Acta Cryst. (1996). A**52**, 365–368

The Map Correlation Coefficient for Optimally Superposed Maps

V. Yu. Lunin* and N. L. Lunina

Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow Region 142292, Russia. E-mail: lunin@impb.serpukhov.su

(*Received 10 February 1995; accepted 22 November 1995*)

Abstract

In judging the effectiveness of methods of solving crystal structures, it is customary to use the mean phase error or the map correlation coefficient as a measure of success. When methods for *ab initio* phase determination are being developed, it is necessary to take into account that the solutions found may correspond to different origin and enantiomorph choices, so a procedure of map alignment must be performed prior to calculation of a phase-similarity measure. A fast-Fourier-transform-based algorithm for map alignment is proposed for different space groups. The possibility of additional alignment by map 'overturning' is discussed for very low resolution protein syntheses. Simple analytical equations defining the permitted origin shift are included.

1. Introduction

To judge how close together two phase sets are, the mean phase difference is usually calculated. To emphasize the role of strong reflections, the weighted-by-modulus values of the mean phase difference or map correlation coefficient

$$C = \int_V \rho_1(\mathbf{r})\rho_2(\mathbf{r})\,\mathrm{d}V_r \Big/ \left(\int_V \rho_1^2(\mathbf{r})\,\mathrm{d}V_r \int_V \rho_2^2(\mathbf{r})\,\mathrm{d}V_r\right)^{1/2} \tag{1}$$

are commonly used measures (Lunin & Woolfson, 1993). The map correlation coefficient (1) may be equally expressed in terms of structure-factor values as

$$C = \alpha \sum_{\mathbf{s}} F_1(\mathbf{s})F_2(\mathbf{s}) \exp\{i[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})]\}$$
$$= \alpha \sum_{\mathbf{s}} F_1(\mathbf{s})F_2(\mathbf{s}) \cos[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})], \tag{2}$$

$$\alpha = \left(\sum_{\mathbf{s}} F_1^2(\mathbf{s}) \sum_{\mathbf{s}} F_2^2(\mathbf{s})\right)^{-1/2}. \tag{3}$$

Here and below, we suppose that all the maps and structure-factor sums are calculated without the F_{000} term.

When the phase problem is solved *ab initio*, *e.g.* by generating random phase sets, it is necessary to realize that the phase sets produced may correspond to different origin and enantiomorph choices, so preliminary alignment of maps (or corresponding phase sets) must be carried out before the phase-difference measure is calculated (Lunin, Urzhumtsev & Skovoroda, 1990; Hašek & Schenk, 1992). As a criterion of best alignment, the map correlation coefficient (1) may be used.

If the space-group symmetry permits only a finite number of different origin choices, all variants of the origin and enantiomorph choice may be easily checked in order to find the best one (Lunin, 1993; Weeks, De Titta, Hauptman, Thuman & Miller, 1994). When arbitrary shifts along one or three ($P1$) axes are allowed, the search for the best alignment is more complicated and it forces one sometimes (Weeks *et al.*, 1994) to look for other, less direct, measures of phase similarity. Nevertheless, as is shown below [and similarly in the problem of a translation search in the molecular replacement method (Crowther & Blow, 1967; Navaza & Vernoslova, 1995)], the map correlation coefficient considered as a function of the origin-shift value has a Fourier-series structure. The fast Fourier transform (Ten Eyck, 1973) may be applied in such cases to calculate a set of correlation values for origin shifts sampled to a fine discrete grid in real space and the best one chosen. When necessary, the values found may be refined by local minimization but in practice a fine enough grid seems to provide sufficient accuracy.

2. Map alignment

2.1. *Group P1*

Let us suppose that a map $\rho_1(\mathbf{r})$ is fixed and we have to fit another map $\rho_2(\mathbf{r})$ to the first one using the map correlation coefficient (2) as the measure of success. As the basic possible transformations, we consider:

(i) origin shifts

$$\rho_2'(\mathbf{r}) = \rho_2(\mathbf{r} + \mathbf{u}); \tag{4}$$

(ii) enantiomorph alternation

$$\rho_2'(\mathbf{r}) = \rho_2(-\mathbf{r}). \tag{5}$$

Equivalent transformations of the phases have the forms

$$\varphi_2'(\mathbf{s}) = \varphi_2(\mathbf{s}) - 2\pi(\mathbf{s}, \mathbf{u}); \tag{6}$$

$$\varphi_2'(\mathbf{s}) = -\varphi_2(\mathbf{s}). \tag{7}$$

The map correlation coefficient for the fixed first and the shifted second map is

$$C(\mathbf{u}) = \alpha \sum_{\mathbf{s}} F_1(\mathbf{s})F_2(\mathbf{s}) \exp\{i[\varphi_2'(\mathbf{s}) - \varphi_1(\mathbf{s})]\}$$
$$= \alpha \sum_{\mathbf{s}} F_1(\mathbf{s})F_2(\mathbf{s}) \exp\{i[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})] \times \exp[-2\pi i(\mathbf{s}, \mathbf{u})]\}. \quad (8)$$

Formula (8) is a Fourier transform, calculated with coefficients

$$H(\mathbf{s}) \exp[i\psi(\mathbf{s})] = F_1(\mathbf{s})F_2(\mathbf{s}) \exp\{i[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})]\}. \quad (9)$$

The fast-Fourier-transform algorithm (Ten Eyck, 1973) allows one to perform efficient calculation of the set of correlation values corresponding to possible shifts sampled on a grid in the unit cell.

When origin shifts are combined with the enantiomorph change (in this case, the first transformation step is the enantiomorph change and the second one the origin shift), they lead to the set of correlation values calculated as Fourier series with the coefficients

$$H(\mathbf{s}) \exp[i\psi(\mathbf{s})] = F_1(\mathbf{s})F_2(\mathbf{s}) \exp\{i[-\varphi_2(\mathbf{s}) - \varphi(\mathbf{s})]\}. \quad (10)$$

The choice of the optimal grid point (and enantiomorph) from the two three-dimensional sets of correlation values is then straightforward.

2.2. *Non-trivial space groups*

When one is working with a non-trivial space group, not all origin shifts preserve map symmetry. Similarly, the enantiomorph change may not be permitted by space-group symmetry. In this case, only the shifts permitted by the space group must be considered when looking for the best alignment. As is shown below, there is no need to calculate three-dimensional Fourier transforms in this case.

If space-group symmetry restricts possible origin shifts to a finite number, only these values must be tried.

Another case is when the possible origin shifts are restricted to a finite number of variants in two dimensions and may be arbitrary in the third direction. For simplicity, let arbitrary shifts be permitted along z axes. Let $\mathbf{U} = \{\mathbf{u}^m\}_{m=1}^M$ be a set of permitted origin shifts in the xy plane: $\mathbf{u}^m = (x^m, y^m, 0)$. [For example, for $P2$, $\mathbf{u}$ contains four shifts (0,0,0), (0,$\frac{1}{2}$,0), ($\frac{1}{2}$,0,0), ($\frac{1}{2}$,$\frac{1}{2}$,0); for $P6$, it consists of one element (0,0,0) only.] Any permitted shift in this case may be expressed as

$$(u_x, u_y, u_z) = (u_x^m, u_y^m, 0) + (0, 0, u_z), \quad (11)$$

where $(u_x^m, y_y^m, 0) \in \mathbf{U}$ and $u_z \in (0, 1)$.

Formula (8) can be reduced in this case to

$$C^m(u_z) = \alpha \sum_l H^m(l) \exp[i\psi^m(l)] \exp(-2\pi i l u_z) \quad (12)$$

with

$$H^m(l) \exp[i\psi^m(l)] = \sum_{h,k} F_1(hkl)F_2(hkl) \times \exp\{i[\varphi_2(hkl) - \varphi_1(hkl)] - 2\pi i(hu_x^m + ku_y^m)\}. \quad (13)$$

So, in this case, it is possible to perform a one-dimensional Fourier transform for every shift from the **U** set (and enantiomorph choice if permitted) and then look for the best correlation value among M (or $2M$ if the enantiomorph may be changed) one-dimensional sets.

Similar formulas may be written if arbitrary origin shifts are possible along other than z axes. The only difference is in the indexes over which the summation in (13) is extended.

3. The use of an asymmetric part of the structure-factor set

It is customary to use an asymmetric part of a structure-factor set when working with non-trivial space groups. We reduce now the summation in (2), (3) and (13) to an asymmetric part of the structure-factor set. Let S be the full set of structure factors used to calculate $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$ maps, let $\Gamma = \{(G_\nu, t_\nu)\}_{\nu=1}^n$ be the corresponding space group and S_{asym} be an asymmetric part of the set S, *i.e.*

(i) all reflections of S may be generated from S_{asym} by space-group and Hermitian symmetry transformations;

(ii) for every $\mathbf{s} \in S_{\text{asym}}$, $\mathbf{s}$ is the only point of the orbit $\{\pm G^T\mathbf{s}\}_{\nu=1}^n$ that belongs to S_{asym}.

If we denote

$$m(\mathbf{s}) \text{ as the number of different reciprocal space points in the orbit } \{\pm G^T\mathbf{s}\}_{\nu=1}^n, \quad (14)$$

then (3) may be reduced to

$$\alpha = \left\{ \sum_{\mathbf{s}\in S_{\text{asym}}} m(\mathbf{s})F_1(\mathbf{s})^2 \sum_{\mathbf{s}\in S_{\text{asym}}} m(\mathbf{s})F_2(\mathbf{s})^2 \right\}^{-1/2}. \quad (15)$$

To reduce summation in (2) and (13) to S_{asym}, it is necessary to take into account how the space group restricts possible origin shifts and systematic absences. It is shown in Appendix *A* that for nonvanishing reflections

$$\exp[-2\pi i(\mathbf{s}, G_\nu\mathbf{u})] = \exp[-2\pi i(\mathbf{s}, \mathbf{u})] \quad (16)$$

for all permitted shifts $\mathbf{u}$ and all $\nu = 1, \ldots, n$. The general reciprocal-space symmetry formula

$$F(G_\nu^T\mathbf{s}) \exp[i\varphi(G_\nu^T\mathbf{s})] = F(\mathbf{s}) \exp\{i[\varphi(\mathbf{s}) - 2\pi(\mathbf{s}, t_\nu)]\} \quad (17)$$

gives in this case

$$F_1(G_\nu^T\mathbf{s})F_2(G_\nu^T\mathbf{s}) \times \exp\{i[\varphi_2(G_\nu^T\mathbf{s}) - \varphi_1(G_\nu^T\mathbf{s})] - 2\pi i(G_\nu^T\mathbf{s}, \mathbf{u}^m)\} = F_1(\mathbf{s})F_2(\mathbf{s})\exp\{i[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})] - 2\pi i(\mathbf{s}, G_\nu\mathbf{u}^m)\} = F_1(\mathbf{s})F_2(\mathbf{s})\exp\{i[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})] - 2\pi i(\mathbf{s}, \mathbf{u}^m)\}, \quad (18)$$

so all the points of the orbit $\{G_\nu^T\mathbf{s}\}_{\nu=1}^n$ participate equally in the sums. This means that instead of (2) we may write

$$C = \alpha \sum_{\mathbf{s}\in S_{\text{asym}}} m(\mathbf{s})F_1(\mathbf{s})F_2(\mathbf{s})\cos[\varphi_2(\mathbf{s}) - \varphi_1(\mathbf{s})], \quad (19)$$

where $m(\mathbf{s})$ is again the number of different points in the orbit $\{\pm G^T\mathbf{s}\}_{\nu=1}^n$. The sum (13) is calculated for particular l, so for $l \neq 0$ only half of the orbit $\{\pm G^T\mathbf{s}\}_{\nu=1}^n$ is present in the sum and we can write the sum for $\mathbf{s} = (hkl) \in S_{\text{asym}}$ as

$$H^m(l)\exp[i\psi^m(l)] = \sum_{\substack{h,k \\ \mathbf{s}\in S_{\text{asym}}}} p(hkl)F_1(hkl)F_2(hkl) \times \exp\{i[\varphi_2(hkl) - \varphi_1(hkl)] - 2\pi i(hu_x^m + ku_y^m)\}, \quad (20)$$

where $p(hkl)$ is the number of different points in the orbit $\{\pm G^T\mathbf{s}\}_{\nu=1}^n$ with fixed l value. For $l = 0$, it is equal to the multiplicity $m(hk0)$ defined in (14). In other cases, $p(\mathbf{s}) = m(\mathbf{s})/2$. [It is worth noting that we use summation (13) for space groups allowing arbitrary origin shift along z axes only.]

4. Overturned maps

It has been mentioned previously (Subbiah, 1993; Lunin *et al.*, 1995) that some very low resolution phasing methods tend sometimes to give as solution an 'overturned' synthesis $-\rho(\mathbf{r})$ instead of the true one. In other words, the solvent region may be found rather that the molecular region. Discussions of the reasons for this are beyond the scope of this paper and we concentrate here on how this additional 'degree of freedom' for permitted map transformation influences the map alignment.

The possibility of changing synthesis sign (or, equivalently, increasing all phases by π) means that we must look for the largest $\max|C(\mathbf{t})|$ value instead of $\max C(\mathbf{t})$. As

$$\max_{\mathbf{t}}|C(\mathbf{t})| = \max\left\{\max_{\mathbf{t}} C(\mathbf{t}), -\min_{\mathbf{t}} C(\mathbf{t})\right\},$$

then we must look for both maximal and minimal correlation values and take as optimal the one possessing the largest amplitude. So the consideration of map overturning as an additional permitted transformation does not necessitate a new correlation-value calculation but only changes what we take as the best correlation value.

This work was partially supported by ISF grant RMZ000 and RFFI grant 94-04-12844. The Fortran programs for optimal map alignment search may be obtained from the authors by e-mail to lunina@impb.serpukhov.su.

APPENDIX A
Permitted origin shifts

Let a map $\rho(\mathbf{r})$ have the symmetry group $\Gamma = \{(G_\nu, \mathbf{t}_\nu)\}_{\nu=1}^n$, *i.e.* $\rho(G_\nu\mathbf{r} + t_\nu) = \rho(\mathbf{r})$ for every $\mathbf{r}$, $\nu = 1, \ldots, n$. Let $\Gamma^t = \{(E, \mathbf{t})\}_{\mu=1}^m$ ($\mathbf{t}_0 = 0, t_\mu \neq 0|_{\text{mod }1}$ for $\mu = 2, \ldots, m$) be its translation subgroup. The subgroup Γ^t consists of identical transformations only ($m = 1$) for primitive lattices. For two vectors $\mathbf{u}$ and $\mathbf{v}$, we will write $u = v|_{\text{mod}\,\Gamma^t}$ if there exists such $(E, \mathbf{t}_\mu) \in \Gamma^t$ that $\mathbf{u} = \mathbf{v} + \mathbf{t}_\mu|_{\text{mod }1}$.

We say that $\mathbf{u}$ is a permitted origin choice if the map $\rho'(\mathbf{r}) = \rho(\mathbf{r} + \mathbf{u})$ has the same space group Γ as $\rho(\mathbf{r})$. Similarly, we say that Γ permits the enantiomorph choice if $\rho'(\mathbf{r}) = \rho(-\mathbf{r})$ has the same group of symmetries Γ. The set of permitted origin choices is defined by the following lemma.

Lemma 1. The function $\rho(\mathbf{r} + \mathbf{u})$ has the same space group as $\rho(\mathbf{r})$ if and only if

$$(G_\nu - E)\mathbf{u} = 0|_{\text{mod}\,\Gamma^t} \text{ for every } \nu = 1, \ldots, n. \quad (21)$$

Proof. Let Γ be the space group for $\rho(\mathbf{r})$ and $\mathbf{u}$ be such that $\mathbf{u} = G_\nu\mathbf{u} - \mathbf{t}_\mu$, where $(E, \mathbf{t}_\mu) \in \Gamma^t$. Then,

$$\begin{aligned}\rho'(G_\nu\mathbf{r} + \mathbf{t}_\nu) &= \rho(G_\nu\mathbf{r} + \mathbf{t}_\nu + \mathbf{u}) \\ &= \rho[G_\nu(\mathbf{r} + \mathbf{u}) + \mathbf{t}_\nu - \mathbf{t}_\mu] \\ &= \rho[G_\nu(\mathbf{r} + \mathbf{u}) + \mathbf{t}_\nu] \\ &= \rho(\mathbf{r} + \mathbf{u}) = \rho'(\mathbf{r}).\end{aligned}$$

So, $\rho'(\mathbf{r})$ has all asymmetries from Γ.

Otherwise, let $\rho'(\mathbf{r})$ have some symmetry from Γ: $\rho'(G_\nu\mathbf{r} + \mathbf{t}_\nu) = \rho'(\mathbf{r})$ for every $\mathbf{r}$, *i.e.* $\rho(G_\nu\mathbf{r} + \mathbf{t}_\nu - \mathbf{u}) = \rho(\mathbf{r} - \mathbf{u})$. Then, denoting $\mathbf{r} - \mathbf{u} = \mathbf{v}$, we have $\rho[G_\nu(\mathbf{v} + \mathbf{u}) + \mathbf{t}_\nu - \mathbf{u}] = \rho(\mathbf{v})$ or $\rho[G_\nu\mathbf{v}_\nu + \mathbf{t}_\nu + (G_\nu - E)\mathbf{u}] = \rho(\mathbf{v})$ for every $\mathbf{v}$. This means that transformation $[G_\nu, \mathbf{t}_\nu + (G_\nu - E)\mathbf{u}] \in \Gamma$ and so its translation vector may differ from $\mathbf{t}_\nu$ for $\mathbf{t}_\mu$ (with some $\mu = 1, \ldots, m$) only. So, $(G_\nu - E)\mathbf{u} = \mathbf{t}_\mu$, where $(E, \mathbf{t}_\mu) \in \Gamma^t$.

Now we are ready to prove (17). For space groups with a primitive lattice [*i.e.* Γ^t containing the identical transformation $(E, 0)$ only], it follows immediately from $(G_\nu - E)\mathbf{u} = 0|_{\text{mod }1}$. If Γ contains transformation $(E, \mathbf{t}_\mu)$, $\mathbf{t}_\mu \neq 0|_{\text{mod }1}$, then $F(\mathbf{s}) \neq 0$ is possible only with $(\mathbf{s}, \mathbf{t}_\mu) = 0$. So $G_\mu\mathbf{u} = \mathbf{u} + \mathbf{t}_\mu$ means (16) again.

Similar considerations show that the necessary and sufficient conditions for function $\rho(-\mathbf{r})$ to have the same space group Γ as $\rho(\mathbf{r})$ are

$$2\mathbf{t}_\nu = 0|_{\text{mod}\,\Gamma^t} \text{ for every } \nu = 1, \ldots, n. \quad (22)$$

For primitive lattices, this means that all the components of translation parts in symmetry equations must be equal to 0 or to 1/2. Inspection of the list of space groups shows that, for groups with nonprimitive lattices and without the inversion center at the origin, the only exceptions from this simple rule are $I4_132$ and $I\bar{4}32$.

References

Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.

Hašek, J. & Schenk, H. (1992). *Acta Cryst.* A**48**, 693–695.

Lunin, V. Yu. (1993). *Acta Cryst.* D**49**, 90–99.

Lunin, V. Yu., Lunina, N. L., Petrova, T. E., Vernoslova, E. A., Urzhumtsev, A. G. & Podjarny, A. (1995). *Acta Cryst.* D**51**, 896–903.

Lunin, V. Yu., Urzhumtsev, A. G. & Skovoroda, T. P. (1990). *Acta Cryst.* A**46**, 540–544.

Lunin, V. Yu. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.

Navaza, J. & Vernoslova, E. (1995). *Acta Cryst.* A**51**, 445–449.

Subbiah, S. (1993). *Acta Cryst.* D**49**, 108–119.

Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.

Weeks, C. M., De Titta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A**50**, 210–220.