Molecular symmetry-constrained systematic search approach to structure solution of the coiled-coil SRGAP2 F-BARx domain

Michael Sporny, Julia Guez-Haddad, David G. Waterman, Michail N. Isupov and Yarden Opatowsky

Molecular symmetry-constrained systematic search approach to structure solution of the coiled-coil SRGAP2 F-BARx domain

Michael Sporny,[a] Julia Guez-Haddad,[a] David G. Waterman,[b,c] Michail N. Isupov[d]* and Yarden Opatowsky[a]*

[a]The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 52900, Israel, [b]STFC Rutherford Appleton Laboratory, Didcot, England, [c]CCP4, Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot, England, and [d]Biosciences, University of Exeter, Exeter, England. *Correspondence e-mail: m.isupov@exeter.ac.uk, yarden.opatowsky@biu.ac.il

SRGAP2 (Slit–Robo GTPase-activating protein 2) is a cytoplasmic protein found to be involved in neuronal branching, restriction of neuronal migration and restriction of the length and density of dendritic postsynaptic spines. The extended F-BAR (F-BARx) domain of SRGAP2 generates membrane protrusions when expressed in COS-7 cells, while most F-BARs induce the opposite effect: membrane invaginations. As a first step to understand this discrepancy, the F-BARx domain of SRGAP2 was isolated and crystallized after co-expression with the carboxy domains of the protein. Diffraction data were collected from two significantly non-isomorphous crystals in the same monoclinic C2 space group. A correct molecular-replacment solution was obtained by applying a molecular symmetry-constrained systematic search approach that took advantage of the conserved biological symmetry of the F-BAR domains. It is shown that similar approaches can solve other F-BAR structures that were previously determined by experimental phasing. Diffraction data were reprocessed with a high-resolution cutoff of 2.2 Å, chosen using less strict statistical criteria. This has improved the outcome of multi-crystal averaging and other density-modification procedures.

1. Introduction

SRGAP2 regulates dendritic spine maturation, neuronal migration and neuronal branching (Charrier *et al.*, 2012; Guerrier *et al.*, 2009) by coordinating the dynamics of the plasma membrane with the actin-based cytoskeleton (Fritz *et al.*, 2015; Pertz *et al.*, 2008; Wong *et al.*, 2001). To achieve this task, SRGAP2 and the other members of the SRGAP family (humans have four members: SRGAP1–4) utilize a unique domain composition, with an amino-terminal extended F-BAR (F-BARx) domain followed by RhoGAP and SH3 domains, mediating membrane association, cytoskeleton remodelling and protein–protein interactions, respectively. Binding studies of SRGAP proteins to their ligands (*e.g.* the Robo1 receptor) and a low-resolution small-angle X-ray scattering (SAXS) structure of full-length SRGAP4 suggest that the RhoGAP and SH3 domains directly interact with the F-BARx domain, and thus the three domains function in a cooperative manner (Guez-Haddad *et al.*, 2015).

The three-dimensional shape of F-BAR (FES-CIP4 homology) domains, as well as of the other members of the BAR superfamily [*i.e.* BAR (Bin/amphiphysin/RVS), I-BAR (inverse BAR) and N-BAR (N-terminal BAR) domains], is directly linked to their function (Frost *et al.*, 2009; Peter *et al.*, 2004; Kessels & Qualmann, 2015). BAR domains that have
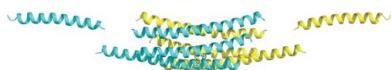
concave membrane-binding 'N-surfaces' associate with membrane invaginations, while I-BARs that have convex N-surfaces associate with protrusions. All BAR domains form elongated antiparallel dimers, in which each protomer contains three primary helices (helices 1–3; Fig. 1a). At the centre of the dimer, where the two protomers overlap, a six-
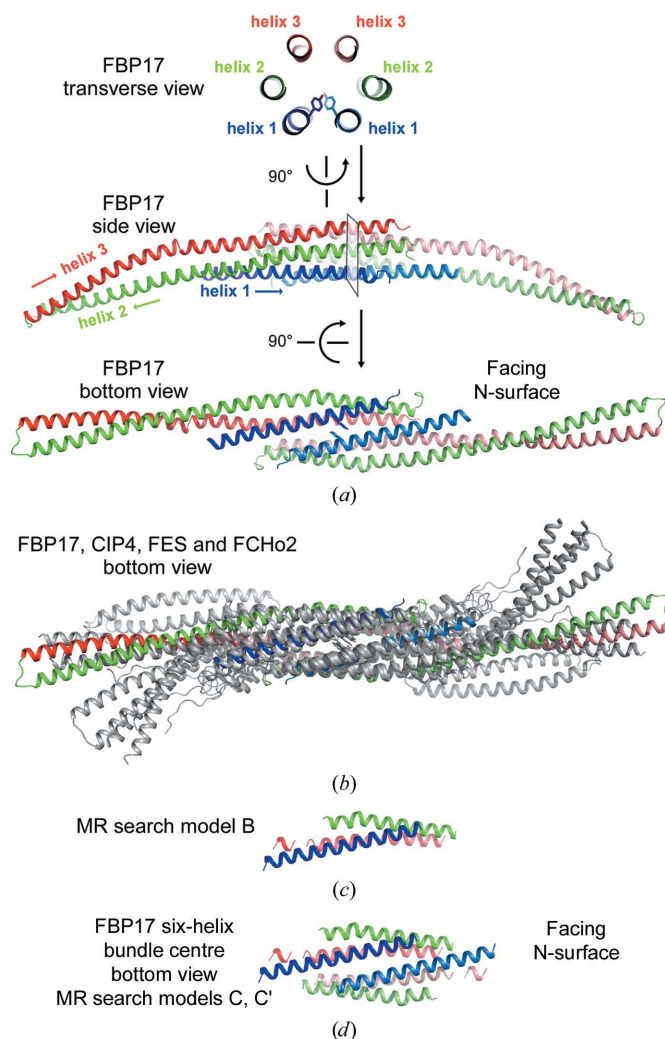

Figure 1
Conservation and variability in F-BAR structures. (a) Transverse (top), side (middle) and bottom views of the canonical F-BAR structure, illustrating the F-BAR domain of FBP17 (PDB entry 2efl) as an example. F-BAR domains are composed of three α-helices and are arranged as homotypic antiparallel dimers. The overlapping centre of the dimer is designated the six-helix core, and the remaining portions of helices 2 and 3 that extend from the six-helix core are the F-BAR arms. Generally, F-BARs bind membranes through their 'N-surface' that includes the N-terminal helix 1. This way, an F-BAR domain with a concave N-surface will associate with cellular membrane invaginations. The sequence conservation between different F-BAR domains is low, with the exception of a specific sequence signature at the helix 1 homotypic interface, with a conserved tyrosine side chain stacked at the centre of the symmetry dyad (seen in the transverse and bottom views, and in Fig. 4e). (b) The superposition of several F-BAR structures demonstrates the structural conservation in the six-helix core region, which is in contrast to the variability of the positions of the F-BAR arms. (c) Based on the structural F-BAR conservation in the core region, search model B does not include the F-BAR arms. (d) Search models C and C' are based on the six-helix core dyad.

helix bundle is formed, leaving two protruding coiled-coil arms at the periphery of the dimer. F-BAR family members that have had their crystal structures determined [*i.e.* FBP17, syndapin, FCHo2, Rgd1p, Imp2 and FES; PDB entries 2efk and 2efl (Shimada *et al.*, 2007), 3hah (Wang *et al.*, 2009), 2v0o (Henne *et al.*, 2007), 4wpe (Moravcevic *et al.*, 2015), 5c1f (McDonald *et al.*, 2016) and 4dyl (Structural Genomics Consortium, unpublished work)] show concave N-surfaces, consistent with their membrane invagination-inducing activity. However, the F-BARx of SRGAP2 is different. It is the largest F-BAR (∼500 residues), mostly owing to a 150-residue 'Fx' carboxy-extension coiled coil. Also, when SRGAP2 is expressed in COS-7 cells it inflicts a robust membrane-protrusions phenotype, the opposite effect to the other F-BARs (Yamazaki *et al.*, 2013; Guerrier *et al.*, 2009). The larger size and inverse membrane-remodelling activity of the SRGAP2 F-BARx probably reflect significant structural deviation from the canonical F-BAR structure. These differences were considered carefully when selecting search models for molecular-replacement (MR) phasing.

A significant proportion (around 40%) of the coiled-coil protein structures determined to date have been phased by experimental methods, as estimated using a PDB keyword search (Rose *et al.*, 2015). For these proteins, MR often fails owing to poor sequence conservation and relative movements of helices in the coiled-coil regions. Moreover, the long α-helical fragments, which are parallel within the single molecule and between neighbouring crystallographically and noncrystallographically related molecules, create multiple identical self- and cross-Patterson vectors, which are highly confusing in standard MR methods (Dauter, 2015). In fact, all of the six primary F-BAR structures mentioned above were determined by experimental phasing, including Imp2, which is relatively similar to Rgd1p, and for which MR was reported to have failed (McDonald *et al.*, 2016).

Recently, significant improvements in methods of structure prediction, coupled with the increased power of density-modification techniques, have greatly increased the success of phasing of coiled-coil proteins by MR methods; for example, by using a remote or theoretical model, as performed by *Rosetta* (Shortle *et al.*, 1998) or *QUARK* (Xu *et al.*, 2012). This is carried out by creating an assembly of most probable coiled models with ideal helices and putting each of these models through an MR/density-modification pipeline such as *AMPLE* (Bibby *et al.*, 2012) or *CCSolve* (Rämisch *et al.*, 2015). Usually, the success of these methods depends on the resolution of the data. For some high-resolution cases the structure could be resolved by density modification, *e.g.* the *SHELXE* phase-extension procedure (Thorn & Sheldrick, 2013), when only a part of the structure is correctly positioned (Thomas *et al.*, 2015). Lower resolution cases, and in particular larger proteins, are usually more challenging since they require better initial positioning of the model helices and more user involvement.

For some cases where the rotation-function (RF) step of MR cannot find a correct orientation, the structures of several nucleic acid (Rabinovich & Shakked, 1984) and protein

molecules (Sheriff *et al.*, 1999) have been solved by an exhaustive six-dimensional MR search, calculating the translation function (TF) for each possible model orientation. Such an evolutionary six-dimensional MR search was implemented in *EPMR* (Kissinger *et al.*, 1999), which semi-randomly explores six-*N*-dimensional space of rotating and translating *N* molecules simultaneously, thus reducing the search calculation times.

The NCS-constrained exhaustive search method (Isupov & Lebedev, 2008) utilizes an approach based on conservation of the quaternary structure between related proteins. Knowledge of the molecular symmetry of the target structure and information derived from the self-rotation function allows a limitation of the number of orientations for which the TF search is conducted and omission of the RF step of the MR method. Provided the conservation of quaternary structure was correctly inferred and the NCS was correctly interpreted, such an approach significantly restricts the search space. This reduces both the calculation times and the probability of finding high-scoring false solution. The NCS-constrained exhaustive search can be conducted with MR programs such as *AMoRe* (Navaza, 1994), *MOLREP* (Vagin & Teplyakov, 2010) and *Phaser* (McCoy *et al.*, 2007). The NCS-constrained exhaustive search has successfully been used for the structure solution of decameric human peroxiredoxin II (Schröder *et al.*, 2000), dodecameric *Bacillus* anti-TRAP protein (Shevtsov *et al.*, 2010) and dimeric *Pseudomonas* 3,6-diketocamphane monooxygenase (Isupov *et al.*, 2015). A related approach of NCS-constrained exhaustive model generation was demonstrated to successfully *ab initio* phase a structure of the helical membrane protein MscL by MR (Strop *et al.*, 2007); however, after test models were generated these were positioned in the unit cell by conventional MR.

The F-BARx domain (residues 1–484) of SRGAP2 can be readily expressed in *Escherichia coli* and purified, but not with high enough homogeneity to be suitable for X-ray crystallography (Wang *et al.*, 2014). Here, we first describe the expression system that we devised in order to overcome this obstacle. We next present the purification, crystallization and X-ray diffraction data-collection procedure that we performed. Finally, we describe the challenges that hampered the standard MR attempts, and the methodology that we used to overcome these difficulties.

2. Experimental

2.1. Design and cloning of SRGAP constructs

The human SRGAP2 full-length (SRGAP2-FL) cDNA clone (KIAA0456) was purchased from ImaGenes GmbH. Human SRGAP4-FL codon-optimized cDNA for bacterial expression (KIAA0131) was synthesized by GenScript. SRGAP4-FL and SRGAP2 F-BARx (residues 1–484) inserts were ligated into a modified pHis-parallel2 vector (Novagen) which includes an amino-terminal hexahistidine tag followed by a TRX fusion protein, a TEV cleavage sequence and a multiple cloning site. The RhoGAP-SH3-C′ term (SRGAP2

residues 487–1071; Fig. 2*b*) insert was ligated into a His tag-deleted pET-28 vector (Novagen). For subcloning, the internal BamHI digestion site in the SRGAP2 cDNA was mutated.

2.2. Expression and purification

All constructs were expressed in *E. coli* BL21 Tuner strain (Novagen) carrying the RIL CodonPlus plasmid, as described in Barak & Opatowsky (2013). For limited proteolysis, TRX-SRGAP4-FL was purified by the following steps: the bacterial lysate supernatant was loaded onto a pre-equilibrated nickel-chelate column (HisTrap, GE Healthcare) with buffer *A* [50 m*M* phosphate buffer pH 8, 400 m*M* NaCl, 5% glycerol, 5 m*M* β-mercaptoethanol (βME)] and then washed and eluted with a buffer *B* gradient (50 m*M* phosphate buffer pH 8, 400 m*M* NaCl, 5% glycerol, 500 m*M* imidazole, 5 m*M* βME). Protein-containing fractions were pooled, incubated with TEV protease [1:50(*v*:*v*)] and dialyzed overnight at 4°C against buffer *A*. The protein was then passed through a nickel-chelate column and further isolated using a size-exclusion column (HiLoad 26/60 Superdex 200, GE Healthcare) pre-equilibrated with buffer *C* (120 m*M* NaCl, 50 m*M* phosphate buffer pH 8, 1 m*M* DTT). The protein was then concentrated using a Vivaspin 20 concentrator (Sartorius). For crystallization, F-BARx was co-expressed with RhoGAP-SH3-C′ term. The proteins were purified on a nickel-chelate column and incubated with TEV protease as described above. F-BARx was then further purified using an increasing salt gradient on an anion-exchange column (Mono Q GL10/100, GE Healthcare) pre-equilibrated with buffer *D* (25 m*M* Tris buffer pH 7.4, 1 m*M* DTT). The F-BARx protein fractions were analyzed using SDS–PAGE, pooled, concentrated to 18 mg ml^{-1} using a Vivaspin 20 concentrator, divided into aliquots and flash-frozen in liquid N$_2$.

2.3. Limited proteolysis of SRGAP4

Chymotrypsin (Sigma–Aldrich) was diluted to 10 µg ml^{-1} in 10 m*M* HEPES pH 7.4, 100 m*M* NaCl buffer and mixed with 2 mg ml^{-1} SRGAP4. The final ratio of protease to protein substrate was 1:1000, and reactions were performed on ice. Proteolysis products were excised from SDS–PAGE and analyzed using LC-ESI-MS/MS.

2.4. Preparation of search models A, B, C and C′

Since all F-BAR domains contain long α-helical stretches, an ideal 30-residue polyalanine helix was generated in *Coot* (Emsley *et al.*, 2010) and used as search model A.

The FBP17 F-BAR structure is the closest homologue of the F-BARx of SRGAP2. It shares only 19% sequence identity over 64% of the length; however, the central region is expected to show structural similarity, since other F-BAR structures that share very little sequence homology have good structural alignment in this area (Fig. 1). The full atomic model of FBP17 (PDB entry 2efl) and the F-BARx domain sequence of SRGAP2 were used to create a sequence-modified model with *MOLREP* (Vagin & Teplyakov, 2010; Lebedev *et al.*, 2008). The resulting model was truncated to leave only the

central regions of helices 1–3 (114-amino-acid search model B; Fig. 1c). Since the asymmetric unit of FBP17 contains a single protomer in a monoclinic space group, model B was merged with its crystallographic symmetry mate to create a six-helix core bundle search model C (Fig. 1d). Model C had a molecular dyad parallel to the crystallographic axis y with the N-surface pointing in the positive direction of y. Since there is an equal probability that the N-surface points in the negative
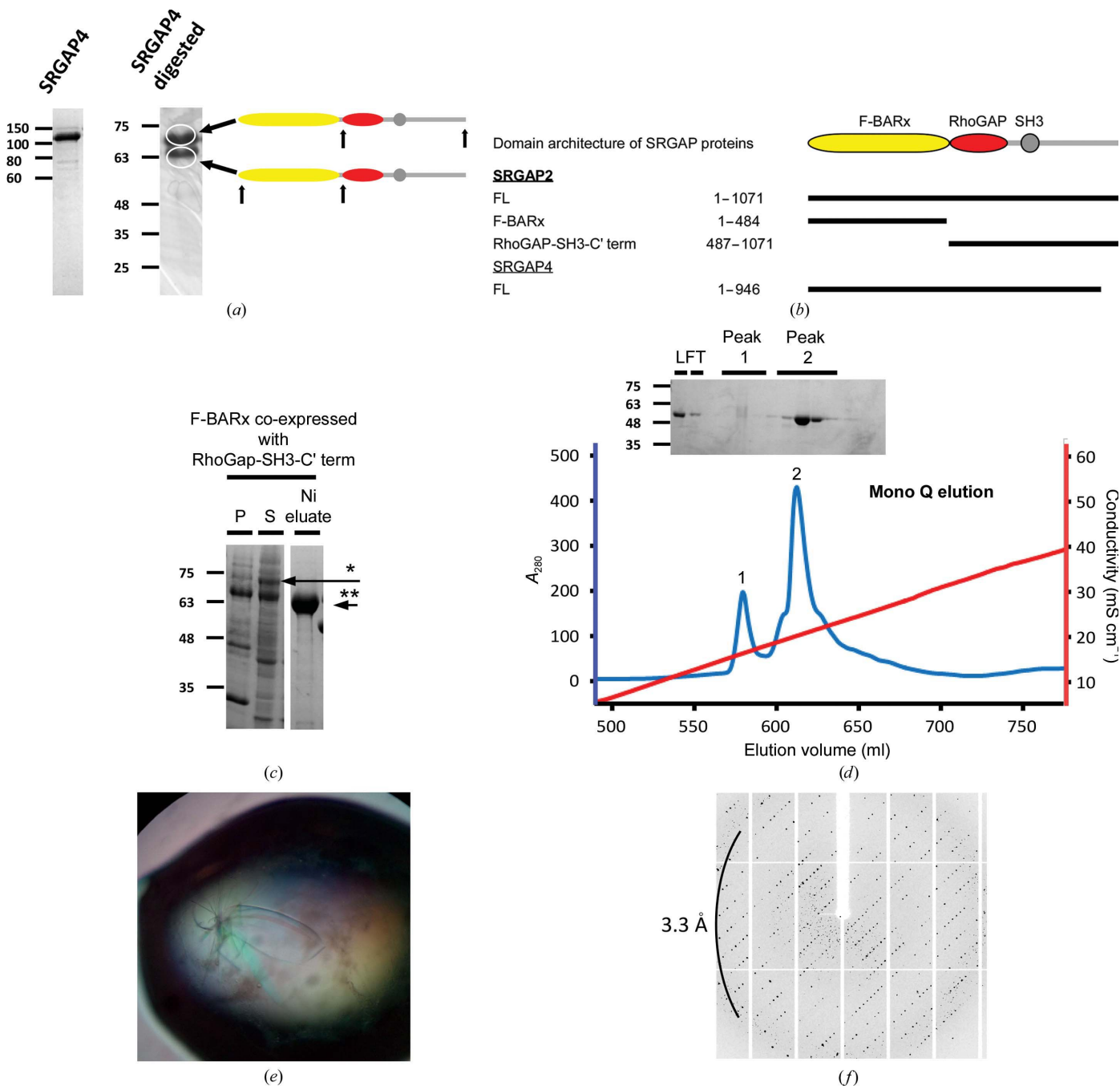

Figure 2
Domain organization, expression and crystallization of SRGAP proteins. (a) Limited proteolysis of full-length SRGAP4. Full-length SRGAP4 was incubated with chymotrypsin for 4 h. Samples were analyzed by SDS–PAGE and identified by LC-ESI-MS to be the amino and carboxy halves of SRGAP4, as indicated. (b) The human SRGAP2, SRGAP4 and SRGAP2 deletion mutants used in this study. In the scheme, the F-BARx, RhoGAP and SH3 domains are indicated and appear as coloured shapes. (c) Expression of F-BARx by co-expression with RhoGAP-SH3-C′ term. E. coli cells expressing F-BARx co-expressed with RhoGAP-SH3-C′ term were lysed and fractionated into insoluble (P) and soluble (S) fractions. The soluble fractions were then purified by metal-chelate nickel chromatography (Ni eluate) and analyzed by SDS–PAGE. Note that co-expression with RhoGAP-SH3-C′ term (marked with double asterisks) allows good solubilization of the F-BARx (marked with a single asterisk). (d) F-BARx Mono Q ion-exchange chromatography and SDS–PAGE analysis of the elution profile. The protein elutes in two peaks, indicating heterogeneity in protein size or conformation. (e) Crystals of F-BARx grown in 0.2 M ammonium citrate tribasic pH 7.25, 22% PEG 1500. Single-crystal dimensions were approximately 0.15 × 0.1 × 0.01 mm. (f) A diffraction image collected on ID29 at ESRF using a PILATUS 6M detector, recorded by long exposure after the data-set collection was completed.

direction of *y* in the crystal, model C was rotated by 180° around the *x* axis using the *CCP*4 (Winn *et al.*, 2011) program *LSQKAB* (Kabsch, 1976) to create model C′.

2.5. Preliminary MR search

As a first attempt, the SRGAP2 F-BARx sequence and data were submitted to the *BALBES*–*ARP*/*wARP* pipeline (Long *et al.*, 2008; Langer *et al.*, 2008) on the CCP4 Online web server (http://www.ccp4.ac.uk/ccp4online). Several search models based on sequence homology did not yield a convincing MR solution. Next, *BALBES* proceeded to use sequence-nonrelated search models and positioned three chains (one of PDB entry 2w74 and two of PDB entry 2w00) with a total size of 494 residues in tight packing (Fig. 3*a*). These structures were not related in sequence and structure to the all-helix target F-BARx, but after five *ARP*/*wARP* rebuild-and-refinement cycles an electron-density map that showed stacks of long helices was produced (Fig. 3*b*). Moreover, twofold antiparallel symmetry reminiscent of the F-BAR hallmark signature was clearly present. Helices were placed using *Coot* (Emsley *et al.*, 2010) and *Buccaneer* (Cowtan, 2006), and after refinement with *PHENIX* (Adams *et al.*, 2010) electron density for amino-acid side chains was visible; however, this could not be matched to the SRGAP2 protein sequence. This model (solution 1) could not be improved by further refinement or density modification.

Next, the SURF option of *MOLREP* was used to conduct a search with the full sequence model (keyword Y) or with the polyalanine model (keyword A). The rotation function was calculated at 3 Å resolution with an integration radius of 18 Å for model A and 25 Å for models B and C, and the translation function (TF) was calculated at 4 Å resolution. For the exhaustive search, only the TF was calculated at 4 Å resolution.

The rotation-function list to 3 Å resolution for model A contained about 70 strong peaks of 6–8σ in height. A related approach has been described previously for *ARCIMBOLDO* (Millán *et al.*, 2015; Rodríguez *et al.*, 2012) and *AMPLE* (Thomas *et al.*, 2015). Five helices (solution 2) of a total of 150 amino acids were consequently positioned by a translation search. In the other case (solution 3), model B was positioned by *MOLREP* with a good score. The dimeric model C could not be confidently positioned in the SRGAP2 unit cell.

2.6. Molecular symmetry-constrained systematic MR search

For the symmetry-constrained search, the *MOLREP* rotation-function table was prepared with the polar angles θ and φ fixed at 90° and −90°, respectively, and χ varied between 1 and 180° with 1° increments. The search was conducted both for the full sequence and the polyalanine models C and C′. The *MOLREP* option NP 200 was used to include all values from the table into a single run. Option PACK N switches off the packing function and was used to allow positioning of a dimer onto the crystallographic dyad. Option NPT 1, which only uses a single peak of the TF, was necessary to position the model on a crystallographic dyad.

(a)

(b)
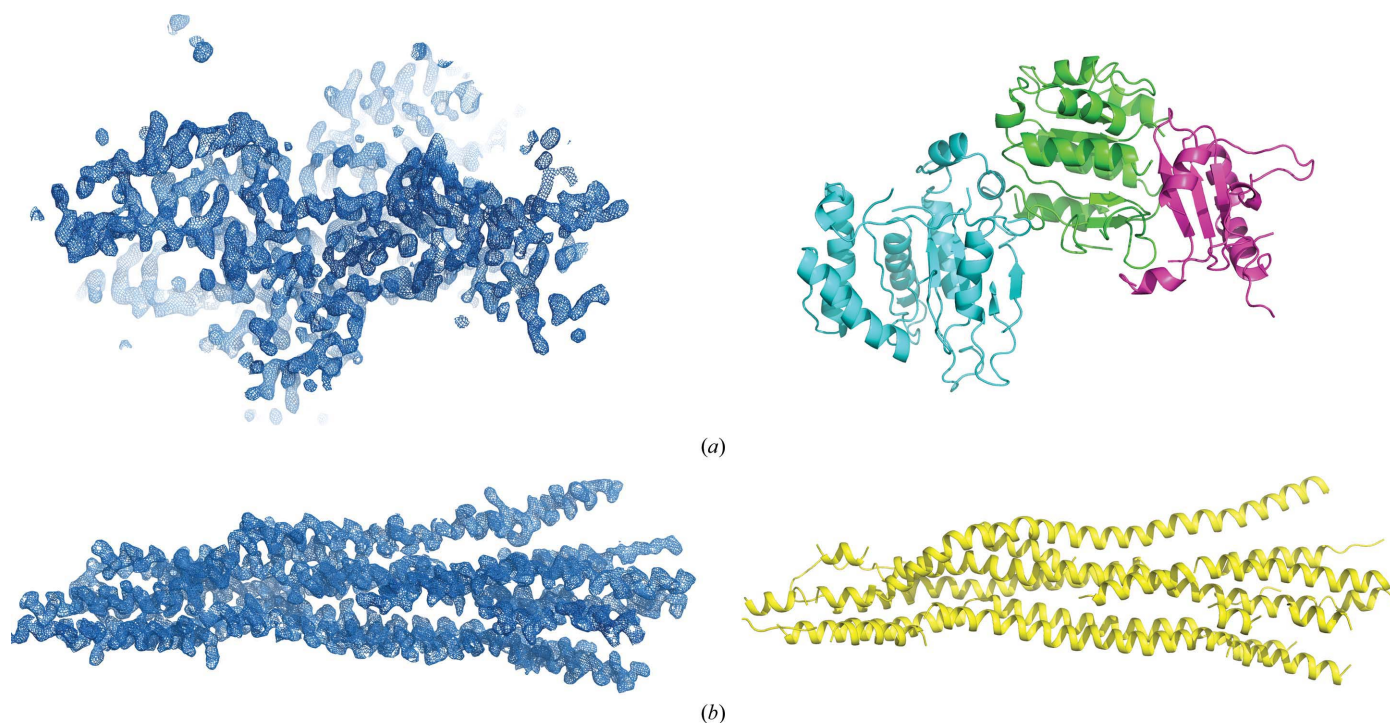
Figure 3
BALBES MR. (*a*) *ARP*/*wARP*-modified electron-density map (left) and the ribbon model of the initial *BALBES* solution (right). The search models had no sequence or structural homology to SRGAP2, yet the resulting *ARP*/*wARP* map showed stacks of α-helices characteristic of F-BAR domains. (*b*) Manual placement of helices, followed by several cycles of refinement and rebuilding, resulted in clearly defined helices organized in a way that is reminiscent of F-BAR structures. However, this structure could not be further refined and did not reveal helix directionality or clear side-chain densities at any point.

Table 1
Summary of the data-processing statistics.

Values in parentheses are for the outer resolution shell.

Crystal	Crystal 1		Crystal 2
Beamline	ID29, ESRF		ID29, ESRF
Wavelength (Å)	0.9763		0.9763
Space group	$C2$		$C2$
Unit-cell parameters (Å, °)	$a = 203.8$, $b = 29.9$, $c = 95.0$, $\alpha = 90.0$, $\beta = 91.9$, $\gamma = 90.0$		$a = 216.9$, $b = 29.6$, $c = 94.7$, $\alpha = 90.0$, $\beta = 92.0$, $\gamma = 90.0$
Solvent content (%)	50.0		52.3
V_M (Å^3 Da^{-1})	2.48		2.60
Processing tool	$DIALS$	XDS	XDS
Resolution range (Å)	101.86–2.20 (2.27–2.20)	47.3–2.89 (3.06–2.89)	47.3–2.94 (3.12–2.94)
Multiplicity	2.8 (2.6)	3.7 (3.7)	3.2 (3.0)
Unique reflections	28869	13202	12658
Completeness (%)	96.2 (94.0)	98.4 (94.5)	94.9 (78.7)
R_{merge} (%)†	15.2 (128.1)	12.1 (73.3)	14.9 (56.2)
$\langle I/\sigma(I) \rangle$	3.3 (0.7)	7.9 (1.9)	5.3 (1.7)
$CC_{1/2}$	0.994 (0.286)	0.996 (0.823)	0.991 (0.821)

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum_{hkl}$ is the sum over all reflections and $\sum_i$ is the sum over i measurements of the reflection.

The program *SOLUTION_CHECK* (Vagin *et al.*, 1998) from the *MoRDa* (Vagin & Lebedev, 2015) molecular-replacement pipeline was used to calculate misset angles and differences in centre-of-mass positions between MR solutions and final structures. The *CCP4* program *CPHASEMATCH* was used to calculate the mean phase difference between models at different stages of refinement and the final refined model. The *PHENIX* package (Adams *et al.*, 2010) was used to calculate correlation coefficients between intermediate models and the final weighted *REFMAC* map.

2.7. Refinement and density modification

All potential solutions for each model were subjected to rigid-body refinement at 4 Å resolution in *MOLREP* and *REFMAC*5 (Murshudov *et al.*, 2011). The resulting models were isotropically refined in *REFMAC*5. The refined models were positioned by *MOLREP* in the unit cell of crystal 2. Multi-crystal averaging was conducted using *DMMULTI* (Cowtan, 2010). The averaged phases were used as input in *REFMAC*5 phased refinement (Pannu *et al.*, 1998). The *SHELXE* phase-extension procedure (Thorn & Sheldrick, 2013) and *ARP/wARP* (Long *et al.*, 2008; Langer *et al.*, 2008) were also used as density-modification techniques. For model rebuilding in *Coot*, *REFMAC* maps calculated using combined phases were used alongside the *SHELXE* map and *ARP/ wARP* maps.

3. Results

3.1. Protein-expression system

Previously (Guez-Haddad *et al.*, 2015), we successfully used *E. coli* for recombinant expression of various human SRGAP2 and SRGAP4 fragments. Of these, the SRGAP2 fragment spanning residues 1–799, which includes the F-BARx, RhoGAP and SH3 domains, and full-length SRGAP4 exhib-

ited good expression levels, high solubility and monodisperse peaks in size-exclusion chromatography. The gel-filtration elution volume and gradient-fixation analyses showed that SRGAP proteins are dimers in solution, consistent with the general principle of F-BAR-mediated dimerization. Here, we used limited chymotrypsin digestion of SRGAP4, followed by LC-ESI-MS analysis, to probe the structural boundaries of the F-BARx domain. This strategy is useful for structural studies in identifying sites with enhanced backbone flexibility (Gao *et al.*, 2005), and has successfully been implemented by us before (Opatowsky *et al.*, 2003; Guez-Haddad *et al.*, 2015). One of the most prominent protein fragments (11 RGLQ...YTQR 486) that resisted a four-hour digestion period spans the F-BARx domain, which includes the canonical N-terminal F-BAR domain and an ~150-amino-acid extension that is predicted to harbour two helices, also known as the F-BAR extension (Fx). Based on the limited proteolysis analysis, we determined the boundaries and ligated the F-BARx domain of SRGAP2 (residues 1–484) into a modified pHis Parallel vector (Novagen), resulting in a Trx-HisTag-TEV-F-BARx fusion protein. The expressed protein was found mostly in the insoluble fraction, and could not be isolated with high homogeneity. Since the larger version of SRGAP2 that also includes the RhoGAP and SH3 domains shows good solubility, we hypothesized that the RhoGAP-SH3 module has a chaperone-like effect on the proper folding and stability of F-BARx. As co-expression of interacting proteins is an established strategy to increase solubility for structural studies (Kerrigan *et al.*, 2011), we co-expressed Trx-HisTag-TEV-F-BARx with the carboxy half of SRGAP2 (residues 487–1071, ligated into a kanamycin-resistant His tag-deleted pET-28 vector; Novagen) in *E. coli*. Under these conditions, large quantities of soluble F-BARx were obtained, separated from the carboxy half of SRGAP2 and purified to homogeneity using consecutive metal-chelate, ion-exchange and size-exclusion chromatography. Importantly, F-BARx retained its solubility better in phosphate buffer than in Tris or HEPES buffers, presumably because of strong electropositive patches on the proteins surface that the phosphate buffer can mask, thereby reducing aggregation.

3.2. Protein crystallization, data collection and processing

F-BARx was screened for crystallization using a mosquito Crystal robot (TTP Labtech) against PEGRx HT, PEG/Ion HT, SaltRx HT, Index HT and Crystal Screen HT (Hampton Research, Aliso Viejo, California, USA) at 293 and 277 K in 96-well hanging-drop clear polystyrene plates. Drop sizes were 0.3 µl, with a 1:1 sample:reservoir screen ratio. A

crystallization hit appeared at 293 K in 0.2 *M* ammonium citrate tribasic pH 7.0, 20%(*w*/*v*) polyethylene glycol (PEG) 3350. Crystallization conditions were refined in 24-well hanging-drop vapour-diffusion plates by varying the pH value and the molecular mass of the precipitant, after which diffraction-grade crystals appeared in 0.2 *M* ammonium citrate tribasic pH 7.25, 22%(*w*/*v*) PEG 1500. Several very thin leaf-shaped crystals were harvested and cryocooled after being soaked in a cryoprotectant solution consisting of 0.2 *M* ammonium citrate tribasic pH 7.25, 35%(*w*/*v*) PEG 1500.

Diffraction data for two F-BARx crystals were measured on the tunable beamline ID29 (de Sanctis *et al.*, 2012) at the European Synchrotron Radiation Facility (ESRF) under standard cryogenic conditions. Data processing of crystals 1 and 2 was carried out using the *XDSAPP* software package (Kabsch, 2010; Krug *et al.*, 2012) to resolutions of 2.89 and 2.94 Å, respectively. Crystal 1 belonged to space group *C*2, with unit-cell parameters *a* = 203.8, *b* = 29.9, *c* = 94.9 Å, β = 91.9°. Crystal 2 was significantly different from crystal 1 (*a* = 216.9 Å), with an *R* factor of 0.41 between the two sets of structure factors according to the *CCP*4 program *SCALEIT*. This allowed phase improvement by multi-crystal averaging.

Matthews probability calculation (Kantardjieff & Rupp, 2003) suggests the presence of one molecule in the asymmetric unit and a solvent content of 54%. For the purpose of density modification the diffraction data from crystal 1 were later reprocessed using both *XDS* and *DIALS* (Waterman *et al.*, 2013) to 2.2 Å resolution. The data-collection statistics are summarized in Table 1.

3.3. Observing 'ghost helices'

In the course of preliminary MR searches, three potential solutions were identified. Several helices were added to the model when potential MR solution 1 was subjected to the *ARP*/*wARP* procedure. After refinement and multi-crystal averaging of both potential solutions 2 and 3, the electron density revealed features which allowed the building of new helices and extensions of the existing helices. Normally, strong helical features in the unassigned electron density are very helpful in building the model, but confusingly, solutions 2 and 3 were distinct from each other and from solution 1, and their refinement statistics did not improve after adding more helices.

Such artificial helical features for coiled-coil proteins are likely to have been observed previously by other groups in the field, but we could not find these described anywhere. Indeed, we have learned that strong helical features were observed in other wrongly positioned coiled-coil models (Ronan Keegan, personal communication).

The term 'ghost helices' reflects their nature, since their position depends on that of the starting model. These 'ghost helices' are likely to be related to significant non-origin peaks in the native Patterson function calculated for the SRGAP2 data (Fig. 4*c*). Such peaks, which are sometimes called 'phantom' (Urzhumtsev *et al.*, 2016), are common for crystals of nucleic acids and coiled-coil proteins. In the SRGAP2 case,

the phantom peaks indicate that the protein contains several helices with axes approximately normal to the crystallographic dyad and nearly parallel to each other and to helices from other asymmetric units.

3.4. Molecular symmetry-constrained search MR

Eventually, we solved the F-BARx structure by applying a molecular symmetry-constrained systematic search approach, for which we used the six-helix core bundle of FBP17 as a search model. The choice of this search model was based on the observation that in spite of the overall variability within F-BAR domain structures that have previously been determined, all have a conserved six-helix core bundle (Figs. 1*a* and 1*b*). Since the asymmetric unit of the SRGAP2 F-BARx crystal contains a single protomer, the molecular dyad of the biological unit dimer must be located on a crystallographic twofold axis.

Macromolecular crystals which have one or more oligomer molecular symmetry axes coinciding with crystallographic axes are very common. Such a crystal organization restrains the possible positions of the oligomer in the unit cell, and can be exploited to simplify MR search/validation or experimental phasing and phase improvement. For example, the virion structure of common cold virus (Rossmann *et al.*, 1985) was situated on a crystallographic threefold axis in the cubic space group *P*2$_1$3. The particle orientation and its centre of mass were found from the self-rotation function, packing considerations and heavy-atom positions. MIR phases were extended from 5 to 3 Å resolution using 20-fold averaging and icosahedral molecular symmetry operators. In a second example, tetrameric tyrosine phenol-lyase (Antson *et al.*, 1992), which has two monomers per asymmetric unit, was phased by MIR. Phases were improved by averaging with NCS operators derived from molecular symmetry 222. In a third case, a tetramer of a remote homologue could be positioned on a crystallographic dyad of alcohol dehydrogenase (Guy *et al.*, 2003) using MAD phases.

The dimer core bundle structure of SRGAP2 (models C and C′; Fig. 4*a*) was positioned by a molecular symmetry-constrained search procedure (Isupov & Lebedev, 2008) that was adapted for the scenario of an oligomer on a crystallographic axis. The six-dimensional search in this case was reduced to two one-dimensional searches. This was achieved by switching off the packing function in *MOLREP* and calculating the TF for all possible rotations of models C and C′ around the crystallographic *y* axis. The φ = 80° rotation proved to be a clear solution (Fig. 4*b*) for the *MOLREP* sequence-modified model C and for a polyalanine model C, although with a reduced contrast. The MR search gave the same solution for data processed with both *XDS* and *DIALS*. The mean phase error calculated for the sequence-modified and poly-alanine model solutions (MR solution in relation to the final refined SRGAP2 model) at 40–3.0 Å (40–5.5 Å) resolution were 87.9° (87.3°) and 87.6° (84.6°), respectively. The correlations calculated between the sequence-modified and poly-alanine MR solutions and the final weighted *REFMAC* map of
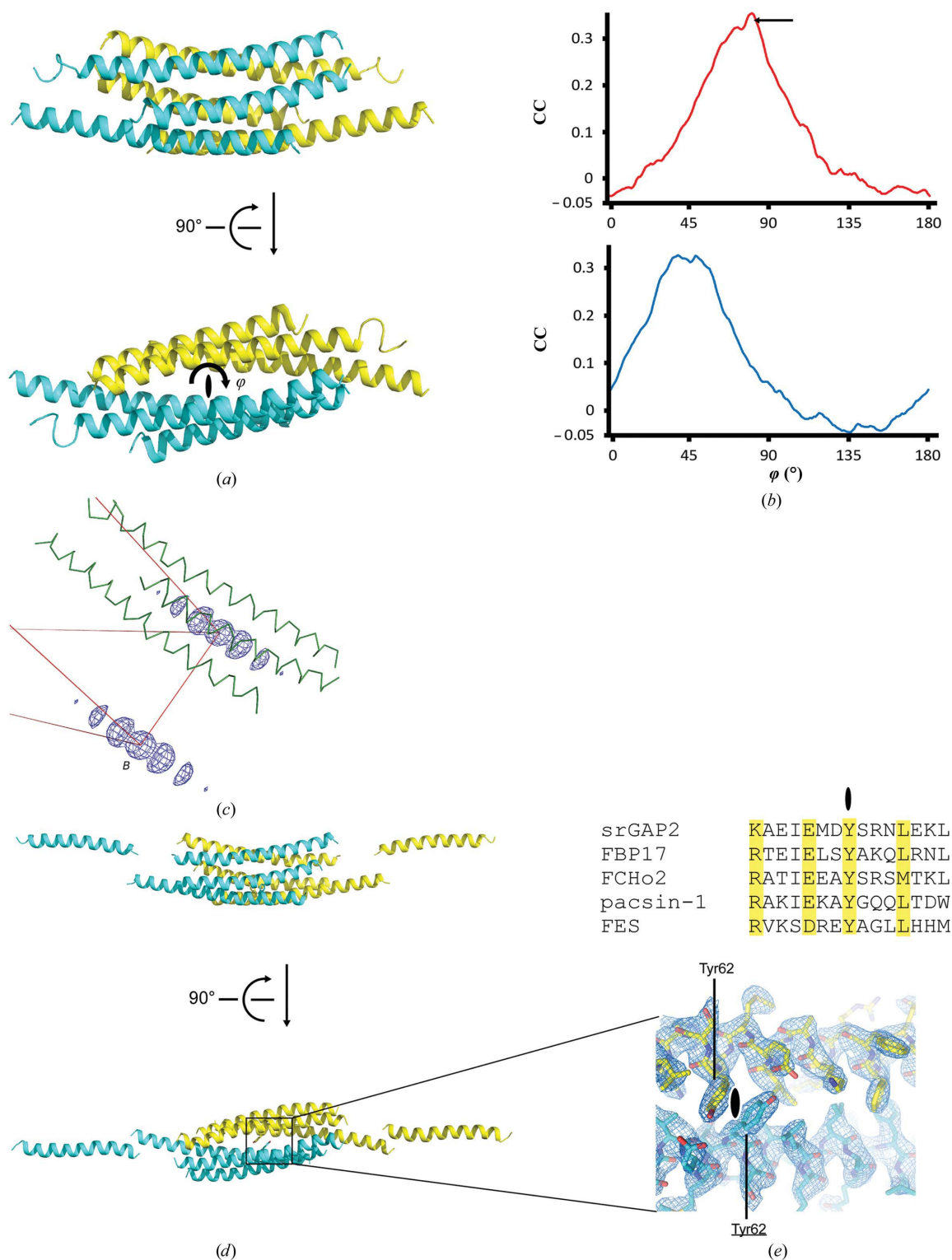
Figure 4
MR solution by the exhaustive molecular-symmetry search approach. (a) Side and top views of the dimeric six-helix core bundle model C located on a crystallographic twofold axis. Helices from different protomers are shown in different colours. (b) TF searches were performed for both sequence-modified models C (top) and C′ (bottom) rotated around the crystallographic twofold axis by the variable angle φ. The correlation coefficient (CC) is plotted as a function of φ. The highest peak at $\varphi = 80°$ for model C corresponds to a correct orientation (marked by an arrow). (c) A diagram showing native Patterson synthesis calculated using SRGAP2 data at 2.7 Å resolution, contoured at 6σ. The exhaustive-search MR solution is shown as a C$^{\alpha}$ trace. The line connecting the peaks is approximately parallel to the axes of the three helices in the core bundle. (d) Side and top views of the partially refined SRGAP2 model for which unambiguous amino-acid assignment for the core bundle helices became feasible. (e) The alignment of the SRGAP2 sequence in the region of the FCH signature motif on the symmetrical helix 1 interface (top). Tyr62 is nearest to the dyad. The $2F_o - F_c$ electron-density map calculated using *REFMAC* combined averaged and partial model phases in the region of the FCH signature motif. The assigned amino-acid side chains are shown with C atoms from the two crystallographic protomers in different colours (bottom).

SRGAP2 were 0.21/0.21 at 3.0 Å resolution and 0.30/0.33 at 5.5 Å resolution, respectively.

3.5. Refinement of the MR solution and density modification

Although the MR solution contrast was higher for the sequence-modified model C, the electron-density maps calculated for the polyalanine model solution revealed more features after partial refinement and multi-crystal density averaging, which is consistent with a smaller mean phase error for this model at low resolution. Therefore, model building and refinement started from the polyalanine MR solution. Rigid-body and positional refinement of the starting model against the 2.89 Å resolution *XDS*-processed data, multi-crystal averaging and phase combination in *REFMAC* suggested the presence of two new helical stretches of about 20 residues in the electron density. This extended model had a mean phase error of 82.3° at 40.0–3.0 Å resolution. After a further two cycles of model rebuilding/refinement the refinement stalled. The electron-density maps were not showing any new features and subsequent analysis showed an increase in mean phase error owing to over-refinement. Inspection of the model showed that residues in the original three helices did not keep their main-chain angles in the course of the initial refinement. With hindsight, *PROSMART* secondary-structure restraints (Nicholls *et al.*, 2012) should have been used in the early stages of *REFMAC* refinement.

To overcome difficulties in refinement, the following two steps were undertaken. Firstly, the core helices were rebuilt as ideal in *Coot*. Secondly, the data for crystal 1 were reprocessed in both *XDS* (Kabsch, 2010) and *DIALS* to a higher resolution of 2.2 Å and a reference resolution of 2.7 Å (Karplus & Diederichs, 2012; Evans & Murshudov, 2013).

The over-refined model could not be rescued by refinement and density modification with higher resolution data. However, when *SHELXE* and multi-crystal averaging were applied to the rebuilt helix model, a significant improvement in the quality of the electron-density maps was observed for higher resolution data. The *DIALS* data had a $CC_{1/2}$ of around 0.3 in the highest resolution shell (Table 1) and were chosen for further refinement, since data processed in *XDS* had a significantly lower $CC_{1/2}$ in the last shell. The rebuilt model containing three core and two additional helices had a mean phase error of 75.4° (correlation of 0.53 to the final *REFMAC* weighted map) at 3 Å resolution after *REFMAC* refinement against reprocessed data.

Model rebuilding was performed conservatively, using maps from *SHELXE* and *REFMAC*, with the PDB output of the *ARP/wARP* procedure used for side-chain assignment. Additional helical features and protein side chains were added only when they were clearly visible in the electron density calculated for combined averaged/partial model phases (Pannu *et al.*, 1998). Three core helices and two additional helices were extended and the amino-acid side chains were fitted into electron density (Figs. 4*d* and 4*e*). Further model rebuilding and refinement is in progress, with the current free *R* factor being just below 30% at 2.2 Å resolution.

To estimate the success of density modification at different stages, the model phases were compared with those from the final SRGAP2 model. The *SHELXE* procedure conducted with both *XDS* and *DIALS* 2.2 Å resolution data sets and the original polyalanine MR solution did not improve the phases, with a resulting mean phase error of 86.2° at 3.0 Å resolution. After the original model refinement, *DMMULTI* multi-crystal averaging and phase combination in *REFMAC*, the mean phase error was 84.7° at 3.0 Å resolution.

The helix-rebuilt model which contained three core and two additional helices had a mean phase error of 75.4° at 3.0 Å resolution. Six cycles of *SHELXE* global auto-tracing at 2.7 Å resolution resulted in no improvement in phases, with a mean phase error of 78.0 and 77.1° for the *XDS* and *DIALS* data, respectively. The phase improvement was significant at 2.2 Å resolution, with a mean phase error of 72.4 and 66.1° at 3.0 Å resolution for the *XDS* and *DIALS* data, respectively, after six cycles of *SHELXE*. *ARP/wARP* failed to build a model at this stage, with the mean phase error increasing to 83.0°.

The mean phase differences between each of the refined potential solutions 1–3 and the final SRGAP2 model are 90°. This confirms that all of these potential solutions were wrong, and that a molecular symmetry-constrained systematic search was the best method for SRGAP2 structure solution.

3.6. Application of the molecular symmetry-constrained search to other F-BAR family members

Finally, we tested whether the molecular symmetry-constrained search approach could be applied to structure solution of the wider family of F-BAR domains, and selected data and models from three representative examples: (i) Rgd1p (PDB entry 4wpe), containing a single protomer per asymmetric unit, for which the molecular symmetry-constrained search is reduced to two searches in one dimension, (ii) Imp2 (PDB entry 5c1f), containing an F-BAR dimer in the asymmetric unit, which requires two four-dimensional searches (a translation search for a table of orientations) and (iii) FCHo2 (PDB entry 2v0o), which has a more complex crystal arrangement in a *C*-centred monoclinic unit cell and contains three protomers in its asymmetric unit, with one of dimers located on a crystallographic dyad.

The three proteins share ~25% identity over ~85% of their sequences, and their structures were solved by experimental phasing. For each case, search models (polyalanine six-helix core bundles) derived from the other two structures were used. The native Patterson synthesis of all three of the above F-BAR structures contained phantom peaks close to the origin, similar to those shown in Fig. 4(*c*).

Since the self-rotation and Patterson functions were difficult to interpret for the FCHo2 data, we concluded that this case is not a good candidate for the molecular symmetry-constrained search.

A molecular symmetry-constrained search (two one-dimensional searches) for Rgd1p with both the Imp2 and FCHo2 search models gave high-contrast solutions. The monomer of Imp2 was positioned with a misset angle of

2.1° and within 0.5 Å of the correct solution according to *SOLUTION_CHECK* (Vagin *et al.*, 1998). This solution could be easily refined without manual model rebuilding. The monomer of FCHo2 was positioned with a significantly higher rotational misset of 7.2° and a displacement of 3.0 Å. This MR solution could not be refined, which is in line with the much higher structural similarity between Rgd1p and Imp2 (in spite of having a comparable sequence identity between the three proteins).

The structure of the Imp2 F-BAR domain (PDB entry 5c1f) was previously solved by experimental phasing with a specific reference to difficulties in MR (McDonald *et al.*, 2016). The asymmetric unit of space group *C*2 contains a dimeric Imp2 molecule with the direction of the molecular dyad (polar angles −67.6, 90.0, 180.0°) clearly visible in the self-rotation function calculated by *MOLREP*. Molecular symmetry-constrained searches for both the FCHo2 and Rgd1p models has clearly identified the correct orientations of both molecules (these had the highest TF scores) but could not position the centre of mass of either dimeric model. The orientations of protomers *A* and *B* of FCHo2 were found with misset angles of 4.5 and 3.1°, respectively, and the Rgd1p protomers had misset angles of 2.3 and 3.8°.

To facilitate the model positioning, the Imp2 data were expanded to space group *P*1 in the same *C*-centred setting using *CCP*4 (Winn *et al.*, 2011), and both FCHo2 and Rgd1p models in the correct orientation were subjected to 90 cycles of *REFMAC*5 jelly-body refinement in this space group. This approach is related to the PC refinement used to discriminate between correct and incorrect orientations of the search model in *X-PLOR* and *CNS* by preliminary refinement in the triclinic unit cell (Brünger, 1996). Preliminary refinement in *P*1 of models in the correct orientation is commonly used for model improvement since there is no fixed origin in the triclinic cell.

No improvement in the FCHo2 model was achieved by the triclinic unit-cell refinement and no translation solution could be found for the preliminary refined model. The Rgd1p model was improved after refinement in *P*1, with misset angles of 1.2 and 4.9° for the two protomers. The search for this refined model in space group *C*2 with varying resolution limits of 3, 4 and 4.5 Å gave a consistent TF solution, bringing the centres of mass of model protomers *A* and *B* within 0.46 and 0.73 Å of their matches in the Imp2 F-BAR structure. *REFMAC* refinement of the identified TF solution in the *C*2 space group, followed by *ARP/wARP*, *SHELXE* and further *ARP/wARP* procedures, without any manual model rebuilding, resulted in a model which refined to a free *R* factor of 42.5% (mean phase difference to the final model of 50.7°) and contained 341 residues out of 616 of Imp2 F-BAR, with 238 residues docked in sequence.

4. Discussion

In this work, we met both biochemical and crystallographic challenges. Both were resolved thanks to our prior knowledge about the structural organization and particular properties of

SRGAP2 and of F-BAR domains in general. Biochemically, the protein fragment of interest, the F-BARx domain of SRGAP2, could not be isolated to high enough homogeneity after recombinant expression using standard approaches, *i.e.* different expression systems and fusion proteins. The crystallographic obstacles to reach a molecular-replacement solution included poor search models, the tendency of coiled coils to generate native Patterson peaks and the appearance of 'ghost helices' in density modification and refinement.

4.1. RhoGAP-SH3 has a chaperone-like effect on F-BARx folding

The observation that intact SRGAP4, and SRGAP2 constructs that include the F-BARx-RhoGAP-SH3 domains, can be isolated to high homogeneity while the F-BARx alone cannot, led us to realise that the RhoGAP-SH3 module may have a chaperone-like effect on the folding and stability of the F-BARx. Therefore, in order to obtain F-BARx suitable for crystallography, we co-expressed it together with the amino half of SRGAP2 that contains the RhoGAP-SH3 domains. Indeed, this strategy had a dramatic positive effect on the amounts and quality of the soluble F-BARx that we could isolate.

4.2. Molecular symmetry facilitates MR search solution and validation

Three unsuccessful attempts to phase SRGAP2 F-BARx by MR produced significantly different potential solutions, none of which could be refined to an R_{free} of below 45%. These solutions had a common, but so far under-reported feature, which was probably related to the difficulties mentioned for MR of coiled-coil proteins by Dauter (2015). The electron density after partial refinement and density modification contained so-called 'ghost helices'. Addition of these helices did not improve the refinement statistics of the structures, which remained distinct from each other.

In the absence of high-resolution data, a different approach was required which would guarantee the correct positioning of a starting partial structure. The conserved structure of the six-helical bundle in F-BAR domain dimers allowed us to adapt a molecular symmetry-constrained systematic search to the case of a molecular dimer located on a crystallographic axis. Only 360 orientations were tested, with the centre of the bundle fixed on a crystallographic dyad. Although only 13% of SRGAP2 F-BARx atoms (polyalanine model) were thus positioned, all additional helical features present in the electron-density maps could now be added to the model with confidence.

We also demonstrate that a molecular symmetry-constrained systematic search can be a useful tool in solution of some other BAR domain proteins. Although Rgd1p, Imp2 and FCHo2 have equivalent sequence identity at around 25%, an FCHo2-based model proved to be less useful than the other two proteins. Superposition of the final structures showed that Rgd1p and Imp2 are significantly closer to each other than to FCHo2, a fact that was not detectable at the primary-structure

level. Therefore, assemblies of superimposed models, such as those used in *Phaser*, *BALBES* or *MoRDa*, would produce better search models for molecular symmetry-constrained search.

Strong phantom peaks in the direction of axes of helical stretches separated by a helical turn distance were present in the native Patterson synthesis calculated with SRGAP2 F-BARx (Fig. 4c). Owing to these features, significant TF peaks were observed not only when the dimeric search model in the correct orientation was positioned on the crystallographic dyad, but also when the model was shifted away from it by a helical turn distance. The *MOLREP* score function (correlation) in some cases was higher for the shifted model. To ensure that the centre of mass of the MR solution is located at the origin, the number of translation-function peaks in *MOLREP* was limited to one, since the highest value of the TF was always observed when the molecular dyad of the model coincided with the crystallographic axis.

4.3. Helical features in electron-density maps

As shown in, for example, Luzzati (1953), the $F_o - F_c$ maps reveal features missing from the model with the maximum weight of 1/2. The limiting value corresponds to a situation when the modelled part of the crystal structure contains no errors and the missing part of the structure is small.

A situation in which molecular replacement produces the correct orientation of the search model, but in an incorrect position, can be considered in similar terms. To simplify things, assume one molecule per asymmetric unit. The true structure (from which the observed diffraction data are derived) can be translated such that molecules in a selected orientation in the true structure overlap with those in the model to a significant extent. Here, the term crystal structure designates an infinite periodic structure, not the asymmetric unit. Hence, the combination of the model phases and measured data leads to the appearance of new artefactual features that do not exist in the true structure. These appearances occur in a position relative to the selected subset of molecules in the model and not to the true origin. If a different reference subset of molecules is selected, in a different orientation, the same considerations apply, but require a different translation; subsets are related by the symmetry operation of the space group, and the number of different subsets equals the number of point-group operations. This consideration shows that each new feature in the difference map is duplicated at different positions, and its weight is no higher than 1/2 divided by the number of crystal point-group operations. Such maps are in most cases non-interpretable, but in low-symmetry space groups, the strongest features, such as helices, may still be recognizable although their locations have nothing to do with the true structure (hence the term 'ghosts').

Translational NCS (Read *et al.*, 2013) presents a special case, in which the maximum weight for missing substructures may be higher than 1/2. The limiting value of 1 is achieved, *e.g.* in an artificial example, where the true structure is represented by a cell which is twice the size, with one of the crystallo-graphic translations being treated as an NCS translation, while every second molecule in each molecular array generated by that translation is removed from the model. It is therefore expected that for the structure with several nearly parallel helices (which can be treated as related by translational NCS) the density for helices not present in the initial model remains interpretable for significantly larger model errors than it would have been for missing parts in a globular protein. This effect proved to be critical for the model-building procedure of SRGAP2 described in this article.

On the other hand, the same effect was apparently responsible for the appearance of misleading 'ghost helices' in initial attempts to solve the SRGAP2 structure, when several helices in the potential MR solutions 1–3 had the correct orientation and relative position but were positioned wrongly in relation to the crystallographic symmetry elements in *C*2. It is worth mentioning here that 'ghost density' was also reported for the data collected from partially disordered crystals, where it was attributed to interference terms between adjacent crystal domains (see, for example, Ponnusamy *et al.*, 2014); our crystals were single crystals and the misleading features in the difference maps were owing to the use of incorrect models for the generation of these maps.

4.4. Refinement and density modification

It appears that the several density-modification techniques used in SRGAP2 model rebuilding and refinement were useful at different stages of the project. In early stages, multi-crystal averaging proved to be a robust phase-improvement procedure, even though only a single unit-cell parameter difference between crystals 1 and 2 was explored. *SHELXE* and *ARP/ wARP* were sensitive to low completeness of the starting model and became useful at later stages, when a more complete model became available and higher resolution data were used.

Even a conservative model-building approach could not prevent secondary-structure breakdown when the positions of the main-chain atoms of residues in helices 1–3 were adjusted in the course of positional refinement. Furthermore, main-chain torsion angles were not preserved, although residues in newly built helical stretches were not affected. Replacement of the starting three helices by ideal helical stretches was required for further refinement and model building.

5. Conclusions

The co-expression of SRGAP2 F-BARx with the C-terminal RhoGAP and SH3 domains of SRGAP2 allowed the production of significant quantities of soluble SRGAP2 F-BARx.

A molecular symmetry-constrained systematic search has successfully been applied for structure solution of the human SRGAP2 F-BARx protein (and two other F-BAR proteins that were previously determined by experimental phasing), using the conserved six-helix bundle core structure of the F-BAR domain family. This has allowed solution and refine-

ment of a medium-resolution structure using a remote model with only 19% sequence identity, which represented 13% (in the polyalanine model) of the asymmetric unit contents. The extension of resolution to include weak anisotropic data was important for successful density modification and refinement of SRGAP2.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Antson, A. A., Strokopytov, B. V., Murshudov, G. N., Isupov, M. N., Harutyunyan, E. H., Demidkina, T. V., Vassylyev, D. G., Dauter, Z., Terry, H. & Wilson, K. S. (1992). *FEBS Lett.* **302**, 256–260.
Barak, R. & Opatowsky, Y. (2013). *Acta Cryst.* F**69**, 771–775.
Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2012). *Acta Cryst.* D**68**, 1622–1631.
Brünger, A. T. (1996). *Methods Mol. Biol.* **56**, 245–266.
Charrier, C., Joshi, K., Coutinho-Budd, J., Kim, J.-E., Lambert, N., de Marchena, J., Jin, W.-L., Vanderhaeghen, P., Ghosh, A., Sassa, T. & Polleux, F. (2012). *Cell*, **149**, 923–935.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Cowtan, K. (2010). *Acta Cryst.* D**66**, 470–478.
Dauter, Z. (2015). *IUCrJ*, **2**, 164–165.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Fritz, R. D., Menshykau, D., Martin, K., Reimann, A., Pontelli, V. & Pertz, O. (2015). *Dev. Cell*, **35**, 78–92.
Frost, A., Unger, V. M. & De Camilli, P. (2009). *Cell*, **137**, 191–196.
Gao, X., Bain, K., Bonanno, J. B., Buchanan, M., Henderson, D., Lorimer, D., Marsh, C., Reynes, J. A., Sauder, J. M., Schwinn, K., Thai, C. & Burley, S. K. (2005). *J. Struct. Funct. Genomics*, **6**, 129–134.
Guerrier, S., Coutinho-Budd, J., Sassa, T., Gresset, A., Jordan, N. V., Chen, K., Jin, W.-L., Frost, A. & Polleux, F. (2009). *Cell*, **138**, 990–1004.
Guez-Haddad, J., Sporny, M., Sasson, Y., Gevorkyan-Airapetov, L., Lahav-Mankovski, N., Margulies, D., Radzimanowski, J. & Opatowsky, Y. (2015). *Structure*, **23**, 1989–2000.
Guy, J. E., Isupov, M. N. & Littlechild, J. A. (2003). *Acta Cryst.* D**59**, 174–176.
Henne, W. M., Kent, H. M., Ford, M. G., Hegde, B. G., Daumke, O., Butler, P. J., Mittal, R., Langen, R., Evans, P. R. & McMahon, H. T. (2007). *Structure*, **15**, 839–852.
Isupov, M. N. & Lebedev, A. A. (2008). *Acta Cryst.* D**64**, 90–98.
Isupov, M. N. *et al.* (2015). *Acta Cryst.* D**71**, 2344–2353.
Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Kerrigan, J. J., Xie, Q., Ames, R. S. & Lu, Q. (2011). *Protein Expr. Purif.* **75**, 1–14.
Kessels, M. M. & Qualmann, B. (2015). *J. Cell Sci.* **128**, 3177–3185.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Krug, M., Weiss, M. S., Heinemann, U. & Mueller, U. (2012). *J. Appl. Cryst.* **45**, 568–572.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 33–39.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Luzzati, V. (1953). *Acta Cryst.* **6**, 142–152.
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.
McDonald, N. A., Takizawa, Y., Feoktistova, A., Xu, P., Ohi, M. D., Vander Kooi, C. W. & Gould, K. L. (2016). *Cell. Rep.* **14**, 534–546.
Millán, C., Sammito, M. & Usón, I. (2015). *IUCrJ*, **2**, 95–105.
Moravcevic, K., Alvarado, D., Schmitz, K. R., Kenniston, J. A., Mendrola, J. M., Ferguson, K. M. & Lemmon, M. A. (2015). *Structure*, **23**, 352–363.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, R. A., Long, F. & Murshudov, G. N. (2012). *Acta Cryst.* D**68**, 404–417.
Opatowsky, Y., Chomsky-Hecht, O., Kang, M. G., Campbell, K. P. & Hirsch, J. A. (2003). *J. Biol. Chem.* **278**, 52323–52332.
Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst.* D**54**, 1285–1294.
Pertz, O. C., Wang, Y., Yang, F., Wang, W., Gay, L. J., Gristenko, M. A., Clauss, T. R., Anderson, D. J., Liu, T., Auberry, K. J., Camp, D. G., Smith, R. D. & Klemke, R. L. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 1931–1936.
Peter, B. J., Kent, H. M., Mills, I. G., Vallis, Y., Butler, P. J., Evans, P. R. & McMahon, H. T. (2004). *Science*, **303**, 495–499.
Ponnusamy, R., Lebedev, A. A., Pahlow, S. & Lohkamp, B. (2014). *Acta Cryst.* D**70**, 1680–1694.
Rabinovich, D. & Shakked, Z. (1984). *Acta Cryst.* A**40**, 195–200.
Rämisch, S., Lizatović, R. & André, I. (2015). *Acta Cryst.* D**71**, 606–614.
Read, R. J., Adams, P. D. & McCoy, A. J. (2013). *Acta Cryst.* D**69**, 176–183.
Rodríguez, D., Sammito, M., Meindl, K., de Ilarduya, I. M., Potratz, M., Sheldrick, G. M. & Usón, I. (2012). *Acta Cryst.* D**68**, 336–343.
Rose, P. W., Prlić, A., Bi, C., Bluhm, W. F., Christie, C. H., Dutta, S., Green, R. K., Goodsell, D. S., Westbrook, J. D., Woo, J., Young, J., Zardecki, C., Berman, H. M., Bourne, P. E. & Burley, S. K. (2015). *Nucleic Acids Res.* **43**, D345–D356.
Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H.-J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. & Vriend, G. (1985). *Nature (London)*, **317**, 145–153.
Sanctis, D. de *et al.* (2012). *J. Synchrotron Rad.* **19**, 455–461.
Schröder, E., Littlechild, J. A., Lebedev, A. A., Errington, N., Vagin, A. A. & Isupov, M. N. (2000). *Structure*, **8**, 605–615.
Sheriff, S., Klei, H. E. & Davis, M. E. (1999). *J. Appl. Cryst.* **32**, 98–101.
Shevtsov, M. B., Chen, Y., Isupov, M. N., Leech, A., Gollnick, P. & Antson, A. A. (2010). *J. Struct. Biol.* **170**, 127–133.
Shimada, A. *et al.* (2007). *Cell*, **129**, 761–772.
Shortle, D., Simons, K. T. & Baker, D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 11158–11162.
Strop, P., Brzustowicz, M. R. & Brunger, A. T. (2007). *Acta Cryst.* D**63**, 188–196.
Thomas, J. M. H., Keegan, R. M., Bibby, J., Winn, M. D., Mayans, O. & Rigden, D. J. (2015). *IUCrJ*, **2**, 198–206.
Thorn, A. & Sheldrick, G. M. (2013). *Acta Cryst.* D**69**, 2251–2256.

Urzhumtsev, A., Urzhumtseva, L. & Baumann, U. (2016). *Methods Mol. Biol.* **1320**, 259–267.

Vagin, A. & Lebedev, A. (2015). *Acta Cryst.* A**71**, s19.

Vagin, A. A., Murshudov, G. N. & Strokopytov, B. V. (1998). *J. Appl. Cryst.* **31**, 98–102.

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

Wang, Q., Navarro, M. V., Peng, G., Molinelli, E., Lin Goh, S., Judson, B. L., Rajashankar, K. R. & Sondermann, H. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 12700–12705.

Wang, H., Zhang, Y., Zhang, Z., Jin, W.-L. & Wu, G. (2014). *Acta Cryst.* F**70**, 123–126.

Waterman, D. G., Winter, G., Parkhurst, J. M., Fuentes-Montero, L., Hattne, J., Brewster, A. S., Sauter, N. K. & Evans, G. (2013). *CCP4 Newsl. Protein Crystallogr.* **49**, 16–19.

Winn, M. D. (2011). *Acta Cryst.* D**67**, 235–242.

Wong, K., Ren, X.-R., Huang, Y.-Z., Xie, Y., Liu, G., Saito, H., Tang, H., Wen, L., Brady-Kalnay, S. M., Mei, L., Wu, J. Y., Xiong, W.-C. & Rao, Y. (2001). *Cell*, **107**, 209–221.

Xu, D., Rahman, N. A. B. D., Othman, R., Hu, P. & Huang, M. (2012). *Proteins*, **80**, 2154–2168.

Yamazaki, D., Itoh, T., Miki, H. & Takenawa, T. (2013). *Mol. Biol. Cell*, **24**, 3393–3405.