Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Analysis of lattice-translocation disorder in the layered hexagonal structure of carboxysome shell protein CsoS1C

Yingssu Tsai,[a] Michael R. Sawaya[a,b,c] and Todd O. Yeates[b,c]*

[a]Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California, USA, [b]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California, USA, and [c]UCLA–DOE Institute for Genomics and Proteomics, Los Angeles, California, USA

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

Received 30 April 2009
Accepted 29 June 2009

PDB Reference: CsoS1C, 3h8y, r3h8ysf.

Lattice-translocation or crystal order–disorder phenomena occur when some layers or groups of molecules in a crystal are randomly displaced relative to other groups of molecules by a discrete set of vectors. In previous work, the effects of lattice translocation on diffraction intensities have been corrected by considering that the observed intensities are the product of the intensities from an ideal crystal (lacking disorder) multiplied by the squared magnitude of the Fourier transform of the set of translocation vectors. Here, the structure determination is presented of carboxysome protein CsoS1C from *Halothiobacillus neapolitanus* in a crystal exhibiting a lattice translocation with unique features. The diffraction data are fully accounted for by a crystal unit cell composed of two layers of cyclic protein hexamers. The first layer is fully ordered (*i.e.* has one fixed position), while the second layer randomly takes one of three alternative positions whose displacements are related to each other by threefold symmetry. Remarkably, the highest symmetry present in the crystal is *P*3, yet the intensity data (and the Patterson map) obey 6/*m* instead of $\overline{3}$ symmetry; the intensities exceed the symmetry expected from combining the crystal space group with an inversion center. The origin of this rare phenomenon, known as symmetry enhancement, is discussed and shown to be possible even for a perfectly ordered crystal. The lattice-translocation treatment described here may be useful in analyzing other cases of disorder in which layers or groups of molecules are shifted in multiple symmetry-related directions.

1. Introduction

Crystal-growth disorders can lead to various problems in the analysis of diffraction data. Types that have been studied in macromolecular crystals include twinning (Yeates, 1997; Parsons, 2003; Helliwell, 2008), in which distinct three-dimensional domains of a crystal are oriented differently but are otherwise well ordered individually, and various types of disorder that occur over shorter length scales in a crystal (*e.g.* from unit cell to unit cell). Different formalisms that are not necessarily mutually exclusive have been used to describe the latter kinds of disorder. Order–disorder, or OD, has been used to describe layered structures in which different layers are related to each other in a stochastic fashion according to operations of various types (Dornberger-Schiff, 1956; Dornberger-Schiff & Grell-Niemann, 1961). The term lattice translocation has been used to describe situations where different groups of molecules in a crystal, which may or may not be in layers, are displaced relative to each other in a stochastic fashion according to one or more translation vectors (Wang, Rho *et al.*, 2005; Bragg & Howells, 1954). Lattice-translocation disorder in macromolecular crystals was first

reported in the X-ray diffraction data of imidazole methemoglobin (Bragg & Howells, 1954) and is beginning to be reported with increasing frequency (Wang, Kamtekar *et al.*, 2005; Trame & McKay, 2001; Tanaka *et al.*, 2008; Zhu *et al.*, 2008). The situation presented in the current work is compatible with both the OD and lattice-translocation formalisms. Here, to be consistent with work on other macromolecular crystals showing similar kinds of disorder, we adopt the lattice-translocation terminology.

Lattice-translocation defects often manifest themselves by diffraction patterns that have streaky or alternating sharp and diffuse reflections (Zhu *et al.*, 2008; Wang, Rho *et al.*, 2005; Hwang *et al.*, 2006). The reflection intensities are also strongly affected by the disorder. In particular, a crystal suffering from a lattice-translocation disorder can be described as a convolution of (i) an ideal crystal with (ii) a function consisting of discrete points separated by the appropriate displacement vectors. The structure factors of such a crystal are therefore a multiplication of the idealized structure factors with the Fourier transform of the constellation of displacement vectors. The method developed by Wang, Kamtekar *et al.* (2005) has been used successfully to correct intensity data from such crystals by undoing the effects of multiplication by the latter transform (Wang, Rho *et al.*, 2005; Tanaka *et al.*, 2008; Kamtekar *et al.*, 2004; Hwang *et al.*, 2006; Zhu *et al.*, 2008). We extend those methods here to work on proteins from the carboxysome.

The carboxysome is a protein-based organelle that is found in cyanobacteria and some chemoautotrophs (reviewed in Yeates *et al.*, 2007, 2008; Cheng *et al.*, 2008). The carboxysome comprises the second part of a carbon-concentrating mechanism (CCM) in these organisms, whose overall purpose is to increase the local concentration of carbon dioxide available to the enzyme ribulose bisphosphate carboxylase/oxygenase (RuBisCO). The first part of the CCM is composed of transmembrane protein pumps that actively transport bicarbonate into the cytosol (Badger & Price, 2003). Bicarbonate then passes through pores in a viral capsid-like carboxysome shell, where it is converted to CO_2 inside the carboxysome by carbonic anhydrase (CA). The CO_2 is then used in reaction with ribulose bisphosphate by RuBisCO (also inside the carboxysome) to form two molecules of 3-phosphoglyerate (3PGA), which exit the shell into the cytosol. The carboxysome shell is assembled mainly from thousands of copies of a few proteins, which are homologous within and between organisms. The crystal structures of various carboxysome shell proteins have been elucidated (Klein *et al.*, 2009; Tanaka *et al.*, 2008; Tsai *et al.*, 2007; Kerfeld *et al.*, 2005) and rough atomic models for the assembly of the carboxysome have been developed (Tanaka *et al.*, 2008). The structures of homologous shell proteins from functionally distinct kinds of bacterial microcompartments have also been reported (Sagermann *et al.*, 2009; Crowley *et al.*, 2008).

The major shell proteins of the carboxysome and related bacterial microcompartments are cyclic homohexamers, which further assemble to form a tightly packed molecular layer that surrounds the enzymes inside. The formation of hexagonal layers apparently contributes to a tendency for some of the shell proteins to exhibit crystal-growth disorders (Kerfeld *et al.*, 2005; Yeates, unpublished data). Here, we report the crystal structure at 2.5 Å resolution of the shell protein CsoS1C from *H. neapolitanus* in the presence of a unique lattice-translocation disorder.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the last shell.

Data-collection statistics	
Resolution (Å)	2.5
Space group	*P*6
Unit-cell parameters	
a (Å)	66.7
b (Å)	66.7
c (Å)	63.0
Radiation source	APS beamline 24-ID-C
Radiation wavelength (Å)	0.9792
Measured reflections	46215
Unique reflections	5538
Completeness (%)	100 (100)
$R_{\rm sym}$†	0.186 (0.359)
$I/\sigma(I)$	13.17 (6.94)
Refinement statistics	
$R_{\rm work}$‡	0.255
$R_{\rm free}$§	0.311
R.m.s.d. bond lengths (Å)	0.019
R.m.s.d. bond angles (°)	1.9
PDB code	3h8y
$R_{\rm work}$¶	0.241
$R_{\rm free}$¶	0.285

† $R_{\rm sym}(I) = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm work} = \sum_{hkl}|F_{\rm obs} - F_{\rm calc}|/\sum_{hkl}F_{\rm obs}$. § $R_{\rm free} = \sum_{hkl}|F_{\rm obs} - F_{\rm calc}|/\sum_{hkl}F_{\rm obs}$, where all reflections belong to a test set of 5%. ¶ Refinement based on uncorrected intensities and multiple partially occupied molecules.

2. Materials and methods

2.1. Cloning and protein purification

The expression vector pProEx-HTb containing the CsoS1C gene was cloned using protocols similar to those described previously for the homologous protein CsoS1A (Tsai *et al.*, 2007). The vector was transformed into *Escherichia coli* BL21-CodonPlus-RP cells (Strategene, Cedar Creek, Texas, USA) for expression. Bacteria were grown at 310 K and protein expression was induced at an OD_{600} of ∼0.8 with isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h. The expressed protein was purified using a gravity Ni–NTA column. The N-terminal histidine tag was cleaved at the tobacco etch virus (TEV) cleavage site using TEV protease. TEV protease was purified as described previously (Kapust *et al.*, 2001).

2.2. Crystallization and structure determination

Csos1C (in native buffer consisting of 0.01 *M* Tris–HCl pH 8.0) was crystallized by hanging-drop vapor diffusion under conditions consisting of 0.1 *M* Tris–HCl pH 8.5, 0.15 *M* sodium citrate, 25% PEG 400, including as an additive 0.01 *M* nickel(II) chloride from a commercial screening kit (Hampton Research, Aliso Viejo, California, USA). A 1:1 ratio of protein to reservoir was used. Crystals formed within a week at room

temperature. Crystals were frozen in liquid nitrogen without additional cryoprotectant. Diffraction data were collected at the Advanced Photon Source (APS) at Argonne National Laboratory in Chicago, Illinois, USA. Strong diffraction was observed to 2.2 Å resolution, but crystal disorder limited subsequent data processing to 2.5 Å. Intensity data were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Molecular replacement was performed using the program *Phaser* (Storoni *et al.*, 2004) using the structure of the highly similar CsoS1A (98% identity) as a search model. The translocation disorder was analyzed as described below. Refinement was performed using *REFMAC* (Murshudov *et al.*, 1997) and *Coot* (Emsley & Cowtan, 2004). Data-collection and refinement statistics are listed in Table 1.

3. Results

3.1. Initial structure analysis

Diffraction data from CsoS1C could be indexed on a hexagonal lattice and processed in space group *P*6. Diffraction was strong to about 2.2 Å resolution, but the ability to obtain accurate intensities was limited by streaking along the **c*** direction (Fig. 1). Situations involving streaked or diffuse diffraction spots have been linked to previous cases of lattice-translocation disorder (Wang, Kamtekar *et al.*, 2005; Zhu *et al.*, 2008) and subsequent analysis proved the present case to be another such example. Diffraction spots showed prominent streaking in some crystal orientations, leading to problems integrating reflection intensities. In some cases, the match between the integration box and the spot dimensions was poor, but spot centroids were still predicted with reasonable accuracy by *DENZO*. Even in orientations where the streaking was not evident within individual frames, it was recognized that the large spread of the reflections along **c*** would introduce errors in measured intensities. Nonetheless, adequate $R_{\rm sym}$ values were obtained (see below). Ultimately, the resolution was limited to 2.5 Å for data processing and structure determination.

The hexagonal **a** and **b** unit-cell spacings (66.7 Å) were very similar to values observed in multiple previous crystal structures of homologous carboxysome shell proteins in which hexamers form tightly packed molecular sheets in crystals, and in the carboxysome by inference (Kerfeld *et al.*, 2005; Tsai *et al.*, 2007). It was therefore anticipated that the structure of CsoS1C would reveal hexamers packed likewise in molecular layers. Molecular replacement revealed a clear solution giving rise to a tightly packed layer, as expected, with a single hexamer on the sixfold axis packed side by side against other hexamers related by lattice translations along **a** and **b**. However, placement of this first layer of molecules, corresponding to the top molecular-replacement solution, left a thick layer of space between layers of molecules related by the unit-cell translation along **c**. This space was just large enough to accommodate another layer of molecules. Electron density was present in this region and showed some features of connectivity, but could not be readily interpreted. A molecular-replacement search for a second layer of molecules yielded a solution with severe steric overlap under *P*6 symmetry. This inconsistency indicated the presence of some kind of disorder, which prompted an in-depth examination.

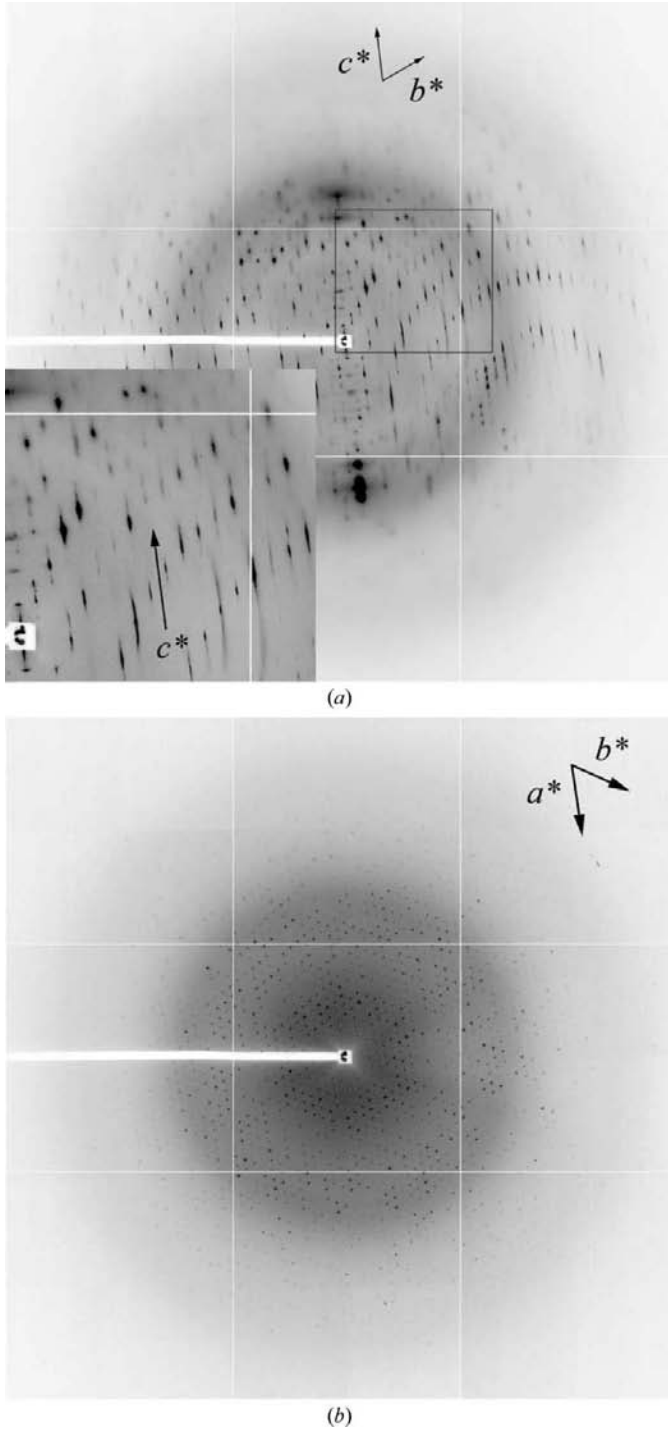

Figure 1
X-ray diffraction patterns from a CsoS1C crystal. (*a*) Disorder is evident in certain crystal orientations. The pattern of streakiness in a subset of the spots is an indication of possible translocation disorder. The insert highlights that the streaky spots are elongated in the **c*** direction. (*b*) An image from the crystal rotated 90° relative to (*a*) shows only sharp Bragg reflections.

3.2. Interpretation of the disorder

The possibility that the crystal might suffer from twinning was examined, since twinning sometimes leads to impossible packing scenarios. The overall statistics of local intensity differences were examined (Padilla & Yeates, 2003). This test provided no indication of twinning. In fact, the observed distribution was shifted slightly away from the trend expected for twinning, as can occur when unusual kinds of translational shifts between molecules lead to classes of strong and weak reflections. The data set was also examined to see if partial twinning (*i.e.* from *P*6 towards *P*622) might be present, but again a statistical analysis (Yeates, 1988) showed no systematic similarity between potentially twin-related reflections. Twinning was therefore ruled out as a cause of the difficulty encountered in interpreting the electron density.

A native Patterson map was examined and found to contain numerous prominent packing peaks (Fig. 2), indicating the presence of multiple layers of hexamers related by noncrystallographic translations. The $w = 1/2$ section of the Patterson unit cell contained six dominant peaks (in general positions) related to each other by sixfold rotational symmetry as required under *P*6. Interpretation of these peaks immediately led to a conflict. Based on the **a** and **b** unit-cell lengths and the size of the hexamer, only a single unique hexamer can be fitted in a layer (at $z = 1/2$ in this instance); the multiple peaks indicated by the native Patterson map could not possibly be accommodated. Furthermore, slightly lower but still prominent peaks (13σ) were present in the native Patterson map at $w = 0$ at positions only 9 Å from the origin. These peaks are closer to the origin than permitted by the dimensions of the molecules; two hexamers separated by that distance, as implied by the packing peaks, could not be present simultaneously in the unit cell. These observations, taken with the diffraction-spot anomalies noted earlier, led to the conclusion that the crystal suffered from a lattice-translocation disorder.

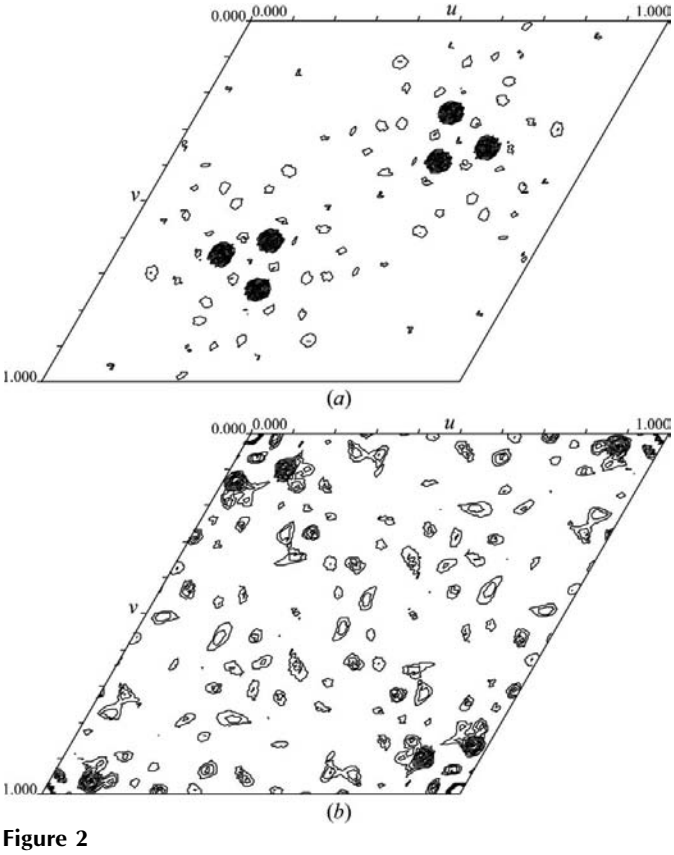

Figure 2
Packing peaks in the native Patterson map of CsoS1C. Map sections are contoured starting at 2σ. (*a*) $w = 1/2$ section. The six prominent symmetry-related peaks have a height of 22σ. (*b*) $w = 0$ section. The six prominent symmetry-related peaks have a height of 13σ.

During the course of the analysis, the correctness of the assigned *P*6 symmetry was examined in more detail. The data had reduced adequately in *P*6 at moderate resolution (18.6% $R_{\rm sym}$ overall out to 2.5 Å), but the $R_{\rm sym}$ values had climbed at the upper resolution limit. Attempts were therefore made to re-reduce the data in the lower symmetries *P*3, *P*2 and *P*1. Lowering the symmetry to *P*3 had no effect, but lowering the symmetry to *P*2 (and *P*1) produced a measurable improvement in $R_{\rm sym}$ which was most noticeable at higher resolution. The overall $R_{\rm sym}$ fell from 18.6% to 9.7%, while the $R_{\rm sym}$ in the outer shell fell from 35.9% to 20.1%. The crystal therefore very nearly obeys *P*6 symmetry, but slight deviations break the symmetry down to *P*2 at higher resolution. The data set reduced in *P*2 was used to examine the potential sources of the symmetry breaking. A cursory round of atomic refinement with just one hexamer present in the unit cell suggested that the observed deviations from perfect *P*6 symmetry could be explained by minor coordinate variations between subunits. Secondly, a native Patterson map based on the *P*2 data set showed small but measurable differences (∼5%) between the heights of the packing peaks at $w = 1/2$, which were necessarily equal under *P*6 but not under *P*2. We judged that the deviations from *P*6 symmetry were small enough that the advantages of lowering the symmetry to *P*2 in order to better account for these deviations would be more than offset by the disadvantages, particularly in view of the other challenges to be faced in interpreting the lattice-translocation disorder. Therefore, except for tests conducted in the very final stages, the lattice-translocation disorder problem was treated in *P*6, as if the hexamers obeyed perfect hexagonal symmetry, and the probabilities or occupancies of lattice translocation in the different directions were equal.

Based on the occurrence of six translation peaks at $w = 1/2$ in the native Patterson map (Fig. 2*a*), an initial interpretation was reached that the lattice translocation could be described as a single hexamer layer at $z = 0$ combined with six translocated hexamer layers at $z = 1/2$, each at 1/6 occupancy, shifted according to vectors obeying the observed *P*6 symmetry. This model turned out to be incorrect. If six copies of the hexamer were (partially) present at $z = 1/2$ following *P*6 symmetry, then the six translocation vectors would constitute a hexagonal set of points (*i.e.* one at the center of each hexamer). With the understanding that multiple layers of molecules would scatter coherently in the crystal, translation vectors between these points would be expected at $w = 0$ in the native Patterson map. Cross-peaks between a hexagonal set of points give rise to three different-length vectors (between adjacent points in the

edge of the hexagon, between next-adjacent points and between opposing points), each present in six symmetry-related copies. No such constellation of peaks was present in the native Patterson at $w = 0$. Instead, only one unique peak was present, along with its sixfold-related peaks (Fig. 2*b*).

A potential resolution of the problem came from the realization that the six packing peaks at $w = 1/2$ are related to each other not only by the assumed sixfold symmetry but also by the inversion center at (1/2, 1/2, 1/2) in the Patterson map. Instead of six vectors, three translocation vectors following 120° or *P*3 symmetry would be sufficient to produce all six packing peaks at $w = 1/2$ (Fig. 3). In this case, a smaller set of cross peaks between these points (of an equilateral triangle) would be generated in the Patterson at $w = 0$. In fact, only one unique peak would be generated, consistent with the observed Patterson; the location of that peak matched the location predicted by choosing three (instead of six) translocation vectors at $z = 1/2$ (Fig. 4).

A model containing three translocation vectors produced excellent agreement with the observed native Patterson map, but led to a further paradox. Such a model would no longer obey *P*6 (or *P*2) symmetry, even when averaging over the possible translocations (Fig. 3). Including only three translocations breaks the crystal symmetry down to *P*3. The crystal twofold symmetry about z is entirely lost, yet the presence of that symmetry was unequivocal in the intensity data. Also, twinning had been ruled out as a source of extra symmetry. A resolution of this final paradox came to light when we realised that the proposed arrangement of molecules is an exceptionally rare situation in which the intensity data and the Patterson map have higher rotational symmetry than is present in the space-group symmetry of the crystal. In the present case, twofold (and hence sixfold) symmetry is present in the calculated intensities and the Patterson but is absent from the crystal symmetry. This was confirmed by calculations as well as analytically. The appendix explains how this effect, which has been referred to as symmetry enhancement (Sadanaga & Takeda, 1968), can occur through an unexpected interplay of the internal symmetry of the hexagonal molecular unit with the (mirror) symmetry of the vectors describing the translational shifts between molecules. We note also that the phenomenon can occur in more general settings regardless of the presence of translocation disorder.

3.3. Correcting for the effects of translocation disorder

Wang, Kamtekar *et al.* (2005) have presented equations that relate the observed intensities to the true crystallographic intensities one would expect from an ideal crystal not exhibiting the translocation disorder. The former quantities are

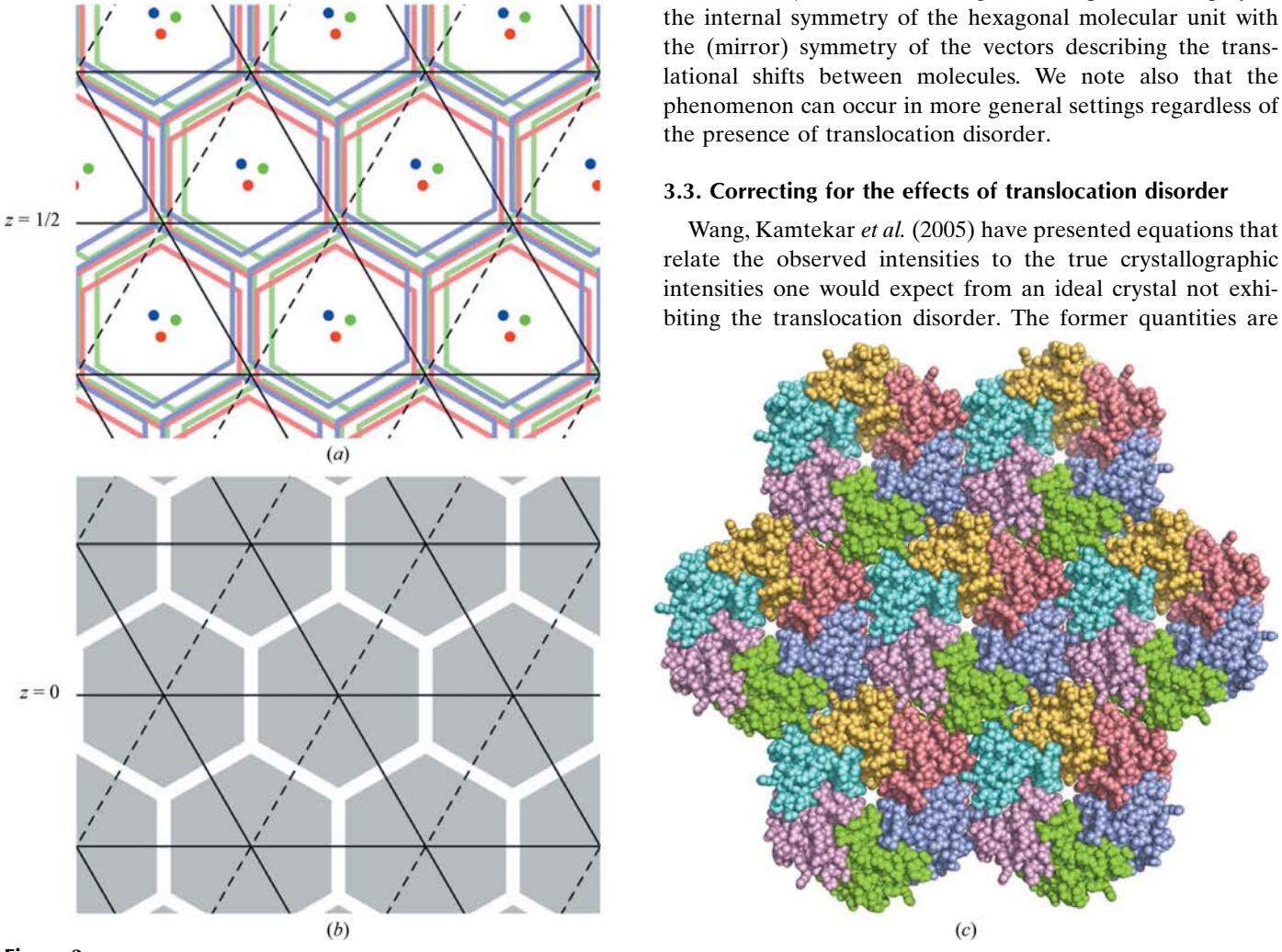

Figure 3
A diagram of CsoS1C hexamers. (*a*, *b*) A diagram of the lattice-translocation model for the CsoS1C crystal. Each hexagon represents a cyclic three-dimensional protein hexamer. The unit-cell boundaries are drawn in black lines. The disordered layers of hexamers at $z = 1/2$ are drawn in (*a*). The three alternate positions (having 1/3 occupancy) are shown in distinct colors. The hexamers are shown in outline to not obscure the alternate positions. The centers of the different hexamers are indicated by dots of the corresponding color. The ordered layer of hexamers at $z = 0$ is drawn in (*b*). Note that there is no axis of twofold symmetry passing through both layers. (*c*) A space-filling diagram of the packing of CsoS1C hexamers in the **ab** crystal plane.

the product of the latter quantities and the intensities one would obtain from the set of discrete points representing the translocation vectors, appropriately weighted. That is, $I_{obs} = I_{true} \cdot I_{trans}$, where I_{obs} are the observed intensities, I_{true} are the crystallographic intensities without disorder and I_{trans} is the squared magnitude of the Fourier transform of the translocation vectors (see Appendix *A*).

Under the treatment provided here, the translocation vectors comprise the origin vector with unit occupancy and three vectors with 1/3 occupancy each at (0.609, 0.256, 0.500), (0.646, 0.390, 0.500) and (0.745, 0.354, 0.500), positions related to each other by *P*3 symmetry. Intensity data were corrected by dividing the observed intensities by correction factors calculated as a transform of the translocation vectors. A potential problem arises when correcting intensities by dividing by a correction factor that may take on small values. In order to mitigate this problem, 5% of the reflections with the smallest correction factors were discarded.

The validity of the correction above, and the model on which it was based, was evaluated in two ways. Firstly, we examined a Patterson map calculated from the corrected intensities. The major packing peak in the native Patterson map in the $w = 1/2$ section was reduced to approximately 10% of its original height. The peak in the $w = 0$ section was reduced to approximately 25% of its original value (Fig. 5). Considerable effort was made to lower these residual peaks further by adjusting the occupancies of the hexamers in the different layers or by introducing hexamers in additional positions, but the packing peaks could not be lowered substantially. Secondly, the validity of the translocation model was evaluated by considering whether the correction function was consistent with the distribution of observed intensities. That was tested by dividing the reflections into bins based only on the magnitude of the correction term. Since reflections were binned without information about their true crystallographic intensities, the average value of the true crystallographic intensities should be essentially the same across different bins of reflections. Consequently, the average value of the observed intensities in each bin should correlate closely with the (nearly constant) value of the correction term for that bin. The linear correlation between correction factor and observed intensities is shown in Fig. 6. Overall the agreement is excellent, although a spike in observed intensities was noted at the upper limit owing to a small number of exceptionally strong reflections, including some of the low-resolution reflections of the form (0, 0, 2*n*).

The corrected intensities were used for atomic refinement. Because the data (reduced in *P*6) and the correction factors

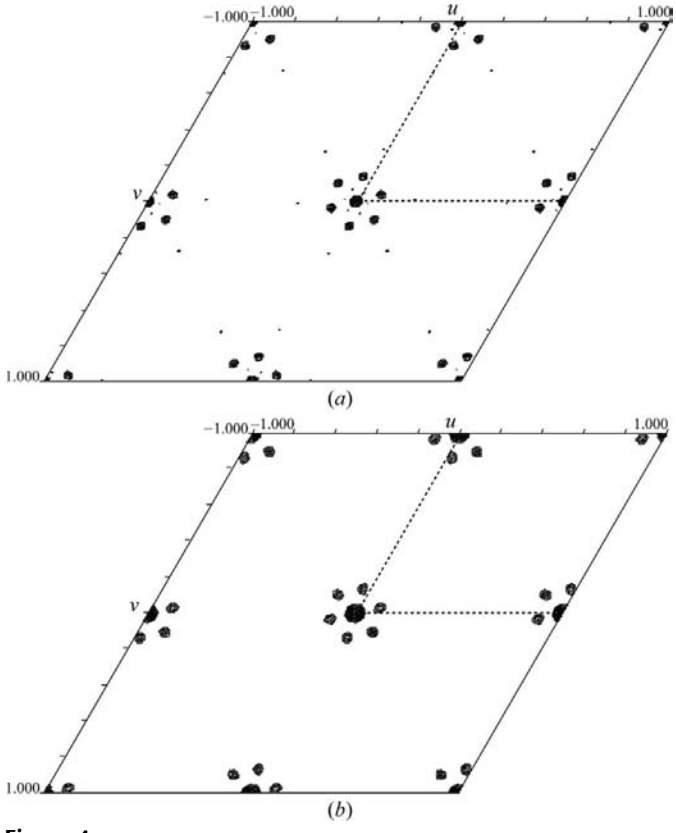

Figure 4
Agreement between Patterson maps based on observed data and on the translocation model. Sections at $w = 0$ are contoured starting at 5σ. Maps are plotted from -1 to 1 in u and v to help to illustrate the sixfold symmetry. Dashed lines denote one unit cell. (*a*) A Patterson map from observed intensities. (*b*) A Patterson map based on intensities calculated from the model with one hexamer at $z = 0$ and three partially occupied hexamers at $z = 1/2$. A similar level of agreement between observed and calculated Patterson maps is observed in the $w = 1/2$ section (not shown).

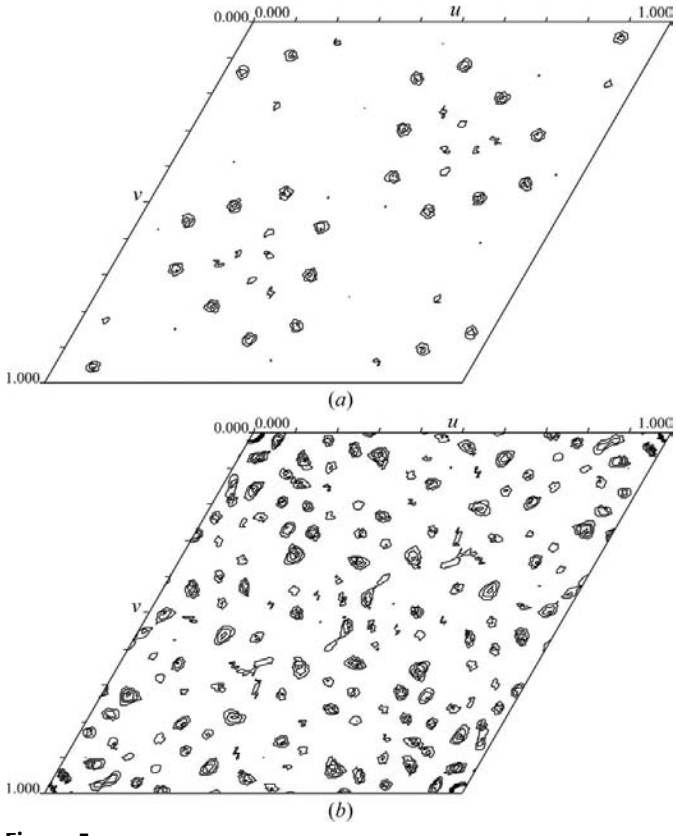

Figure 5
A Patterson map for CsoS1C based on intensities corrected for translocation. Maps are contoured starting at 2σ. (*a*) $w = 1/2$ section. Note that the prominent peaks present in the uncorrected Patterson (see Fig. 2*a*) are less than 3σ. (*b*) $w = 0$ section. Note that the prominent peak present in the uncorrected Patterson (see Fig. 2*b*) has diminished to 3σ.

both obey *P*6 symmetry, so do the corrected intensities. Therefore, refinement could be carried out with one protein chain in the asymmetric unit in *P*6. The final model gave values of 0.255 for R_{work} and 0.311 for R_{free} (Table 1), which were judged to be acceptable given the complications produced by the disorder. The final model contained amino-acid residues 4–98 of a total of 98 amino acids. The difference map contained density in the pore region of the hexamer that could not be interpreted owing to its proximity to the local sixfold symmetry axis. Few differences were observed between the refined model and the original molecular search model; so, to validate the model a core region of the structure was deleted to calculate an omit map. Residues 47–49 were omitted and the subsequent map was examined (Fig. 7). The difference map verified the location and conformation of the deleted residues.

As discussed in Zhu *et al.* (2008), an alternate approach to refinement in the presence of translocation disorder is to avoid applying a correction term to the observed intensities and instead take the disorder into account as part of the atomic model. This can be performed by placing all the molecules in their translocated (typically overlapping) positions in the unit cell. This approach avoids the problems associated with dividing by correction factors and also allows the refinement of potential structural differences between multiple molecules, although at the expense of additional refinable parameters. For the present work, we performed this type of refinement in order to see whether a final improvement could be obtained by relaxing the *P*6 symmetry, which was imperfect as discussed earlier. Here, we refined one hexamer at $z = 0$ and three overlapping hexamers at $z = 1/2$. To avoid an unacceptable increase in the number of parameters, noncrystallographic symmetry restraints were applied to protein chains within the same hexamer. The same R_{free} flags as those used under *P*6 were retained throughout the analysis. Visualization of electron-density maps was complicated by the multiple overlapping models. As a result, it was not possible to model water molecules or to make sensible manual adjustments to the structure. Variations in the occupancies of the three hexamers at $z = 1/2$ were examined, but equal occupancies of 1/3 resulted in the best R_{work} and R_{free} values (0.241 and 0.285, respectively). The statistics for the model refined in this way were slightly better than the statistics for the single model refined against the corrected intensity data, although the atomic coordinates differed very little. Coordinates for a model consisting of one CsoS1C monomer refined using corrected data were deposited, along with structure factors from uncorrected and corrected data processed in *P*6 to 2.5 Å.

As a final check for the importance of accounting for the translocation disorder, attempts were made to refine atomic models without taking translocation into account, for example by refining against uncorrected intensities without incorporating multiple overlapping molecules in the model. These tests led to *R* values that were some 20% higher than when the translocation was taken into account.

The final structure of CsoS1C revealed only minor differences compared with the highly similar structure of CsoS1A. Water molecules (in the case of refinement of a single molecule against corrected intensities) were modeled in the pore region of the hexamer. However, some density in the center of the pore could not be modeled well. Based on the crystal-

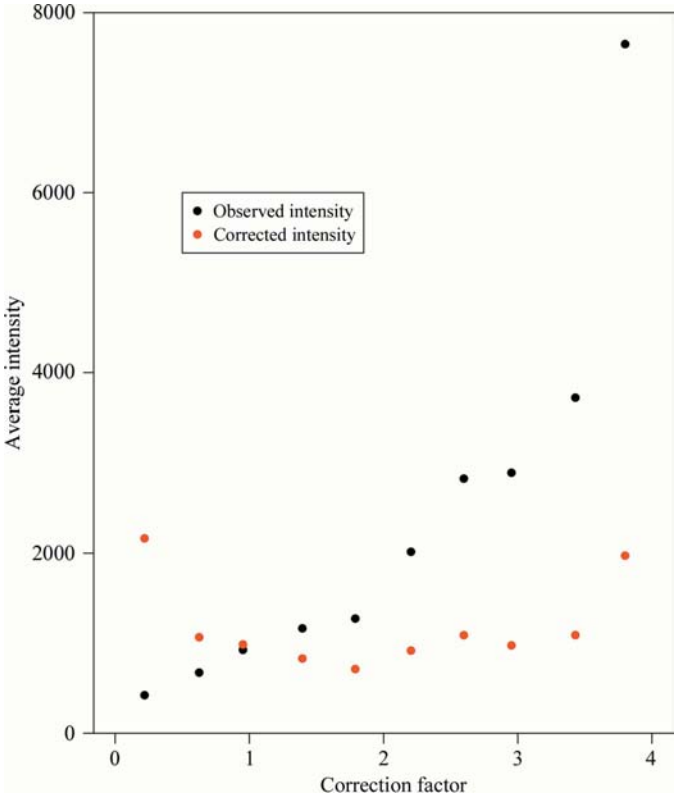

Figure 6
Graph of intensity *versus* translocation correction factor. Observed and corrected intensities were divided into ten bins based on the value of the calculated correction factor. The results (black circles) confirm that reflections requiring larger correction factors have systematically larger observed intensities, as expected. After correction (red circles), the distribution of intensities becomes much more uniform. Note that reflections with small (denominator) corrections, corresponding to the outlying red data point on the left, were not included in the set of corrected intensities used for atomic refinement (see text).

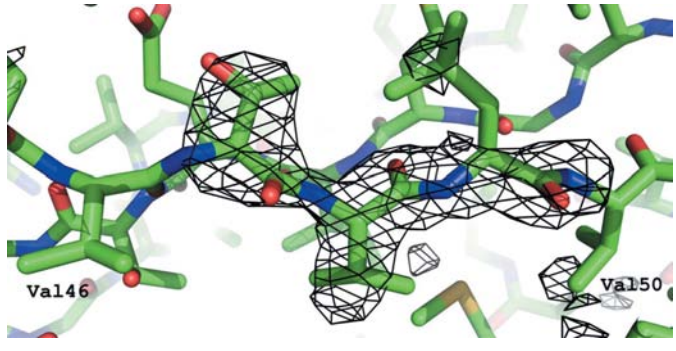

Figure 7
A difference map of deleted residues ($F_{obs} - F_{calc}$, φ_{OMIT}) contoured at 3σ. Residues 47–49 were omitted from the model. Here, the F_{obs} are the square root of intensities corrected for the translocation effect. The map around the original model indicates proper fitting of the model to the observed diffraction data and supports the interpretation of the translocation disorder.

lization conditions, it is possible that chloride or citrate ions could be present in the pore. The model of CsoS1C contains only one amino-acid-residue difference relative to CsoS1A. Residue 97 is a glutamine in CsoS1C and a glutamate in CsoS1A. This amino-acid difference did not produce any significant structural differences in the C-terminal region of the protein. The root-mean-square deviation between the CsoS1C and CsoS1A monomers was 0.29 Å. The minimum pore radius of CsoS1C was similar to the pore radius of CsoS1A (2.0 Å). The shape complementarity of the interface between CsoS1C hexamers was 0.62, similar to the shape complementarity of the analogous CsoS1A interface (Tsai *et al.*, 2007), which is in the range of antibody–antigen interfaces.

4. Discussion and conclusions

The present work illustrates a type of translocation disorder that can occur with layered protein structures. Cases of disorder such as that reported here are likely to occur when there is rotational symmetry in a molecular layer and when the energetics are not favorable for placing adjacent layers of molecules directly on top of each other. When a subsequent layer of molecules prefers to be shifted relative to the previous layer by some displacement that is not crystallographic in nature and the layer has rotational symmetry, then there could be a strong tendency for shifts to occur in multiple different directions between layers, leading to complex types of lattice translocation.

Our experience illustrates the complexities involved in dissecting the true nature of the underlying translocations. The problem of determining the correct set of translocation vectors is highly parallel to the problem of solving the structure of a small molecule using Patterson methods (Harker, 1936; Nordman & Nakatsu, 1963; Patterson, 1934). Given the complexities, it is important to examine the translocation model in various ways to test agreement with observations. Patterson maps based on the translocation model should produce the peaks observed in the native Patterson maps and the magnitudes of the correction terms calculated from the translocation model should correlate with the observed intensities. If the translocation model is correct, then atomic refinements can be carried out either by correcting the observed intensity data (taking care to exclude reflections that would have to be divided by small correction terms) or by including overlapping molecules at partial occupancies in the refinement model. Correcting intensities is useful for calculating and interpreting electron-density maps (which contain overlapping molecules in the absence of correction), while including all translocated copies of the molecule can offer advantages in refinement, for example if slight differences in orientation or configuration cause translocated molecules to not be exactly related by pure translation.

The present structure also illustrates the peculiar phenomenon of symmetry enhancement, in which the symmetry of the intensity data exceeds that expected based on the space-group symmetry of the crystal. While this situation may remain exceptionally rare in macromolecular studies, it emphasizes that the correct structure can have lower symmetry than indicated by the observed intensity data (Sadanaga & Takeda, 1968; Sadanaga *et al.*, 1973).

Finally, while our interpretation of the disorder in the present structure is consistent with the observed data, it might not be unique. Other models of disorder, perhaps even more complex, are not ruled out by our treatment. The strong streaking of the diffraction spots along **c*** suggests that other models could be considered in which the unit-cell repeat along **c** would be longer and span more layers of molecules than the present model. More complex models might allow the propagation of shifts in multiple possible directions between successive layers, rather than the simple alternating model presented here in which disordered layers are preceded and followed by an ordered unshifted layer. In the end, we judged such structural models, which would be considerably more elaborate than the one described here, to be unwarranted.

APPENDIX *A*
Symmetry enhancement wherein the Patterson function acquires higher rotational symmetry than the crystal space group

Consider a situation in which the contents of the unit cell constitute one cyclic hexamer at $z = 0$ and three identical hexamers at $z = 1/2$, all with their local axes of sixfold symmetry oriented along the z direction. The three hexamers at $z = 1/2$ are shifted in x and y relative to those at $z = 0$ by vectors that are related to each other by 120° differences, so that the constellation of vectors describing the positions of all the hexamers obeys threefold rotational symmetry (Fig. 3). Although $P3$ symmetry is preserved, such a crystal does not obey $P2$ or $P6$ symmetry. Surprisingly, however, owing to a peculiarity in the way the symmetry of the hexamer combines with the symmetry of the translational shifts, the diffraction intensities (and the Patterson map) exhibit twofold and sixfold rotational symmetry, though no such relationship is obeyed by the electron density of the crystal or by the complex structure factors it produces.

Let $\mathbf{F}_{\rm hex}$ represent the set of complex-valued structure factors arising from the single hexamer positioned at the origin. Let $\mathbf{F}_{\rm trans}$ represent the set of complex-valued structure factors arising from a Fourier transform of the constellation of points describing the positions of all the hexamers (weighted according to occupancies). The complete contents of the unit cell are described by the single hexamer convoluted with the constellation of translation vectors. As noted above, this crystal arrangement has $P3$ symmetry, but not $P2$ or $P6$ symmetry (Fig. 3). According to the convolution theorem, the structure factors for the whole crystal are given by the product of the separate transforms: $\mathbf{F}_{\rm xtal} = \mathbf{F}_{\rm hex}\cdot\mathbf{F}_{\rm trans}$. When a function is obtained by (point by point) multiplication of two others, its symmetry is the intersection of the symmetry of the two individual functions, meaning that the product function only obeys symmetry that is obeyed by both individual functions. While $\mathbf{F}_{\rm hex}$ obeys twofold and sixfold symmetry (*e.g.* relating

h, k, l and $-h, -k, l$ reflections), $\mathbf{F}_{\rm trans}$ does not and so neither do the complex-valued structure factors from the entire crystal ($\mathbf{F}_{\rm xtal}$); only the $P3$ symmetry obeyed by both separate functions is preserved in $\mathbf{F}_{\rm xtal}$. The absence of twofold symmetry in $\mathbf{F}_{\rm xtal}$ is consistent with the absence of twofold symmetry in the crystal (Fig. 3). It is understood that all the sets of structure factors obey Friedel's law, but by itself this does not contribute to the rotational symmetry under consideration here.

The situation is different when intensities are considered. Now, taking the intensity to be the product of the complex-valued structure factor with its complex conjugate: $I_{\rm xtal} = \mathbf{F}_{\rm xtal}\cdot\mathbf{F}^*_{\rm xtal} = (\mathbf{F}_{\rm hex}\cdot\mathbf{F}_{\rm trans})\cdot(\mathbf{F}_{\rm hex}\cdot\mathbf{F}_{\rm trans})^* = \mathbf{F}_{\rm hex}\cdot\mathbf{F}^*_{\rm hex}\cdot\mathbf{F}_{\rm trans}\cdot\mathbf{F}^*_{\rm trans}$. Therefore, $I_{\rm xtal} = I_{\rm hex}\cdot I_{\rm trans}$, where $I_{\rm hex}$ represents the set of intensities that would be obtained from one hexamer, while $I_{\rm trans}$ represents the set of intensities that would be obtained from the constellation of translation vectors. Whereas the structure factors $\mathbf{F}_{\rm trans}$ do not obey twofold or sixfold rotational symmetry, the intensities $I_{\rm trans}$ do obey this symmetry. This is a consequence of the mirror symmetry obeyed by the translation vectors in the present case, combined with Friedel's law; mirrors are present at $z = 0$ and 1/2 because the translocation vectors all lie precisely on those planes. Because $I_{\rm hex}$ and $I_{\rm trans}$ both obey twofold symmetry, so does $I_{\rm xtal}$. Therefore, the structure-factor magnitudes from the crystal obey $P2$ and $P6$ symmetry, although the complex-valued structure factors (namely their phases) do not. Note that if the occupancies of the three hexamers at $z = 1/2$ are not equal, the threefold symmetry is broken in the translation vectors and in the complete crystal. However, as long as the constellation of translation vectors retains mirror symmetry, then the intensities will obey $P2$ symmetry, though the $P6$ symmetry may be broken.

Finally, we note that the unusual symmetry effect demonstrated here does not rely on lattice translocation. In the structure discussed in this paper, the hexamers at $z = 1/2$ are each only partially (1/3) occupied, as they represent alternate positions arising from a lattice-translocation disorder. However, the same symmetry effect can arise just as well from a situation where there is no disorder and no partial occupancies. All that is required is a constellation of cyclic symmetric hexamers arranged according to translation vectors that obey $3/m$ symmetry (also denoted as $\bar{6}$). In such a case the intensities and the Patterson map will obey $6/m$ symmetry, though the crystal lacks sixfold rotational symmetry. Or, if the translation vectors obey simple mirror symmetry m, the intensities and Patterson map will obey $2/m$ symmetry, though the crystal may lack twofold rotation symmetry. The phenomenon observed here is similar in key respects to a structure of silicon carbide, type 10H, reported by Ramsdell & Kohn (1951), in which a structure interpreted in space group $P3m$ gave rise to intensities and a Patterson with $6/mmm$ symmetry instead of the expected $3/mm$. This rare phenomenon has been termed symmetry enhancement. As far as we are aware, this is the first instance of a macromolecular crystal (excluding cases of twinning) in which the intensities have a higher symmetry than expected based on the crystal space group.

We thank Gordon Cannon and Sabine Heinhorst at the University of Southern Mississippi for the vector containing CsoS1C. We thank Jason Navarro and Duilio Cascio at UCLA and the staff at APS Beamline 24-ID-C for their help. We thank Garib Murshudov for making available the version of *REFMAC*5 capable of refining overlapping molecules. This work was supported by NSF Grant MCB-0843065.

References

Badger, M. R. & Price, G. D. (2003). *J. Exp. Bot.* **54**, 609–622.
Bragg, W. L. & Howells, E. R. (1954). *Acta Cryst.* **7**, 409–411.
Cheng, S., Liu, Y., Crowley, C. S., Yeates, T. O. & Bobik, T. A. (2008). *Bioessays*, **30**, 1084–1095.
Crowley, C. S., Sawaya, M. R., Bobik, T. A. & Yeates, T. O. (2008). *Structure*, **16**, 1324–1332.
Dornberger-Schiff, K. (1956). *Acta Cryst.* **9**, 593–601.
Dornberger-Schiff, K. & Grell-Niemann, H. (1961). *Acta Cryst.* **14**, 167–177.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Harker, D. (1936). *J. Chem. Phys.* **4**, 381–390.
Helliwell, J. R. (2008). *Crystallogr. Rev.* **14**, 189–250.
Hwang, W. C., Lin, Y., Santelli, E., Sui, J., Jaroszewski, L., Stec, B., Farzan, M., Marasco, W. A. & Liddington, R. C. (2006). *J. Biol. Chem.* **281**, 34610–34616.
Kamtekar, S., Berman, A. J., Wang, J., Lazaro, J. M., de Vega, M., Blanco, L., Salas, M. & Steitz, T. A. (2004). *Mol. Cell*, **16**, 609–618.
Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D. & Waugh, D. S. (2001). *Protein Eng.* **14**, 993–1000.
Kerfeld, C. A., Sawaya, M. R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M. & Yeates, T. O. (2005). *Science*, **309**, 936–938.
Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S.,Cannon, G. C. & Kerfeld, C. A. (2009). *J. Mol. Biol.* doi:10.1016/j.jmb.2009.03.056.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nordman, C. E. & Nakatsu, K. (1963). *J. Am. Chem. Soc.* **85**, 353–354.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.
Parsons, S. (2003). *Acta Cryst.* D**59**, 1995–2003.
Patterson, A. L. (1934). *Phys. Rev.* **46**, 372–376.
Ramsdell, L. S. & Kohn, J. A. (1951). *Acta Cryst.* **4**, 111–113.
Sadanaga, R., Ohsumi, K. & Matsumoto, T. (1973). *Proc. Jpn. Acad.* **49**, 816–819.
Sadanaga, R. & Takeda, H. (1968). *Acta Cryst.* B**24**, 144–149.
Sagermann, M., Ohtaki, A. & Nikolakakis, K. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 8883–8887.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C. & Yeates, T. O. (2008). *Science*, **319**, 1083–1086.
Trame, C. B. & McKay, D. B. (2001). *Acta Cryst.* D**57**, 1079–1090.
Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A. & Yeates, T. O. (2007). *PLoS Biol.* **5**, e144.
Wang, J., Kamtekar, S., Berman, A. J. & Steitz, T. A. (2005). *Acta Cryst.* D**61**, 67–74.
Wang, J., Rho, S.-H., Park, H. H. & Eom, S. H. (2005). *Acta Cryst.* D**61**, 932–941.
Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. & Shively, J. M. (2008). *Nature Rev. Microbiol.* **6**, 681–691.
Yeates, T. O., Tsai, Y., Tanaka, S., Sawaya, M. R. & Kerfeld, C. A. (2007). *Biochem. Soc. Trans.* **35**, 508–511.
Zhu, X., Xu, X. & Wilson, I. A. (2008). *Acta Cryst.* D**64**, 843–850.