Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Resolving the space-group ambiguity of crystals of tomato fruit polygalacturonase

Susan Heffron,[a] Stephan Watkins,[a] Rhonda Moeller,[a] Aise Huma Taban,[b] Rafal Butowt,[b] Dean DellaPenna[c] and Frances Jurnak[a]*

[a]University of California, Irvine, USA, [b]University of Nevada, Reno, USA, and [c]Michigan State University, East Lansing, USA

Correspondence e-mail: jurnak@uci.edu

Received 25 June 2003
Accepted 13 August 2003

The first polygalacturonase from a plant, tomato fruit PG2, has been crystallized and data have been collected to a resolution of 1.87 Å. The autoindexing program strongly favors one of the primitive orthorhombic cells. A plausible molecular-replacement solution for two molecules in the asymmetric unit has been found for data assigned to space group $P2_12_12_1$. Although the numerical criteria and the electron-density maps are reasonable for this solution, manually adjusted models do not refine to an R factor below 0.48. Visual inspection of hkl Bragg planes does not reveal a breakdown in mm symmetry. Nevertheless, the correct space group has been determined to be $P2_1$, with similar unit-cell parameters, a β angle of 90.04° and four molecules in the asymmetric unit. The $R_{\rm sym}$ of 0.053 for data processed in $P2_12_12_1$ is very similar to the $R_{\rm sym}$ of 0.047 for the same data processed in $P2_1$. Comparisons of the intermediate results using the $P2_12_12_1$ and $P2_1$ data sets are provided and the subtle indications of an initial erroneous space-group assignment are discussed.

1. Introduction

Polygalacturonases (PGs) are enzymes that hydrolyze α-1,4 linkages in the polygalacturonic acid component of plant cell walls (Kester & Visser, 1990). PGs are secreted by fungal and bacterial organisms that infect plants by macerating the cell wall (Collmer & Keen, 1986). PGs are also produced by plants during various stages of growth and development, including pollen maturation, abscission, dehiscence, fruit ripening and in rapidly expanding plant tissues (Hadfield & Bennett, 1998). Elevated production of PGs is an event common to the ripening process of many fruits and vegetables and is accompanied by increased solubility and decreased polymerization of the pectic components of cell walls. The best studied plant polygalacturonase is tomato fruit polygalacturonase (PG2), which is produced at high levels during tomato fruit ripening. The *PG2* gene encodes a 457 amino-acid primary translation product, of which 71 amino acids are removed post-translationally from the N-terminus and 13 amino acids from the C-terminus in the mature tomato fruit PG2. Purified PG2 is a mixture of two isoforms, PG2A and PG2B, derived from the same polypeptide but having different amounts of glycosylation.

Over the years, the collection and processing of X-ray diffraction data have become increasingly automated, so much so that autoindexing programs are commonly used to determine the space group of a crystal. When pseudosymmetry is present, the diffraction pattern may mimic a higher symmetry space group. A visual inspection of actual precession photographs is useful in detecting minor violations in symmetry

patterns. With automated space-group determination, the statistics for a large number of reflections may obliterate significant differences between a few pseudosymmetry-related reflections. Thus, an incorrect space-group assignment might not be suspected at an early stage of analysis. This report details the space-group ambiguity encountered for PG2 crystals and examines the results at each stage of the analysis for indications of an initial error in the space-group assignment.

2. Experimental methods

2.1. Crystallization

Tomato fruit polygalacturonase (EC 3.2.1.15) isoenzymes 2A and 2B were collectively purified from ripe tomato fruit pericarp following the method of Zheng *et al.* (1992). Both isoenzymes were glycosylated, with isoenzyme PG2B having a greater degree of glycosylation than PG2A. Crystals of an equimolar mixture of PG2A and PG2B were grown at 277 K by the sitting-drop vapor-diffusion method using a reservoir containing 0.1 *M* sodium acetate pH 6.0, 60%(*v*/*v*) ammonium sulfate and 2%(*v*/*v*) methylpentanediol. The protein droplet contained equal volumes of the reservoir solution and a solution of 12 mg ml^{-1} PG2 in 10 m*M* *N*-(2-hydroxyethyl)-piperazine-*N*′-(2-ethanesulfonic acid) pH 7.5, 25 m*M* sodium chloride and 0.1 m*M* dithiothreitol. In preliminary crystallization experiments the largest crystals appeared in four weeks. Using microseeding techniques (Fitzgerald & Madsen, 1986), the time for crystal growth was reduced to one week.

2.2. X-ray data collection

For data collection, a crystal was quickly passed through a solution of 30% glycerol and 70% reservoir before freezing in a stream of nitrogen at 95 K. A complete X-ray diffraction data set was measured from a single crystal at Stanford Synchrotron Radiation Laboratory beamline 7-1 at a wavelength of 1.08 Å. The crystal diffracted to a resolution of 1.87 Å. Autoindexing and data integration were carried out using *MOSFLM* (Leslie, 1992; Steller *et al.*, 1997). Reflections were scaled and merged with *SCALA* (Evans, 1997) and intensities were converted to structure-factor amplitudes using *TRUNCATE* (French & Wilson, 1978). To visually inspect the symmetry of the diffraction, pseudo-precession images were examined with *HKLVIEW* (Collaborative Computational Project, Number 4, 1994). To reindex the data set from $P2_12_12_1$ to other primitive orthorhombic space groups, as well as to reindex into the monoclinic space group, the *CCP*4 programs *REINDEX* and *SORTMTZ* (Collaborative Computational Project, Number 4, 1994) were used.

After the correct monoclinic space group was determined, *XPREP* version 6.13 (Bruker–Nonius) was tested to see whether it could have been used to select the space group correctly.

2.3. Molecular replacement

2.3.1. Search models. To construct a search model for molecular replacement, a starting model using the PG2 sequence (Swiss-Prot accession No. P05117) was obtained from the SWISS-MODEL web server (Peitsch, 1996). The model included only residues 110–329. Additions and adjustments to the model were made manually in *O* (Jones *et al.*, 1990) based on a comparison of the *CLUSTALW* multiple-sequence alignment (Thompson *et al.*, 1994) and the parallel β-helical structures of PG from *Aspergillus niger* (PDB code 1czf; Van Santen *et al.*, 1999) and *Erwinia carotovora* (PDB code 1bhe; Pickersgill *et al.*, 1998). Loops were modeled after one or the other structure, depending upon the sequence similarity and length of the corresponding region in PG2. Two loop regions had minimal similarity to either PG structure and were omitted from the model. The final PG2 search model consisted of residues 18–75, 85–110 and 126–345. The model was subjected to 60 cycles of energy minimization in *CNS* (Brünger *et al.*, 1998). For searches in *BEAST*, a second search model was superimposed on the first and used simultaneously with it. This second model was constructed by converting the *A. niger* PG structure to a polyalanine model using *MOLEMAN* (Kleywegt *et al.*, 2001).

2.3.2. Molecular-replacement strategies. The structure was phased by molecular-replacement methods using the programs *EPMR* (Kissinger *et al.*, 1999), *CNS* and *BEAST* (Read, 2001). All three programs were used in space group $P2_12_12_1$, while only *BEAST* was employed in $P2_1$. Molecular-replacement searches were executed using data between 4 and 15 Å in *CNS* and *EPMR* and using data between 3.5 and 50 Å in *BEAST*.

In $P2_12_12_1$, rotation searches were first carried out with *EPMR* and *CNS*. The list of top rotation peaks from each was subsequently used as input to the translation search in *CNS*, which included Patterson correlation refinement of the orientations (Brünger, 1990) followed by the translation search for each entry on the list. To search for the second molecule, the first molecule was fixed and the translation search was carried out again with the same rotation list. In *BEAST* (in both $P2_12_12_1$ and $P2_1$), the rotation search was carried out first with a coarse grid, followed by fine-grid rotation searches around the top peaks. Likewise, translation was first performed on a coarse grid, followed by a fine-grid combination of rotation and translation searches ('six-dimensional refinement') around the top solution. To search for the second or subsequent molecule with *BEAST*, both the rotation searches and translation searches were performed with the first molecule fixed.

2.4. Preliminary refinement

Crystallographic refinement was carried out in *CNS*, including rigid-body refinement, conjugate-gradient minimization, simulated annealing (Brünger *et al.*, 1990) and *B*-factor refinement. Density-modified maps with solvent flipping as well as σ_A-weighted electron-density maps (Read, 1986) were produced with *CNS*. For refinement in $P2_12_12_1$, 5% of reflections were set aside for cross-validation. For refinement in $P2_1$, the original test set of reflections was kept and then extended with the *CCP*4 programs *CAD* and

Table 1
MOSFLM autoindexing.

No.	Penalty	Lattice	*a* (Å)	*b* (Å)	*c* (Å)	α (°)	β (°)	γ (°)	Possible space groups
9	460	t*P*	89.55	89.55	157.26	90.0	90.0	90.0	*P*4, $P4_1$, $P4_2$ *etc.*
8	459	m*C*	127.93	127.93	157.26	90.0	90.0	90.0	*C*2
7	459	m*C*	127.93	127.93	157.26	90.0	90.0	90.0	*C*2
6	1	m*P*	76.73	157.26	102.36	90.0	90.0	90.0	*P*2, $P2_1$
5	**1**	**o*P***	**76.73**	**102.36**	**157.26**	**90.0**	**90.0**	**90.0**	***P*222, *P*222$_1$, *P*2$_1$2$_1$2 *etc.***
4	1	m*P*	102.36	76.73	157.26	90.0	90.0	90.0	*P*2, $P2_1$
3	0	a*P*	76.73	102.36	157.26	90.0	89.9	90.0	*P*1
2	0	m*P*	76.73	102.36	157.26	90.0	90.1	90.0	*P*2, $P2_1$
1	0	a*P*	76.73	102.36	157.26	90.0	90.1	90.0	*P*1

UNIQUEIFY to constitute 5% of reflections in the lower symmetry space group.

3. Results and discussion

The autoindexing routine in *MOSFLM* clearly indicated that the data were consistent with a primitive orthorhombic space group, *P*222, $P222_1$, $P2_12_12$ or $P2_12_12_1$, with a penalty value of only 1, as shown in Table 1. The data were initially integrated, scaled and merged as $P2_12_12_1$, with refined unit-cell parameters a = 76.47, b = 101.91, c = 156.67 Å. The data statistics are shown in Table 2. After scaling, the data had an $R_{\rm sym}$ of 0.053, with 96.6% completeness. The list of axial reflections indicated that the systematic absences were consistent with $P2_12_12_1$ symmetry. A summary of the axial reflections is given in Table 3. The symmetry of the diffraction images, an example of which is shown in Fig. 1, appeared to be consistent with *mm* symmetry. There was no apparent indication of an error in the space-group assignment at the autoindexing stage of data processing for the PG2 data set.

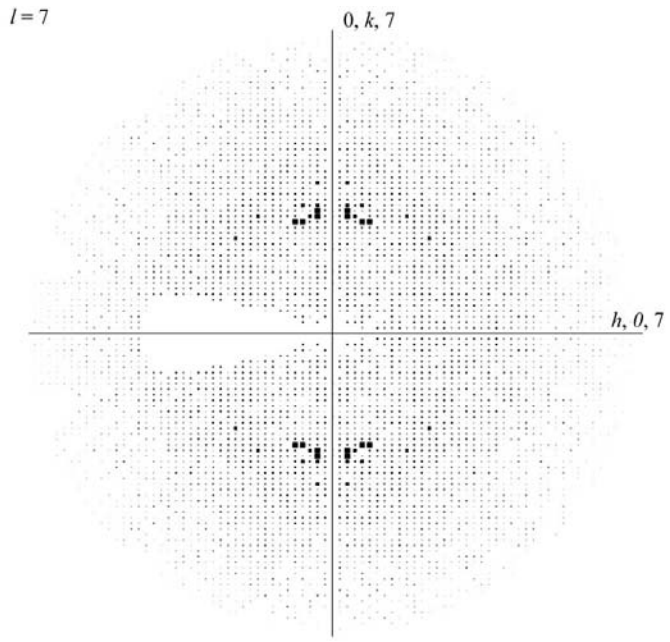

Figure 1
Pseudo-precession image of zone (*h*, *k*, 7) generated from PG2 data processed in space group $P2_1$ demonstrates the apparent *mm* symmetry resulting from the pseudosymmetry of the crystal. (The image was produced using *HKLVIEW*.)

Table 2
Crystal parameters and data statistics.

Values in parentheses are for the highest resolution bin.

	$P2_12_12_1$	Reindexed $P2_1$	Reprocessed $P2_1$
Unit-cell parameters			
a (Å)	76.47	76.47	76.51
b (Å)	101.91	101.87	101.99
c (Å)	156.67	156.59	156.75
β (°)	90.00	90.00	90.04
Molecules per AU	2	4	4
Resolution limits (Å)	1.87–32.61 (1.87–1.92)	1.87–32.61 (1.87–1.97)	1.87–48.22 (1.87–1.92)
No. reflections measured	333105	345692	344986
No. unique reflections	98087	179444	178771
Completeness (%)	96.6 (82.4)	91.0 (80.9)	90.3 (71.1)
$R_{\rm sym}$†	0.053 (0.324)	0.048 (0.259)	0.047 (0.266)
$R_{\rm meas}$‡	0.062 (0.388)	0.063 (0.349)	0.063 (0.359)
$I/\sigma(I)$	9.5 (2.2)	9.4 (2.8)	10.7 (2.8)
Multiplicity	3.4 (2.9)	1.9 (1.8)	1.9 (1.7)

† $R_{\rm sym} = \sum(|I - \langle I\rangle|)/\sum I$. ‡ $R_{\rm meas}$ is a redundancy-independent $R_{\rm merge}$ (Diederichs & Karplus, 1997).

Table 3
Axial reflection statistics.

	*h*00		0*k*0		00*l*	
	h = 2*n* + 1	*h* = 2*n*	*k* = 2*n* + 1	*k* = 2*n*	*l* = 2*n* + 1	*l* = 2*n*
No. reflections	16	14	19	19	40	38
No. $I > 3\sigma$	1	12	0	14	0	27
No. $I > 2\sigma$	2	14	0	16	0	28
Avg. *I*	46	16660	28	17182	−38	15835
Avg. $I/\sigma(I)$	0.38	12.45	0.26	13.06	−0.14	11.00

The Matthews coefficient (Matthews, 1968) calculated using *CNS* resulted in a reasonable value for $V_{\rm M}$ of 3.69 Å^3 Da^{-1} and a solvent content of 66% for two molecules per asymmetric unit in space group $P2_12_12_1$. The native Patterson also indicated two molecules per asymmetric unit in $P2_12_12_1$. The top non-origin peaks, (0.000, 0.050, 0.000) and (0.047, 0.000, 0.025) in fractional coordinates, had peak heights of 15.27 and 12.08 standard deviations above the mean, respectively. The respective peaks had vector lengths of 5.1 and 5.3 Å, which were in good agreement with the expected vector length of 4.5–5.5 Å arising from the repetitive nature of the parallel β-helix. Vectors between main-chain atoms of neighboring 'rungs' of the parallel β-helix have similar length and an orientation that is nearly parallel to the axis of the parallel β-helix.

The molecular-replacement results using *EPMR* suggested only one set of four rotation solutions, all of which were crystallographically related to each other in $P2_12_12_1$. Similarly, *CNS* found only one rotation and translation solution, which packed well and was significantly above the mean. The single solution had Euler angles θ_1 = 108.14, θ_2 = 74.52, θ_3 = 349.26° after PC refinement and an initial rotation score of 0.136. The

Table 4
Molecular-replacement results (*BEAST*).

	α (°)	β (°)	γ (°)	*X* trans.	*Y* trans.	*Z* trans.	Rotational score	Translational score	Next highest translation
$P2_12_12_1$									
Soln 1	100.74	75.17	161.33	0.958	0.260	0.411	105.3	118.2	100.1
Soln 2†	213.27	18.39	190.96	0.378	0.156	0.462	198.1	148.8	138.4
$P2_1$									
Soln 1	100.97	75.27	161.97	0.449	0.000	0.162	142.2	164.6	148.1
Soln 2†	281.24	75.27	161.23	0.038	0.474	0.661	284.5	378.4	312.1
Soln 3‡	215.91	17.27	185.31	0.826	0.940	0.207	515.6	589.0	497.1
Soln 4§	35.38	18.19	186.16	0.664	0.538	0.709	704.0	784.6	616.9

† Solution 2 scores were calculated with solution 1 fixed. ‡ Solution 3 scores were calculated with solutions 1 and 2 fixed. § Solution 4 scores were calculated with solutions 1, 2 and 3 fixed.

translation vector (33.92, 27.57, 63.76) had a score of 0.241. *CNS* could not identify a second solution that packed well with the first solution. Using *BEAST*, two plausible molecular-replacement solutions were found, as shown in Table 4. The signal-to-noise ratio of the top solution was 4.7. The log-likelihood gain (LLG) of 105.3 for the rotation solution increased to 118.2 for the translation solution. Although it is not obvious owing to the different Euler angle conventions, this *BEAST* solution is actually equivalent to the *CNS* solution with a different choice of origin. A second plausible solution emerged from *BEAST* as the top rotation-search peak when the first molecule was fixed. Neither *CNS* nor *EPMR* molecular-replacement searches had identified this orientation. A translation search of the second orientation, with the first molecule fixed, yielded a translation solution with a signal-to-noise ratio of 5.3. The LLG of 148.8 for the translation solution was less than the LLG of 198.1 for the rotation solution alone. Although the decrease in LLG was unusual, the second molecule packed very well with the first one, with no discernible molecular overlaps. Both orientations were consistent with the native Patterson peaks.

After an initial round of rigid-body refinement, conjugate-gradient minimization and individual restrained *B*-factor refinement, the model had an $R_{\rm work}$ of 0.50 and an $R_{\rm free}$ of 0.50. $R_{\rm free}$ increased with the application of simulated annealing and *ARP*/*wARP* was unable to build the missing parts of the structure. In the warpNtrace mode, the final model consisted of only 89 residues and many water molecules, yielding an $R_{\rm free}$ of 0.52. Despite these discouraging signs, σ_A-weighted electron-density maps looked reasonable, with clear protein–solvent boundaries. Therefore, manual reconstruction of the model was initiated using both σ_A-weighted and density-modified maps. However, after five rounds of refinement and model adjustment the *R* factor remained at 0.48. At this point, it was obvious that there was an error in the analysis.

The systematic absences were re-examined for a possible error in the screw-axis assignment. As shown in Table 3, there were only two small violations, both in odd-numbered *h*00 reflections, one having an intensity $I > 3\sigma$ and the other having $I > 2\sigma$. The data were reindexed in $P2_12_12$ with the unit-cell axes permutated appropriately and the molecular-replacement searches with *BEAST* were repeated. Two plausible molecular-replacement solutions were found for the reindexed space group $P2_12_12$, with scores that were slightly better than the original solution, but crystallographic refinement with this solution was also unsuccessful. Molecular replacement was also attempted, without success, with data reindexed in each of the other alternative primitive orthorhombic space groups with all possible permutations of the unit-cell axes.

In order to test for twinning, the data were reindexed in space group $P2_1$, assuming a β angle of 90.0°. Although the merging *R* factors for the highest resolution bin were somewhat better in $P2_1$, the overall statistics from *SCALA* were similar to those for the data originally processed in $P2_12_12_1$, as shown in Table 2. In the cumulative intensity distribution plot from *TRUNCATE*, the curves of the actual values matched the theoretical distribution well; thus, twinning was eliminated as a possible problem.

Finally, the possibility that the crystal might be in the lower symmetry space group $P2_1$ was considered. Only 100° of data had been collected, yet the data set, when reindexed in $P2_1$, was 91.0% complete, allowing the analysis to proceed. Although there were three possible permutations of the unit-cell parameters in the monoclinic space group, the first option selected retained the same assignment of *a*, *b* and *c* as in $P2_12_12_1$ and proved to be the correct one. Four molecular-replacement solutions were readily identified with *BEAST* and are shown in Table 4. Rigid-body refinement of these four copies in *CNS* yielded an $R_{\rm work}$ of 0.45 and an $R_{\rm free}$ of 0.44. Forty cycles of conjugate-gradient minimization followed by simulated annealing with a starting temperature of 5000 K and then 50 cycles of restrained individual *B*-factor refinement yielded a structure with an $R_{\rm work}$ of 0.40 and an $R_{\rm free}$ of 0.41. Maps showed clear electron density for the missing parts of the protein. Model rebuilding with *ARP*/*wARP* in warpNtrace mode resulted in a model consisting of 19 segments of main chain, representing a total of 1407 residues, plus multiple water molecules. By visual inspection in *O*, these main-chain segments were assigned to the four molecules in the $P2_1$ asymmetric unit. Side-chain docking in *ARP*/*wARP* resulted in a model of the four copies of PG2 consisting of chain *A* residues 18–344, chain *B* residues 3–368, chain *C* residues 17–362 and chain *D* residues 2–368. After individual restrained *B*-factor refinement, this model had an $R_{\rm work}$ of 0.30 and an $R_{\rm free}$ of 0.30. The significant decrease in *R* factors indicated that the correct space group and molecular-replacement solution had been found.

An examination of the correct solution in the monoclinic space group reveals that the pseudosymmetry axis is a perfect 2_1 axis but is tilted from the *x* axis by 0.02° and offset by approximately 1.2 Å along *y* from the crystallographic axis in $P2_12_12_1$. Two of the molecules in the asymmetric unit are equivalent to the molecular-replacement solutions identified

by *BEAST* in the incorrect orthorhombic space group. The additional two molecules in the monoclinic asymmetric unit are related to the first two by the pseudo-2_1 screw axis parallel to the *a* unit-cell edge. Fig. 2 shows a comparison of the molecular-replacement solutions in the monoclinic and orthorhombic space groups. The four molecules of the $P2_1$ molecular-replacement solution are shown as ribbons, while the two molecules of the $P2_12_12_1$ solution and a crystallographic symmetry-generated copy of each are shown as threads superimposed on the $P2_1$ asymmetric unit. Chain 1 of the $P2_12_12_1$ solution, shown as light orange threads, superimposes exactly on chain *A* of the $P2_1$ solution, while chain 2, shown as yellow threads, superimposes almost exactly on chain *B*. The crystallographically related chain 2′ (gold threads) in $P2_12_12_1$ superimposes upon chain *D* in the monoclinic asymmetric unit but is shifted along the axis of the parallel β-helix by an average of 4.6 Å or approximately one 'rung'. The superposition results suggest that most of the main-chain atoms of the parallel β-helix core of chain 2′ are correctly placed by orthorhombic symmetry on the pseudo-symmetry-related copy, chain *D*, in $P2_1$. In contrast, the crystallographically related chain 1′ (red–orange threads) in $P2_12_12_1$ is offset by approximately 5.1 Å from chain *C* in the monoclinic asymmetric unit in such a way that few if any of the atoms are in the correct positions. The superposition results explain why the molecular-replacement solution for one molecule (chain 2) in $P2_12_12_1$ was so much easier to identify than the second molecule (chain 1). Indeed, only *BEAST*, with its novel feature for fixing the first molecule in both the rotation and translation searches, was able to identify a second plausible solution. Overall, the four chains (*A*, *B*, *C* and *D*) are packed together more tightly, making additional favourable, albeit asymmetric, contacts in $P2_1$ compared with the packing in $P2_12_12_1$. The glycosylation sites are well resolved in the electron density, revealing that chains *A* and *C* correspond to the PG2A isozyme and chains *B* and *D* correspond to the PG2B isozyme. The oligosaccharide groups are quite distant from the asymmetric intermolecular contacts and do not appear to have influenced the breakdown of orthorhombic symmetry. Although the repetitive nature of the parallel β-helix structure may have accentuated the difficulty in recognizing that the initial space-group assignment was incorrect, ambiguity of a space-group assignment caused by pseudo-orthorhombic symmetry in a monoclinic space group is not unique to this type of topology but has occasionally occurred in other structures as well (Isupov *et al.*, 1996).

After the space group was correctly deduced, the original X-ray images were reprocessed in space group $P2_1$. The refined unit-cell parameters in $P2_1$ were nearly identical to the data processed in $P2_12_12_1$, but with a β angle of 90.04°. The statistics for the reprocessed $P2_1$ data are compared with those for the reindexed data, as well as for the data originally processed in $P2_12_12_1$, in Table 2.

To determine whether another automated method might have detected the true symmetry of the crystal, the data were analysed with *XPREP*. The intensity data with unmerged Friedel pairs were reduced in $P2_1$ and then subjected to the 'search for higher metric symmetry' option in *XPREP*. The program reported an R_{sym} of 0.04 for primitive orthorhombic symmetry. It recommended $P2_12_12_1$ as the space group and listed systematic absence statistics consistent with this space group.

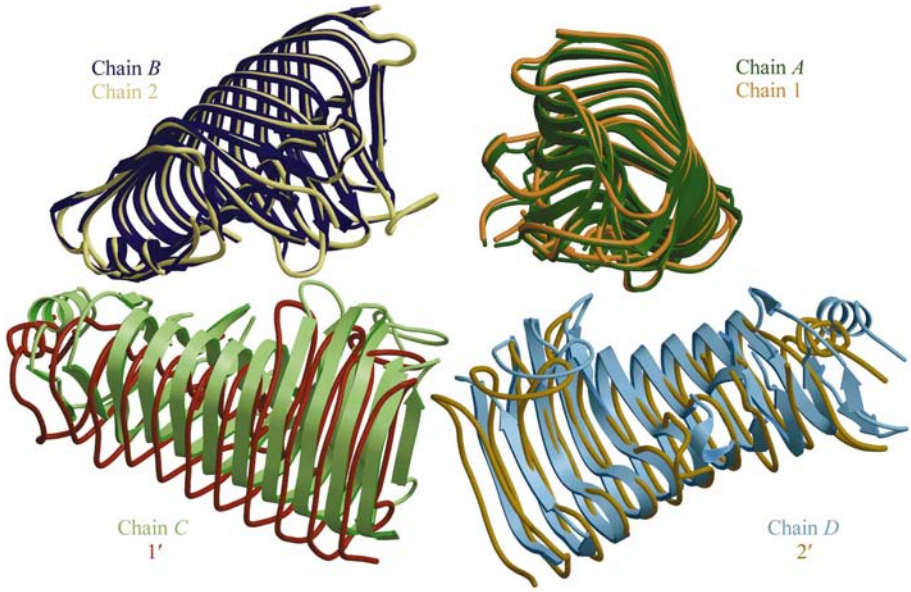

Figure 2
Comparison of the molecular-replacement solutions for PG2 in $P2_1$ and in $P2_12_12_1$. Each copy of the protein shown includes only those residues present in the search model; namely, residues 18–75, 85–110 and 126–345. The four copies of PG2 in $P2_1$ are illustrated as ribbons, with chain *A* shown in dark green and its pseudosymmetry copy, chain *C*, in light green and with chain *B* shown in dark blue and its pseudosymmetry copy, chain *D*, in light blue. The corresponding copies in $P2_12_12_1$ are represented by threads, with chain 1 shown in light orange and its crystallographically related copy, 1′, shown in red–orange and with chain 2 shown in yellow and its crystallographically related copy, 2′, shown in gold. As can be seen, chains 1 and 2 of the $P2_12_12_1$ molecular-replacement solution superimpose almost exactly on chains *A* and *B* of the $P2_1$ solution. However, their $P2_12_12_1$ symmetry-generated copies, 1′ and 2′, are offset slightly from the positions of chains *C* and *D* of the correct molecular-replacement solution in $P2_1$. This image was produced with *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997).

4. Conclusions

In summary, the correct monoclinic space group and unit-cell parameters for PG2 crystals were not distinguishable from the higher symmetry orthorhombic cell until the crystallographic refinement stage of the structure determination. Neither automated space-group determination procedures nor visual inspection of the symmetry were able to resolve the ambiguity. The first indications of an incorrect space-group assignment occurred at an intermediate stage of the analysis. Two

molecular-replacement programs failed to find a second PG2 copy expected from crystal-density calculations and from native Patterson vectors. In *BEAST*, which did successfully identify two copies, the LLG for the second copy was lower for the translation solution than for the rotation solution, a subtle indication of a possible problem. Moreover, *ARP*/*wARP* failed to construct huge segments of the model even though the electron-density maps appeared to be interpretable. Together, the problems did not immediately suggest an error in the space-group assignment, as these observations could easily be attributed to an incorrect molecular-replacement solution. Ultimately, it was the failure of the *R* factors to decrease upon refinement of the $P2_12_12_1$ model that indicated that alternative errors in the analysis needed to be considered. Once other primitive orthorhombic space groups and twinning were eliminated then lower symmetry space groups had to be considered. Reindexing assuming a β angle of 90° rather than reprocessing was sufficient to determine that the monoclinic space group was correct for PG2, but this may not always be the case for other proteins.

Pseudosymmetry close to exact crystallographic symmetry is not common, but when it occurs it complicates the structural analysis. Given the current trend to use only automated data-processing programs to determine space groups, errors arising from pseudosymmetry may become a more significant problem. To minimize potential problems, the best approach is to collect enough data, if possible, to process the data in a lower symmetry space group. The extra data may be essential to a structure solution if the true space group is a lower symmetry one, but in any case the extra multiplicity will tend to improve the quality of the data. Moreover, to avoid undue effort in the analysis, the possibility that a crystal belongs to a lower symmetry space group should be considered sooner rather than later when a molecular-replacement solution will not refine but the electron-density maps appear to be of reasonable quality.

The research was supported by the US Department of Agriculture (Grant No. 02-03560 to FJ). The research was conducted in part at the Stanford Synchrotron Radiation Laboratory, which is operated by the Office of Basic Energy Science of the US Department of Energy.

References

Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Brünger, A. T., Krukowski, A. & Erickson, J. (1990). *Acta Cryst.* A**46**, 585–593.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Collmer, A. & Keen, N. T. (1986). *Annu. Rev. Phytopathol.* **24**, 383–409.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Evans, P. R. (1997). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
Fitzgerald, P. M. D. & Madsen, N. B. (1986). *J. Cryst. Growth*, **76**, 600–606.
French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–525.
Hadfield, K. A. & Bennett, A. B. (1998). *Plant Physiol.* **117**, 337–343.
Isupov, M. N., Obmolova, G., Butterworth, S., Badet-Denisot, M.-A., Badet, B., Polikarpov, I., Littlechild, J. A. & Teplyakov, A. (1996). *Structure*, **4**, 801–810.
Jones, T. A., Bergdoll, M. & Kjeldgaard, M. (1990). *Crystallographic and Modeling Methods in Molecular Design*, edited by C. Bugg & S. Ealick, pp. 189–195. New York: Springer–Verlag.
Kester, H. C. M. & Visser, J. (1990). *Biotech. Appl. Biochem.* **12**, 150–160.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, ch. 17.1. Dordrecht: Kluwer Academic Publishers.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**.
Matthews, B. W. (1968). *J. Mol. Biol.*, **33**, 491–497.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Peitsch, M. C. (1996). *Biochem. Soc. Trans.* **24**, 274–279.
Pickersgill, R., Smith, D., Worboys, K. & Jenkins, J. (1998). *J. Biol. Chem.* **273**, 24660–24664.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Steller, I., Bolotovsky, R. & Rossmann, M. G. (1997). *J. Appl. Cryst.* **30**, 1036–1040.
Santen, Y. van, Benen, J. A. E., Schröter, K.-H., Kalk, K. H., Armand, S., Visser, J. & Dijkstra, B. W. (1999). *J. Biol. Chem.* **274**, 30474–30480.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Zheng, L. S., Heupel, R. & DellaPenna, D. (1992). *Plant Cell*, **4**, 1147–1156.