Phasing of Quasi-Racemic Protein Diffraction

David J. Schuller
3 February 2012
230 possible space groups for 3D crystals

65 of these are available to chiral molecules.

These may include crystallographic axes with 2,3,4,6-fold symmetry* about simple rotation axes or screw axes.

The rest include inversion symmetry: center of inversion, mirror, glide plane. These are available only to non-chiral molecules – and racemic mixtures.

* Other symmetries (5,7,11) can be a part of the asymmetric unit, but do not constitute part of the crystalline space group symmetry.
Why protein crystals favour some space-groups over others.

“the maximum number of degrees of freedom, D=8, is achieved only in space-group P \( \bar{1} \). Therefore, we predict that P \( \bar{1} \) will be the most frequently observed space-group for racemic protein mixtures”

Small molecules: space filling

Macromolecules: Rigid body degrees of freedom
Table 1 The number of rigid-body degrees of freedom, \( D \), and the observed occurrence of monomeric protein crystals among the 65 crystallographic space groups

<table>
<thead>
<tr>
<th>Symmetry group</th>
<th>S</th>
<th>L</th>
<th>C</th>
<th>( D^1 )</th>
<th>Freq</th>
<th>%²</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P2_{1212} )</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>88</td>
<td>36.1</td>
</tr>
<tr>
<td>( P2_1 )</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>27</td>
<td>11.1</td>
</tr>
<tr>
<td>( C2 )</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>15</td>
<td>6.1</td>
</tr>
<tr>
<td>( P4_{321} )</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>14</td>
<td>5.7</td>
</tr>
<tr>
<td>( P3_{121} )</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>4.9</td>
</tr>
<tr>
<td>( C22_{12}, P2_{1212} )</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>3.7</td>
</tr>
<tr>
<td>( P3_{21}, P6_{122} )</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>2.9</td>
</tr>
<tr>
<td>( P1 )</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>2.9</td>
</tr>
<tr>
<td>( P6_{522} )</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>2.5</td>
</tr>
<tr>
<td>( P4_{1212} )</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>—</td>
</tr>
<tr>
<td>( I22_2 )</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>( I_{2121} )</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>( I4, P6_{1}, R3 )</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>—</td>
</tr>
<tr>
<td>( P4_{212} )</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>( P3_{1}, P4_{1}, P4_{3} )</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>( P3_{2}, P6_{1}, P6_{3}, P6_{5} )</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>—</td>
</tr>
</tbody>
</table>
Examples of racemic crystals

RNA

r(CUGGGGCUGG) & r(CCGCCUGGG)
Space group P $\bar{I}$, phased by MR
6 PDB entries:
2G32, 2GPM,
2GQ4, 2GQ5,
2GQ6, 2GQ7
Examples of racemic crystals

DNA

d(CGCGCG)
Z-DNA, space group P $\bar{1}$
Phased with MR.
1 PDB entry: 1VTU
Examples of racemic crystals

**PNA (Peptide nucleic acid)**


Self-complementary, (H-CGTACG-NH2)

PDB entry 1PUP, Space group P Ī

Phased with SIR (5-bromo-uracil)

“The PNA crystallized in a centrosymmetric spacegroup which is possible because PNA – in contrast to natural nucleic acids – contains no asymmetric centres.”
Examples of racemic crystals

Proteins

Space group $P \bar{1}$:

(no entry) rubredoxin (45 a.a.)
Zawadzke & Berg (1993)
3ODV kaliotoxin (38 a.a.)
3E7R plectasin (40 a.a.)
3BOI snow flea antifreeze (81 a.a.)
1KRL DL-monellin (44, 50 a.a.)
3ALI designed peptide Alpha-1 (13 a.a.)
3TRW villin headpiece (35 a.a.)

Space group $I - 4 C 2$:

3TRY villin headpiece (35 a.a.)
“These results demonstrate that the growth of centrosymmetric crystals of racemic synthetic proteins is, indeed, possible. It should be noted that other outcomes are possible. Racemic proteins could spontaneously resolve into crystals of enantiomerically pure materials or could crystallize with the two enantiomers related by symmetry operations other than an inversion center.”

“For crystals of many small molecules, extremely high quality electron density maps can be obtained. This is especially true for crystals that are centrosymmetric, a feature which requires that the phase of each term in the expansion be exactly 0 or exactly 180° rather than potentially taking on any value from 0 to 360°, as is the case for noncentrosymmetric crystals.”

“it may well be possible to solve centrosymmetric protein structures by direct methods”
“It might be noted, however, that racemic crystals complicate the use of anomalous scattering data in determining the structure of the protein. If a protein crystal contains heavy atoms such as mercury, iodine or selenium the Friedel-related reflections \( F_0(hk\ell) \) and \( F_0(-h-k-l) \) will differ slightly in intensity and these differences can be used to obtain information on the phase angles. In a centrosymmetric crystal, however, \( F_0(hk\ell) \) and \( F_0(-h-k-l) \) have equal amplitudes and no phase information can be obtained. 

... If, however, measurements are made at two or more different wavelengths there will be a change in the anomalous scattering of the heavy atoms which in turn will alter the amplitude of \( F_0(hk\ell) \), ...”

I.e. SAD data provides no phasing for centrosymmetric crystals. MAD data could be used as a weak case of MIR.
Quasi-racemic crystallization

Just like racemic crystallization – but with slight changes in one of the enantiomers. The hope is that the mixture will still crystallize with quasi-centrosymmetry.

The slight changes can be anomalous scatterers, which can be used for phasing.

Since at least one of the enantiomers is created by total chemical synthesis, the additional challenge of inserting the changes is small.
**Quasi-racemic protein crystals**

Villin headpiece subdomain (35 a.a.)
D.E. Mortenson, K.A. Satyshur, I.A. Guzei, K.T. Forest and S.H. Gellman

PDB:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>MR</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>3TJW</td>
<td>F222</td>
<td>MR</td>
<td>quasi-racemic</td>
</tr>
<tr>
<td>3TRV</td>
<td>P 1</td>
<td>ab initio</td>
<td>quasi-racemic</td>
</tr>
<tr>
<td>3TRW</td>
<td>P İ</td>
<td>MR</td>
<td>racemic</td>
</tr>
<tr>
<td>3TRY</td>
<td>I -4 C 2</td>
<td>MR</td>
<td>racemic</td>
</tr>
</tbody>
</table>

Stated purpose was to test use of Quasi-racemic crystallization for Evaluating substitution of Pentafluorophenylalanine (F₅Phe) for Phe.
Quasi-racemic crystals

Snow flea antifreeze protein (81 a.a.)

PDB:
3BOG  P 1  MAD
3BOI  P 1  (Phased from 3BOG)
2PNE  P 2_1  MR

Replaced Asn^{11} with Seleno-Cys alkylated with bromoacetamide in one enantiomer for phasing
Single-wavelength phasing strategy for quasi-racemic protein crystal diffraction data


Used existing data from snow flea antifreeze protein (3BOG)

Combined constraints of SAD with constraints of pseudo-centrosymmetry to gain additional phasing benefit.

Fig 1a: crystal unit cell is nearly centrosymmetric, except for the anomalous scatterers.
Fig 1b: 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 ($f^{\prime\prime}$) 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 ($0^\circ$ or $180^\circ$). 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 $FP$. 

Fig 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. 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.

Fig 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.

<table>
<thead>
<tr>
<th>Phasing Strategy</th>
<th>Map CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWAQR</td>
<td>SAD best centric</td>
</tr>
<tr>
<td>“” &quot;&quot; followed by DM</td>
<td><strong>0.841</strong></td>
</tr>
<tr>
<td>SAD: MLPHARE</td>
<td>0.483</td>
</tr>
<tr>
<td>SAD: Phaser</td>
<td>0.510</td>
</tr>
<tr>
<td>SAD: Phaser followed by DM</td>
<td>0.564</td>
</tr>
</tbody>
</table>

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 (Fcalc) 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.
Fig 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, 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.
Summary

* Racemic crystallization works; examples of RNA, DNA, PNA, peptide/proteins. Crystallization may be easier than for individual enantiomers, but the sample set is too small to draw a general conclusion.

* Sometimes a molecule crystallizes better in complex with another molecule, regardless of whether that other molecule is its opposite enantiomer.

* Obstacles include difficulty and expense of total chemical synthesis, proper folding, and post-translational modification. Most of these do not scale well.

* As predicted by Wukovits & Yeates, P ̅I is the most populous centrosymmetric space-group, so far.

* Racemic crystals can be phased with MR, isomorphous replacement or direct methods. Anomalous does not work well. Direct methods do not scale well to larger size or poorer resolution.

* Quasi-racemic reintroduces the possibility of anomalous phasing while maintaining the benefits of racemic crystallization.

* Attempt by Sawaya et al. To combine SAD and quasi-centrosymmetric constraints for single-wavelength phasing appears successful.