PBR: a heavy-atom refinement and phasing procedure to reduce phase bias when heavy-atom derivatives contain common sites

**E. Chabrière,[a] M.-H. Charon[a]
and F. M. D. Vellieux[b]**

[a]Laboratoire de Cristallographie et de Cristallogenèse des Protéines, Institut de Biologie Structurale J.-P. Ebel CEA-CNRS, 41 Avenue des Martyrs, 38027 Grenoble CEDEX 01, France, and [b]Laboratoire de Biophysique Moleculaire, Institut de Biologie Structurale J.-P. Ebel CEA-CNRS, 41 Avenue des Martyrs, 38027 Grenoble CEDEX 01, France

A procedure, called PBR (phase-bias reduction), has been developed to properly refine heavy-atom derivatives and to generate less biased heavy-atom phases when these derivatives contain common heavy-atom sites. Two independent events are obtained by splitting the refinement and phasing calculations into two stages, the first in which one of the derivatives having common sites is used together with the native amplitudes and the second in which both derivatives with common sites are used simultaneously, with one of them being used as the native data set. Improved centroid phases and the corresponding figures of merit are obtained by phase combination. This procedure has been used in the structure determination of the iron-cluster-containing protein pyruvate–ferredoxin oxidoreductase. When the common heavy-atom sites are properly treated by the PBR procedure, the resulting calculated centroid phases are improved with respect to classical heavy-atom refinement centroid phases where all derivatives are refined together. This leads to improved electron-density distributions, since anomalous difference Fourier maps calculated with the PBR-refined centroid phases and corresponding figures of merit show more clearly the positions of the iron sites.

1. Definitions

$\mathbf{F}_P$, native structure factor, of measured amplitude $|\mathbf{F}_P|$; $\mathbf{F}_{PH1}$ and $\mathbf{F}_{PH2}$, structure factors of the two derivatives PH1 and PH2, of measured amplitudes $|\mathbf{F}_{PH1}|$ and $|\mathbf{F}_{PH2}|$, respectively; $\mathbf{F}_{Hx}$, structure factors for the heavy-atom constellations Hx; $\Delta\mathbf{F}_H$, difference vector between $\mathbf{F}_{H1}$ and $\mathbf{F}_{H2}$.

2. Introduction

When the structure of an unknown crystallized macromolecule needs to be determined by heavy-atom methods [multiple isomorphous replacement (MIR) or multiple anomalous dispersion (MAD)], a necessary step in the process is the refinement of the heavy-atom positions, occupancies and temperature factors (see, for example, Blundell & Johnson, 1976) followed by the computation of centroid phases and associated figure-of-merit values (Blow & Crick, 1959). A problem that frequently arises is that the heavy-atom derivatives have common sites, including cases where different chemical species share the same sites. In such cases, if all derivatives are used together for heavy-atom refinement and centroid-phase calculation, the resulting figures of merit are grossly overestimated and the resultant map may not be as interpretable as the mean figure-of-merit value might suggest. This can be easily seen from the phase circles (Fig. 1a) and

associated Argand vector diagram: if the lack of closure is used to measure the distribution of errors, the phase and the phase probability distribution derived from $\mathbf{F}_{H1}$ are very close to those derived from $\mathbf{F}_{H2}$. If phasing is carried out using all

(a)

(b)

Figure 1
Phase circles corresponding to (a) 'classical' heavy-atom refinement and phasing, in which all derivatives are used jointly, and (b) PBR, the phase-bias reduction procedure in which refinement and phasing are split into two, as described in §3. In this diagram, the subscripts n and d indicate whether the corresponding amplitude data set is used as native or derivative amplitudes; four circles have to be drawn: two for $\mathbf{F}_{Pn}$ and $\mathbf{F}_{PH1d}$, and two for $\mathbf{F}_{PH1n}$ and $\mathbf{F}_{PH2d}$. The large dots indicate where the phase circles intersect, hence showing the maxima of the phase probability distributions. These phase circles are for the general case where $\mathbf{F}_{H1}$, $\mathbf{F}_{H2}$ and thus $\Delta\mathbf{F}_H$ are not collinear.

derivatives together, the resulting phase probability distribution is, as is usually the case, bimodal. As the two events $\mathbf{F}_{H1}$ and $\mathbf{F}_{H2}$ are not independent, this leads to the overestimation of the figure of merit corresponding to the centroid phase (Blow & Crick, 1959). The purpose of this paper is to show how this problem can be solved and to give an application in the structure determination of the iron-cluster-containing protein pyruvate–ferredoxin oxidoreductase (PFOR; Chabrière *et al.*, 1999).

3. Materials and methods

When two heavy-atom derivatives, PH1 and PH2, contain common sites, the vector difference[1] $\Delta\mathbf{F}_H$ between the heavy-atom structure factors $\mathbf{F}_{H1}$ and $\mathbf{F}_{H2}$ is small. This difference arises from the differences in occupancies, temperature factors, number of electrons involved and number of heavy-atom sites concerned. The same argument can also be applied to MAD data if we consider $\mathbf{F}_H$ as the structure factor corresponding to the difference in anomalous contribution between data collected at two wavelengths.

We have

$$\mathbf{F}_{PH1} = \mathbf{F}_P + \mathbf{F}_{H1}, \tag{1}$$

$$\mathbf{F}_{PH2} = \mathbf{F}_P + \mathbf{F}_{H2}, \tag{2}$$

$$\Delta\mathbf{F}_H = \mathbf{F}_{H2} - \mathbf{F}_{H1} \text{ (or } \Delta\mathbf{F}_H = \mathbf{F}_{H1} - \mathbf{F}_{H2}), \tag{3}$$

$$\mathbf{F}_{PH2} = \mathbf{F}_{PH1} + \Delta\mathbf{F}_H. \tag{4}$$

Assuming, as is usually done (for example, in the computation of difference Fourier maps), that the phases $\alpha_{PH1} \simeq \alpha_P$, then we obtain two independent derivatives using ($\mathbf{F}_P$, $\mathbf{F}_{H1}$, $\mathbf{F}_{PH1}$) and ($\mathbf{F}_{PH1}$, $\Delta\mathbf{F}_H$, $\mathbf{F}_{PH2}$) separately. In our case, we represent $\mathbf{F}_{H1}$ by the two-dimensional vector (A, B), and $\mathbf{F}_{H2}$ by the vector ($A + \delta A$, $B + \zeta B$), δA and ζB being small compared to A and B. The scalar product of $\mathbf{F}_{H1}$ and $\mathbf{F}_{H2}$ is ($A^2 + A\delta A + B^2 + B\zeta B)/(A^2 + B^2)^{1/2} \cdot [(A + \delta A)^2 + (B + \zeta B)^2]^{1/2}$, thus showing that $\mathbf{F}_{H1}$ is nearly collinear with $\mathbf{F}_{H2}$. When we use $\mathbf{F}_{H1}$ and $\Delta\mathbf{F}_H$, the scalar product between these two is ($A\delta A + B\zeta B)/(A^2 + B^2)^{1/2} \cdot (\delta A^2 + \zeta B^2)^{1/2}$, thus showing that $\mathbf{F}_{H1}$ and $\Delta\mathbf{F}_H$ are not collinear if $\delta A/A \neq \zeta B/B$. Hence (see Fig. 1b), the phase probability distribution functions derived from $\mathbf{F}_{H1}$ and $\Delta\mathbf{F}_H$ are different and independent (this can also be seen from the corresponding Argand vector diagram), and the resulting combined centroid phase probability distribution is more unimodal than in the case where all derivatives are used jointly (Fig. 1a). Since the two centroid phase probability distributions now refer to independent events, they can be combined by adding the corresponding Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970; Hendrickson, 1971).

These facts can be exploited to refine properly heavy-atom parameters by phase refinement and to generate appropriate

---

[1] In this paper, the term 'vector' is used to represent structure factors in the complex plane. This is because we only use the addition properties of complex numbers, which are identical to those of vectors.
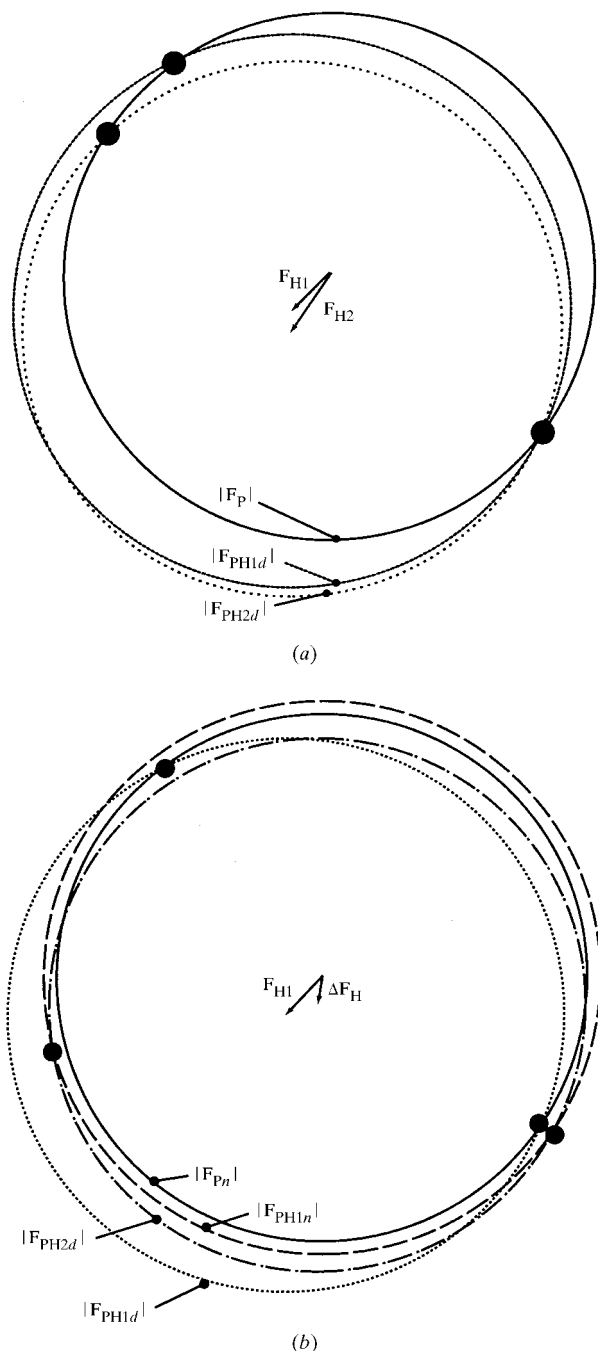
centroid phases and corresponding figure-of-merit values. This procedure has the advantage of not requiring a single line of programming as it makes use of existing computer programs.

First, one of the derivatives with shared sites and with structure-factor amplitude $|\mathbf{F}_{PH1}|$ is used together with the other derivatives with unshared sites and the native data for heavy-atom refinement; a first set of Hendrickson–Lattman probability coefficients is then computed. Then, the second derivative with shared sites of amplitude $|\mathbf{F}_{PH2}|$ is used with the first derivative amplitude $|\mathbf{F}_{PH1}|$ given as the native amplitude, for refinement of either the real or the anomalous occupancy and temperature factor (and possibly refinement of the small differences in positions); a second set of Hendrickson–Lattman coefficients is then computed. The figure-of-merit values and corresponding centroid phases to be used for map calculation are then obtained by phase combination, carried out by addition of the two sets of coefficients.

The procedure described in this paper has been used in the structure determination of the iron-cluster-containing protein PFOR, for which two crystal forms were obtained (Pieulle *et al.*, 1999). Initial heavy-atom phasing was carried out using the multiple isomorphous replacement method for one of the crystal forms and multiple anomalous dispersion for the second crystal form.

For MIR phasing (first crystal form), four heavy-atom derivatives were obtained by crystal soaking, including two K_2PtCl_4 derivatives which have common sites (Chabrière *et al.*, 1999); these are referred to hereafter as Pt1 and Pt2. In the application of the phase-bias reduction (PBR) procedure, the amplitudes from three heavy-atom derivatives (including Pt1 only) and the native amplitudes were used with the *CCP4* maximum-likelihood heavy-atom refinement program *MLPHARE* (Otwinowski, 1991; Collaborative Computational Project, Number 4, 1994) to refine the heavy-atom parameters; the first set of Hendrickson–Lattman coefficients was then generated. Pt1 and Pt2 were then used jointly (with Pt1 used as the native amplitudes) for refinement and computation of the second set of coefficients. These two sets were then added with the program *SIGMAA* (Read, 1986) to generate combined figure-of-merit values and corresponding centroid phases.

For MAD phasing in the second crystal form, four wavelengths were used around the iron edge in order to utilize the anomalous dispersion of the Fe atoms contained in the four [Fe–S] clusters: λ_1 corresponds to an energy lower than the iron edge (f'' minimum, remote high wavelength), λ_2 corresponds exactly to the iron edge (f' minimum, inflection point in absorption), λ_3 corresponds to an energy slightly higher than this edge (f'' maximum, absorption peak) and λ_4 (the remote low wavelength) corresponds to an energy much higher than the edge (f'' and f' high). In the application of the classical procedure, we used the data set collected at wavelength λ_2 as the native data set, and the data sets collected at the three other wavelengths as derivatives, termed the λ_1, λ_3 and λ_4 derivatives. We also used the anomalous contribution of λ_2 data in the λ_2 derivative. Diffusion factors corresponding to wavelengths λ_3 and λ_4 are close to each other, so that if data collected at wavelength λ_2 is used as the reference 'native' data set, the difference in anomalous contribution between λ_3 and λ_2 is very close to that obtained between λ_4 and λ_2. This leads to non-independent λ_3 and λ_4 derivatives. To solve this problem with the PBR procedure in a manner similar to the MIR case, we used separately the λ_1, λ_3 and λ_2 derivatives with the λ_2 data set as native data set and the λ_4 derivative using the data set collected at wavelength λ_3 as the native data set, for two runs of heavy-atom refinement and computation of Hendrickson–Lattman coefficients. Resulting phase sets were then combined as previously with the program *SIGMAA*.

4. Results

For both heavy-atom phasing procedures (MIR and MAD), analysis of the results was carried out in two different ways: first, the two sets of centroid phases obtained by application of the two procedures ('classical' and PBR) were compared using the program *SFTOOLS* (B. Hazes, unpublished work) to the phases given by the crystallographic model obtained at 2.3 Å resolution, which gave an R factor of 19.6% (Chabrière *et al.*, 1999).

The results obtained with PFOR during heavy-atom refinement and phasing at 5.0 Å resolution, where each [4Fe–4S] cluster behaves as a single scattering centre, can be described as follows (Table 1).

For the MIR data, when all four derivatives are used simultaneously for heavy-atom refinement and centroid phase calculation, the mean figure-of-merit value (Blow & Crick, 1959) is 0.48, whereas the correct cosine of the phase difference to the true phase is 0.282 (Table 1). The resulting electron-density distribution, computed with coefficients of the form $m|\mathbf{F}_P|\exp(i\alpha_{centroid})$, is of poor quality and is not compatible with such a high mean figure-of-merit value and corresponding phase error (map not shown). This clearly indicates improper treatment of the derivatives having common sites during heavy-atom refinement and phasing and leads to overestimated figure-of-merit values.

When the phasing procedure was repeated using the PBR procedure, by splitting the calculations in two because of the presence of common sites for Pt1 and Pt2, the mean figure-of-merit value was 0.39 for the three derivatives (including Pt1 only) used together with the native amplitudes, and 0.28 for Pt1 and Pt2 used together (with Pt1 as the native set of amplitudes). The resulting mean combined figure-of-merit value was 0.44, which is closer to its correct value of 0.295. In this case (Table 1), the phase error has decreased slightly from 73.6 to 72.8°, so that there is minimal phase improvement. The resulting electron-density map was of improved quality, and was used for the structure determination by iterative solvent flattening and cross-crystal-forms averaging (Chabrière *et al.*, 1999).

The improvement obtained by the use of the PBR procedure is also indicated by computing the statistical correlation coefficient between the phases thus obtained and those of the crystallographic model (Table 1*b*): the global correlation was 0.196 with the PBR procedure *versus* a figure of 0.183 obtained

research papers

Table 1
Comparison of the heavy-atom phasing results obtained with PFOR at 5.0 Å resolution for MIR and MAD data using 'classical' heavy-atom refinement and phasing and the PBR procedure which reduces phase bias when heavy-atom derivatives have common sites.

The 'classical' procedure is *MLPHARE* heavy-atom refinement and phasing in which all derivatives are used jointly with the native amplitudes; the PBR procedure is the phase-bias reduction procedure described here.

(*a*) Figure-of-merit values resulting from the two procedures used.

Method	Procedure	$\langle$fom1$\rangle$†	$\langle$fom2$\rangle$‡	$\langle$fom$\rangle$§	cos α_{DIF}¶ at 5 Å (at 3.3 Å)
MIR	Classical			0.480	0.282 (0.137)
	PBR	0.390	0.280	0.440	0.295 (0.143)
MAD	Classical			0.581	0.483
	PBR	0.485	0.445	0.578	0.470

(*b*) Linear correlation coefficient (in various resolution ranges) between heavy-atom phases and the phases obtained from the refined model of PFOR.

Resolution (Å)	MIR Classical	MIR PBR	MAD Classical	MAD PBR
20.00–10.31	0.33	0.28	0.44	0.50
10.31–8.38	0.27	0.26	0.44	0.44
8.38–7.38	0.20	0.26	0.35	0.38
7.38–6.73	0.17	0.21	0.31	0.33
6.73–6.27	0.17	0.20	0.30	0.32
6.27–5.91	0.18	0.15	0.28	0.28
5.91–5.62	0.18	0.22	0.25	0.25
5.62–5.38	0.10	0.10	0.21	0.25
5.38–5.18	0.16	0.15	0.26	0.23
5.18–5.00	0.09	0.12	0.24	0.26
Overall	0.183	0.196	0.301	0.312

† Mean figure-of-merit value of the first set of derivatives. ‡ Mean figure-of-merit value computed from the second set of derivatives. § Mean figure-of-merit value, either global in case of classical refinement and centroid-phase calculation, or computed by phase combination in the PBR procedure. ¶ Mean cosine of the phase difference to the true phases, taken for these calculations as the phases generated from the final refined model at 2.3 Å resolution (Chabrière *et al.*, 1999).

Table 2
Comparison of the heights of the peaks corresponding to the iron clusters observed in the anomalous difference maps.

These values are expressed in numbers of standard deviations above the mean value of the map.

Procedure	Cluster max†	Cluster min‡	Protein§
Classical	9.4	6.8	4.2
PBR	10.8	7.7	4.1

† Peak height corresponding to the iron cluster with the highest density, obtained from an $m(|\mathbf{F}^+| - |\mathbf{F}^-|)\exp(i\alpha_{centroid})$ anomalous difference Fourier map. ‡ As cluster max, except that it corresponds to the iron cluster with the lowest electron-density value. § Peak height corresponding to the first non-Fe atom of the protein.

The relative peak heights corresponding to the iron clusters (Table 2), obtained with PBR-combined centroid phases and figures of merit, are higher than those obtained using centroid phases and figure-of-merit values which result from the 'classical' run of *MLPHARE* in which all four derivatives are used together (with peak heights of 10.8 and 9.4, respectively, *i.e.* a 15% increase in peak height; this can be compared to the corresponding increase in correlation). The contrast between the protein density and that of the iron clusters also increases for all iron positions (Table 2). This therefore indicates that the PBR procedure proposed here is effective in reducing the phase bias arising from the presence of heavy-atom derivatives having common sites, thereby providing clearer maps.

The PBR procedure may also have a more general applicability, *i.e.* to obtain several heavy-atom derivatives from a single heavy-atom reagent by soaking crystals for different periods of time or by using several concentrations of reagent, thus obtaining common sites with different effective occupancies.

References

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. London: Academic Press.
Chabrière, E., Charon, M. H., Volbeda, A., Pieulle, L., Hatchikian, E. C. & Fontecilla-Camps, J. C. (1999). *Nature Struct. Biol.* In the press.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
Hendrickson, W. A. (1971). *Acta Cryst.* **B27**, 1472–1475.
Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* **B26**, 136–143.
Otwinowski, Z. (1991). *Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Pieulle, L., Chabrière, E., Hatchikian, C., Fontecilla-Camps, J. C. & Charon, M.-H. (1999). *Acta Cryst.* **D55**, 329–331.
Read, R. J. (1986). *Acta Cryst.* **A42**, 140–149.

when the 'classical' refinement procedure was used, *i.e.* a 7.1% increase in correlation.

In our test of the PBR procedure using MAD data, the linear correlation coefficient between phases is also seen to improve, with a 3.7% increase in correlation (Table 1*b*). The unweighted phase difference to the true phases (Table 1*a*) might not be an optimal measure of phase improvement owing to anomalous contributions which give systematic shifts in the phases.

Next, for further analysis of the effectiveness of the PBR procedure in the case of MIR data, we also computed anomalous difference maps with coefficients of the form $m(|\mathbf{F}^+| - |\mathbf{F}^-|)\exp(i\alpha_{centroid})$ in order to establish which of these two figure-of-merit and associated centroid phase sets gave the highest peaks corresponding to the iron clusters: both sets were used together with anomalous amplitude data from the protein to compute anomalous difference Fourier maps. Note that the anomalous signal arising from the Fe atoms was not used during heavy-atom refinement and phasing.