A challenging interpretation of a hexagonally layered protein structure

Michael C. Thompson and Todd O. Yeates*

Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA 90095, USA

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

The carboxysome is a giant protein complex that acts as a metabolic organelle in cyanobacteria and some chemoautotrophs. Its outer structure is formed by the assembly of thousands of copies of hexameric shell protein subunits into a molecular layer. The structure determination of a CcmK1 shell protein mutant (L11K) from the β-carboxysome of the cyanobacterium *Synechocystis* PCC6803 led to challenges in structure determination. Twinning, noncrystallographic symmetry and packing of hexameric units in a special arrangement led to initial difficulties in space-group assignment. The correct space group was clarified after initial model refinement revealed additional symmetry. This study provides an instructive example in which broken symmetry requires a new choice of unit-cell origin in order to identify the highest symmetry space group. An additional observation related to the packing arrangement of molecules in this crystal suggests that these hexameric shell proteins might have lower internal symmetry than previously believed.

1. Introduction

Bacterial microcompartments (MCPs) are giant protein complexes that function as metabolic organelles in many bacteria (Kerfeld *et al.*, 2010; Cheng *et al.*, 2008; Yeates *et al.*, 2008). Their exterior protein shells are reminiscent of large viral capsids, as thousands of copies of paralogous shell proteins assemble into a polyhedral molecular layer (Tanaka *et al.*, 2008; Yeates *et al.*, 2010, 2011). Within the MCP interior is a series of sequentially acting metabolic enzymes, which are able to function more efficiently as a result of their colocalization within a contained microenvironment. The β-carboxysome is a well studied MCP involved in the carbon-fixation reactions of the cyanobacterial Calvin cycle (Badger & Price, 2003; Fig. 1). This MCP sequesters the enzymes carbonic anhydrase and RuBisCO, thereby increasing the local concentration of carbon dioxide in the vicinity of RuBisCO, an enzyme with notoriously poor catalytic efficiency and substrate selectivity.

The role of the MCP in cellular metabolism requires its proteinaceous shell to act as a semi-permeable barrier, allowing the passage of substrates, products and cofactors while restricting the efflux of metabolic intermediates. The MCP shell is able to simultaneously achieve these two contradictory functions as a result of its unique architecture (Kerfeld *et al.*, 2005; Tanaka *et al.*, 2008; Yeates *et al.*, 2010, 2011). Shell proteins belonging to the BMC family (Pfam family PF00936; Punta *et al.*, 2011) first assemble into cyclic homohexamers, which bear small pores at their centers. These hexameric units further assemble into two-dimensional hexagonal layers (Dryden *et al.*, 2009), which form the flat facets of the polyhedral MCP shell. The result of this assembly process is a tightly packed molecular sheet, which is perforated only by small pores at the centers of the hexamers (Fig. 1). The specific chemical properties of these pores (*i.e.* shape, size, electrostatic potential *etc.*) govern their permeability to various small molecules.

Consistent with the form of their natural biological assemblies, BMC shell proteins often crystallize in hexagonally packed layers (Tsai *et al.*, 2007; Tanaka *et al.*, 2009; Crowley *et al.*, 2010; Samborska & Kimber, 2012). Layered structures are generally prone to various crystal-growth pathologies (Zwart *et al.*, 2008), and this has proven to be especially true of BMC shell proteins, where twinning and lattice-

translocation disorders have often been observed (Tanaka *et al.*, 2008; Tsai *et al.*, 2009). The present study involved another such case in which a specific shift of hexagonal layers in a crystal led to a challenging space-group interpretation. This case demonstrates that when broken crystallographic symmetry leads to doubling of a unit-cell axis, it is sometimes necessary to consider an alternate unit-cell origin in order to identify the highest symmetry space group. Additionally, the alternating conformations of individual shell protein monomers in the crystal form used for this study provide potential connections to biological function.

2. Materials and methods

2.1. Cloning, overexpression and purification of recombinant protein

The wild-type DNA sequence for residues 1–91 of the CcmK1 protein from the cyanobacterium *Synechococystis* sp. PCC6803 was cloned into the pET-22b expression vector incorporating a C-terminal hexahistidine tag consisting of amino acids –LeuGluHis$_6$. Specific details of the cloning protocol have been described previously (Tanaka *et al.*, 2009). The L11K mutation was introduced by site-directed mutagenesis using the QuikChange method (Stratagene).
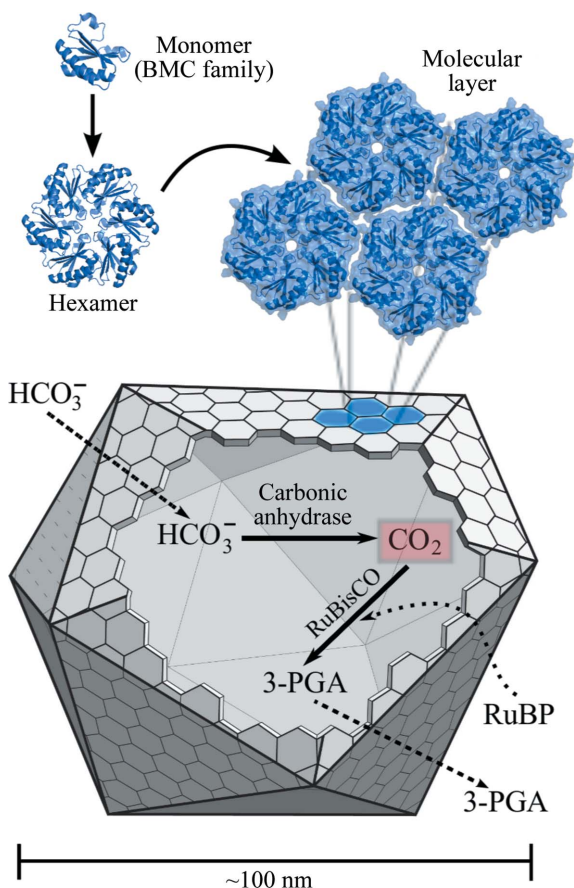

Figure 1
A model for bacterial microcompartment (MCP) structure and function in the carboxysome. The illustration highlights the assembly of shell protein monomers into hexamers, as well as the edgewise association of hexamers to form the tightly packed facets of the polyhedral MCP shell. Within the shell, a reaction scheme depicts the metabolic events that occur in the lumen of the carboxysome. The critical intermediate, CO_2, is highlighted in red. In the reaction scheme, solid black lines depict enzymatic reactions, while dashed lines indicate transport events. (RuBP, ribulose 1,6-bisphosphate; 3-PGA, 3-phosphoglyceric acid; RuBisCO, ribulose-1,6-bisphosphate carboxylase/oxygenase.)

Table 1
Diffraction data and refinement statistics.

Values in parentheses are for the highest resolution shell.

X-ray wavelength (Å)	1.54
Resolution range (Å)	41.24–1.60 (1.64–1.60)
Unit-cell parameters (Å, °)	$a = b = 70.0$, $c = 56.2$, $\alpha = \beta = 90$, $\gamma = 120$
Space group	$P6_3$
Total reflections	353118 (7262)
Unique reflections	20387 (1293)
Multiplicity	17.3 (5.6)
Completeness (%)	98.0 (84.6)
$\langle I/\sigma(I)\rangle$	34.07 (5.55)
$R_{r.i.m.}$† (%)	5.7 (26.8)
$CC_{1/2}$	100.0 (96.4)
Wilson B factor (Å^2)	25.0
R_{work}‡ (%)	18.0
R_{free}‡ (%)	19.5
No. of atoms	1441
Protein residues	182
Water molecules	51
Average B factor (Å^2)	
Protein	19.7
Solvent	26.5
R.m.s.d., bonds (Å)	0.012
R.m.s.d., angles (°)	1.32
Ramachandran plot§ (%)	
Favored	97.8
Allowed	2.2
Outliers	0.0
MolProbity¶ clashscore	4.66

† Diederichs & Karplus (1997). ‡ R_{work} and R_{free} are given by the following equation, computed for the working and test sets of reflections, respectively: $R = \sum_{hkl} ||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$. The values of F_c used in the calculations include scattering contributions from the riding H atoms. § Lovell *et al.* (2003). ¶ Chen *et al.* (2010).

The sequence of the mutated plasmid was verified by dideoxy chain-termination sequencing.

We expressed recombinant protein using transformed *Escherichia coli* BL21 (DE3) Rosetta cells (Novagen). During the exponential phase of cell growth in selective Luria–Bertani (LB) broth, 1 m*M* isopropyl β-D-1-thiogalactopyranoside was added to induce protein expression for 4 h at 310 K. Cells were pelleted by centrifugation for 15 min at 5000*g* and were then lysed by sonication. The lysis buffer consisted of 20 m*M* Tris buffer, 300 m*M* sodium chloride at pH 8.0 with a protease-inhibitor additive (Sigma–Aldrich), 10 m*M* MgCl$_2$, 1 mg ml^{-1} lysozyme and 100 units ml^{-1} of both DNase and RNase. We clarified the cell lysate by centrifugation at 30 000*g* for 30 min and then used a HisTrap nickel-affinity column (GE Healthcare) to purify the protein from the clarified lysate. The bound protein was eluted with lysis buffer containing 300 m*M* imidazole and was then dialyzed against a buffer consisting of 20 m*M* Tris, 100 mM sodium chloride at pH 8.0. This single purification step resulted in a highly pure protein sample, as demonstrated by SDS–PAGE.

2.2. Protein crystallization

Following purification, the protein was concentrated to approximately 10 mg ml^{-1}. We crystallized the CcmK1 L11K mutant by hanging-drop vapor diffusion in mother liquor consisting of 1.26 *M* ammonium sulfate, 0.15 *M* sodium chloride buffered at pH 9.5 with 0.1 *M* *N*-cyclohexyl-2-aminoethanesulfonic acid. Crystallization drops were prepared by mixing a 1:1 ratio of the concentrated protein solution and the mother liquor in a total volume of 1 µl using a Mosquito pipetting robot (TTP LabTech). The drops were sealed above 100 µl reservoirs and allowed to equilibrate at 296 K. Our initial crystals diffracted poorly and we found that the addition of 10 m*M* guanidinium chloride to the mother liquor prior to mixing it

with the protein drop produced a large, high-quality crystal that diffracted well.

2.3. X-ray data collection and processing

We collected single-crystal X-ray diffraction data using a Rigaku FR-E+ rotating-anode X-ray source operating at the Cu $K\alpha$ wavelength (1.54 Å) equipped with VariMax HF optics and an R-AXIS HTC detector. Crystals were harvested and cooled directly in the liquid-nitrogen cryostream without additional cryoprotection and were subsequently maintained at cryogenic temperature (100 K) throughout the course of the data collection. We indexed and integrated the reflection data to 1.6 Å resolution using *XDS*, performed scaling with *XSCALE* and converted the intensities to structure-factor amplitudes using *XDSCONV* (Kabsch, 2010). The free set of reflections was assigned using *phenix.reflection_file_converter* (Adams *et al.*, 2010) such that reflections related by lattice symmetry would belong to the same set. Information regarding data collection and processing is presented in Table 1. We note that the data are strong out to the 1.6 Å resolution limit used in our X-ray experiment; the geometry of the detector setup precluded the collection of higher resolution data.

2.4. Analysis of symmetry and twinning

Automatic data indexing by *XDS* (as described above) revealed that the crystal lattice was primitive hexagonal. During the early stages of data processing, we evaluated different Laue symmetries and also evaluated a self-rotation function using *MOLREP* (Vagin & Teplyakov, 2010; Potterton *et al.*, 2003), which suggested that the diffraction pattern obeyed 622 rotational symmetry (Fig. 2a). We also ran the *phenix.xtriage* program (Adams *et al.*, 2010) to check for crystal pathologies and noncrystallographic symmetry (NCS). The *phenix.xtriage* analysis revealed a strong peak in the native Patterson map (Patterson, 1934), as discussed subsequently. In addition to revealing translational NCS, the *phenix.xtriage* analysis also suggested the presence of hemihedral twinning based on the results of the *L*-test (Fig. 2c; Padilla & Yeates, 2003).

2.5. Phasing by molecular replacement

In order to calculate initial phases, we used the method of molecular replacement (MR) with the structure of the CcmK2 BMC shell protein (PDB entry 3dnc; Tanaka *et al.*, 2009) as a search model. The program *Phaser* (McCoy *et al.*, 2007) found a solution containing four monomers in a *P*3 asymmetric unit (two hexamers per unit cell). This solution was used as a starting point for model building and refinement of our CcmK1 L11K structure.

2.6. Model building and refinement

As an initial step towards generating our final model, we subjected the MR solution to simulated-annealing torsion-angle refinement with *phenix.refine* (Adams *et al.*, 2010; Afonine *et al.*, 2012) in order to minimize potential model phase bias resulting from the MR procedure. We began refining our model in space group *P*3. However, during the course of the refinement process we noticed that our model was also consistent with the higher-symmetry space group $P6_3$ under a different choice of origin, as discussed subsequently. At this point we selected two of the four chains from our structure in *P*3, which corresponded to a $P6_3$ asymmetric unit under the appropriate choice of origin. For subsequent model refinement we merged our diffraction data to $P6_3$ using *XSCALE*, taking care to conserve the R_{free} flags from our original *P*3 data. Our early stages of refinement produced a relatively high-quality model, although some regions of the resulting electron-density map appeared fragmented or featureless and were not consistent with the model in these areas. In order to resolve these problematic regions of the model, we generated an electron-density map in which questionable regions of the model were not included in the phase calculation. Using this map, we were able to manually fix incorrectly built regions of the model with *Coot* (Emsley *et al.*, 2010). Finally, we refined the model to convergence against merged structure-factor amplitudes with automatic weight optimization and a TLS model for atomic displacement parameters [ten groups per chain, as determined by a combination of automated *TLSMD* analysis (Painter & Merritt, 2006a,b) and visual analysis of secondary-structural elements], as well as twin refinement with
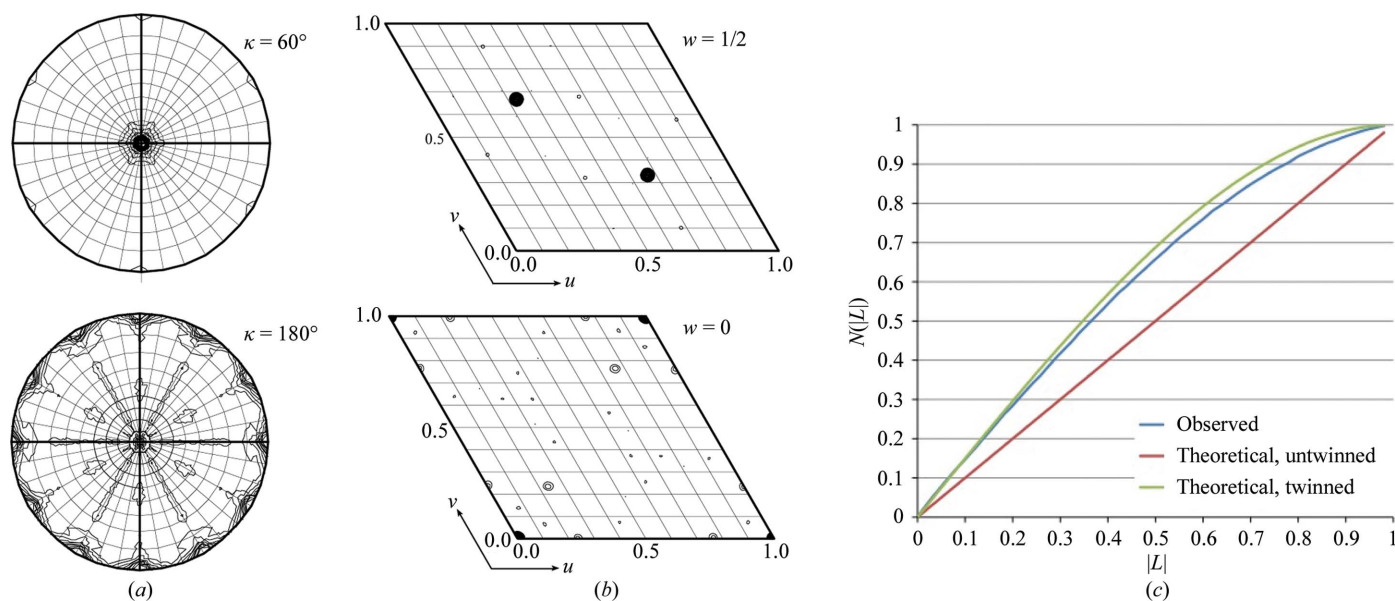

Figure 2
Analysis of symmetry and twinning. (a) Sections of the self-rotation function ($\kappa = 180°$ and $\kappa = 60°$) indicate that the underlying point-group symmetry of the crystal is 622. (b) Sections of a native Patterson map ($w = 0$ and $w = 1/2$) calculated from observed intensities show a prominent (47.5% of the origin) packing peak at $\langle uvw \rangle = (1/3, 2/3, 1/2)$. (c) Intensity statistics showing that the CcmK1 L11K crystal specimen is hemihedrally twinned. $N(|L|)$ is the cumulative distribution of $|L|$ (Padilla & Yeates, 2003).

algebraic detwinning, twin operator $(k, h, -l)$ and a twin fraction of 0.4. H atoms were added to the model using *phenix.reduce* (Word *et al.*, 1999) and were included in riding positions throughout refinement for the purpose of including their scattering contributions, but their positions were not independently refined against the X-ray data or included in the deposited coordinates. We performed the final refinement step using *phenix.refine* within version 1.8.2-dev1334 of the *PHENIX* suite (Adams *et al.*, 2010; Afonine *et al.*, 2012).

Our final model includes two protein monomers in the asymmetric unit, including residues 3–94 of 99 in chain *A* and 3–92 of 99 in chain *B* as well as 51 water molecules and a single sulfate ion. Regions of the $2mF_o - DF_c$ and $mF_o - DF_c$ maps near the C-termini of the protein chains were somewhat noisy, indicating that the C-terminal hexahistidine tags might be partially ordered. However, the features were not clear enough to justify extending the model into this density. Following model refinement, we evaluated the quality of the structure and compared the two crystallographically independent protein molecules in the unit cell using the *PHENIX* graphical interface (Adams *et al.*, 2010). The atomic model has been deposited in the Protein Data Bank as PDB entry 4liw.

3. Results and discussion

The CcmK1 L11K shell protein mutant from *Synechocystis* sp. PCC6803 was expressed and purified from *E. coli* and crystallized by hanging-drop vapor diffusion (see §2). Diffraction data were collected to 1.6 Å resolution. The unit cell was determined to be hexagonal ($a = b = 70.0$, $c = 56.2$ Å) and data reduction suggested the Laue symmetry to be $P6/mmm$. Diffraction was strong in all directions, Bragg peaks were sharp and no obvious signs of disorder were evident. However, a very strong native Patterson peak (47.5% of the origin height) was noted at fractional coordinates $\langle uvw \rangle = (1/3, 2/3, 1/2)$ (Fig. 2b). Although this peak appears to be a special position, it

does not correspond to any centering operation in an alternate unit cell. Instead, such a translation interchanges the locations of the sixfold and threefold symmetry axes in *P*6.

3.1. Initial space-group identification (*P*3)

Some of the essential features of the final structure could be discerned at the outset from the geometry and symmetry of the unit cell based on prior experience with crystal structures of similar proteins from the BMC shell protein family. In many crystal structures, BMC shell proteins pack in hexagonal layers with a unit-cell spacing along **a** and **b** in the range between 65 and 70 Å (Tsai *et al.*, 2007; Tanaka *et al.*, 2009; Crowley *et al.*, 2010; Samborska & Kimber, 2012). In their natural assembly state, the hexamers pack side by side and in the same (*i.e.* upward-facing) orientation, to give an essentially solid molecular layer, except for narrow pores at the centers of the hexamers. The thickness of one layer of proteins is generally just under 30 Å at its thickest point. In the present case, the unit-cell parameter $c = 56.25$ Å dictated that two layers of hexamers would be present within one unit cell. However, a challenge arose immediately from the observed Laue symmetry and the strong translational NCS peak in the native Patterson at $w = 1/2$. The translational NCS peak required the two layers in the unit cell to be oriented likewise (*i.e.* both face up). However, this arrangement would not give the 622 rotational symmetry (apparent Laue symmetry $P6/mmm$) observed in the diffraction data. This conflict was resolved by realising that the specimen was almost perfectly twinned by hemihedry. An evaluation of the overall intensity statistics, using local differences, gave a value for $\langle |L| \rangle$ of 0.394 (Fig. 2c). The theoretical expected value of $\langle |L| \rangle$ is 0.500 for untwinned data and 0.375 for perfectly twinned data (Padilla & Yeates, 2003). This finding allowed the two layers in the unit cell to share the same molecular orientation; the additional twofold rotational symmetry in the diffraction pattern arises from hemihedral twinning.
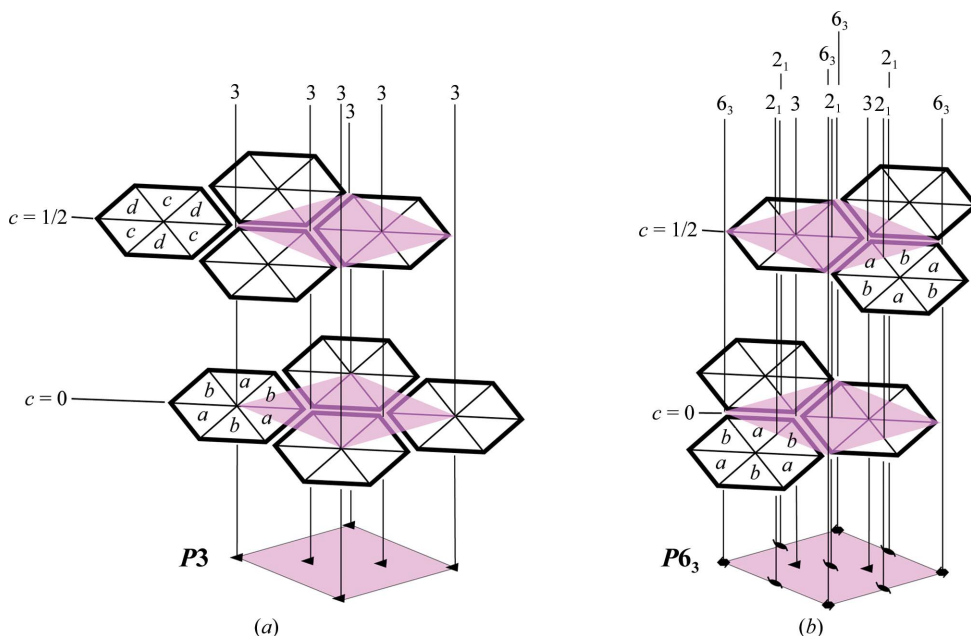
Given the two layers in the same orientation, what remained was to establish the shift between them. The relative shift was dictated by the native Patterson peak at (1/3, 2/3, 1/2). In this arrangement, the axis of local sixfold symmetry in each layer is coincident with a local axis of three-fold symmetry (where the corners of three hexamers meet) in the other layer (Fig. 3a). As a result, the sixfold symmetry axes through the centers of the hexamers in one layer are broken by the other layer, leading to a system of equally spaced threefold crystallographic axes of symmetry, as in space group *P*3. The apparent noncrystallographic translational relationship between the two layers (shown by the native Patterson peak), together with the breakage of the sixfold symmetry, led to a space-group assignment of *P*3 and to a molecular-replacement solution consistent with this symmetry. The asymmetric unit contained two copies of a third of a hexamer, one in each layer: a total of four polypeptide chains (Fig. 3a).

Figure 3
Space-group assignment and packing arrangement of the two hexameric layers in the unit cell. (a) An initial assignment of space group *P*3, with one hexameric layer having its center at the origin and the other layer having its center at (1/3, 2/3, 1/2), leading to an essentially correct structure but with incompletely assigned symmetry. The four crystallographically independent subunits based on this space-group assignment are labeled *a*–*d*. (b) The correct assignment of *P*6₃ symmetry after shifting the initial structure to a different origin. Note that the 6₃ screw axis contains within it a pure threefold axis and a 2₁ screw axis of symmetry. Only two crystallographically independent subunits (*a* and *b*) are present. Symmetry elements are illustrated with their conventional symbols.

Despite the absence of sixfold rotational symmetry in the $P3$ space group, sixfold symmetry was present in the calculated intensities as a result of the local sixfold symmetry of the hexamers and the special translational shift between them (in agreement with the sixfold symmetry of the observed intensities; Iwasaki, 1972; Sadanaga & Ohsumi, 1979). After confirming the packing analysis above by molecular replacement, using CcmK2 (PDB entry 3dnc) as the search model, we were able to refine an atomic model in space group $P3$ that had good final statistics ($R_{work} = 0.171$ and $R_{free} = 0.199$).

3.2. Identification of higher symmetry ($P6_3$) by shifting the unit-cell origin

During the refinement process, careful analysis of the refined model in space group $P3$ illuminated an element of symmetry that had been overlooked. A single layer of molecules in our crystal supports $P6$ symmetry, with the origin at the center of a hexamer. Our initial space-group assignment for the crystal ($P3$) assumed that although the $P6$ unit cell doubled along c and the sixfold symmetry was broken owing to the translational shift between layers of molecules, the origin would remain at the center of a hexamer. We failed to recognize that although the translational shift breaks a pure sixfold crystallographic axis, it allows a 6_3 axis. Consequently, our initial MR search was performed in space group $P3$ rather than in any member of the $P6_x$ family. Upon closer analysis, we realised that an alternate choice for the origin of the unit cell would allow a higher symmetry to be assigned without significant modification of the structure.

The correct space group conformed to $P6_3$ (in its standard setting) only when the origin of the unit cell was placed on a threefold axis of symmetry (contained within the 6_3 screw axis) that passes through points in both layers where three hexamers meet at a corner (Fig. 3b). The two layers are then related by a 2_1 screw axis (also contained within the 6_3 screw axis) through the origin. Under the space-group assignment of $P6_3$, the asymmetric unit contains just one-third of a hexamer, or two adjacent subunits whose slightly different conformations give the hexamers threefold, but not sixfold, symmetry.

The NCS packing peaks between the two layers (at $w = 1/2$ in the native Patterson map) arise from a combination of the 2_1 crystallographic axis with local twofold NCS axes through the centers of the hexamers (along c), which are slightly broken by subtle deviations between the two monomers in the asymmetric unit. The combination of two rotational symmetry elements to produce this translational NCS introduces another source of confusion in indentifying the presence of 6_3 and 2_1 screw axes in $P6_3$: these screw axes are typically identified by systematically absent $(0, 0, l = odd)$ reflections, which would also be absent as a result of a translational NCS operation with $z = 1/2$.

The correct space-group assignment was further confirmed retrospectively by rerunning molecular replacement in all space groups having Laue symmetry $P6/m$, where $P6_3$ was readily identified as the correct space group. Additionally, a *post hoc* analysis of the $P3$ coordinates with the *LABELIT* software (Poon *et al.*, 2010) also identified the higher symmetry $P6_3$ unit cell, further confirming our manual analysis. With this correct space group (and choice of origin), and application of the necessary twin law ($k, h, -l$), a final model was successfully refined (Table 1).

3.3. Broken crystallographic symmetry often requires a new choice of unit-cell origin

The space-group complication in the present study shares some similarities with another recently described case of symmetry breaking. While studying a new crystal form of human carbonic anhydrase II, it was found that alternating protein conformations led to a unit-cell doubling relative to a previously characterized simpler crystal form (Robbins *et al.*, 2010b). Unexpectedly, however, the $P2_1$ space-group symmetry of the simpler crystal form was also broken under the most obvious choice of origin for the doubled unit cell, leading to apparent disorder. The correct structure, which was fully ordered, was obtained in retrospect when it was found that shifting the origin of the doubled unit cell led to recovery of the required symmetry elements in $P2_1$ (Robbins *et al.*, 2010a). Likewise, in our study, the shifting of a second hexagonal layer relative to the first layer breaks the $P6$ symmetry that would have been present in a single layer, leaving what appears to be $P3$ symmetry under the original choice of unit cell (Fig. 3a). The assignment of $P3$ symmetry is not incorrect, and in our case no apparent disorder resulted from refining the structure in $P3$, but it is incomplete; higher symmetry ($P6_3$) is in fact present and is recognized under a different choice of origin (Fig. 3b). Cases such as these serve as reminders that when symmetry breaking gives rise to a new crystal form, identifying the highest possible symmetry in the new form may require a different choice of origin in order to match the correct space group in its standard setting. This requires either careful manual intervention or a reanalysis by computational methods able to automatically evaluate new space-group possibilities (Lebedev & Isupov, 2012; Poon *et al.*, 2010).

3.4. Symmetry breaking may play a role in shell protein function

The final model of CcmK1 L11K is mainly consistent with previous structures of homologous BMC shell proteins. However, the packing arrangement of hexamers in this crystal provides potential insight into the natural symmetry of these cyclic homo-oligomers. If we assume that the hexamers are sixfold symmetric, and that the hexamers in one layer prefer to not pack directly on top of hexamers from adjacent layers, then there would be six energetically equivalent translational shifts between hexamers in adjacent layers. These six translational shifts admit two distinct positions for the second layer relative to the first. The energetic equivalence of those two outcomes would lead to growth of a crystal exhibiting a lattice-translocation disorder, with hexamers occupying mutually exclusive (partially occupied) positions within one layer of molecules. This lattice-translocation disorder would produce strong packing peaks in a native Patterson map at $w = 0$, similar to a case previously observed in crystals of another BMC shell protein (Tsai *et al.*, 2009). The lack of packing peaks in the $w = 0$ section (Fig. 2b) showed that our structure does not exhibit such a pathology. This indicates that the hexamers add to the crystal in a way that breaks their sixfold symmetry. Deviations from sixfold symmetry, although subtle, were indeed observed in the refined atomic coordinates. These deviations include small differences in rotameric configurations and backbone torsion angles between the two crystallographically independent protein molecules in the unit cell. These rotamer and torsion-angle differences cause a departure from perfect β-sheet geometry in chain A versus chain B, as well as a small difference in the conformation of the C-terminal helix in each chain, which appears to be a 3_{10}-helix in chain A but not in chain B. The breakage of sixfold symmetry also brings up an element of potential biological interest. The broken symmetry implies, by exchange of alternating conformations, a certain amount of dynamic motion in this family of proteins, which may relate to their roles in molecular transport. The role of symmetry breaking in these systems is the subject of ongoing analysis.

short communications

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.

Badger, M. R. & Price, G. D. (2003). *J. Exp. Bot.* **54**, 609–622.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Cheng, S., Liu, Y., Crowley, C. S., Yeates, T. O. & Bobik, T. A. (2008). *Bioessays*, **30**, 1084–1095.

Crowley, C. S., Cascio, D., Sawaya, M. R., Kopstein, J. S., Bobik, T. A. & Yeates, T. O. (2010). *J. Biol. Chem.* **285**, 37838–37846.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Mol. Biol.* **4**, 269–275.

Dryden, K. A., Crowley, C. S., Tanaka, S., Yeates, T. O. & Yeager, M. (2009). *Protein Sci.* **18**, 2629–2635.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Iwasaki, H. (1972). *Acta Cryst.* A**28**, 253–260.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kerfeld, C. A., Heinhorst, S. & Cannon, G. C. (2010). *Annu. Rev. Microbiol.* **64**, 391–408.

Kerfeld, C. A., Sawaya, M. R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M. & Yeates, T. O. (2005). *Science*, **309**, 936–938.

Lebedev, A. A. & Isupov, M. N. (2012). *CCP4 Newsl. Protein Crystallogr.* **48**, 53–61.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.

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.

Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.

Painter, J. & Merritt, E. A. (2006a). *Acta Cryst.* D**62**, 439–450.

Painter, J. & Merritt, E. A. (2006b). *J. Appl. Cryst.* **39**, 109–111.

Patterson, A. (1934). *Phys. Rev.* **46**, 372–376.

Poon, B. K., Grosse-Kunstleve, R. W., Zwart, P. H. & Sauter, N. K. (2010). *Acta Cryst.* D**66**, 503–513.

Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.

Punta, M. *et al.* (2011). *Nucleic Acids Res.* **40**, D290–D301.

Robbins, A. H., Domsic, J. F., Agbandje-McKenna, M. & McKenna, R. (2010a). *Acta Cryst.* D**66**, 950–952.

Robbins, A. H., Domsic, J. F., Agbandje-McKenna, M. & McKenna, R. (2010b). *Acta Cryst.* D**66**, 628–634.

Sadanaga, R. & Ohsumi, K. (1979). *Acta Cryst.* A**35**, 115–122.

Samborska, B. & Kimber, M. S. (2012). *Structure*, **20**, 1353–1362.

Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C. & Yeates, T. O. (2008). *Science*, **319**, 1083–1086.

Tanaka, S., Sawaya, M. R., Phillips, M. & Yeates, T. O. (2009). *Protein Sci.* **18**, 108–120.

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.

Tsai, Y., Sawaya, M. R. & Yeates, T. O. (2009). *Acta Cryst.* D**65**, 980–988.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999). *J. Mol. Biol.* **285**, 1735–1747.

Yeates, T. O., Crowley, C. S. & Tanaka, S. (2010). *Annu. Rev. Biophys.* **39**, 185–205.

Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. & Shively, J. M. (2008). *Nature Rev. Microbiol.* **6**, 681–691.

Yeates, T. O., Thompson, M. C. & Bobik, T. A. (2011). *Curr. Opin. Struct. Biol.* **21**, 223–231.

Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 99–107.