption of $\varphi_\mathbf{e}$'s uniformly dispersed: $P'_e(\mathbf{u})$ is calculated and then multiplied by $P_\Phi(\mathbf{u})$; by fast Fourier transform (FFT) of the modified Patterson map, the moduli $|F'_h|^2$ are obtained.

8. Experimental tests

The new point of view introduced by the concept of the CS suggests new phasing algorithms, faster than those exploited in papers I–VI, and potentially useful for the crystal structure solution. In this section, we will use the experimental data of the protein M-FABP [Zanotti *et al.* (1992); space group $P2_12_12_1$, chemical composition $C_{667}N_{170}O_{216}S_3$, $Z = 4$, data resolution up to 2.14 Å, 7589 measured reflections] for a pilot study of the algebraic and probabilistic relationships provided by the crystallography of the CS's. We will use four protocols.

Protocol 1. The experimental moduli and the published phases of the (*eee*) reflections are used to estimate the moduli and the phases of the reflections with different parity. The results will be compared with the corresponding outcomes provided by the probabilistic relationships obtained in paper V .

In accordance with §7, $\rho'_\mathbf{e}(\mathbf{r})$ is calculated in U_C; zeroing the map in seven of the eight subcells and inverting it by FFT provide the simultaneous estimate of reflections of any parity. For the 219665 $F_\mathbf{h}$ reflections (with $\mathbf{h} \neq \mathbf{e}$) (this large number is justified by the fact that we are operating in $P1$), the average error is $\langle|\Delta\varphi|\rangle = 28^\circ$ and the discrepancy index R is 0.32, where

$$R = \sum_\mathbf{h} ||F_\mathbf{h}|_\text{true} - |F_\mathbf{h}|_\text{est}| \Big/ \sum_\mathbf{h} |F_\mathbf{h}|_\text{true}.$$

F_true is the true (calculated from the published crystal structure) structure factor and F_est is its estimated value.

17 s is the c.p.u. time necessary (Dell-Precision 530 Xeon 1.7 GHz workstation) for completing the entire process. On the contrary, the estimation of only 31472 (*ddd*) reflections, based on the formulas (CPR1) and (CPR2) of paper V, requires 1753 s: in this case, $\langle|\Delta\varphi|\rangle = 32^\circ$ and $R = 0.33$ (see Table 2 of paper V).

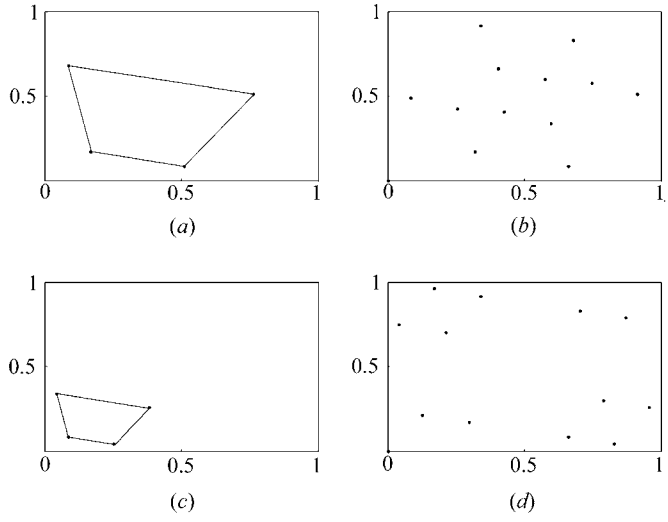

Figure 5
$\rho(\mathbf{r})$ and $P(\mathbf{u})$ are sketched in (a) and (b) for a simple two-dimensional OS. The corresponding functions for the CS are shown in (c) and (d).

Protocol 2. In order to check how the error in the phases $\varphi_\mathbf{e}$ degrades the reliability of the $F_\mathbf{h}$ estimates (with $\mathbf{h} \neq \mathbf{e}$), we assign to them the MIR phase values (with average error equal to 63° for 7355 reflections). The same algorithm described for protocol 1 (we invert the positive part of the map) provides $\langle|\Delta\varphi|\rangle = 66^\circ$ and $R = 0.53$. If equations (CPR1) and (CPR2) are applied, we obtain $\langle|\Delta\varphi|\rangle = 70^\circ$ and $R = 0.54$.

Protocol 3. We try to estimate the moduli $|F_\mathbf{h}|^2$, with $\mathbf{h} \neq \mathbf{e}$, from the moduli $|F_\mathbf{e}|^2$. In accordance with §7, the Patterson function $P'_\mathbf{e}(\mathbf{u})$ is calculated and then modified *via* (15). The Fourier inversion of the map will provide the $|F_\mathbf{h}|^2$ estimates. The discrepancy index R is 0.42, much better than the value 0.53 (see Table 6 of paper V) obtained *via* the probabilistic approach. The difference between the two results may be ascribed to the mathematical difficulties, described in paper V, of obtaining a sensitive expression for $\langle|F_\mathbf{h}||\{|F_\mathbf{e}|\}\rangle$ from the complicated distribution $P(|F_\mathbf{h}|, \{|F_\mathbf{e}|\})$ [see equation (13) of paper V).

Protocol 4. We explore now some algorithms potentially useful for the solution of the phase problem. If the measured intensities of the rational index reflections are not available for an OS (as in a standard diffraction experiment), the simplest algorithm for solving the phase problem for a CS is the following (Miao & Sayre, 2000).

(i) A trial (random, in the absence of prior information) set of phases $\{\varphi_\mathbf{e}\}$ is associated with the measured set $\{|F_\mathbf{e}|_\text{obs}\}$, and the electron-density function $\rho'_\mathbf{e}(\mathbf{r})$ is calculated.

(ii) Since (see §7)

$$\rho'(\mathbf{r}) = \rho'_e(\mathbf{r}) \cdot \Phi'(\mathbf{r}), \tag{19}$$

the density outside the subcell to which the structure is confined is driven to zero.

(iii) The $F_\mathbf{h}$ estimates are obtained by FFT (after having imposed the positivity condition on the map). The $F_\mathbf{e}$ values to be used in the next calculations are constituted by the measured $|F_\mathbf{e}|_\text{obs}$ moduli and by the calculated phases. For reflections with index $\mathbf{h} \neq \mathbf{e}$, the structure factors to use in the

next calculations coincide, in modulus and phase, with the estimates provided by the FFT.

(iv) A new electron-density map is calculated, and steps (ii)–(iii) are applied cyclically.

We used the above algorithm under different conditions.

Case (a): $\{\varphi_{\mathbf{e}}\}$ is a random set of phases, the moduli $\{|F_{\mathbf{h}}|\}$ with $\mathbf{h} \neq \mathbf{e}$ are those derived by the inverse FFT (*i.e.* they are not those calculated by the known atomic positions). We used a large variety of techniques for driving to zero the electron density outside the subcell to which the structure is confined. The algorithm never converged to significant phases (at least in several hundreds or few thousands of cycles).

Case (b): $\{\varphi_{\mathbf{e}}\}$ is a random set of phases, the moduli $\{|F_{\mathbf{h}}|\}$ with $\mathbf{h} \neq \mathbf{e}$ are supposed to be measured (actually, we have simulated this experimental situation by using the calculated non-Bragg reflections from the known atomic positions). In this case, at step (iii) of the algorithm, after the application of the inverse FFT, the $|F_{\mathbf{h}}|$ so-obtained are replaced by the 'observed' ones. The algorithm converges in a few thousands of cycles (about 1800 cycles in the best run) to the average phase error $\langle|\Delta\varphi|\rangle = 35^\circ$ for the 7589 Bragg reflections. This finding confirms the theoretical possibility of solving *ab initio* protein structures, just knowing the intensities of non-Bragg reflections.

Case (c): $\{\varphi_{\mathbf{e}}\}$ is the set of MIR phases, the moduli $\{|F_{\mathbf{h}}|\}$ with $\mathbf{h} \neq \mathbf{e}$ are supposed to be measured, as in (*b*). Even in this case, at step 3 of the algorithm, the $|F_{\mathbf{h}}|$ with $\mathbf{h} \neq \mathbf{e}$ obtained by the inverse FFT are replaced by the 'observed' values. The algorithm reduces the phase error from 63° for 7355 reflections to 31° for 7595 reflections in 20 cycles;

Case (d): $\{\varphi_{\mathbf{e}}\}$ is the set of MIR phases, the moduli $\{|F_{\mathbf{h}}|\}$ with $\mathbf{h} \neq \mathbf{e}$ are those obtained by the inverse FFT. The algorithm does not converge to significant phases (at least in several hundreds or a few thousands of cycles).

The outcome of the above tests may be interpreted as follows: the amount of structural information contained in the set $\{F_{\mathbf{e}}\}$ is entirely transferred into the set $\{F_{\mathbf{h}}\}$. In particular, the moduli $F_{\mathbf{h}}$ and the phases $\varphi_{\mathbf{h}}$ critically depend on the moduli and on the phases of **e** reflections. Random $\varphi_{\mathbf{e}}$ values (case *a*) will produce random $\varphi_{\mathbf{h}}$ values and structurally insignificant $|F_{\mathbf{h}}|$ moduli: in the absence of some supplementary information, the phasing process would not converge. Analogously, the absence of a supplementary constraint cannot lead MIR phases (case *d*) to better values. The assumed prior knowledge of the $|F_{\mathbf{h}}|$'s constitutes (cases *b* and *c*) a supplementary source of information independent of the $\varphi_{\mathbf{e}}$ values and leads the phasing process to success.

The above considerations suggest that estimates (experimental or statistical) of the $|F_{\mathbf{h}}|$ moduli, which are independent of the current phase values, are a necessary (in some cases not sufficient) condition for the success of the phasing process. We therefore planned three supplementary tests:

Case (e): $\{\varphi_{\mathbf{e}}\}$ is the set of MIR phases, the moduli $\{|F_{\mathbf{h}}|\}$ with $\mathbf{h} \neq \mathbf{e}$ are estimated as in protocol 1 (with discrepancy index $R = 0.32$). This set of non-Bragg reflections, being determined by Fourier inversion of the CS map, contains structural information up to the experimental data resolution of 2.14 Å. In this case, at step (iii) of the algorithm, after the FFT application, the estimated values of the moduli $|F_{\mathbf{h}}|$ are restored. By imposing the positivity condition, the algorithm reduces the phase error from 63 to 36° in 45 cycles. On the contrary, in the cases (*b*) and (*c*) discussed above, using the non-Bragg reflection moduli calculated by the known atomic coordinates, they contain a larger amount of information.

Case (f): $\{\varphi_{\mathbf{e}}\}$ is the set of MIR phases, the moduli $\{|F_{\mathbf{h}}|\}$ with $\mathbf{h} \neq \mathbf{e}$ are estimated as in protocol 3 (with the discrepancy index $R = 0.42$). Again, at step (iii) of the algorithm, the estimated values of the $|F_{\mathbf{h}}|$ moduli are restored. The algorithm does not converge.

Case (g): $\{\varphi_{\mathbf{e}}\}$ is the set of MIR phases, the moduli $\{|F_{\mathbf{h}}|\}$ with $\mathbf{h} \neq \mathbf{e}$ are supposed to be measured, as assumed in (*b*). Even in this case, at step (iii) of the algorithm the $|F_{\mathbf{h}}|$ with $\mathbf{h} \neq \mathbf{e}$ obtained by the FFT are restored with the 'observed' values. The only difference with respect to case (*b*) concerns the centring of the CS cell: the total number of non-Bragg reflections strongly changes as a function of different centring, owing to the systematic absences. We have found that the algorithm converges always, excepted the case of the *F*-centred CS cell. It is worth noting that in this case we have just one additional non-Bragg reflection [*i.e.* the (*ooo*)'s] in the CS cell for each Bragg reflection [*i.e.* the (*eee*)'s] of the OS cell.

To summarize the above results: the test (*e*) suggests that, even in the presence of a non–negligible error in the moduli $|F_{\mathbf{h}}|$, the phasing process may succeed. The test (*f*) indicates that estimates more accurate than those provided by the inversion of the modified Patterson map, obtained on the assumption of $\varphi_{\mathbf{e}}$'s uniformly dispersed, are necessary. Finally, the test (*g*) seems to indicate that at non-atomic resolution of the experimental data (such as in the considered example) it is indispensable to have the ratio 'number of non-Bragg reflections/number of Bragg reflections' larger than one in order to compensate for the lack of information on the measured standard reflections.

9. Conclusions

In this paper, we have established the crystallography of the confined structures. We have shown that: (*a*) the statistical properties of the structure factors of the CS are identical with the properties stated in papers I–VI for the rational index reflections of an OS; (*b*) the properties of the Patterson maps of the CS may be used for estimating the moduli of the non-measured structure factors; (*c*) sufficiently good estimates of such moduli constitute a supplementary source of information useful for solving the phase problem in macromolecular crystallography; (*d*) the present estimates of such moduli (by inversion of a modified Patterson map based on the assumption of $\varphi_{\mathbf{e}}$'s uniformly dispersed) are not too far from the usefulness threshold.

The above results are also very important for the experimental aspects connected to the method. Indeed, the experimental measurements of the $|F_{\mathbf{h}}|$ moduli (with $\mathbf{h} \neq \mathbf{e}$) are expected to be much less reliable than for the ordinary

structure factors because: (i) the intensities of non-Bragg reflections are very faint; (ii) they are continuously distributed in the reciprocal space for few-cell structures and, therefore, decomposition techniques could be necessary to partition a measured intensity into its components.

Consequently, it may be argued that supplementary efforts would be necessary to make the measures of experimental non-Bragg structure-factor moduli or its theoretical estimations closer to the usefulness threshold.

The authors are grateful to Antonio Cervellino for very useful discussions.

References

Giacovazzo, C. & Siliqi, D. (1998). *Acta Cryst.* A**54**, 957–970.
Giacovazzo, C., Siliqi, D., Altomare, A., Cascarano, G. L., Rizzi, R. & Spagna, R. (1999). *Acta Cryst.* A**55**, 322–331.
Giacovazzo, C., Siliqi, D., Carrozzini, B., Guagliardi, A. & Moliterni, A. G. (1999). *Acta Cryst.* A**55**, 314–321.
Giacovazzo, C., Siliqi, D. & Fernández-Castaño, C. (1999). *Acta Cryst.* A**55**, 512–524.
Giacovazzo, C., Siliqi, D., Fernández-Castaño, C., Cascarano, G. L. & Carrozzini, B. (1999). *Acta Cryst.* A**55**, 984–990.
Giacovazzo, C., Siliqi, D., Fernández-Castaño, C. & Comunale, G. (1999). *Acta Cryst.* A**55**, 525–532.
Miao, J. & Sayre, D. (2000). *Acta Cryst.* A**56**, 596–605.
Miao, J., Sayre, D. & Chapman, H. N. (1998). *J. Opt. Soc. Am.* A**15**, 1662–1669.
Mishnev, A. F. (1993). *Acta Cryst.* A**49**, 159–161.
Mishnev, A. F. (1996). *Acta Cryst.* A**52**, 629–633.
Ramachandran, G. N. (1969). *Mater. Res. Bull.* **4**, 525–534.
Sayre, D. (1952) *Acta Cryst.* **5**, 843.
Sayre, D., Chapman, H. N. & Miao, J. (1998). *Acta Cryst.* A**54**, 232–239.
Shannon, C. E. (1949). *Proc. Inst. Radio Eng. NY*, **37**, 10–41.
Zanotti, G., Fogale, F. & Capitani, G. (1996). *Acta Cryst.* A**52**, 757–765.
Zanotti, G., Scapin, G., Spadon, P., Veerkamp, J. H. & Sacchettini, J. C. (1992). *J. Biol. Chem.* **267**, 18541–18550.