Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural characterization of an MJ1267 ATP-binding cassette crystal with a complex pattern of twinning caused by promiscuous fiber packing

Yu-Ren Yuan, Oskana Martsinkevich and John F. Hunt*

Department of Biological Sciences, 702A Fairchild Center, MC2434, Columbia University, New York, NY 10027, USA

Correspondence e-mail: hunt@sid.bio.columbia.edu

Received 11 April 2002
Accepted 15 October 2002

PDB Reference: twinned trigonal crystal form of the MJ1267 ATP-binding cassette, 1g9x, r1g9xsf.

ATP-binding cassettes represent the motor domains in ABC transporters, a superfamily of integral membrane-protein pumps that couple the hydrolysis of ATP to transmembrane solute translocation. A crystal of a Mg-ADP complex of the MJ1267 ATP-binding cassette was obtained that produced a diffraction pattern characterized by pathological streaking of the spots in the $a^* \times b^*$ plane. While the Laue symmetry of the diffraction pattern was $P\bar{3}1m$, the crystal was determined to be twinned based on intensity statistics, molecular-replacement analysis and difference Fourier analysis of an engineered single-site methylmercury derivative. The unit cell contains three similar 3_1 fibers, with two of them related by primarily translational non-crystallographic symmetry (NCS) and the third related to the first two by approximate twofold screw operations whose rotational components are very similar to the twinning operator. The promiscuous packing of these 3_1 fibers, which make both parallel and antiparallel interactions in the primary crystal lattice, can explain the twinning tendency based on the ability of the twin-related lattices to interact with one another while making only one slightly sub-optimal intermolecular contact per unit cell in the boundary region. The promiscuous fiber packing can also explain the streaking in the diffraction pattern based on the ability to form a variety of different lattices with similar inter-fiber packing interactions. The crystal structure was refined as a twin in space group $P3_1$ using the program *CNS*, yielding a free R factor of 28.9% at 2.6 Å and a refined twin fraction of 0.50. The structure shows a rigid-body rotation of the ABC-transporter-specific α-helical subdomain (ABCα subdomain) in MJ1267 compared with the conformation observed for the same protein in a $C2$ crystal lattice; this observation suggests that the ABCα subdomain is flexibly attached to the F1-type ATP-binding core of the ATP-binding cassette when Mg-ADP is bound at the active site.

1. Introduction

ATP-binding-cassette transporters, or ABC transporters, represent the molecular architecture most commonly used by cells to couple the hydrolysis of ATP to the transmembrane transport of ligands (Higgins, 1992; Paulsen *et al.*, 1998; Davidson, 2002). Representatives of this superfamily of molecular pumps, as well as various soluble mechanoenzymes which employ the same motor domains, are found in all organisms from bacteria to man. ABC transporter components and the related soluble mechanoenzymes comprise between 1 and 5% of the open reading frames in bacterial genomes (Dassa *et al.*, 1999; Linton & Higgins, 1998; other unpublished observations), and there are approximately 130

proteins containing ATP-binding cassettes in the initial release of the human genome sequence (Klein *et al.*, 1999; other unpublished observations). ABC transporters mediate the transmembrane translocation of an exceedingly diverse array of chemical species, ranging from small organic compounds such as amino acids and chemotherapeutic drugs to entire polypeptide chains. These molecular pumps play a central role in a variety of human diseases, including cancer and leukemia, where overexpression of an ABC transporter mediates multidrug resistance in advanced tumors (Gottesman *et al.*, 1996), as well as in cystic fibrosis (Riordan *et al.*, 1989) and adrenoleukodystrophy (Mosser *et al.*, 1993), which are among the myriad of genetic diseases caused by mutations in ABC-transporter proteins.

ABC transporters have a conserved oligomeric organization, comprising a pair of characteristic ATP-binding cassettes (ABCs) plus a pair of α-helical transmembrane (TM) domains (Higgins, 1992; Gottesman *et al.*, 1996; Davidson, 2002). However, both the ABCs and the TMs can occur as either heterodimers or homodimers, with different organizations being observed in different transporters. The organization of the polypeptide chains in an intact ABC-transporter complex is also variable and can comprise anything from one chain containing all four domains to four separate chains each containing one of the domains. The ABCs share strong sequence homology in all transporters and soluble mechanoenzymes across the phylogenetic spectrum, accounting for the name of the superfamily, but the transmembrane domains generally show a relatively low level of sequence homology other than between closely related transporters with similar substrate specificity (Higgins, 1992; Paulsen *et al.*, 1998; Saurin *et al.*, 1999).

Several crystal structures of ABCs have now been determined in various nucleotide-bound forms, including the HisP cassette from *Salmonella typhimurium* (Hung *et al.*, 1998) and the MJ0796 cassette from *Methanococcus jannaschii* (Smith *et al.*, 2002) with ATP bound, the human TAP cassette (Gaudet & Wiley, 2001) and the MJ1267 (Karpowich *et al.*, 2001) and MJ0796 (Yuan *et al.*, 2001) cassettes from *M. jannaschii* with Mg-ADP bound, and the MJ1267 cassette as well as the MalK cassette from *Thermococcus litoralis* (Diederichs *et al.*, 2000) without nucleotide bound. The basic architecture of the ABC comprises three subdomains (Karpowich *et al.*, 2001): (i) a nucleotide-binding core subdomain with identical topology and strong structural homology to that in F1 ATPase (Abrahams *et al.*, 1994), (ii) an antiparallel β-sheet subdomain (ABCβ subdomain) that packs onto the surface of the Walker A (P-loop) helix (Walker *et al.*, 1982) and holds the base and ribose of the nucleotide in an unusually solvent-exposed position and (iii) an α-helical subdomain (ABCα subdomain) that forms one flank of the ABC and contains the 'LSGGQ' signature sequence that is a stereotypical feature of this superfamily of ATPases (Higgins, 1992; Yoshida & Amano, 1995). The F1-type ATP-binding core plus the ABCβ subdomain constitute a rigid nucleotide-binding structure whose conformation appears to be tightly conserved in all structures and in all nucleotide-bound states. However, the ABCα subdomain appears to be flexibly attached to the other subdomains in the absence of nucleotide (Diederichs *et al.*, 2000). Comparison of ATP-bound and ADP-bound structures of MJ0796 (Smith et al., 2002) shows that the presence of the γ-phosphate of ATP in the active site clamps the ABCα subdomain in close apposition to the active site, based on hydrogen bonding between the γ-phosphate and a phylogenetically invariant glutamine that is located in a linker connecting the ABCα subdomain to the nucleotide-binding core of the cassette. This result confirms inferences drawn from comparison of the ATP-bound structure of HisP with the Mg-ADP-bound structures of MJ0796 and MJ1267 (Karpowich *et al.*, 2001; Yuan *et al.*, 2001) and is also supported by structural studies of the remotely related soluble mechanoenzyme Rad50, which has had its structure determined in both the presence and absence of Mg-AMPPNP (Hopfner *et al.*, 2000).

In our efforts to characterize the structural mechanics of ABC transporters based on crystallographic studies of isolated ATP-binding cassettes, the first crystal that we obtained of the Mg-ADP-bound MJ1267 ATP-binding cassette (Karpowich *et al.*, 2001) yielded a diffraction pattern with pathological streaking in the $a^* \times b^*$ plane of a trigonal diffraction pattern. Furthermore, analyses of the resulting intensity statistics suggested that the dominant crystal lattice was merohedrally twinned (Yeates, 1988; Yeates & Fam, 1999).

Twinning occurs when multiple lattice domains with different molecular structures and/or orientations exist in a single morphological crystal (Yeates, 1997; Yeates & Fam, 1999). Merohedral twinning refers to the situation where all of the different structural domains are well ordered crystal lattices with identical parameters so that the diffraction peaks from the different domains superimpose perfectly. Hemihedral twinning refers to the situation where exactly two lattice domains with different structures contribute to the observed diffraction pattern. With any form of merohedral twinning, direct measurement of the intensities from any individual domain is not possible from a single crystal. Although undistorted intensity data can be recovered if multiple crystals are available with different twinning fractions, this is not possible with 'perfect' merohedral twinning in which each of the domains always occurs in an equal fraction, as is apparently the case with the trigonal crystal form of MJ1267. Furthermore, merohedral twinning produces a systematic loss of weak peaks in the diffraction pattern, as two weak reflections are unlikely to superimpose upon one another after application of the twinning operation (which is unrelated to the symmetry of the individual crystal domains); this effect allows merohedral twinning to be detected based on the analysis of intensity statistics (Yeates, 1988; Yeates & Fam, 1999). Although a formalism has been presented for the solution of a crystal structure with perfect merohedral twinning using multiple isomorphous replacement (Yeates & Rees, 1987), no new structure has yet been solved in this way. However, relatively simple methods exist for the solution of such a structure using molecular replacement (Redinbo & Yeates, 1993; Yeates, 1997).

Once an untwinned crystal form of Mg-ADP-bound MJ1267 was obtained and solved using multiple-wavelength anomalous diffraction techniques, the structure of the protein in the twinned crystal could be evaluated. In this paper, we report the use of molecular replacement to characterize the packing of MJ1267 in the twinned trigonal crystal lattice, as well as the refinement of the twinned structure using the method of Redinbo & Yeates (1993) as implemented in the program *CNS* (Brünger *et al.*, 1998). This refinement shows significant rotation of the ABCα subdomain compared with the nucleotide-binding core of MJ1267, suggesting that the ABCα subdomain is also flexibly attached when Mg-ADP is bound at the active site of cassette. Based on examination of the intermolecular packing in the trigonal lattice, the structural basis for the perfect hemihedral twinning of the dominant lattice in this crystal form can be understood based on the promiscuous parallel and antiparallel packing of very similar fibers of MJ1267 molecules with 3_1 helical symmetry. The pathological streaking of the diffraction pattern in the $a^* \times b^*$ plane of the crystal is likely to represent another manifestation of twinning which can be explained on the basis of the same promiscuous fiber-packing interactions.

2. Experimental

2.1. Protein expression, purification and crystallization

The MJ1267 purification protocol was described in detail by Karpowich *et al.* (2001). The purified protein was concentrated to 36 mg ml^{-1} in 10 m*M* Tris–HCl pH 8.8; 20 m*M* Mg-ADP was added before mixing in a 1:1 ratio with 16%(*w*/*v*) PEG 3350 (Sigma), 30%(*v*/*v*) ethylene glycol, 20%(*v*/*v*) glycerol, 4 m*M* $MgCl_2$, 0.2 m*M* DTT, 200 m*M* MES pH 5.9. After adding microseeds (obtained under similar conditions but at higher PEG concentrations), crystals formed in this solution within 48 h during hanging-drop or sitting-drop vapor diffusion at 310 K against a well containing 15–20%(*w*/*v*) PEG 3350, 15%(*v*/*v*) ethylene glycol, 10%(*v*/*v*) glycerol, 0.1 m*M* DTT, 100 m*M* MES pH 5.9. A series of 11 single-site cysteine mutations, including N31C and N109C, were made in the otherwise cysteine-free MJ1267 protein using the QuickChange kit (Stratagene). The wild-type protein either with or without selenomethionine labeling [using the method of LeMaster and Hendrickson (Hendrickson *et al.*, 1990) with the addition of the KAO vitamin mix (Sigma)] and the subset of underivatized cysteine mutants that still crystallized all yielded exclusively twinned trigonal crystals under the standard crystallization conditions. Mass spectrometry indicated that most of the cysteine mutants could be derivatized with methylmercury in the protein stock solution, but no evidence of derivatization was ever observed in the diffraction data after soaking pre-formed crystals in the same reagent. Therefore, the cysteine mutants were pre-derivatized with methylmercury at the engineered cysteine site (using a protocol described by Karpowich *et al.*, 2001) prior to co-crystallization. The methylmercury-derivatized N31C protein yielded an isomorphous derivative of the twinned trigonal crystal form, but the methylmercury-derivatized N109C protein yielded untwinned monoclinic crystals, by epitaxial nucleation from trigonal seed crystals, that enabled the structure of MJ1267 to be determined using multiwavelength anomalous diffraction (Karpowich *et al.*, 2001).

Table 1
Crystallographic data-collection and refinement statistics.

Standard definitions were used for all of the parameters (Drenth, 1994). Refinement was performed on data from a selenomethionine-containing crystal collected at 12 900 eV on beamline X4A at the NSLS. Data merging was performed using *SCALEPACK* (Otwinowski & Minor, 1997) including all individual observations with $I \geq -3\sigma(I)$. Anomalous pairs were treated independently for calculation of $R_{\rm sym}$ (and multiplicity) but were merged for calculation of $\langle I/\sigma(I)\rangle$. The last shell is 77% complete for reflections with $I \geq 2\sigma(I)$ after merging. The completeness and refinement statistics are based on data with $F \geq \sigma(F)$. The multiplicity was calculated using only reflections accepted for merging (*i.e.* ignoring rejects). The refinement and geometric statistics come from *CNS* (Brünger *et al.*, 1998), while the *B*-factor and Ramachandran analyses were performed with *MOLEMAN* (Kleywegt & Jones, 1996) and *PROCHECK* (Laskowski *et al.*, 1993), respectively.

(*a*) Data-collection statistics. Space group $P3_1$. Unit-cell parameters 117.31 × 117.31 × 42.19 Å, 90, 90, 120°.

	Overall	Last shell
Resolution (Å)	20.0–2.60	2.76–2.60
Error scale factor	1.60	1.60
Estimated error	0.05	0.05
χ^2	2.00	1.83
Rejected observations (%)	8.4	5.4
$R_{\rm sym}$ (%)	6.9	19.1
Multiplicity	1.50	1.43
$\langle I/\sigma(I)\rangle$	15.6	7.08
Completeness (%)	92.6	90.7

(*b*) Refinement statistics.

	Detwinned refinement	Untwinned refinement
$R_{\rm free}$ (%)	28.8	35.7
$R_{\rm work}$	19.8	35.1
Bond-length r.m.s.d. (Å)	0.012	0.027
Bond-angle r.m.s.d. (°)	1.8	2.3
Bonded *B* r.m.s.d. (Å^2)	2.2	1.3
Non-bonded *B* r.m.s.d. (Å^2)	7.5	6.1
Ramachandran distribution		
Core (%)	78.6	90.3
Allowed (%)	20.4	8.2
Generously allowed (%)	1.0	1.5
Model contents		
Protein residues	759	759
Ligands	3 Mg-ADP	3 Mg-ADP
Waters	164	66

2.2. Crystallographic data collection and analysis

Crystals were taken directly from the mother liquor and frozen in liquid propane in preparation for data collection at approximately 100 K. Data from wild-type and N31C crystals (with and without methylmercury derivatization) were collected on an R-AXIS IV area detector using an RU-H3R generator equipped with focusing mirrors. The data used for refinement (Table 1) were obtained from a selenomethionine-labeled wild-type protein crystal and were collected using an R-AXIS IV area detector on beamline X4A at the National

Synchrotron Light Source. Diffraction data were integrated and merged with the programs *DENZO* and *SCALEPACK*, respectively (Otwinowski & Minor, 1997). The merging statistics from all crystals were similar to those from the synchrotron data set shown in Table 1. Although the limiting resolution of the home-source data was lower, its completeness was higher and its redundancy was much higher (data not shown). The cumulative intensity distribution and self-rotation function were calculated using the programs *TRUNCATE* and *POLARRFN* from *CCP*4 (Collaborative Computational Project, 1994), respectively. The $\langle I^2\rangle/\langle I\rangle^2$ analysis was performed using the web server at http://www.doe-mbi.ucla.edu/Services/Twinning (Yeates, 1988, 1997; Yeates & Fam, 1999). Computational estimates of twin fraction based on intensity statistics were performed using the same web server and the program *CNS* (Brünger *et al.*, 1998).

2.3. Molecular-replacement and difference Fourier calculations

Molecular-replacement (MR) calculations were performed using the *CCP*4 version (Collaborative Computational Project, 1994) of the program *AMoRe* (Navaza, 1994) with normalized structure factors (E values). The search model comprised a refined 1.9 Å model of MJ1267 derived from the untwinned monoclinic crystal form of the methylmercury derivative of the N109C mutant of the protein including all ligands (Karpowich *et al.*, 2001). For both MR and difference Fourier calculations, structure factors were calculated from the intensity data without detwinning.

2.4. Structure refinement

The molecular-replacement solution from *AMoRe*, comprising three molecules in the asymmetric unit (model '1+, 2− and 3+' in Tables 2 and 3), was refined in space group $P3_1$ using the program *CNS* (Brünger *et al.*, 1998) with Engh and Huber parameters (Engh & Huber, 1991) and either ignoring the twinning or 'detwinning' the data according to the method of Redinbo & Yeates (1993) (Table 1). The $R_{\rm free}$ set (Brünger, 1992) comprised 10% of the reflections which were chosen at random in one half of the asymmetric unit in reciprocal space and then expanded to cover the full asymmetric unit so that twin-related reflections were systematically either included or excluded in the $R_{\rm free}$ set. Refinement comprised a single cycle of positional refinement, individual B-factor refinement and automated water addition including all of the data to 2.60 Å. NCS restraints were not used during any of the refinement steps so that the structures determined for the three molecules in the asymmetric unit are independent. No manual rebuilding was performed as the detwinned electron-density maps appeared to suffer from strong model bias (see discussion below). The results of the detwinned refinement, along with the twinned structure factors, were depoisted in the Protein Data Bank under accession code 1g9x.

3. Results and discussion

3.1. Crystal diffraction properties

Initially, crystals of Mg-ADP-bound MJ1267 were obtained in the form of clusters of tiny rods resembling a sea urchin. Optimization of these conditions led to the formation of substantially larger rods that suffered from a severe tendency to fray at their tips, giving the appearance of a broccoli floret in extreme cases. Further optimization trials led to the identification of conditions yielding trigonal rods of approximately 40 × 40 × 250 µm in size with only minimal morphological abnormalities (Fig. 1*a*) involving slight irregularities at their tips and a stepped morphology on their facets visible exclusively under high magnification. However, these crystals consistently yielded a pathological diffraction pattern (Figs. 1*b* and 1*c*) characterized by pronounced streaking in the reciprocal-space planes perpendicular to the long axis of the rod (*i.e.* the $a^* \times b^*$ planes in Figs. 1*b* and 1*c*).

The dominant lattice in this diffraction pattern was readily indexed by *DENZO* (Otwinowski & Minor, 1997) in a trigonal crystal system with unit-cell parameters of approximately 117 × 117 × 42 Å. Integration of the diffraction data was

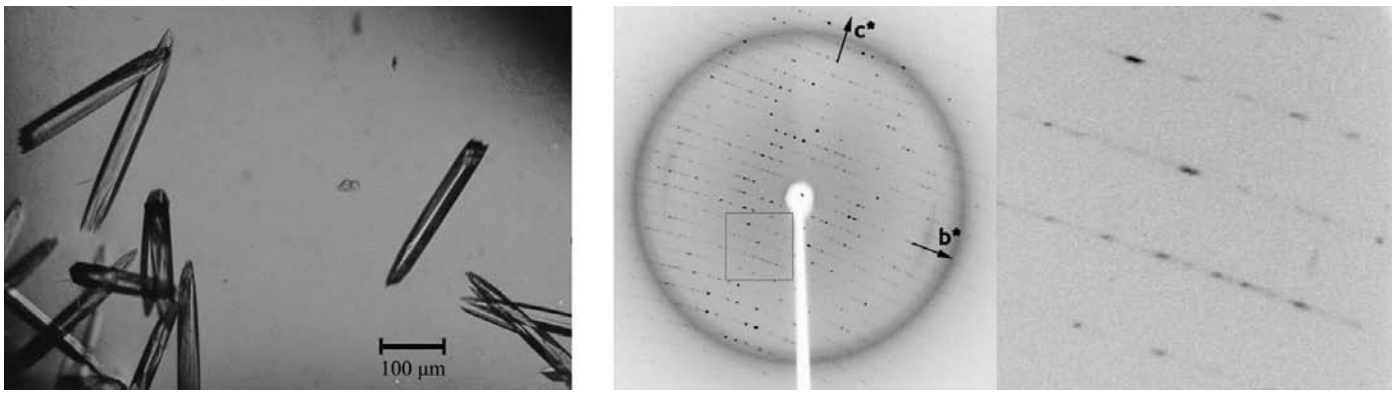


Figure 1
The trigonal crystal form of MJ1267. (*a*) Photomicrograph of typical crystals in approximately 18%(*w*/*v*) PEG 3350, 15%(*v*/*v*) ethylene glycol, 10%(*v*/*v*) glycerol, 0.1 m*M* DTT, 10 m*M* Mg-ADP, 100 m*M* MES pH 5.9. (*b*) A 1° oscillation frame from a selenomethionine-labeled crystal of the wild-type protein collected at 12 900 eV on beamline X4A at the NSLS (displayed with the program *XDISP*; Otwinowski & Minor, 1997). (*c*) A magnified view of the region of the diffraction image in (*a*) delineated by the black square.

performed using an elongated guard region such that the streaks in the $a^* \times b^*$ planes were systematically excluded from the background calculation for each spot. The data merged equally well assuming a Laue symmetry of $P\bar{1}$ or $P\bar{3}$ and with only slightly poorer statistics assuming a Laue symmetry of $P\bar{3}1m$ (not shown). However, the merging statistics were somewhat remedial in all cases. Table 1(*a*) shows a representative example of the statistics obtained after merging the data from a crystal in space group $P3_1$ (*i.e.* assuming Laue symmetry of $P\bar{3}$). While an overall $R_{\rm sym}$ of 6.9% from 20.0 to 2.6 Å resolution is reasonable, the mean overall value of χ^2 was 2.0, and 8.4% of the reflections were rejected. Improvements in merging statistics (including a reduction in the number of rejections) could be obtained by post-refinement of crystal mosaicity in *SCALEPACK* (Otwinowski & Minor, 1997). However, as the refined mosaicity ranged from 2.5° to 3.5° in different crystals, this procedure led to severe reductions in completeness arising from overlaps when data integration was conducted using the refined mosaicity. As a compromise between loss of quality and loss of completeness, integration and merging were conducted with the mosaicity fixed at 1.6°, consistently yielding data of the type shown in Table 1(*a*). However, much higher redundancy levels were generally obtained than indicated in Table 1(*a*) as these data came from an SeMet crystal collected above its anomalous absorption edge, so that in this singular case anomalous pairs were treated independently during merging.

After processing the diffraction data in space group $P3_1$, the $\kappa = 180°$ section of the self-rotation function (Fig. 2*a*) yielded a strong peak at an azimuthal angle of 30° which was equal in height to the origin peak. This result indicated that the symmetry of the diffraction pattern was $P\bar{3}1m$ (*i.e.* that there was twofold crystallographic symmetry around the b^* axis of the diffraction pattern), suggesting space-group symmetry of $P312$ or $P3_112/P3_212$. However, analysis of intensity statistics (Figs. 2*b* and 2*c*) suggested that the crystal suffered from hemihedral twinning (Yeates, 1988; Yeates & Fam, 1999). The cumulative intensity distribution (Collaborative Computational Project, Number 4, 1994) over a broad resolution range

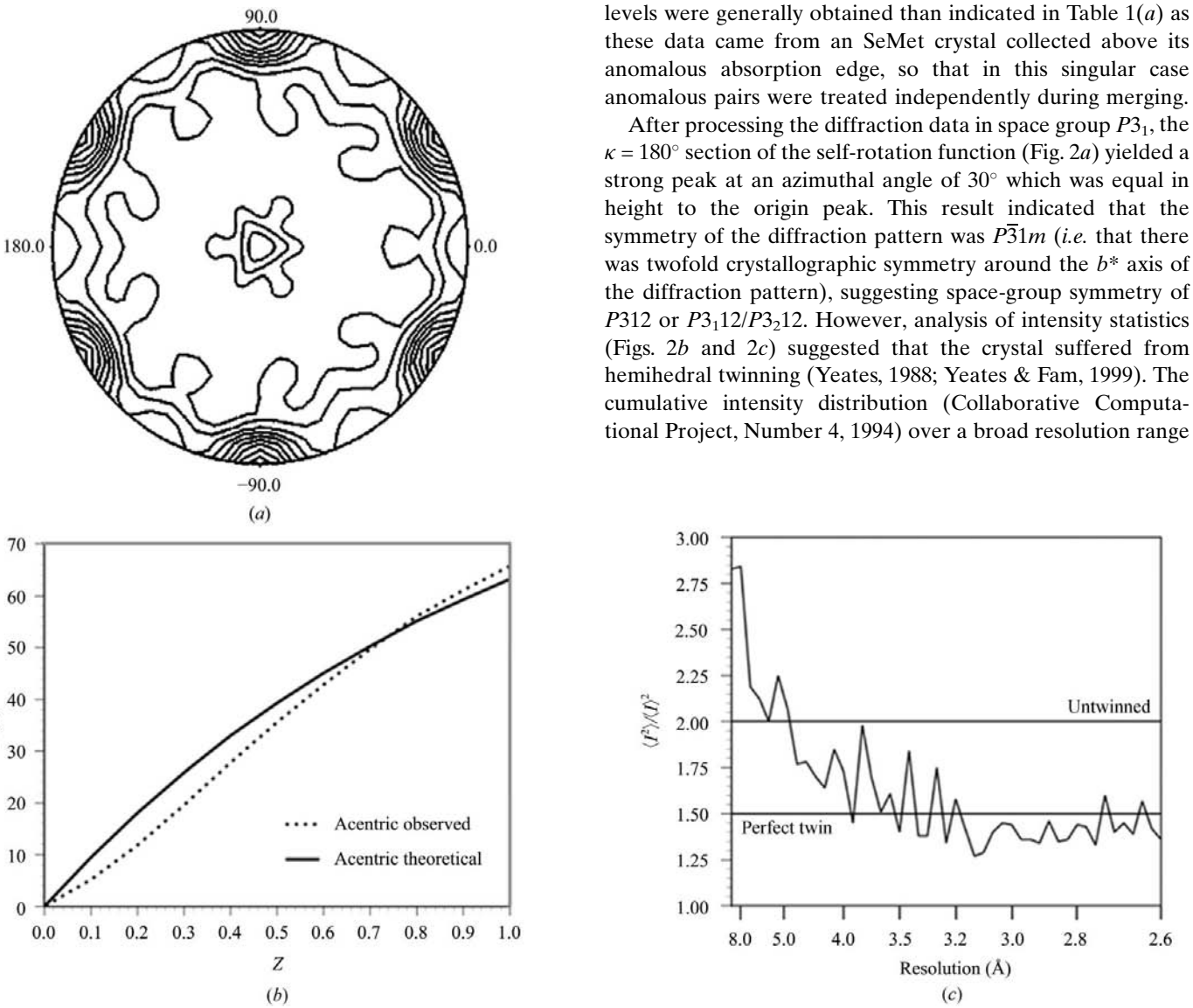


Figure 2
Analysis of the diffraction data from the twinned trigonal crystals. The data was processed in space group $P3_1$. All the data shown here were obtained from a selenomethionine-labeled crystal of the wild-type protein and were collected at 12 900 eV on beamline X4A at the NSLS, although equivalent results were obtained with other crystals. (*a*) The $\kappa = 180°$ section of the self-rotation function for the crystal calculated using the program *POLARRFN* from *CCP*4 (Collaborative Computational Project, Number 4, 1994). Data from 20.0 to 2.60 Å resolution were included in the calculation. The peak at an azimuthal angle of 30° (and the symmetry-related peaks at azimuthal angle intervals of 60°) are equal in height to the origin peak and are caused by the twinning operation around the b^* axis. (*b*) The cumulative intensity distribution for the crystal in the resolution range 20.0–2.60 Å calculated using the program *TRUNCATE* from *CCP*4 (Collaborative Computational Project, Number 4, 1994). The plot gives the percentage of reflections with normalized intensity $(I/\langle I\rangle) \leq$ the fraction Z. The theoretical curve assumes normal intensity statistics for an untwinned specimen (Wilson, 1949). (*c*) Yeates plot (Yeates, 1988; Yeates & Fam, 1999) showing $\langle I^2\rangle/\langle I\rangle^2$ as a function of resolution from 20.0 to 2.60 Å. The values for untwinned and perfectly twinned specimens again assume normal intensity statistics.

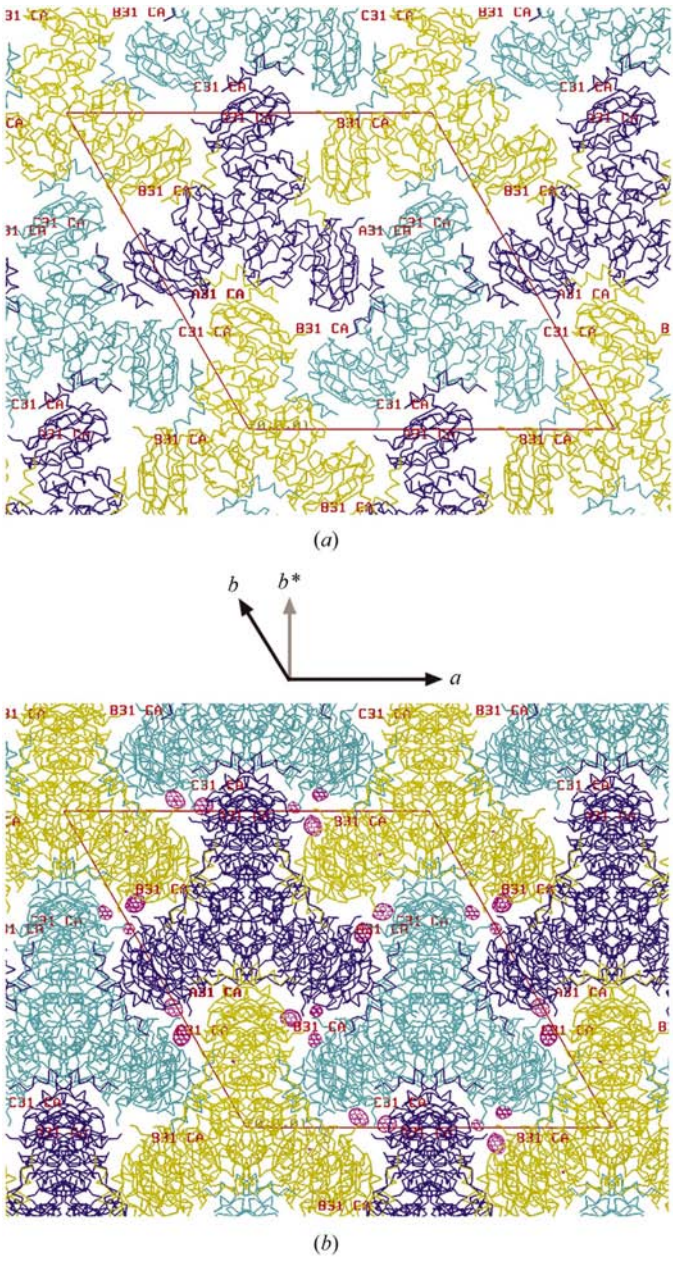


Figure 3
Packing diagrams for the twinned trigonal crystal structure. The red lines show the unit-cell boundary. (*a*) Molecular packing in one of the twin domains in the dominant lattice. The '1+' (subunit *A*) and '3+' (subunit *C*) molecules are shown in blue and cyan, respectively, while the '2−' (subunit *B*) molecule is shown in light (yellowish) green. The three molecules related by 3_1 screw symmetry at each of these positions are all shown in projection along the *c* axis (which is perpendicular to the plane of the page). Thus, each trimer represents a 3_1 fiber which projects perpendicular to the plane of the page. (*b*) Difference Fourier map of the methylmercury derivative of the N31C mutant of MJ1267 superimposed on a packing diagram including both of the twin-related sub-lattices. The twinning operation is seen to superimpose each of the individual 3_1 fibers in the untwinned lattice upon itself but in an antiparallel orientation. The difference Fourier map shown in magenta was calculated using data from 15.0 to 5.0 Å resolution with phases from the optimal molecular-replacement solution comprising three NCS-related molecules (designated '1+, 2−, 3+' in Tables 2 and 3) but without taking into account the twinning. Thus, difference Fourier analysis using phases from an untwinned model representing one of the two twin domains reveals heavy-metal sites from molecules in both of the twin domains.

Table 2
Molecular-replacement analysis in space group $P3_1$.

In space group $P3_1$, each molecule placed in the asymmetric unit gives rise to a helical fiber with 3_1 symmetry along the *c* axis. Molecules 1, 2 and 3 are related to molecules 1′, 2′ and 3′ by the twinning operation (*i.e.* 180° rotation around *b**). Based on the analysis presented in this table, the primary lattice contains two 3_1 fibers in one orientation at positions 1 and 3 and a third 3_1 fiber related by an approximately 180° rotation around *b** at position 2. The search model comprised a refined 1.9 Å model of MJ1267 derived from the monoclinic crystal form obtained with the methylmercury derivative of the N109C mutant of the protein (Karpowich *et al.*, 2001). Calculations were performed based on *E* values (*i.e.* normalized structure factors) using the program *AMoRe* (Navaza, 1994) in *CCP*4 (Collaborative Computational Project, Number 4, 1994) with data from 20.0 to 3.2 Å resolution.

(*a*) Fiber molecular rotations and translations (after rigid-body refinement).

	α (°)	β (°)	γ (°)	TA	TB	TC
+						
Position 1	63.56	96.95	247.50	0.4420	0.5288	0.0006
Position 2′	63.83	96.93	247.45	0.1153	0.8671	0.3585
Position 3	63.75	96.88	247.78	0.7882	0.2079	−0.0014
−						
Position 1′	116.43	83.33	67.53	0.0870	0.5290	−0.0003
Position 2	116.15	83.05	67.45	0.7529	0.8671	0.6413
Position 3′	116.35	83.21	67.79	0.4207	0.2084	0.0019

(*b*) Molecular-replacement statistics.

Fiber orientation(s)	Fiber position(s)	Correlation coefficient	*R* factor
+	1	25.1	44.3
+, −	1, 2	37.0	41.2
+, +	1, 3	32.3	43.7
+, −, +	1, 2, 3	46.4	39.0
+, +, +	1, 3, 2′	24.6	45.2
+, −, +, +	1, 2, 3, 2′	38.0	41.4
+, −, +, −	1, 2, 3, 1′	40.6	40.8
−	1′	25.1	44.3
−, +	1′, 2′	37.0	41.2
−, −	1′, 3′	32.4	43.7
−, +, −	1′, 2′, 3′	46.5	39.0
−, −, −	1′, 3′, 2	24.5	45.1
−, +, −, −	1′, 2′, 3′, 2	37.1	41.5
−, +, −, +	1′, 2′, 3′, 3	39.2	41.3

(Fig. 2*b*) showed only minor distortions, reflecting a small reduction in the proportion of weak reflections compared with expectations based on normal intensity statistics (Wilson, 1949). Given the fact that the streaking in the $a^* \times b^*$ planes could itself have produced background tending to increase the apparent intensity of weak reflections, the cause of these minor distortions was unclear. However, analysis of a Yeates plot (Fig. 2*c*), showing $\langle I^2\rangle/\langle I\rangle^2$ as a function of resolution, gave results indicative of perfect hemihedral twinning (Yeates, 1988; Yeates & Fam, 1999). The low-resolution region of this plot showed pathologically high values, but the normal intensity statistics that represent the theoretical basis of this analysis are only rigorously valid at high resolution and this region of the Yeates plot showed essentially the canonical value for a hemihedral twin. The pathological values observed in the low-resolution region of the Yeates plot probably result from pseudo-symmetry in the primary lattice which causes a highly non-random distribution of electron density in the cell at low resolution (see Fig. 3 below). The extreme irregularity

in the intensity statistics in different resolution ranges is likely to account for the relatively normal appearance of the cumulative intensity distribution when all of the data is combined and analyzed in a single bin (as in Fig. 2*b*).

The lattice symmetry in trigonal space groups permits merohedral twinning based on twofold rotation around the *b** axis (Yeates, 1997). In this context, the data from the Yeates plot led us to conclude that the dominant lattice in the crystal probably suffered from perfect hemihedral twinning around the *b** axis, which would mean that the point-group symmetry of the individual twin domains in the primary real-space lattice was $P3_1$.

3.2. Molecular-replacement calculations on the dominant lattice

Molecular-replacement (MR) calculations (Navaza, 1994; Collaborative Computational Project, Number 4, 1994) were performed on the reduced diffraction data from the dominant lattice using a search model comprising the refined structure of the same protein determined from the untwinned monoclinic crystal form obtained from the methylmercury derivative of the N109C mutant of the protein (see §2 for details). Consistent with the conclusion that the trigonal crystals suffered from hemihedral twinning around the *b** axis, MR calculations in space groups $P3_112$ or $P3_212$ did not yield any statistically significant solutions (not shown). However, equivalent calculations in space group $P3_1$ immediately yielded a strong and unambiguous solution (Table 2 and Fig. 3*a*). In this space group, the cross-rotation function gave exactly two statistically significant solutions (Table 2*a*) which were related to one another by the twinning operator (*i.e.* an approximately 180° rotation around *b**). Subsequent translation-function calculations (Table 2*b*) yielded a solution comprising three molecules in the asymmetric unit (seen in projection along the *c* axis in Fig. 3*a*), corresponding to a Matthews number of 1.89 Å^3 Da^{-1} and a solvent content of 35%.

Two of the molecules in this MR solution are related to one another by exclusively translational NCS and the third is related to the others by twofold screw operations whose rotational components are the same as the twinning operator. Combining the 3_1 crystallographic symmetry with the translational symmetry of the lattice, each molecule in the asymmetric unit gives rise to a helical fiber which runs continuously through the crystal along its *c* axis (*i.e.* extending out of the page as viewed in Fig. 3*a*). Therefore, a single twin domain in the dominant lattice comprises three similar helical fibers which make both parallel and antiparallel intermolecular packing interactions.

In addition to having a reasonable packing density, the MR solution in space group $P3_1$ shows convincing crystal packing interactions between the three molecules (Fig. 3*a*) with no significant intermolecular overlap. However, given the complexity of the crystal, several additional computational analyses were performed to verify the MR solution. First, the translation-function calculations were repeated trying to add molecules in each of the two twin-related rotational orientations at each step in the iterative process of adding molecules to the unit cell (bottom of Table 2). An identical MR solution is obtained when the polarity of all three helical fibers are reversed simultaneously, as expected based on the twofold symmetry around the *b** axis in the twinned diffraction pattern. However, starting with one fiber in either orientation, there is an improvement in the correlation coefficient and *R* factor when the next fiber is added in either orientation, but a considerably better improvement when the second fiber is added in the antiparallel orientation (*e.g.* compare '+, −' and '+, +' in Table 2). This result makes sense if there are two antiparallel packing interactions but only a single parallel packing interaction in the lattice, as more intermolecular chords in Patterson space will be matched when placing fibers in the antiparallel orientation. There is an additional improvement in correlation coefficient and *R* factor when two parallel fibers and one antiparallel fiber are placed in the unit cell (*e.g.* '+, −, +' in Table 2), but the quality of the solution degrades dramatically when three parallel fibers are placed in the unit cell (*e.g.* '+, +, +' in Table 2). These results support the validity of the MR solution comprising two parallel fibers and one antiparallel fiber (shown in Fig. 3*a*).

A second check on the validity of the MR solution was performed using isomorphous difference Fourier analysis of a methylmercury derivative of the N31C mutant of MJ1267 (Table 3). This cysteine mutation was engineered into the otherwise cysteine-free protein. Phasing the structure with partial MR solutions comprising just a single molecule in the asymmetric unit yields statistically outstanding difference Fourier peaks at the N31C positions in all three molecules in the corresponding complete MR solution (*i.e.* including the equivalent positions in the two molecules not used in the phase calculation), while phasing the structure with the

Table 3
Signal-to-noise [$\rho/\sigma(\rho)$] of mercury peaks in isomorphous difference Fourier maps phased with different protein models.

Isomorphous difference Fourier maps were calculated at 15.0–5.0 Å using structure factors from crystals containing the N31C mutant of MJ1267 with and without methylmercury bound to the engineered cysteine. The peak height is given for the N31C position corresponding to each of the six protein molecules in the twinned crystal lattice. In column 1, the protein model is defined by the position and orientation of each molecule in the unit cell using the terminology defined in Table 2. Therefore, model '1+' contains a single molecule in the '+' orientation at position 1, while model '1+, 2−, 3+' contains two molecules in the '+' orientation at positions 1 and 3 and also a single molecule in the '−' orientation at position 2.

	Position and orientation of molecule containing the mercury site						
Model contents	1+	2−	3+	1′−	2′+	3′−	Max. noise
1+	12.28	9.09	6.15	—	3.73	9.55	3.72
2−	10.10	13.67	8.34	5.09	—	4.96	3.97
3+	5.03	10.57	11.89	7.13	4.45	—	3.79
1′−	—	3.98	9.49	12.41	10.27	5.44	3.95
2′+	6.31	—	4.56	10.05	14.10	7.90	3.63
3′−	7.82	3.91	—	5.11	10.66	11.89	4.04
1+, 2−, 3+	13.09	16.95	13.08	5.23	—	5.94	3.88
1′−, 2′+, 3′−	6.40	—	5.81	13.64	17.02	11.89	<3.50

complete MR solutions comprising all three molecules yields statistically stronger peaks at the same three positions (Table 3 and Fig. 3*b*). Therefore, the results of difference Fourier analyses on an engineered heavy-metal derivative confirm the validity of the MR solution comprising three molecules which form two parallel and one antiparallel 3_1 fiber in an individual twin domain of the crystal. It is noteworthy that phases from the partial and complete MR solutions for one lattice show difference Fourier peaks at some of the N31C sites in the twin-related lattice. This result is observed because both lattices contain at least one parallel and one antiparallel 3_1 fiber so that the molecular-replacement solution corresponding to either one of these structures has phasing power in both of the twin-related lattices. Thus, difference Fourier analyses using phases derived from the molecular-replacement solution in one lattice will have some tendency to also show heavy-atom positions from the twin-related lattice, although this result would not be observed in the general case, where the molecular coordinates from one lattice do not necessarily have phasing power in the twin-related lattice.

3.3. Detwinned refinement of the crystal structure and protein conformational changes

The crystal structure was refined as a twin (left column in Table 1*b*) using the formalism of Redinbo and Yeates (Redinbo & Yeates, 1993; Yeates, 1997) as implemented in the program *CNS* (Brünger *et al.*, 1998). The model comprised three molecules in the asymmetric unit, corresponding to two parallel fibers and one antiparallel fiber in each twin domain in the primary crystal lattice (*i.e.* the '+, −, +' solution in Table 2). This refinement strategy yielded a working *R* factor of 19.8% and a free *R* factor of 28.8% for data from 20.0 to 2.60 Å resolution after a single round of positional refinement, individual *B*-factor refinement and automated water addition (Table 1*b*). Refinement of the twin fraction in *CNS* led to estimates in the range 0.45–0.50, consistent with the estimate from the analysis of intensity statistics. Because these values did not differ significantly from 0.5, the structure was treated as a perfect hemihedral twin in the refinement reported in Table 1(*b*) (*i.e.* with the twin fraction fixed at 0.5). The *R* factors and molecular geometry of the model refined in this way were as good as those obtained when the twin fraction was optimized during refinement (data not shown). In contrast, the quality of the refinement was considerably worse when the same molecular-replacement solution was refined without twinning, yielding working and free *R* factors of 35.1 and 35.7%, respectively (Table 1*b*).

The detwinned $2F_o - F_c$ maps from the model refined as a hemihedral twin were of very good quality (Fig. 4). However, the electron density appeared to be highly biased by the model based on several criteria, including an absence of interpretable difference density. Detwinned simulated-annealing omit maps showed reasonable features for the most electron-dense features in the structure (*i.e.* the phosphates and ribose of the nucleotide in Fig. 4), but were not informative in other regions (data not shown). Therefore, model bias effectively prevented rebuilding and iterative improvement of the model. As a result, the quality of the Ramachandran plot for the model from the detwinned refinement is slightly remedial, with only 78.6% of the residues in core regions. However, there are no residues in disallowed regions and only 1.0% in generously allowed regions and the overall quality of the molecular geometry is relatively good based on the low root-mean-square deviations of the bond lengths and angles and the bonded and non-bonded *B* factors (Table 1*b*). The model bias in the maps is likely to be caused by the relatively low resolution and low quality of the diffraction data combined with the fact that the F_c terms make a direct contribution to the electron density in the least-squares formalism of Redinbo & Yeates (1993).

Nonetheless, the validity of the detwinned refinement is supported by improvement in the free *R* factor during this procedure and also by the nature of the changes that occur in the protein structure compared with the search model. Fig. 5(*a*) shows a superposition of one of the refined monomers with the search model, based on least-squares alignment of the central β-sheet and the attached Walker A (P-loop) helix in the ATP-binding core. This alignment shows that the major conformational change that takes place during refinement is a movement of the α-helical subdomain. Least-squares alignment of the α-helical subdomain (data not shown) indicates that the movement involves a 6.5° rigid-body rotation of this subdomain relative to the ATP-binding core of the

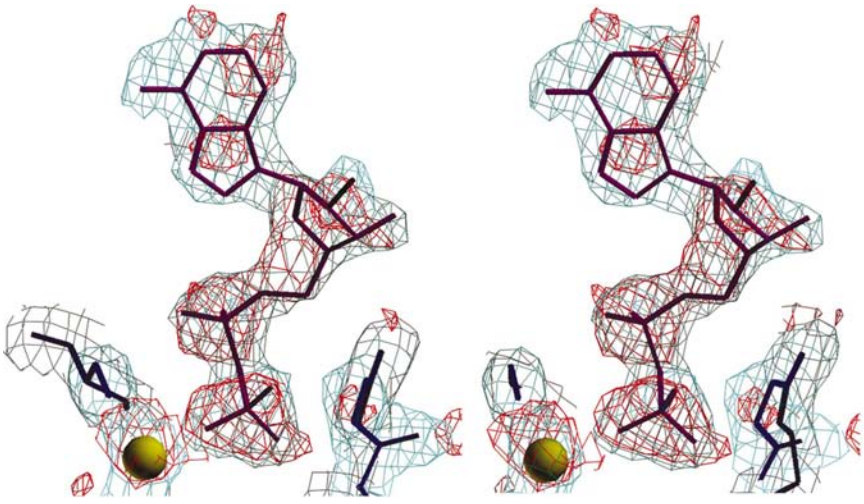


Figure 4
Detwinned electron-density maps (Jones *et al.*, 1991) calculated in *CNS* (Brünger *et al.*, 1998) using the method of Redinbo and Yeates (Redinbo & Yeates, 1993; Yeates, 1997). The cyan contours show the $2F_o - F_c$ map from the refined structure. The red contours show a simulated-annealing omit map (*i.e.* an $F_o - F_c$ map) calculated at 2500 K after removing Mg-ADP from the structure. The yellow sphere represents the Mg cofactor of the ADP molecule.

cassette (leading to a 3 Å displacement of the LSGGQ transporter signature sequence). In the crystal structure of the MalK ATP-binding cassette, a much larger rigid-body rotation of the α-helical subdomain was observed on comparing the two NCS-related molecules (Diederichs *et al.*, 2000). This observation has led to the conclusion that the α-helical subdomain is flexibly attached to the core of the cassette under some circumstances; comparison of the ATP-bound and ADP-bound conformations of several ATP-binding cassettes has led to the hypothesis that this rotation is coupled to the occupancy of γ-phosphate in the active site (Smith *et al.*, 2002; Karpowich *et al.*, 2001; Yuan *et al.*, 2001; Hopfner *et al.*, 2000).

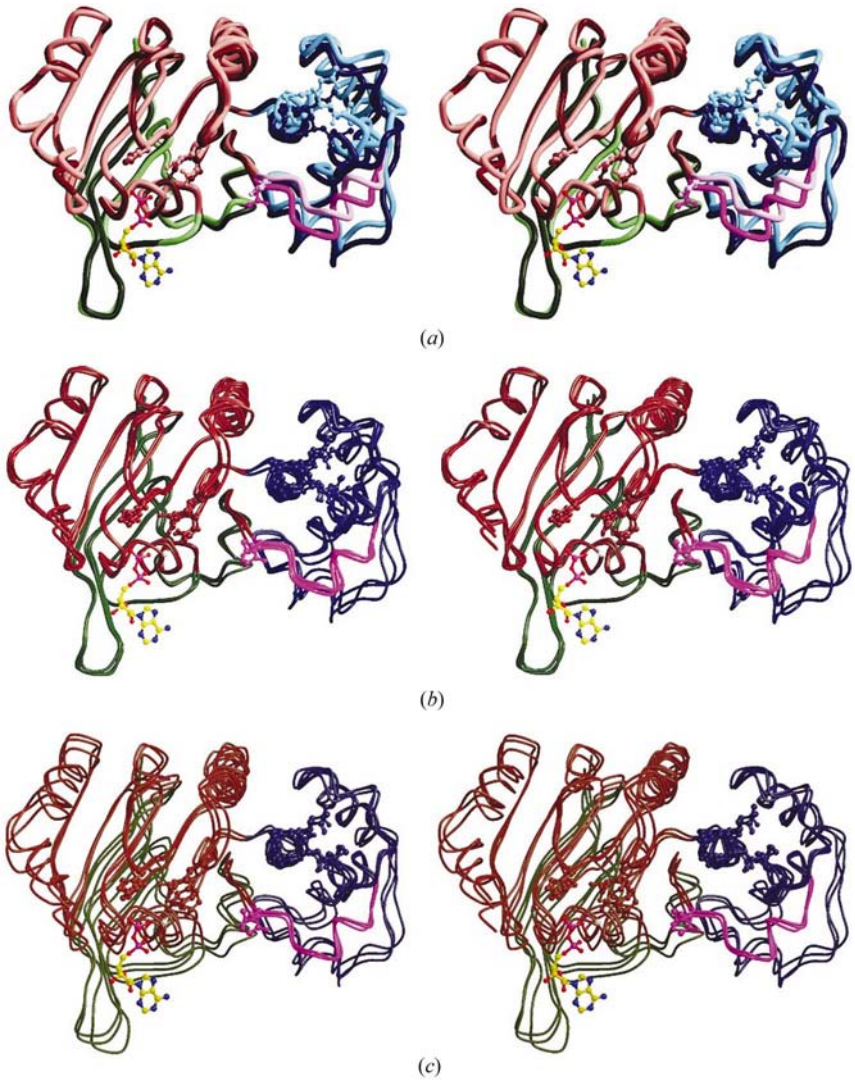


Figure 5
Stereo pairs showing refined structures of the Mg-ADP-bound MJ1267 protein. The molecules are color-coded according to domain organization (Karpowich *et al.*, 2001), with green showing the antiparallel β-sheet subdomain (ABCβ), red showing the F1-type ATP-binding core, blue showing the α-helical subdomain (ABCα) and magenta showing the γ-phosphate linker which connects the ABCα subdomain to the ATP-binding core. The images were produced using *MOLSCRIPT* and *RASTER3D* (Kraulis, 1991; Merritt & Bacon, 1997). (*a*) Comparison of the structure of one of the NCS-related molecules in the twinned trigonal crystal (darker colors) with the structure of the same protein in the untwinned monoclinic crystal obtained from the methylmercury derivative of the N109C mutant of MJ1267 (lighter colors) (Karpowich *et al.*, 2001). The protein has Mg-ADP bound at the active site in both crystal structures. The molecules were aligned based on least-squares superposition of the β-strands and P-loop helix from the F1-type ATP-binding core. Subunit *A* (the '1+' molecule) is shown from the twinned trigonal crystal structure. (*b*) Comparison of the three refined NCS-related molecules in the asymmetric unit of the twinned trigonal crystal after least-squares superposition of the the β-strands and P-loop helix from the F1-type ATP-binding core. (*c*) Comparison of the three refined NCS-related molecules in the asymmetric unit of the twinned trigonal crystal after least-squares superposition of the α-helices in the ABCα subdomain.

Thus, there is a clear precedent for the kind of conformational change that takes place in the MJ1267 cassette during the detwinned refinement. However, the nucleotide-binding state of the cassette in the MalK crystals was ambiguous (Diederichs *et al.*, 2000). The observation of a similar conformational change when comparing two different Mg-ADP-bound structures (*i.e.* a refined monomer from the twinned trigonal crystal form and the search model from the monoclinic crystal form) indicates that the α-helical subdomain is flexibly attached to the ATP-binding core even when Mg-ADP is bound at the active site of the cassette.

More minor conformational differences are observed between the three NCS-related monomers in the detwinned structure (Figs. 5*b* and 5*c*). These differences involve the α-helical subdomain, the extreme C-terminus of the cassette (on the left as shown in Fig. 5) and the Walker B-loop (one of the segments of the ATP-binding core immediately adjacent to the α-helical subdomain, near the center of the molecule as shown in Fig. 5). Some of the conformational differences in the α-helical subdomains in the NCS-related monomers arise from minor rigid-body rotations (compare the core alignment and α-helical subdomain alignment shown in Figs. 5*b* and 5*c*, respectively). However, some backbone conformational changes are also observed in the α-helical subdomain in α-helix 1′ and the loop at its C-terminus (see Karpowich *et al.*, 2001 for nomenclature). These segments of the α-helical subdomain and the extreme C-terminus of the cassette represent the regions of highest backbone *B* factor in the 1.6 Å structure of the untwinned monoclinic crystal form of MJ1267, so that conformational differences in these two regions would be expected in different crystal forms. Indeed, conformational changes in these regions and also in the Walker B-loop were previously observed in the untwinned nucleotide-free cubic crystal form of MJ1267 (Karpowich *et al.*, 2001). Therefore, although manual rebuilding of the twinned trigonal crystal structure was

not possible owing to model bias, the nature of the detailed conformational differences produced by the detwinned refinement procedure support its validity.

3.4. Promiscuous 3_1 fiber packing in the dominant lattice

An analysis of the intermolecular packing in one twin domain in the dominant lattice is presented in Fig. 6 and Table 4. Each of the refined NCS-related monomers is located at a similar position relative to one of the three different 3_1 axes present in the asymmetric unit of the lattice, producing three different 3_1 fibers, each of which has closely similar internal intermolecular packing interactions (data not shown). Two of these fibers involve a very similar rotational orientation of the monomer (designated '+' and shown in dark blue and cyan), while the third fiber involves an orientation that is rotated by approximately 180° around the b^* axis [designated '−' and shown in light (yellowish) green]. In spite of the opposite orientations of the monomers in the '+' and '−' fibers, the overall molecular envelope of all three fibers is very similar irrespective of their polarity (Figs. 3 and 6*a*).

However, the packing interactions between all three fibers are distinct (Table 4 and Fig. 6). Owing to lattice symmetry, all inter-fiber contacts in a single twin domain are identical to one of the three interactions shown in projection in Fig. 6(*a*). One of these contacts occurs between the two parallel fibers with similar monomer orientation, while the other two each occur between one of these fibers and the antiparallel fiber with the opposite orientation. The protein epitopes involved in these packing contacts are listed in Table 4. While the interactions between the parallel fibers do involve some of the same protein epitopes mediating the interactions between the antiparallel fibers, they also involve completely different protein epitopes (*i.e.* residues 56–57, 115–120 and 228–230). The two different interactions between the antiparallel fibers are called '1+/2−' and '2−/3+' in Table 4. These interactions are substantially more similar to one another in that two pairs of epitopes make equivalent interactions in both cases. However, the '2−/3+' interaction also involves a third pair of epitopes (*B*136–*B*140 and *C*246–*C*252) that are slightly too far apart to make packing contacts in the '1+/2−' interface. The plasticity of the intermolecular packing interactions holding

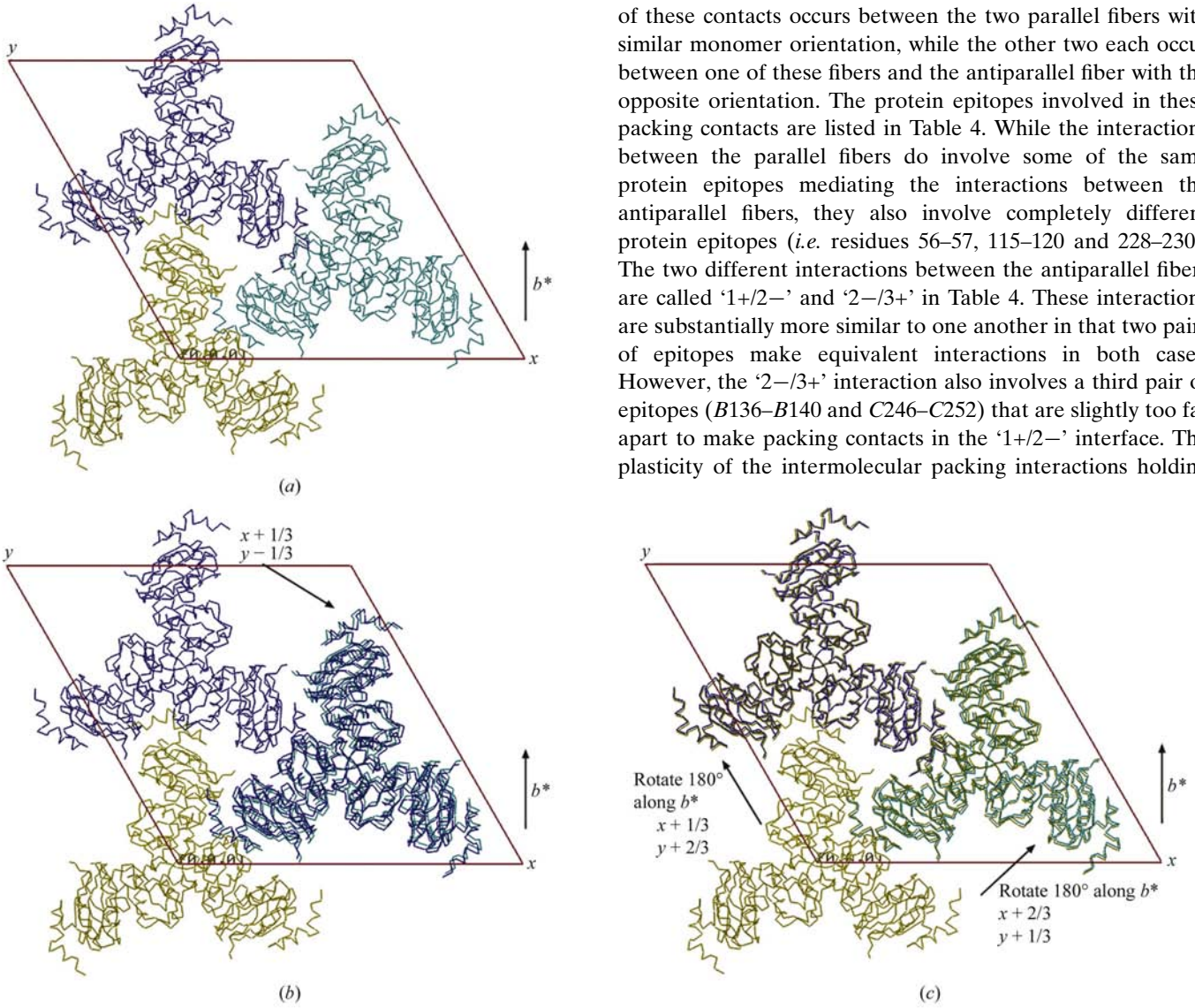


Figure 6
Analysis of the symmetry relationships between the three NCS-related molecules in one of the twin domains in the dominant lattice. Each of the three different 3_1 fibers in the unit cell is constrained by crystallographic symmetry to have its screw axis at exactly (0, 0), (1/3, 2/3) or (2/3, 1/3) in terms of fractional coordinates in the $a \times b$ plane. Translation of the individual 3_1 fibers by these allowed increments permits the conformation of the different 3_1 fibers to be compared. (*a*) The three fibers that comprise the contents of a single unit cell color-coded as in Fig. 3. (*b*) Superposition of the '1+' fiber (blue) on the '3+' fiber (cyan) after translation by $(x + 1/3, y - 1/3, 0)$ without rotation. (*c*) Superposition of the '2−' fiber (light green) on both the '1+' and '3+' fibers after rotation by 180° around b^* (*i.e.* the twinning operation) plus appropriate translations. This analysis shows that the '1+' and '3+' fibers have slightly different conformations in the dominant $P3_1$ lattice, while the '2−' fiber has a conformation that is intermediate between the other two but closer to that of the '1+' fiber. It also shows a close correspondence between the twinning operator and the rotational component of two NCS operators.

Table 4
Intermolecular interactions in the primary crystal lattice.

The analysis was performed on the refined model using the program *CONTACT* from *CCP*4 (Collaborative Computational Project, Number 4, 1994) and assuming a distance cutoff of 3.8 Å for van der Waals interactions. The positions and orientations of the three molecules are defined in Table 2. The *A*, *B* and *C* subunits in the PDB model correspond to the 1+, 2− and 3+ molecules, respectively. See Karpowich *et al.* (2001) for definition of the secondary-structural elements.

Molecules	Source residues	Source secondary structure	Target residues	Target secondary structure
1+/2−	*A*149–*A*158	'LSGGQ'	*B*238–*B*243	α core 6+
	*A*198–*A*202	Loop between α core 3–4 and β core 4	*B*254–B256	α core 6++
2−/3+	*B*149–*B*158	'LSGGQ'	*C*238–*C*243	α core 6+
	*B*136–*B*140	α ABCα 2	*C*246–*C*252	α core 6++
	*B*198–*B*202	Loop between α core 3–4 and β core 4	*C*254–*C*256	α core 6++
1+/3+	*A*228–*A*230	β-turn between β core 5 and β core 6	*C*152–*C*155	'LSGGQ'
	*A*242–*A*248	Loop between α core 6+ and α core 6++	*C*115–*C*120	β-turn between α ABCα 1 and α ABCα 1′
	*A*254–*A*256	α core 6++	*C*56–*C*57	Loop between α core 1–2 and β ABCβ 3

the protein crystal together allows this small but significant difference in interaction geometry which will be seen to play a critical role in the twinning process.

The different packing contacts made by the two parallel fibers require the conformations of these two fibers to be somewhat different. This fact can be seen by translating the parallel fiber shown in dark blue so that its crystallographic 3_1 axis coincides with that of the one shown in cyan (Fig. 6*b*). Following superposition in this manner, small but significant differences are observed in the location of the polypeptide backbone (lower right in Fig. 6*b*) caused by the conformational differences in the monomers combined with small differences in the translational position of the monomers relative to the 3_1 symmetry axes. A similar exercise can be performed to compare the conformation of the antiparallel fiber (shown in light green). Specifically, after rotating this fiber by 180° around the *b** axis and translating it so that its 3_1 axis coincides with either one of the parallel fibers, its backbone position corresponds very closely to that of the parallel fiber shown in dark blue and somewhat less closely to that of the parallel fiber shown in cyan (upper left and lower right of Fig. 6*c*). The conformational differences observed in these superpositions are smaller than those observed after superimposing the two parallel fibers, indicating that the conformation of the antiparallel fiber is intermediate between that of the two parallel fibers. However, the very tight superposition of the antiparallel fiber and the parallel fiber shown in dark blue indicates that the molecular conformation of these two fibers is very similar, except for their rotation by 180° around the *b** axis (*i.e.* by the twinning operation).

The conformational differences between the two parallel fibers explains the preference for packing a combination of two parallel and one antiparallel fiber in one twin domain of the dominant lattice. If the two parallel fibers were identical in conformation, it would mean that any given fiber could make either parallel or antiparallel interactions with each of the three neighboring fibers in the trigonal lattice. In this case, a random mixture of parallel and antiparallel fibers would be likely to result. The very clear preference for two parallel fibers and one antiparallel fiber that is apparent in the molecular-replacement calculations (Table 2) requires that each fiber have a distinct molecular conformation that makes optimal packing interactions only with the specific combination of parallel and antiparallel fibers shown in Fig. 6(*a*). However, the relatively small conformational differences between the fibers combined with the inherent plasticity in the conformation of the protein molecule and also in the packing interactions between protein molecules apparently allows more promiscuous inter-fiber interactions to occur in the lattice with some frequency. These interactions are very likely to be responsible for the pathological diffraction properties of the crystal.

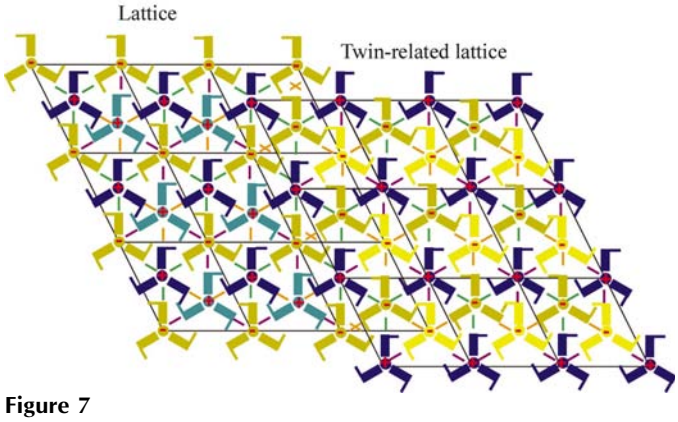


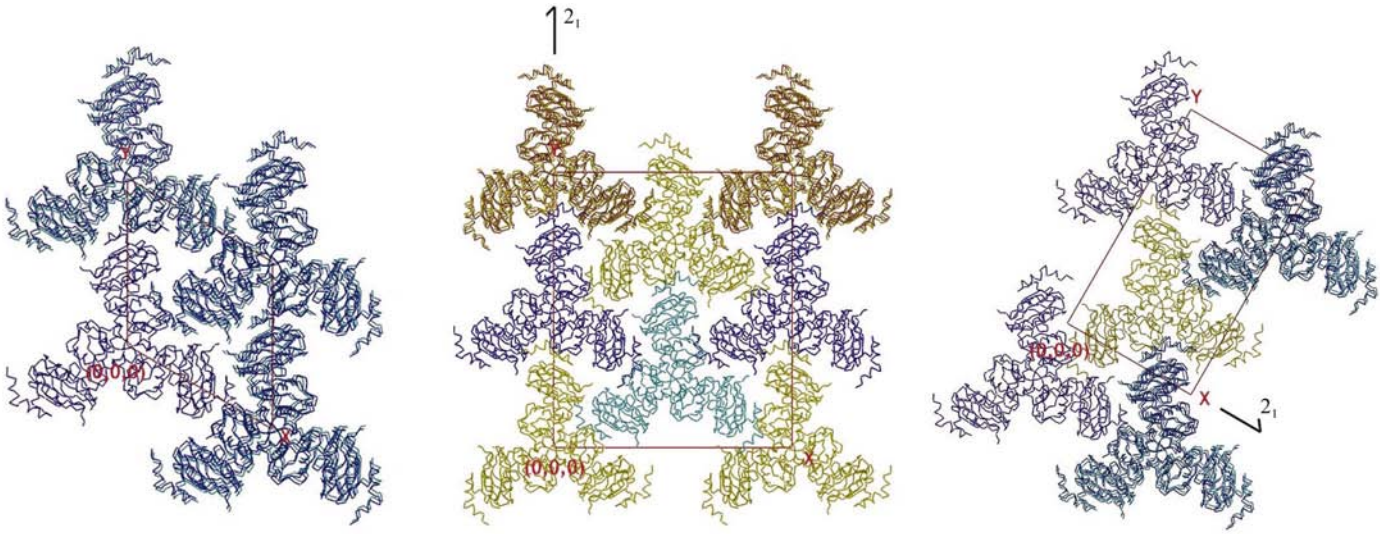
Figure 7
Schematic diagram of the packing interactions in the twin-related domains of the dominant lattice. The NCS-related 3_1 fibers in the first twin domain on the left are shown in colors roughly corresponding to those used in Figs. 3 and 6 (dark blue for the '1+' fiber, cyan for the '3+' fiber and light green for the '2−' fiber). Parallel interactions between the '1+' and '3+' fibers are represented by orange lines, while antiparallel interactions between either the '1+' and '2−' fibers or the '2−' and '3+' fibers are represented by green and purple lines, respectively. The analysis presented in Fig. 6 indicates that the NCS relationship between the '1+' and '2−' fibers is closely similar to the twinning operation. On this basis, the '1′−' and '2′+' fibers in the twin-related domain on the right are closely similar to the '2−' and '1+' fibers, so that they can appropriately be shown in light green and dark blue, respectively. As the '3+' fiber is not as closely related to the '2−' fiber by the twinning operation, its twin-related fiber ('3′−') represents a new structure that is shown in light yellow in the twin-related domain on the right. Superimposing the closely similar dark blue and green fibers in the twin-related lattices shows that a boundary can be formed between the twin-related domains with only one slightly sub-optimal packing interaction in each unit cell (indicated by an orange cross). This interaction is sub-optimal because it involves parallel packing between two identical '2−' fibers rather than the optimal parallel packing interaction, which occurs between the slightly different '1+' and '3+' fibers (as shown in the lower left panel of Fig. 6).

3.5. Promiscuous 3_1 fiber packing can explain hemihedral twinning

Fig. 7 presents a schematic analysis of potential packing interactions at the interface between two twin domains in the dominant lattice. In one of these domains, the antiparallel fiber (labeled '—' and shown in light green) has a very similar conformation to one of the parallel fibers (labeled '+' and shown in dark blue); they differ primarily in their orientations, which are related by a 180° rotation that is identical to the twinning operation. Because of this relationship, it is possible to construct an interface between two domains of the lattice related by the twinning operation that involves only one sub-optimal intermolecular contact per unit cell (indicated by the orange crosses in Fig. 7). This construction assumes that the conformations of the light green antiparallel fiber and the dark blue parallel fiber are identical so that the twinning operation converts one into the other and *vice versa*. Making this assumption, an interface can be constructed between the twin domains in which all interactions are identical to those within a single twin domain, except for one interaction per unit cell between two conformationally identical antiparallel fibers (labeled '—' and shown in light green) at the boundary between the two domains. This interaction is not ideal in that the parallel fibers that interact with each other in the dominant lattice have slightly different conformations (as shown in Fig. 6*b*). However, a boundary can be formed between twin domains with a small number of such interactions and it is not surprising that this occurs with a finite frequency given the general promiscuity of the inter-fiber interactions. Some degree of molecular plasticity is likely to be required to allow such sub-optimal interactions to form, but the promiscuous interactions between the fibers in the dominant lattice show that such plasticity is possible.

3.6. Promiscuous 3_1 fiber packing can also explain streaking in the $a^* \times b^*$ planes of the diffraction pattern

While the computational analyses presented above support the premise that the Bragg peaks from the trigonal MJ1267



(*a*) (*b*) (*c*)

Space group	$P3_1$	$P2_1$	$P2_1$
NCS	–	(+½, +⅔, 0) translation 3_1 screw axis	3_1 screw axis
Cell dimensions (Å,°)			
a	67.73	117.31	117.31
b	67.73	135.45	67.73
c	42.19	42.19	42.19
α	90	90	90
β	90	90	90
γ	120	90	90

Figure 8
Other lattices with different spacings in the $a \times b$ plane can be produced with only slightly sub-optimal packing interactions between 3_1 fibers. The red lines indicate the unit-cell boundaries in the various hypothetical lattices. Fibers are color-coded as in Figs. 3 and 6. At positions of sub-optimal packing, a second fiber is also shown which would have optimal packing interactions based on the interactions observed in the dominant lattice. Three parallel interactions between identical fibers would allow formation of a $P3_1$ unit cell containing a single molecule in the asymmetric unit (*a*); this lattice has smaller unit-cell parameters in the $a \times b$ plane but an unchanged unit-cell parameter along the *c* axis. A single parallel interaction between identical fibers could allow formation of a $P2_1$ lattice with six NCS-related molecules in the asymmetric unit (*b*); in this case, the 3_1 screw axis would represent NCS, but the unit-cell parameter along the *c* axis would remain unchanged. Finally, a single parallel interaction between identical fibers could also allow formation of a $P2_1$ lattice with three molecules in the asymmetric unit related a non-crystallographic 3_1 screw axis (*c*); the unit-cell parameter along the *c* axis once again remains unchanged in this lattice.

crystals are dominated by a hemihedrally twinned lattice comprising two parallel and one antiparallel 3_1 fibers, the streaking observed in the $a^* \times b^*$ plane in reciprocal space (Figs. 1*b* and 1*c*) suggests that there are additional complexities to the lattice structure. If the lattice contained exclusively twin-related domains of large spatial extent, sharp and unstreaked Bragg peaks would be observed. In this case, there would be comparatively minimal interference between the diffracted waves coming from the two lattices (assuming that the spatial extent of the lattice domains is larger than the coherence length of the X-rays) and the total diffraction intensities would be closely approximated as the sum of the intensities from the individual twin domains. Twin-related domains of small spatial extent would produce blurring of the Bragg peaks owing to convolution of the molecular transform with the transform of the finite lattice, and in this case there would also be a greater contribution of inter-lattice interference to the resultant diffraction intensities. Because the twin-related lattices are essentially in phase along the *c* axis, this blurring or streaking effect would be limited to the $a^* \times b^*$ plane. The reasonably successful refinement of the structure of the trigonal MJ1267 crystal as a perfect twin suggests that the Bragg diffraction is dominated by lattice domains of large enough spatial extent that inter-lattice interference is not a major contributor to the observed intensities. Nonetheless, a minority population of twin domains of smaller spatial extent could potentially contribute to the streaking observed in the $a^* \times b^*$ plane in the diffraction pattern from these crystals (Figs. 1*b* and 1*c*).

However, more complexity in the lattice structure is also possible given the promiscuous packing of parallel and antiparallel 3_1 fibers observed in the dominant lattice. If the fibers were to pack together with totally random orientations uninfluenced by the orientations of their neighbors, a smaller $P3_1$ lattice would result (with unit-cell parameters $67 \times 67 \times 42$ Å, equivalent to those in the lattice shown in Fig. 8*a* below) which would yield sharp Bragg peaks with intensities corresponding to the superposition of the structures in the two orientations because of the static but random disorder in the lattice. Because such peaks are not observed in the diffraction pattern, large tracts of a lattice with this structure cannot occur in the MJ1267 crystals. However, a similar pattern of inter-fiber packing interactions but with an intermediate level of orientational correlation between adjacent fibers (*i.e.* a level between totally random and the regular pattern observed in the dominant lattice) would lead to a more complex diffraction pattern characterized by streaking in the $a^* \times b^*$ plane. This streaking would not occur along the c^* axis because the consistent alignment of the fiber axes with the *c* axis of the real-space lattice produces a consistent unit-cell repeat in that dimension.

A qualitative explanation for this streaking pattern can be obtained by considering several alternative regular lattice structures that can be formed based on the packing of parallel and antiparallel 3_1 fibers of MJ1267 (Fig. 8). All of these alternative lattices require a parallel packing interaction to be made between fibers of identical conformation, *i.e.* exactly the kind of intermolecular interaction that is likely to occur at the boundary between twin domains (as shown schematically in Fig. 7). Positioning four parallel fibers of identical conformation at the positions of the adjacent 3_1 screw axes in the dominant lattice gives an alternative $P3_1$ lattice with an identical cell spacing along the *c* axis as in the dominant lattice but with substantially shorter cell spacings along the *a* and *b* axes (Fig. 8*a*). Making parallel packing interactions of this kind at more widely separated 3_1 screw positions in the dominant lattice can produce two different kinds of alternative lattices with $P2_1$ symmetry (Figs. 8*b* and 8*c*). These lattices are also characterized by an identical cell spacing along the *c* axis but different cell spacings along their *a* and *b* axes. However, the preferred packing pattern observed in the dominant lattice indicates that the packing interactions between parallel fibers of identical conformation that are required to form these three alternative lattices are somewhat unfavorable. Consistent with this inference, large tracts of the alternative lattices do not occur in the crystal because sharp diffraction spots are not observed at the alternative spacings in the $a^* \times b^*$ plane. However, lattice zones of the kinds shown in Fig. 8 involving a small number of contiguous unit cells would produce strong smearing of the diffraction in the $a^* \times b^*$ plane, as observed in the diffraction from the trigonal MJ1267 crystals, owing to convolution of the molecular transform with the transform of a lattice of small spatial extent in the $a \times b$ plane. Statistical correlations in the packing between parallel and antiparallel 3_1 fibers could be modeled as random arrays of small lattice zones like these so that the qualitative features of the diffraction expected in this more complex situation can be understood on the basis of these simplified examples.

In summary, if intermolecular interactions occur between parallel fibers with identical conformation with any significant frequency, all of the pathologies in the diffraction properties of the trigonal crystal form of MJ1267 can be explained. Presumably, the energy of the packing interactions in the lattice characterized by molecular replacement is lower than that of the other lattices, accounting for its domination of the diffraction pattern. However, the observation of hemihedral twinning of the dominant lattice combined with the streaking in the $a^* \times b^*$ plane of the diffraction pattern suggests that a parallel packing interaction between fibers of identical conformation is energetically accessible, even if it is less favorable than the more complicated packing interactions between fibers of non-identical conformation that are observed in the dominant lattice.

3.7. Conclusions

In spite of the highly pathological diffraction properties of the trigonal crystal form of MJ1267, it was possible to reliably characterize the protein packing in the dominant lattice using molecular-replacement analysis in conjunction with difference Fourier analysis of heavy-metal derivatives at engineered cysteine positions. The dominant lattice suffers from hemihedral twinning, which results from promiscuous packing

interactions between parallel and antiparallel fibers with 3_1 symmetry. These promiscuous inter-fiber interactions are also likely to be the cause of the systematic streaking of the diffraction pattern in the $a^* \times b^*$ plane. The preference for one particular packing arrangement involving two parallel fibers and one antiparallel fiber can be explained by small but significant differences in the conformations of the two parallel fibers, which are possible because of the conformational flexibility of the protein combined with the inherent plasticity in the intermolecular interactions that hold the lattice together. However, while this plasticity allows formation of the dominant lattice that was characterized by molecular replacement, it probably also allows the slightly variant packing interactions that cause both the hemihedral twinning and the streaking in the $a^* \times b^*$ plane of the diffraction pattern.

The authors wish to acknowledge productive discussions with Phillip Thomas, Wayne Hendrickson, Nathan Karpowich and Paul Smith as well as assistance with protein production from Linda Millen and support during synchrotron data collection from Craig Ogata and Randy Abramowitz of the NSLS. We also thank an anonymous reviewer for helpful suggestions for the improvement of the text.

References

Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. (1994). *Nature (London)*, **370**, 621–628.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
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.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dassa, E., Hofnung, M., Paulsen, I. T. & Saier, M. H. (1999). *Mol. Microbiol.* **32**, 887–889.
Davidson, A. L. (2002). *J. Bacteriol.* **184**, 1225–1233.
Diederichs, K., Diez, J., Greller, G., Muller, C., Breed, J., Schnell, C., Vonrhein, C., Boos, W. & Welte, W. (2000). *EMBO J.* **19**, 5951–5961.
Drenth, J. (1994). *Principles of Protein X-ray Crystallography*. New York: Springer–Verlag.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Gaudet, R. & Wiley, D. C. (2001). *EMBO J.* **20**, 4964–4972.
Gottesman, M. M., Pastan, I. & Ambudkar, S. V. (1996). *Curr. Opin. Genet. Dev.* **6**, 610–617.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Higgins, C. F. (1992). *Annu. Rev. Cell Biol.* **8**, 67–113.
Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P. & Tainer, J. A. (2000). *Cell*, **101**, 789–800.
Hung, L. W., Wang, I. X., Nikaido, K., Liu, P. Q., Ames, G. F. & Kim, S. H. (1998). *Nature (London)*, **396**, 703–707.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Karpowich, N., Martsinkevich, O., Millen, L., Yuan, Y.-R., Dai, P. L., MacVey, K., Thomas, P. J. & Hunt, J. F. (2001). *Structure*, **9**, 571–586.
Klein, I., Sarkadi, B. & Varadi, A. (1999). *Biochim. Biophys. Acta*, **1461**, 237–262.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Linton, K. J. & Higgins, C. F. (1998). *Mol. Microbiol.* **28**, 5–13.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Mosser, J., Douar, A. M., Sarde, C. O., Kioschis, P., Feil, R., Moser, H., Poustka, A. M., Mandel, J. L. & Aubourg, P. (1993). *Nature (London)*, **361**, 726–730.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Paulsen, I. T., Sliwinski, M. K. & Saier, M. H. (1998). *J. Mol. Biol.* **277**, 573–592.
Redinbo, M. R. & Yeates, T. O. (1993). *Acta Cryst.* D**49**, 375–380.
Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z. Zielenski, J., Lok, S., Plavsic, N., Chou, J., Drumm, M. J., Iannuzzi, M. C., Collins, F. S. & Tsui, L. (1989). *Science*, **245**, 1066–1073.
Saurin, W., Hofnung, M. & Dassa, E. (1999). *J. Mol. Evol.* **48**, 22–41.
Smith, P. C., Karpowich, N. K., Millen, L., Moody, J. E., Rosen, J., Thomas, P. J. & Hunt, J. F. (2002). *Mol. Cell*, **10**, 139–149.
Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). *EMBO J.* **1**, 945–951.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Yeates, T. O. & Fam, B. C. (1999). *Structure Fold Des.* **7**, R25–R29.
Yeates, T. O. & Rees, D. C. (1987). *Acta Cryst.* A**43**, 30–36.
Yoshida, M. & Amano, T. (1995). *FEBS Lett.* **359**, 1–5.
Yuan, Y. R., Blecker, S., Martsinkevish, O., Millen, L., Thomas, P. J. & Hunt, J. F. (2001). *J. Biol. Chem.* **276**, 32313–32321.