Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Single-wavelength phasing strategy for quasi-racemic protein crystal diffraction data

**Michael R. Sawaya,[a,b] Brad L. Pentelute,[c] Stephen B. H. Kent[c] and Todd O. Yeates[a,d]***

[a]UCLA–DOE Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA, [b]Howard Hughes Medical Institute, USA, [c]Department of Chemistry, University of Chicago, Chicago, IL 60637, USA, and [d]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

Correspondence e-mail: yeates@mbi.ucla.edu

Received 27 August 2011
Accepted 22 November 2011

Racemic protein crystallography offers two key features: an increased probability of crystallization and the potential advantage of phasing centric diffraction data. In this study, a phasing strategy is developed for the scenario in which a crystal is grown from a mixture in which anomalous scattering atoms have been incorporated into only one enantiomeric form of the protein molecule in an otherwise racemic mixture. The structure of a protein crystallized in such a quasi-racemic form has been determined in previous work [Pentelute *et al.* (2008), *J. Am. Chem. Soc.* **130**, 9695–9701] using the multi-wavelength anomalous dispersion (MAD) method. Here, it is shown that although the phases from such a crystal are not strictly centric, their approximate centricity provides a powerful way to break the phase ambiguity that ordinarily arises when using the single-wavelength anomalous dispersion (SAD) method. It is shown that good phases and electron-density maps can be obtained from a quasi-racemic protein crystal based on single-wavelength data. A prerequisite problem of how to establish the origin of the anomalous scattering substructure relative to the center of pseudo-inversion is also addressed.

## 1. Introduction

Racemic protein crystallography is the practice of crystallizing and determining the structure of a protein from a racemic mixture (Berg & Goffeney, 1997; Matthews, 2009; Yeates & Kent, 2012). This requires chemically synthesizing the target protein in an enantiomeric form mirroring the natural biological hand (*i.e.* from D-amino acids); the natural enantiomer can be either synthesized or purified from a biological source. Racemic protein crystallography is motivated by two advantages. Firstly, theoretical work has predicted that proteins will crystallize with unusual ease in racemic form (Wukovitz & Yeates, 1995). This ease of crystallization, as well as a specific prediction that $P\bar{1}$ would be the dominant space group, has been borne out by several recent experimental studies (Pentelute *et al.*, 2008; Mandal, Pentelute, Tereshko, Kossiakoff *et al.*, 2009; Mandal, Pentelute, Tereshko, Thammavongsa *et al.*, 2009; Banigan *et al.*, 2010). $P\bar{1}$ is one of the 165 achiral space groups (*i.e.* those containing centers of inversion, mirror or glide planes or rotary inversions) that become available to protein molecules only in racemic mixtures. Secondly, most of the achiral space groups (92, including $P\bar{1}$) are centrosymmetric, which offers the prospect of having to phase only centric diffraction data. This was a primary motivation when Zawadzke and Berg crystallized the first protein (rubredoxin) in racemic form (Zawadzke & Berg, 1993). Since then, a number of racemic crystal structures have been determined by

direct phasing methods (Patterson *et al.*, 1999; Mandal, Pentelute, Tereshko, Kossiakoff *et al.*, 2009; Mandal, Pentelute, Tereshko, Thammavongsa *et al.*, 2009; Banigan *et al.*, 2010; Pentelute *et al.*, 2010). The centrosymmetric nature of the crystals in these studies presumably contributed to their successful phasing, and a recent racemic case has pushed the size and resolution limits for successful direct-methods applications (Banigan *et al.*, 2010).

Experimental phasing strategies (*i.e.* those involving heavy-atom or anomalous scattering information) have also been employed to solve the structures of racemic macromolecular crystals. Unique opportunities present themselves here. Pentelute *et al.* (2008) developed a line of attack in which anomalous scattering Se atoms were incorporated into only the natural hand of the molecule (*via* a modified L-amino acid) and not the mirror enantiomer. This general strategy potentially leads to quasi-racemic crystals in which the protein molecules pack in a centrosymmetric arrangement but the anomalous scattering substructure is not centrosymmetric. This latter feature is critical because it avoids the loss of anomalous signal that would accompany a centrosymmetric anomalous scattering substructure; Friedel mates have equal intensities for a centrosymmetric crystal. On the other hand, the quasi-racemic strategy sacrifices the centric phasing advantage of a truly centrosymmetric crystal. In the work by Pentelute *et al.* (2008), a single Se atom was incorporated into the natural enantiomer of the 81-amino-acid snow flea antifreeze protein, the quasi-racemic mixture was crystallized in space group $P1$ (pseudo-$P\bar{1}$) and phases (which were not centric) were obtained using the multi-wavelength anomalous dispersion (MAD) strategy (Hendrickson, 1990).

In the present work, we take a further step along the line taken by Pentelute *et al.* (2008). We show that as long as the phases of a quasi-racemic crystal remain approximately centric this information can be exploited to obtain accurate experimental phases from single-wavelength data, a feat which is only possible using the quasi-racemate strategy.

## 2. Methods

### 2.1. Data processing and SAD phasing

Unmerged $I^+$ and $I^-$ reflection-intensity measurements from quasi-racemic crystals of the snow flea antifreeze protein were obtained from previous work (Pentelute *et al.*, 2008). The data collected at the anomalous peak wavelength for selenium extended to 1.2 Å resolution, with an overall $R_{\rm merge}$ value of 6.8% and a redundancy of 8.0 (Supplementary Table S1[1]). SAD phases and Hendrickson–Lattman coefficients were calculated with the programs *MLPHARE* (Winn *et al.*, 2011) and *Phaser* (McCoy *et al.*, 2007). The coordinates of the two Se atoms which constitute the anomalous substructure were taken from the structure deposited in the Protein Data Bank (PDB entry 3bog).

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: GX5197). Services for accessing this material are described at the back of the journal.

### 2.2. Phase probability calculations

In this study, approaches to phasing (evaluated in §3) combine information from (i) SAD data, *i.e.* using anomalous intensity differences between Friedel pairs at a single X-ray wavelength, and (ii) the expectation that the native protein phases should be nearly centric. The Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970) obtained from SAD phasing were used to identify the two ambiguous phase choices exactly satisfying the SAD measurements for each reflection, to evaluate the probabilities of alternate centric phase choices, and as a basis for combining SAD phase information with the expectation of nearly centric phases.

In an analysis where the centric expectation was incorporated as a probabilistic term, this was accomplished by adding to the third Hendrickson–Lattman coefficient ($C$). This corresponds to multiplying the phase probability distribution by the term $\exp[C\cos(2\varphi)]$. A question arises regarding what value to assign the coefficient $C$ in order to represent a probability distribution with a desired average deviation from centricity. In the absence of analytical expressions for integrals of the required terms {namely $\varphi\exp[C\cos(2\varphi)]\,{\rm d}\varphi$ and $\exp[C\cos(2\varphi)]\,{\rm d}\varphi$}, the problem was treated by numerical integration. In the quasi-racemic protein crystal test case studied here the refined model phases deviated from centricity by an average of 24°. Based on numerical integration, a 24° average deviation corresponds to a value of approximately 1.4 for the Hendrickson–Lattman coefficient $C$.

### 2.3. Map calculation and density modification

Electron-density maps were calculated using the *CCP*4 program *FFT* (Winn *et al.*, 2011). Density modification (solvent flattening, histogram matching and multi-resolution modification) was performed using the program *DM* (Cowtan, 1994). Correlation coefficients between maps were calculated using *get_cc_mtz_mtz* from the *PHENIX* suite of programs (Adams *et al.*, 2010).

## 3. Results

By itself, the SAD phasing approach leads to an ambiguous choice between two possible correct phases for each reflection. In practice, these phase ambiguities are resolved by either additional diffraction data sets at different X-ray wavelengths (MAD; Hendrickson, 1990), by iterative density-modification procedures (Zhang *et al.*, 2001; Wang, 1985; Bricogne, 1976; Rossmann *et al.*, 1992) or, in special cases where the anomalous scattering is very large, by statistical weighting (Hendrickson & Teeter, 1981). The new element introduced in the present study is to exploit the approximate centrosymmetry of a quasi-racemic crystal to break the phase ambiguity in SAD data (Fig. 1). In practice, these two sources of phase information will not agree precisely owing to errors in the SAD data and deviations of the underlying atomic structure from perfect centrosymmetry. Multiple approaches present themselves for reconciling the two sources of phase information. To discern the best phasing approach, we made use of the 1.2 Å

resolution experimental data collected at the anomalous peak wavelength for selenium from a quasi-racemic crystal of snow flea antifreeze protein (Pentelute *et al.*, 2008). The *P*1 (pseudo-$P\bar{1}$) unit cell contains two L- and two D-protein molecules and two Se atoms (in the L-protein) for use in SAD phasing. The SAD 'best' (statistically weighted) phases served as a baseline for evaluating potential improvements that might be obtained by incorporating the centric expectation. As surrogates for the true phases, we adopted the model phases calculated from the refined structure, since the structure is well resolved at atomic resolution. The average error (*i.e.* the deviation from model phases) for the SAD best phases was 52° over all reflections whose figure of merit (FOM) was greater than 0.1 and 42° for reflections with FOM greater than 0.5.

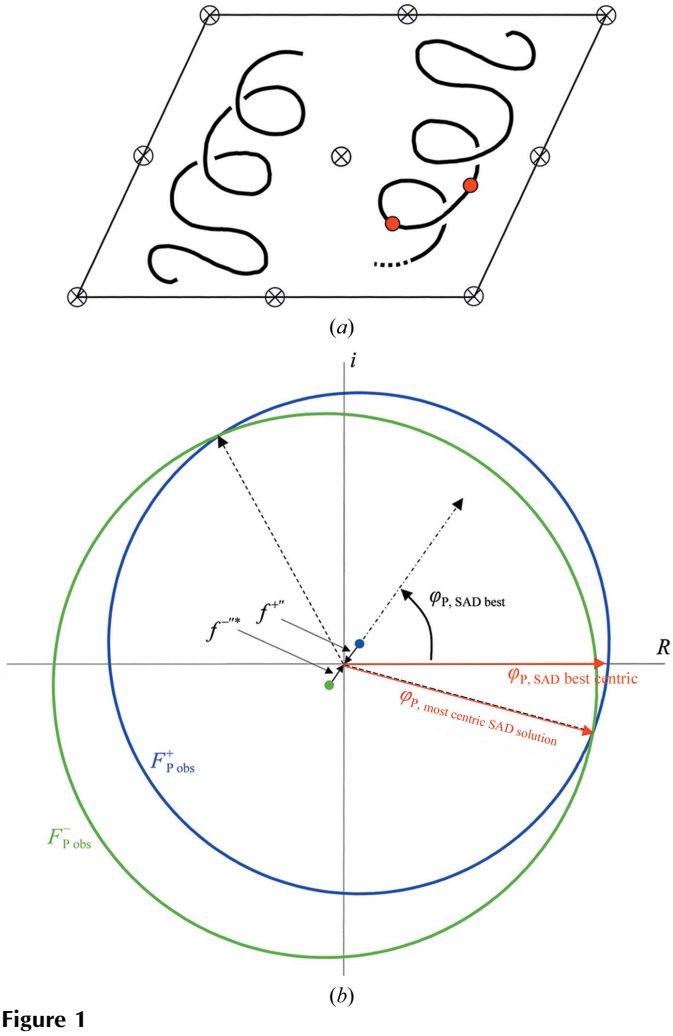


**Figure 1**
Diagram of a quasi-racemic macromolecular crystal and a Harker diagram for phasing. (*a*) A cartoon of a crystal containing both enantiomeric hands of a protein molecule, with anomalous scattering atoms (red) incorporated into only one hand. The dotted region is intended to convey the possibility that conformational differences might exist between the two molecules, which will, in addition to the differences contributed by anomalous scattering atoms, break the center of symmetry in the crystal. The arrangement shown is in space group *P*1 (pseudo-$P\bar{1}$). (*b*) A phase diagram illustrating information from single-wavelength anomalous dispersion (SAD) data. Protein structure-factor phase circles for the plus (blue) and minus (green) reflections of a Friedel pair are offset by vectors related to the anomalous scattering contributions ($f''$) that they contain. Black dashed lines indicate the two possible ambiguous phase choices based on SAD data, assuming perfect data. The black dashed–dotted line indicates the SAD 'best' phase obtained from a statistical weighting of the possible phases, which is used in a typical SAD analysis. The red vectors indicate choices for the protein phase that take into account the expectation that the protein structure factor for a quasi-racemic crystal should be approximately centric (0° or 180°). The two vectors shown describe distinct strategic choices discussed in the text. The phase diagram shown is simplified by omitting the dispersive ($f'$) contribution to the total structure factor, which amounts to considering this contribution to be part of the protein structure factor $F_P$.

### 3.1. Approaches for incorporating the centric expectation

We tested different approaches for exploiting the expectation that the correct protein phases should be approximately centric. In the first approach, we took the near-centricity of the protein phases to be a strong constraint, with ambiguity between the two choices (0° and 180°) being broken by the SAD information. That is, we chose either 0° or 180° as the correct phase based on which choice had the higher

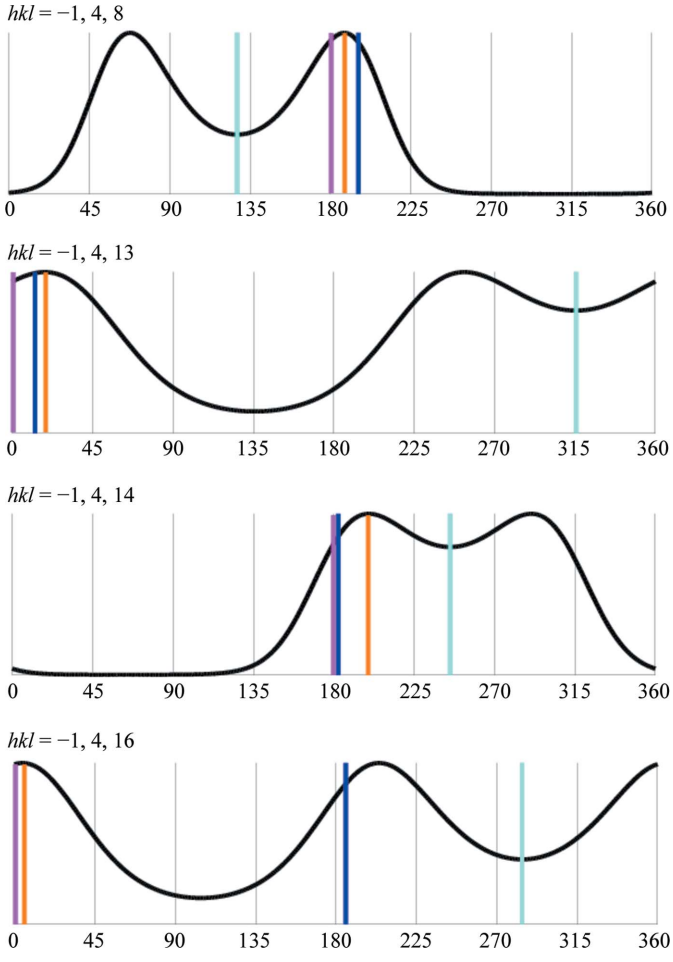


**Figure 2**
SAD phase probability distributions for selected reflections. Here and in subsequent figures the diffraction data are from the quasi-racemic crystal of the snow flea antifreeze protein (Pentelute *et al.*, 2008). The 'SAD best centric' phase choice is more often closer to the 'true' (model) phase than is the conventional 'SAD best' phase. The conventional 'SAD best' phase is indicated in cyan and the 'SAD best centric' phase is indicated in purple. The 'most centric SAD solution' phase is indicated in orange. The model or 'true' phase is indicated in dark blue.

probability according to the SAD data. We refer to phases chosen using this criterion as 'SAD best centric' phases. The first Hendrickson–Lattman coefficient $A$ governs the 'SAD best centric' phase; the more probable centric phase choice is 0° if $A > 0$ and 180° if $A < 0$. Fig. 2 illustrates the comparison of phases for several reflections and Fig. 3 shows the overall agreement between these phases and the model phases. The average deviation from the model phases is 42° for all reflections with FOM > 0.1 and 33° for all reflections with FOM > 0.5. Encouragingly, these errors are approximately 10° less than the deviation between the SAD best phases and the model phases. Furthermore, there is evidence that this 10° improvement is likely to underestimate the real phase improvement provided by the centric approximation. A considerable share of the deviation between the SAD best centric phases and the model phases correlates with the deviation of the model phases from centric values. This correlation leads us to believe that the true phases deviate less from centrosymmetry than the model phases indicate. Some of this deviation probably reflects random divergence of the unconstrained model phases during protein structure refinement. A comparison of the SAD best centric phases to model phases for reflections where the latter phase is nearly centric shows an especially dramatic effect on phase accuracy when the centricity information is included in the phasing (Fig. 3*b*).

The second approach examined was to identify the two ambiguous phase choices that satisfy the SAD data and then choose the one that is closest to being centric. This amounts to taking the SAD data as a strong constraint and using the centric approximation as a secondary criterion to break the ambiguity. We refer to this choice as the 'most centric SAD' phase. The average deviation from model phases was 48° for reflections with FOM > 0.1 and 39° for reflections with FOM > 0.5. This choice of phase is therefore better than the ordinary SAD best phase but considerably poorer than the SAD best centric phase examined in the first approach (Fig. 3). This result presumably reflects the relative weakness of the SAD phasing in the present test case; the overall figure of merit based on SAD phasing alone was 0.26.

The two approaches described above represent alternative limiting assumptions regarding which source of phase information should dominate. A choice that balances those two extreme options could provide an improvement. We examined the quality of the phases obtained by taking a simple average of the SAD best centric phase and the most centric SAD phase. This did not result in a substantial improvement over the SAD best centric phase; the overall phase error compared with model phases was within 1° of the values obtained by simply choosing the SAD best centric phase. A more statistically sound approach for unifying the available phase information would be to combine the phase probability distributions from the SAD data and from the centric approximation. However, including the latter contribution requires a prior estimate of how much the correct phases are expected to deviate overall from centricity. This would be difficult to establish in practice prior to knowledge of the structure, but in the present test case an estimate can be obtained from the model phases. As noted above, the model phases deviate from centric values by 24° on average. The centric expectation, taking into account its expected deviation, was used to introduce an additional contribution into the SAD Hendrickson–Lattman coefficients (see §2). The best statistically weighted phase and FOM were then obtained by numerical integration. The agreement between these 'combined' phases and the final model phases was much better than the ordinary SAD best phases. However, the improvement was within about 1° of that provided by the simpler SAD best centric phase; the combined probability phases deviated on average by 42° from the model phases for reflections with FOM > 0.1.

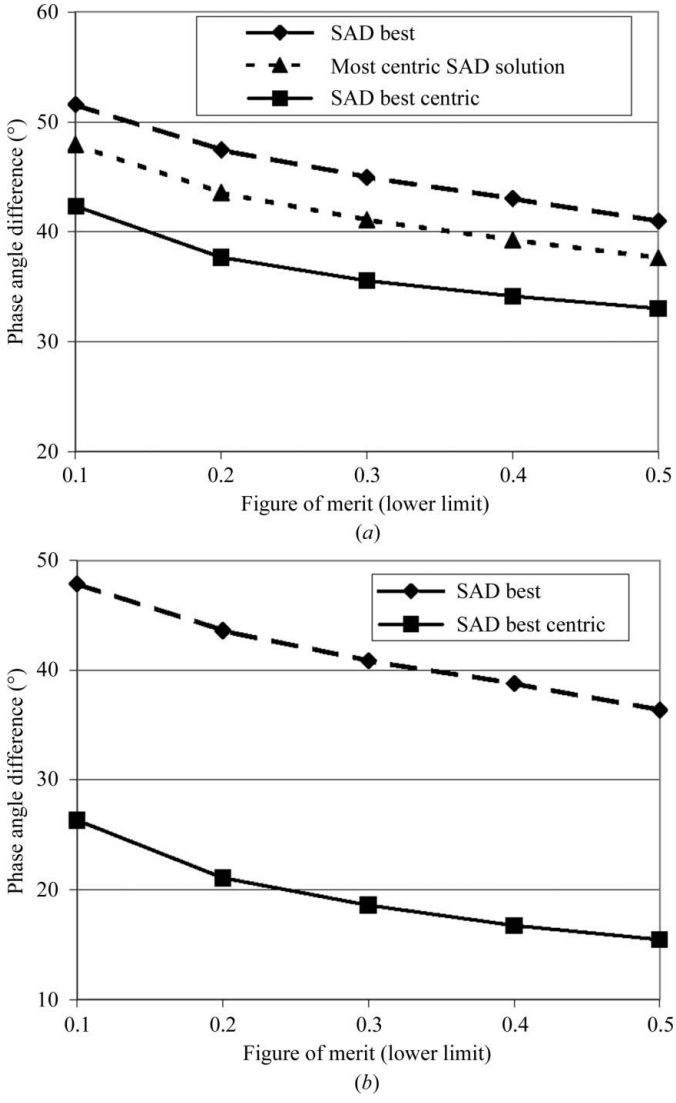


**Figure 3**
Average deviation of experimentally determined phases from the final model phases. (*a*) The 'SAD best centric' phase is the centric phase (0° or 180°) with the greater probability based on the SAD (single-wavelength anomalous dispersion) data. The 'most centric SAD solution' is the phase (of the two ambiguous choices from SAD) that is closest to being centric (0° or 180°). The 'SAD best' phase (the probability-weighted average phase) represents the standard choice for an ordinary (*i.e.* not quasi-racemic) SAD experiment. (*b*) A similar comparison restricted to reflections for which the model phase is within 20° of centric.

We sought to evaluate the significance of the improvement provided by the SAD best centric phases by comparing the electron-density maps obtained with these phases *versus* conventional SAD best phases. For the map calculated with the SAD best centric phases we also recalculated a new FOM according to the probabilistically weighted sum of the two centric phase choices. The resulting map showed readily interpretable features and relatively good connectivity, particularly in view of its derivation from single-wavelength data (Fig. 4). It is considerably more interpretable than the conventional SAD map. We considered whether the relative interpretability of the new map could be evaluated quantitatively by employing automated model-building programs. This is not possible at the present time; current programs do not have the required ability to identify and distinguish biological *versus* nonbiologically handed electron-density features that exist generally within any arbitrarily chosen asymmetric unit of a racemic or quasi-racemic crystal. Instead, we evaluated the correlation coefficient between the final model electron-density map and maps calculated with different phase sets. The map calculated with SAD best centric phases had a correlation coefficient of 0.62 with the model map. For comparison, traditional SAD phases gave a map whose correlation coefficient with the model map was 0.48 when the map was calculated using SAD phases from the program *MLPHARE* and 0.51 when the SAD phases were obtained using the program *Phaser* (Table 1).

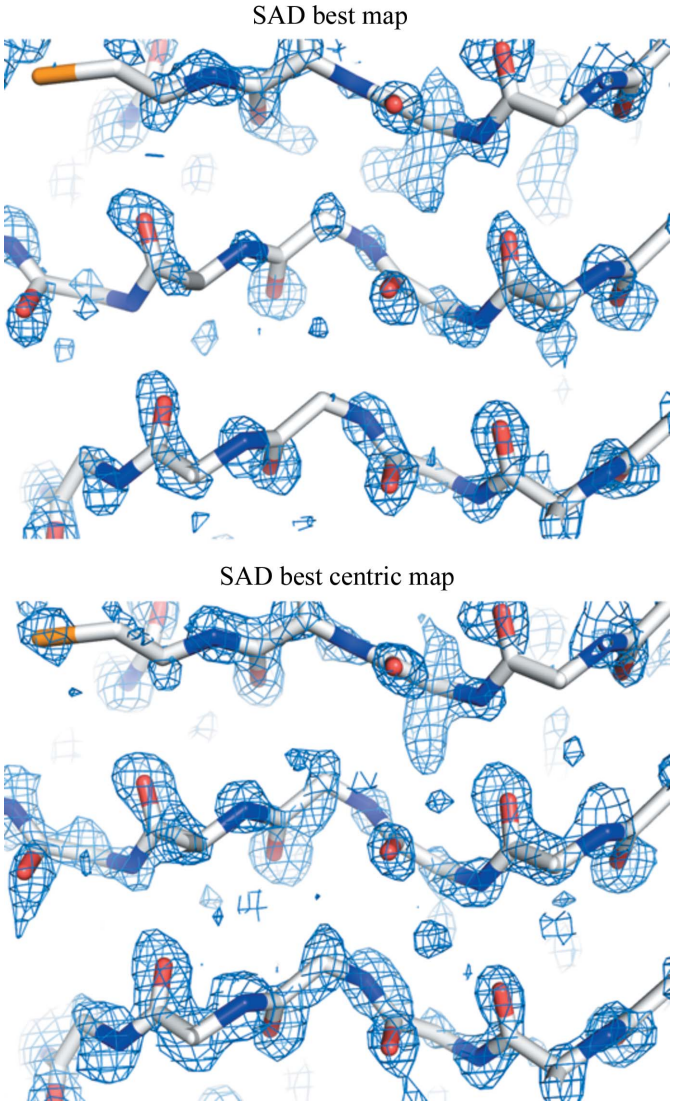


**Figure 4**
An electron-density map calculated using 'SAD best centric' phases. The electron-density map displays higher connectivity and is more easily interpreted than the map calculated using conventional 'SAD best' phases. The maps were calculated using data to 1.2 Å resolution and contoured at 1.4$\sigma$. The sticks represent the refined coordinates of snow flea antifreeze protein (PDB entry 3bog). The maps are displayed using *PyMOL* (DeLano, 2002).

**Table 1**
Agreement between electron-density maps calculated with different phasing strategies.

The reported correlation is a comparison to the final refined snow flea antifreeze protein ($F_{calc}$) electron-density map. DM refers to density modification. SAD refers to phasing by traditional single-wavelength anomalous dispersion using either of two programs, *MLPHARE* or *Phaser*. SWAQR refers to the strategy described in the present study.

| Phasing strategy | Map correlation coefficient |
|---|---|
| SWAQR (SAD best centric) | 0.623 |
| SWAQR (SAD best centric) followed by DM | 0.841 |
| SAD: *MLPHARE* | 0.483 |
| SAD: *Phaser* | 0.510 |
| SAD: *Phaser* followed by DM | 0.564 |

As a further test, we performed density modification on electron-density maps derived from traditional SAD phases or from SAD best centric phases (Table 1 and Supplementary Fig. S1). Density modification of a map based on SAD phases gave a final correlation coefficient of 0.56 compared with the model map. The corresponding value was 0.86 after density modification of a map based on the SAD best centric phases. The close agreement of this map with the model further emphasizes the utility of the single-wavelength quasi-racemic phasing approach.

### 3.2. The relative origin problem

Regardless of the approach employed to combine SAD information with the centric expectation, a problem regarding relative origins must be addressed. The analysis above revolves around the idea of phases being centric (0° or 180°), but this assumes that the crystal coordinate system is centered on a center of inversion symmetry. The problem is that the quasi-racemic anomalous scattering substructure is not centrosymmetric, since it exists in just one hand of the protein molecule. No inversion center is implied by the anomalous substructure and therefore its origin is indeterminate. In a quasi-racemic space group other than $P1$/pseudo-$P\bar{1}$, for example $P2_1$/pseudo-$P2_1/c$, the origin might be indeterminate along only one direction, but the general problem remains. In order to expect the protein phases to be (nearly) centric, the anomalous substructure must be chosen (or shifted) in a way that places the pseudo inversion center of the quasi-racemic crystal at the origin of the coordinate system.

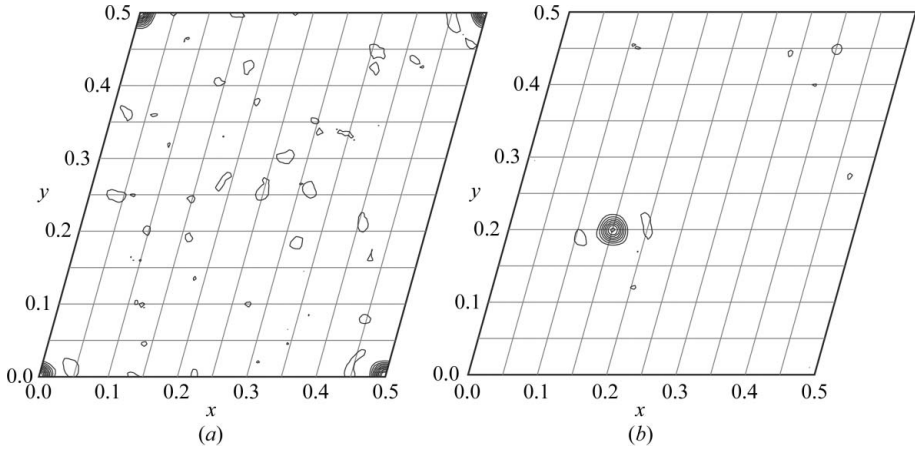


**Figure 5**
Determination of the heavy-atom substructure shift required to bring the pseudo center of inversion of the crystal to the origin. The correct shift is based on the expectation that the protein phases should be approximately centric for a quasi-racemic crystal. The two-dimensional contour plot shows the average deviation of the more nearly centric phase (among the two ambiguous choices for each reflection based on SAD data) from 0° or 180°. Each point describes a candidate coordinate shift of the heavy-atom substructure origin in the *xy* plane. (*a*) For the snow flea antifreeze protein (Pentelute *et al.*, 2008), the structure was reported with the pseudo center of inversion at the origin. The correct shift for the anomalous substructure from its reported position is therefore (0, 0, 0). The existence of equally valid origins at shifts of 1/2 is illuminated by the plot, as expected. (*b*) A similar search for the correct origin after first shifting the heavy-atom substructure by (0.15, 0.20, 0); the required origin shift is evident.

We show how the origin of the pseudo inversion symmetry can be found (before the protein structure is known) by shifting the anomalous substructure systematically and judging whether the resulting (phase-shifted) SAD phases are nearly centric. Diffraction data from the snow flea antifreeze protein were used to test such a calculation. In space group $P\bar{1}$ the correct shift requires a three-dimensional search (from 0 to 1/2 in all three directions, owing to equivalent origins). A two-dimensional section of such a search is shown in Fig. 5(*a*). At each choice of the anomalous substructure shift, we calculated which of the two possible SAD phases for each reflection was nearest to centric and evaluated the average angular deviation of this phase from centricity over all reflections with FOM > 0.2. The correct origin shift is clear (Fig. 5*a*). In the present test case the correct shift is (0, 0, 0); this is expected since the final atomic coordinates were reported in a reference frame that retrospectively placed the molecules according to the pseudo-$P\bar{1}$ symmetry. As a further test of the calculation, we artificially shifted the reported anomalous substructure by fractional coordinates (0.15, 0.2, 0) and repeated the calculation. The correct position for the centric origin was evident at the shifted position (Fig. 5*b*).

## 4. Discussion

We have explained here how the expectation of approximate centricity can be used to unambiguously phase diffraction data from quasi-racemic protein crystals using anomalous scattering data at a single wavelength (SAD). This is a notable distinction compared with other anomalous scattering phasing strategies. With ordinary (chiral) protein crystals, multi-wavelength anomalous data sets are required to resolve phase ambiguities experimentally (assuming isomorphous replacement approaches are not being used). In the absence of multi-wavelength data, the phase ambiguity from SAD must be broken by complicated methods of inference, namely density-modification algorithms such as solvent flattening and/or NCS averaging. The advantages of being able to make a potentially unambiguous experimental determination of phase from single-wavelength data seem clear. We propose the acronym SWAQR to refer to the single-wavelength anomalous quasi-racemate approach to phasing.

We have investigated alternate approaches to combining the SAD information and the centric approximation. In theory, a fully statistical combination of these two sources of information is best. However, missing information challenges this approach. In the test case examined here, excellent results, which were very nearly as good as the results from a fully statistical approach, were obtained with a simple strategy. Strictly centric phases were chosen, with the decision between 0° and 180° being based on which was more likely according to the SAD phase probability distribution. As a conservative estimate, the phases obtained in this way are at least 10° better than ordinary SAD phases. The phases produced a high-quality electron-density map. An alternate, likewise simple, approach to phase selection that rests more strongly on the experimental SAD phasing information and less strongly on the centric approximation was also tested. This approach produced phases that were inferior to the strictly centric approach. However, the relative merits of different approaches are likely to weigh differently in different situations, such as where the SAD phase information is highly reliable and the centric approximation is strongly broken.

The optimum solution is to combine the information from the SAD data and from the centric approximation in a statistically sound fashion. We performed such an analysis here, although it was necessary to rely on the known structure to estimate the expected deviation from centricity in advance. In principle, different choices for this expected deviation could be used to calculate and evaluate a series of electron-density maps. A more robust application of the phase-recombination approach would require the development of a statistical method to estimate the average deviation of the true phases from centricity in advance. Finally, we have explained the existence of a technical challenge to using the SWAQR approach, which relates to relative origins, and have shown using the test case available how this problem can be solved.

Notwithstanding the challenges of protein synthesis, the present study suggests a general route to phasing quasi-racemic protein diffraction data. Single-wavelength data are

sufficient to obtain unambiguous phases as long as the anomalous scattering has a strong $f''$ component at the X-ray wavelength employed. It should be generally possible to meet this demand using the protein-synthesis approach; Se, Br and I are all readily incorporated into derivatized amino acids (*e.g.* selenomethionine, selenocysteine, 4-bromophenylalanine and 4-iodophenylalanine). The successful use of these amino acids for protein phasing has been described in the literature (Hendrickson, 1990; Pentelute *et al.*, 2008; Khakshoor *et al.*, 2010; Xie *et al.*, 2004). The use of iodine (*via* 4-iodophenylalanine, for example) would seem to be a particularly useful possibility, as iodine has a strong $f''$ component (6.9 electrons) at the Cu $K\alpha$ wavelength commonly produced by in-house (rotating-anode) X-ray generators. In view of the tendency for racemic proteins to crystallize easily, the strategies proposed here could lead to straightforward structure determination for proteins that can be prepared by chemical synthesis.

We thank Dr Valya Tereshko for providing the snow flea antifreeze protein diffraction data and Professor Tony Kossiakoff for critical reading of the manuscript. This work was supported by the BER program of the DOE Office of Science.

## References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Banigan, J. R., Mandal, K., Sawaya, M. R., Thammavongsa, V., Hendrickx, A. P., Schneewind, O., Yeates, T. O. & Kent, S. B. H. (2010). *Protein Sci.* **19**, 1840–1849.
Berg, J. M. & Goffeney, N. W. (1997). *Methods Enzymol.* **276**, 619–627.
Bricogne, G. (1976). *Acta Cryst.* A**32**, 832–847.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Hendrickson, W. A. (1990). *Science*, **254**, 51–58.
Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* B**26**, 136–143.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Khakshoor, O., Lin, A. J., Korman, T. P., Sawaya, M. R., Tsai, S. C., Eisenberg, D. & Nowick, J. S. (2010). *J. Am. Chem. Soc.* **132**, 11622–11628.
Mandal, K., Pentelute, B. L., Tereshko, V., Kossiakoff, A. A. & Kent, S. B. H. (2009). *J. Am. Chem. Soc.* **131**, 1362–1363.
Mandal, K., Pentelute, B. L., Tereshko, V., Thammavongsa, V., Schneewind, O., Kossiakoff, A. A. & Kent, S. B. H. (2009). *Protein Sci.* **18**, 1146–1154.
Matthews, B. W. (2009). *Protein Sci.* **18**, 1135–1138.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Patterson, W. R., Anderson, D. H., DeGrado, W. F., Cascio, D. & Eisenberg, D. (1999). *Protein Sci.* **8**, 1410–1422.
Pentelute, B. L., Gates, Z. P., Tereshko, V., Dashnau, J. L., Vanderkooi, J. M., Kossiakoff, A. A. & Kent, S. B. H. (2008). *J. Am. Chem. Soc.* **130**, 9695–9701.
Pentelute, B. L., Mandal, K., Gates, Z. P., Sawaya, M. R., Yeates, T. O. & Kent, S. B. H. (2010). *Chem. Commun.* **46**, 8174–8177.
Rossmann, M. G., McKenna, R., Tong, L., Xia, D., Dai, J., Wu, H., Choi, H. K. & Lynch, R. E. (1992). *J. Appl. Cryst.* **25**, 166–180.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–113.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wukovitz, S. W. & Yeates, T. O. (1995). *Nature Struct. Biol.* **2**, 1062–1067.
Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G. & Schultz, P. G. (2004). *Nature Biotechnol.* **22**, 1297–1301.
Yeates, T. O. & Kent, S. B. H. (2012). *Annu. Rev. Biophys.*, doi:10.1146/annurev-biophys-050511-102333.
Zawadzke, L. E. & Berg, J. M. (1993). *Proteins*, **16**, 301–305.
Zhang, K. Y. J., Cowtan, K. D. & Main, P. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 311–324. Dordrecht: Kluwer Academic Publishers.