Structure of CYRI-B (FAM49B), a key regulator of cellular actin assembly

Elise Kaplan, Rachael Stone, Peter J. Hume, Nicholas P. Greene and Vassilis Koronakis*

Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, United Kingdom. *Correspondence e-mail: vk103@cam.ac.uk

In eukaryotes, numerous fundamental processes are controlled by the WAVE regulatory complex (WRC) that regulates cellular actin polymerization, crucial for cell motility, cell–cell adhesion and epithelial differentiation. Actin assembly is triggered by interaction of the small GTPase Rac1 with CYFIP1, a key component of the WRC. Previously known as FAM49B, CYRI-B is a protein that is highly conserved across the Eukaryota and has recently been revealed to be a key regulator of Rac1 activity. Mutation of CYRI-B or alteration of its expression therefore leads to altered actin nucleation dynamics, with impacts on lamellipodia formation, cell migration and infection by intracellular pathogens. In addition, knockdown of CYRI-B expression in cancer cell lines results in accelerated cell proliferation and invasiveness. Here, the structure of *Rhincodon typus* (whale shark) CYRI-B is presented, which is the first to be reported of any CYRI family member. Solved by X-ray crystallography, the structure reveals that CYRI-B comprises three distinct α-helical subdomains and is highly structurally related to a conserved domain present in CYFIP proteins. The work presented here establishes a template towards a better understanding of CYRI-B biological function.

1. Introduction

Actin filament dynamics are central to a myriad of essential cellular processes such as cell migration, division and intracellular trafficking (Pollard & Cooper, 2009; Rottner *et al.*, 2017; Rottner & Schaks, 2019). Actin assembly is nucleated by cellular machines such as the ubiquitous Arp2/3 complex, which drives the generation of the branched actin networks underlying processes including lamellipodia formation and cell motility (Buracco *et al.*, 2019). Arp2/3 activity is regulated by nucleation-promoting factors (NPFs), which thus provide spatial and temporal control of these processes (Campellone & Welch, 2010; Rotty *et al.*, 2013). The best-characterized NPFs are those belonging to the Wiskott–Aldrich syndrome protein (WASP) family, which stimulate Arp2/3 via C-terminal VCA (verprolin-homology, central and acidic regions) domains (Pollitt & Insall, 2009; Alekhina *et al.*, 2017). The WASP family members WAVE1, WAVE2 and WAVE3 are central to cell motility and protrusion formation (Kurisu & Takenawa, 2009; Rottner & Schaks, 2019), and each functions as part of a heteropentameric complex termed the WAVE regulatory complex (WRC; Chen *et al.*, 2010, 2014). In the resting state, the WAVE VCA domain is hidden in the WRC structure (Ismail *et al.*, 2009; Chen *et al.*, 2010). Multiple signals have been reported that can trigger exposure of the
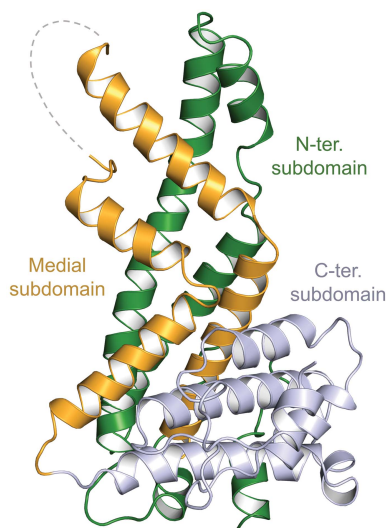
WAVE VCA domain and consequent activation of Arp2/3, the best-characterized of these being interaction with the small GTPase Rac1 (Miki *et al.*, 1998; Ismail *et al.*, 2009; Chen *et al.*, 2010), which is crucial for cell migration (Steffen *et al.*, 2013). There are two possible sites of Rac1 interaction on the WRC. The original 'A' ('adjacent') site is located in the N-terminal part of CYFIP1 (Chen *et al.*, 2010), which is a key component of the WRC, while the newly identified 'D' ('distant') site is situated at the C-terminal extremity of the protein (Chen *et al.*, 2017). While two Rac1 molecules can simultaneously bind to the WRC *in vitro* (Chen *et al.*, 2017), only the 'A' site was found to be essential for its activation *in vivo* (Schaks *et al.*, 2018). Several studies reported that like CYFIP1, FAM49B, a well conserved eukaryotic protein, interacts with the active form of Rac1 (Shang *et al.*, 2018; Fort *et al.*, 2018). FAM49B was consequently renamed CYRI-B for CYFIP-related Rac1 interactor (Fort *et al.*, 2018). CYRI-B binding has been reported to block various Rac1-dependent signalling cascades in the cell, thereby controlling multiple critical cellular functions including T-cell activation (Shang *et al.*, 2018), membrane protrusion, chemotaxis and cell migration (Fort *et al.*, 2018; Whitelaw *et al.*, 2019). By negatively regulating Rac1 signalling, CYRI-B also reduces the entry of several intracellular bacterial pathogens into both phagocytic and non-phagocytic host cells, and indeed is targeted for ubiquitin-mediated destruction by the action of an injected *Salmonella* virulence protein (Yuki *et al.*, 2019). CYRI-B has also been suggested to play a role in certain cancers. Thus, the expression of CYRI-B in pancreatic ductal adenocarcinoma cells is downregulated in the tumour microenvironment, leading to increased cell proliferation and invasion (Chattaragada *et al.*, 2018). Furthermore, knockdown of CYRI-B expression increases cell proliferation in colorectal and liver cancer cell lines (Long *et al.*, 2019), suggesting that CYRI-B may act as a tumour suppressor.

Here, we present the structure of *Rhincodon typus* (whale shark) CYRI-B, which is the first structure of any member of the CYRI protein family. Solved by X-ray crystallography at 2.4 Å resolution using the SAD method, the structure reveals three distinct, entirely α-helical, subdomains and high structural similarity to a conserved domain of CYFIP1. Our work provides a basis to better understand the diverse and important roles of the CYRI protein family in eukaryotic cells.

2. Materials and methods

2.1. Cloning, expression and purification of CYRI-B (FAM49B)

DNA fragments encoding CYRI-B (FAM49B) from *Anthurium amnicola*, *Bison bison bison*, *Crocodylus porosus*, *Homo sapiens*, *Lingula anatina*, *Ornithorhynchus anatinus*, *Orussus abietinus*, *Rhincodon typus*, *Tetranycus urticae* and *Tupaia chinensis* were synthesized (IDT) and cloned into pET-11a using Gibson assembly (Gibson *et al.*, 2009). The fragments were designed to include an N-terminal solubility-enhancement tag (SET2, Stratagene), a hexahistidine tag (His

tag) and a Tobacco etch virus (TEV) protease cleavage site upstream of the CYRI-B coding sequence.

Escherichia coli Rosetta (DE3) cells (Novagen) bearing the appropriate expression plasmid were grown at 37°C in 2×YT medium supplemented with 50 µg ml^{-1} ampicillin and 20 µg ml^{-1} chloramphenicol. When the cells reached an OD$_{600}$ of 0.8, the temperature was reduced to 18°C and protein expression was induced with 0.1 m*M* IPTG. After 16 h at 18°C, the bacteria were harvested by centrifugation at 6000*g* for 10 min at 4°C and the pellet was resuspended in 40 ml lysis buffer composed of 20 m*M* Tris pH 7.4, 150 m*M* NaCl, 0.25 m*M* Bond-Breaker TCEP Solution (Thermo Scientific) and one tablet of EDTA-free protease-inhibitor cocktail (Roche). The bacteria were lysed by cell disruption (Constant Systems) at 207 MPa and the lysate was clarified by centrifugation for 1 h at 60 000*g* and 4°C. The clarified lysate was then incubated with 2 ml Ni–NTA agarose (Qiagen) for 1 h at 4°C. The resin was washed with 20 ml wash buffer (20 m*M* Tris pH 7.4, 500 m*M* NaCl, 0.25 m*M* TCEP) supplemented with 10 m*M* imidazole before elution with the same buffer containing only 150 m*M* NaCl and supplemented with 250 m*M* imidazole. The eluate was dialyzed against 5 l dialysis buffer (20 m*M* Tris pH 7.4, 150 m*M* NaCl, 5 m*M* β-mercaptoethanol). The N-terminal His tag was removed using His-tagged TEV protease produced from plasmid pTH24:TEVSH as described in van den Berg *et al.* (2006). Cleavage reactions utilized 250 µ*M* CYRI-B and 185 µ*M* TEV protease in a total volume of 10 ml and were incubated for 4 h at 22°C. Free His tags, uncleaved CYRI-B protein and TEV protease were removed by several passages over 2 ml nickel resin (Bio-Rad Profinity). Recombinant CYRI-B was concentrated to 10 mg ml^{-1} using a 10 kDa molecular-weight cutoff Amicon ultra centrifugal filter (Millipore), supplemented with 10% glycerol and stored at −80°C.

For preparation of selenomethionine (SeMet)-derivatized CYRI-B, the metabolic inhibition protocol was used (Van Duyne *et al.*, 1993). *E. coli* Rosetta (DE3) cells carrying the CYRI-B expression plasmid were grown in M9 minimal medium supplemented with 50 µg ml^{-1} carbenicillin, 0.2% glucose, 2 m*M* MgSO$_4$, 0.1 m*M* CaCl$_2$ and 0.001% thiamine to an OD$_{600}$ of 0.8. At this point, the culture was supplemented with 60 mg ml^{-1} SeMet, 50 mg ml^{-1} leucine, isoleucine and valine and 100 mg ml^{-1} lysine, phenylalanine and threonine. The temperature was reduced to 18°C and the cells were grown for a further 15 min before the induction of protein expression with 0.1 m*M* IPTG. After a further 16 h, the cells were pelleted by centrifugation at 6000*g* for 6 min. The SeMet-labelled CYRI-B was then purified and the His tag was removed by TEV proteolysis as described for the native protein.

2.2. Analytical size-exclusion chromatography

Size-exclusion chromatography (SEC) was utilized to confirm the oligomeric state of the native and SeMet-derivatized CYRI-B proteins. SEC was performed at room temperature using an ÄKTA pure FPLC system (GE

Healthcare) equipped with a Superdex 75 10/300 GL column in 20 mM Tris pH 7.5, 150 mM NaCl, 5 mM β-mercaptoethanol. A total of 0.5 mg native protein or 0.3 mg of the less soluble SeMet derivative was applied onto the column in a final volume of 100 µl.

2.3. Native and SDS–PAGE gels

The purity and oligomeric state of the recombinant CYRI-B proteins were analyzed using 7% native and 10% SDS–PAGE under non-reducing and reducing conditions, respectively. Protein bands were visualized using Instant Blue (Expedeon).

2.4. Crystallization

Crystallization trials were set up in 2-Drop MRC plates (Molecular Dimensions) using a Mosquito robot (TTP Labtech). Crystals were grown using the sitting-drop vapour-diffusion method. Initial screening was performed using the commercial crystallization screens Structure Screen, MIDAS*plus*, JCSG-*plus* (Molecular Dimensions) and SaltRx (Hampton Research). Subsequent optimization utilized homemade screens.

Native CYRI-B protein at 10 mg ml^{-1} was mixed with the reservoir solution in a 1:2 ratio in a final volume of 1 µl over a reservoir of 80 µl. Crystals were obtained at 15°C in 0.2 M ammonium acetate, 0.1 M MES pH 6.5, 30%(w/v) glycerol ethoxylate using crystals obtained in 0.1 M MES pH 6.5, 7–10% PEG 8000, 20% PEG 1000 as seeds. Single orthorhombic prisms of approximately 300 $\times$ 40 $\times$ 15 µm in size appeared the following day. Crystals were cryoprotected by the reservoir liquor and cooled in liquid nitrogen.

The SeMet-substituted CYRI-B was less soluble than the native protein and was consequently only concentrated to 2.6 mg ml^{-1} in a buffer consisting of 20 mM Tris pH 7.5, 150 mM NaCl, 5 mM β-mercaptoethanol. Crystallization was carried out by mixing the protein with a precipitant solution composed of 0.2 M ammonium acetate, 0.1 M bis-Tris pH 5.5, 25% PEG 3350 in a 1:2 ratio in a final volume of 1 µl. Crystals were obtained after one day with the same seed stock as used for the native protein but were slightly smaller (approximate size 50 $\times$ 20 $\times$ 10 µm). They were cryoprotected with the reservoir solution supplemented with 10%(v/v) glycerol. Images of protein crystals together with a typical diffraction pattern are shown in Supplementary Figs. S1(a) and S1(b).

2.5. Data collection and structure determination

A single-wavelength anomalous diffraction (SAD) data set was collected for SeMet-labelled CYRI-B on beamline I03 at Diamond Light Source (DLS). A total of 720 images (0.04 s exposure, 0.5° oscillation) were collected on a PILATUS3 6M detector from a single crystal by illumination at 12 900 eV (0.9611 Å). The data were indexed and integrated with *iMosflm* (Battye *et al.*, 2011) in space group $P2_12_12_1$ and scaled with *AIMLESS* (Evans & Murshudov, 2013) to a maximum resolution of 2.50 Å. The last 120 images were excluded owing to radiation damage, as judged from the batch $I/\sigma(I)$ and R_{merge} statistics. Anomalous scatterers were located with the

HySS (*Hybrid Substructure Search*; Grosse-Kunstleve & Adams, 2003) module of *Phenix* (Adams *et al.*, 2010; Liebschner *et al.*, 2019). The 15 heavy-atom sites were input into *Phaser* (McCoy *et al.*, 2007) for SAD phasing followed by density modification with *Parrot* (Cowtan, 2010) and model building with *Buccaneer* (Cowtan, 2006). The overall figure of merit for the final SAD phases was 0.48 and increased to 0.57 after the first density modification. After each refinement step with *REFMAC* (Murshudov *et al.*, 2011), model inspection and building were performed with *Coot* (Emsley *et al.*, 2010). The data were finally reprocessed at 2.40 Å resolution and the model was refined with cycles of manual modification with *Coot* and refinement first with *REFMAC* and then with *Phenix* for the last cycles, including two TLS groups and automatic X-ray/stereochemistry weights. The final model has R_{work} and R_{free} factors of 0.2048 and 0.2415, respectively.

The native data set was collected on beamline I04 at DLS. 720 images (0.05 s exposure, 0.5° oscillation) were collected on a PILATUS 6M-F detector. The data were processed with *iMosflm* in space group $P2_12_12_1$ and scaled with *AIMLESS*, giving a data set composed of 13 266 unique reflections to a maximum of 2.37 Å resolution. The structure was solved by molecular replacement with *Phaser* using the model obtained from the SeMet-derivatized data set. The model was then improved by several rounds of manual building with *Coot* and refinement first with *REFMAC* and then with *Phenix* for the last cycles, using two TLS groups and automatic X-ray/stereochemistry constraints. The final model has R_{work} and R_{free} factors of 0.2251 and 0.2772, respectively.

During refinement, eight ordered water molecules were manually added to both the SeMet-derivatized and the native structures. In both data sets, the electron density was too weak to build residues 171–175. Both models were validated using tools from *Coot*, *RAMPAGE* (Lovell *et al.*, 2003) and *PROCHECK* (Laskowski *et al.*, 1993). Data-collection and structure-refinement statistics are listed in Table 1. The coordinates and structure factors of the native and SeMet-derivatized proteins have been deposited in the Protein Data Bank (PDB) with accession codes 6yjk and 6yjj, respectively.

2.6. Electrostatic potential calculations and figure preparation

Electrostatic potential profiles were calculated and visualized in *PyMOL* using the *APBS* plugin (version 2.1; Baker *et al.*, 2001; Jurrus *et al.*, 2018) and PQR files generated by the *PDB2PQR* server (Dolinsky *et al.*, 2004, 2007). The surface potential was set to ± 4.0 kT e^{-1}. All figures were generated with *PyMOL* version 2.2.2.

3. Results and discussion

3.1. Structure of CYRI-B (FAM49B), an α-helical protein

To facilitate structural studies, we cloned and expressed CYRI-B (FAM49B) from *H. sapiens* and nine homologues. Most of the proteins either did not express well (*O. anatinus* and *T. chinensis*), were unstable after purification (*A. amnicola*,

B. bison bison and *O. abietinus*) or failed, so far, to crystallize (*L. anatina*, *C. porosus*, *T. urticae* and *H. sapiens*). However, we successfully produced crystals of whale shark (*R. typus*) CYRI-B. After optimization and screening of multiple crystals, we were able to obtain diffraction data at 2.37 Å resolution. We attempted to solve the structure by molecular replacement using the DUF1394 domain of CYFIP1 (PDB entry 3p8c; Chen *et al.*, 2010), as it shares 21% sequence identity with whale shark CYRI-B and is predicted to be structurally similar (Yuki *et al.*, 2019). However, these efforts were unsuccessful.

We therefore produced SeMet-substituted protein, which allowed us to determine the structure of CYRI-B by SAD phasing. The native and SeMet-derivatized protein crystals have similar unit-cell dimensions and belong to the same orthorhombic space group, $P2_12_12_1$. They were solved at resolutions of 2.37 and 2.40 Å, respectively. Data-collection and refinement statistics are given in Table 1 and representative electron densities for the native and SeMet-substituted
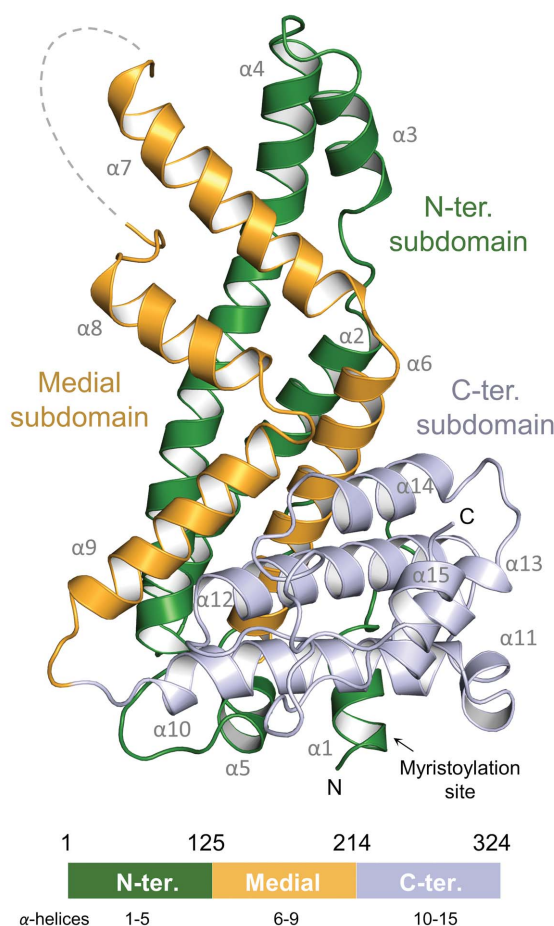
Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer shell.

	SeMet CYRI-B (PDB entry 6yjj)	Native CYRI-B (PDB entry 6yjk)
Data collection		
Temperature (K)	100	100
Wavelength (Å)	0.9611	0.9795
Crystal parameters		
Space group	$P2_12_12_1$	$P2_12_12_1$
a, b, c (Å)	40.4, 72.8, 107.3	39.8, 72.2, 107.8
α, β, γ (°)	90, 90, 90	90, 90, 90
Reflection data†		
Resolution (Å)	72.82–2.40 (2.49–2.40)	60.01–2.37 (2.46–2.37)
Unique reflections	13006 (1329)	13266 (1362)
R_{merge}	0.112 (0.965)	0.122 (0.962)
$\langle I/\sigma(I)\rangle$	12.8 (2.1)	10.1 (2.0)
$CC_{1/2}$	0.998 (0.782)	0.998 (0.590)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
Multiplicity	9.5 (8.2)	10.4 (7.5)
Wilson B factor (Å^2)	44.2	47.1
Refinement‡		
Resolution (Å)	53.6–2.40	60.01–2.37
No. of reflections	12959	13207
R_{work}/R_{free}	0.2048/0.2415	0.2251/0.2772
R.m.s.d., bond lengths (Å)	0.005	0.006
R.m.s.d., bond angles (°)	0.70	0.84
Model composition		
Protein atoms	2570	2573
Waters	8	8
Other	0	0
Model B factors (Å^2)		
Protein atoms	63.2	74.5
Waters	42.5	45.5
Ramachandran statistics§		
Favoured (%)	97.5	97.5
Allowed (%)	2.5	2.5
Outliers (%)	0.0	0.0

† Reflection data are as reported by *AIMLESS* (Evans & Murshudov, 2013). ‡ Refinement statistics are as reported by *PHENIX* (Adams *et al.*, 2010; Liebschner *et al.*, 2019). § Ramachandran statistics are as reported by *RAMPAGE* (Lovell *et al.*, 2003).

Figure 1
Crystal structure of CYRI-B. Cartoon representation of native CYRI-B from *R. typus* showing the N-terminal (green), medial (yellow) and C-terminal (blue) subdomains. The myristoylation site of the protein (Fort *et al.*, 2018) is indicated. A schematic linear organization of the protein subdomains is represented underneath.

structures are shown in Supplementary Movies S1 and S2, respectively. In both cases, a single molecule is present in the asymmetric unit. The two models are nearly identical and could be aligned with a C^α r.m.s.d. of 0.29 Å for 320 atoms.

The structure of CYRI-B (FAM49B) is shown in Fig. 1. It reveals a protein comprised solely of α-helices which can be divided into three linear subdomains. The N-terminal subdomain extends to residue 125 and consists of five α-helices, three of which form a 60 Å elongated α-helical hairpin that runs the entire length of the protein. The medial subdomain extends from residues 126 to 214 and exhibits a 90° broken antiparallel helical hairpin that covers one face of the N-terminal subdomain. The two antiparallel helices α6 and α9 split at the lowest part to surround the extended α4 helix of the N-terminal subdomain. Finally, the C-terminal subdomain (residues 215–324) consists of six α-helices which associate into a globular bundle located at the base of the structure. This bundle caps the lower half of the medial subdomain and is elevated above helices α1 and α5 of the N-terminal subdomain.

In both the native and the SeMet-derivatized structures, the flexible loop composed of residues 171–175 was too disordered to build and is therefore missing from the final models. This loop connects helices $\alpha7$ and $\alpha8$ of the medial subdomain. A non-native serine, residual from the purification strategy, is present at the N-terminus of the structure.

CYRI-B possesses five cysteines: one is located in the N-terminal region, while the C-terminal subdomain has four. In the C-terminal region, the two cysteines Cys231 and Cys253 from helices $\alpha10$ and $\alpha12$ are located too far apart (>3 Å) to form an internal disulfide bond in our structure (Supplementary Fig. S2a). However, it is likely that under non-reducing conditions these two helices are stabilized by a disulfide. Gly2, which is N-myristoylated *in vivo* (Fort *et al.*, 2018), is located at the base of the structure (Fig. 1) and is freely accessible.

Biochemical analyses of *R. typus* CYRI-B were conducted to confirm the oligomeric state of the protein found by X-ray crystallography. As presented in Fig. 2(a), the native and SeMet-derivative proteins migrate as a single band on both native and SDS–PAGE gels. Under denaturing conditions, the proteins migrate with an apparent molecular mass of $\sim$30 kDa, which is close to the theoretical value of 37 kDa for CYRI-B. Further analysis of the purified proteins by size-exclusion chromatography (SEC) revealed a single peak which elutes close to 30 kDa on a calibrated gel-filtration column (Fig. 2b).

The gel electrophoresis and SEC results corroborate the crystallographic data, confirming the monomeric state of CYRI-B in solution.

3.2. Comparison with human CYRI-B and relatives

3.2.1. Whale shark and human CYRI-B are almost indistinguishable. CYRI proteins are ubiquitous and highly conserved, as illustrated by the alignment of sequences from multiple species (Supplementary Fig. S3a). For example, whale shark and human CYRI-B share a sequence identity of 93% (Supplementary Fig. S3b). Thus, over the 324 residues of both proteins, 12 substitutions are conservative, five are semi-conservative and only seven lead to a modification of the amino-acid physico-chemical properties (Fig. 3a). Most of these substitutions localize to residues 10–18, which are located in a loop connecting helices $\alpha1$ and $\alpha2$ within the N-terminal subdomain (Fig. 3b). This region is poorly conserved in CYRI-B proteins (Supplementary Fig. S3a) and is outside the presumed Rac1-binding region (Fort *et al.*, 2018). A few other mutations are located at the apex of this subdomain. The medial and C-terminal subdomains are almost entirely conserved between the two species, containing only two and three conservative substitutions, respectively. As a result, the surface-charge distributions of whale shark CYRI-B and the human homology model are near-identical (Fig. 3c and Supplementary Movie S3).

3.2.2. Comparison of CYRI-A and CYRI-B isoforms. Two isoforms of CYRI proteins exist in many species: CYRI-A and CYRI-B. Some cells express only CYRI-B, such as Jurkat T cells, while others express both (mouse T cells) (Shang *et al.*, 2018). The two proteins have a sequence identity of approximately 80% across multiple species, including whale shark and human (Supplementary Fig. S3b), yet are functionally distinct. While a multitude of roles have been ascribed to CYRI-B, a clear function for CYRI-A has not been defined.

We compared the two isoforms of the whale shark protein and mapped the 13 substitutions which lead to modification of the amino-acid physico-chemical properties onto our CYRI-B structure (Fig. 4a). Most of them are located in the N-terminal subdomain, some on the flexible loop located between helices $\alpha1$ and $\alpha2$ and several along the extended $\alpha4$ helix. Some of these substitutions modify the surface electrostatic potential (Fig. 4b). Thus, CYRI-A is predicted to have a negatively charged patch at the bottom of the structure, where Gly109 in the CYRI-B isoform is replaced by a glutamate in the CYRI-A protein. Conversely, the presence of a glutamate (Glu73), absent in CYRI-A, leads to the formation of a negative patch at the apex of CYRI-B. Sequence analysis reveals that these differences are conserved in CYRI proteins from other species (Fig. 4c), but until more is known about the function of CYRI-A, the significance of these differences remains unclear.
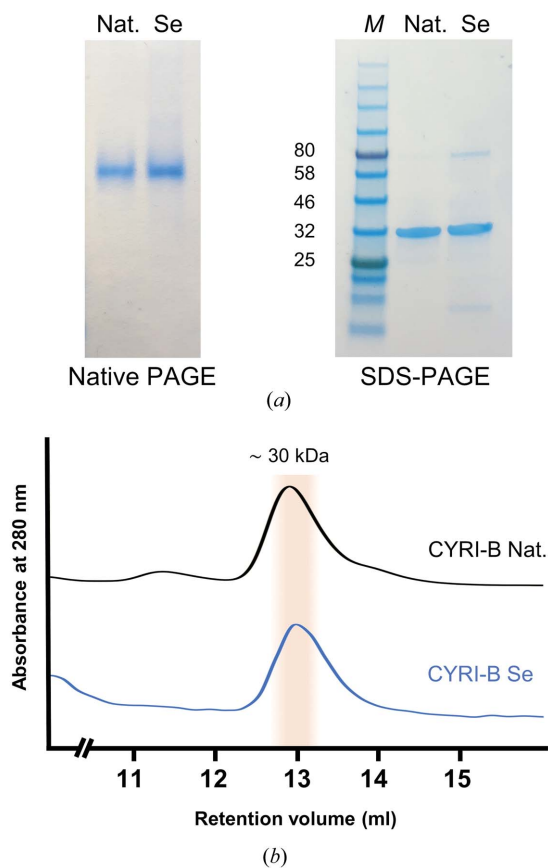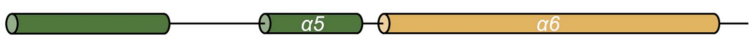

Figure 2
Oligomeric state of CYRI-B. (a) Native PAGE (left) and SDS–PAGE (right) gels of native (Nat.) and SeMet-derivatized (Se) CYRI-B proteins. Molecular masses of protein standards (lane *M*) are shown in kDa. (b) Size-exclusion chromatography profiles of the indicated proteins.
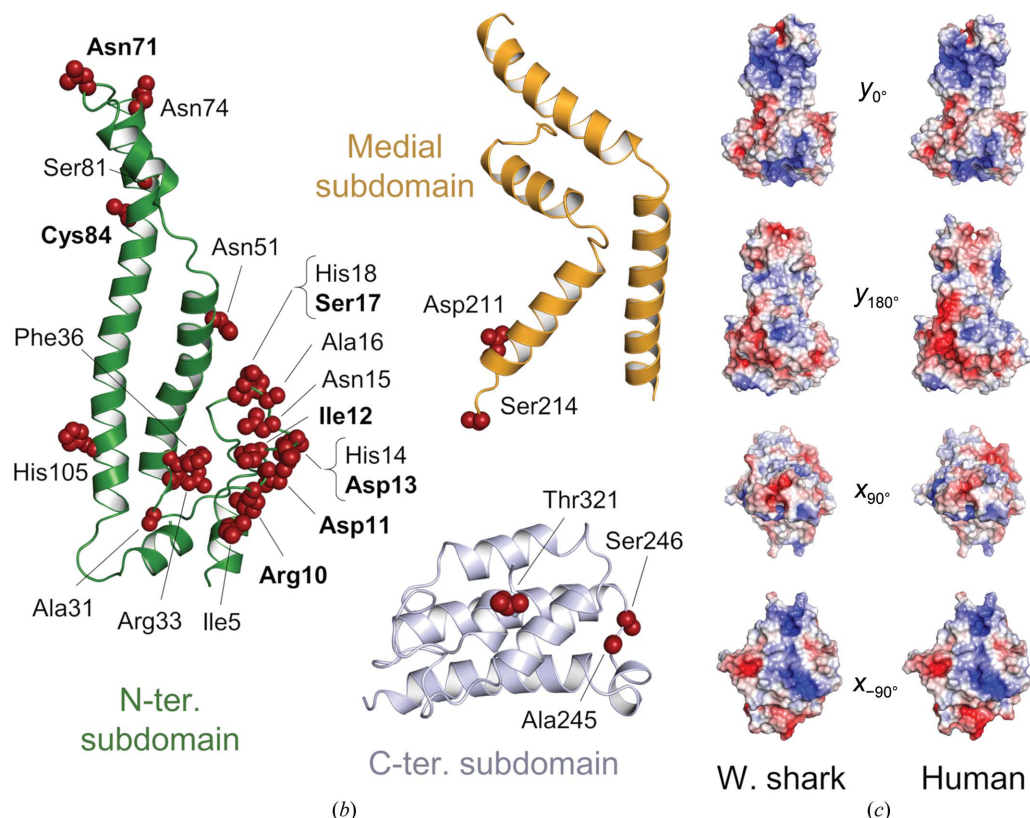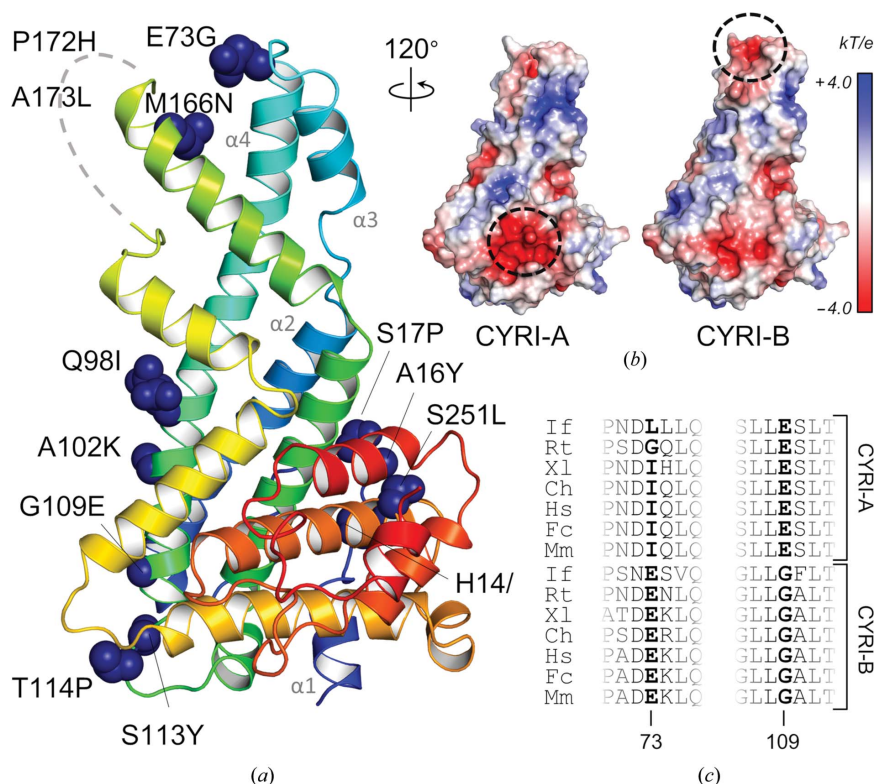
Figure 3
Comparison of human and whale shark CYRI-B reveals high sequence similarity. (*a*) Alignment of whale shark and human protein sequences, where divergent amino acids are displayed in red and identical residues are represented by dots. The secondary structure of whale shark CYRI-B is shown above the sequence using the same colour code as in Fig. 1. The alignment was generated with the NCBI *BLAST* program (Altschul *et al.*, 1997, 2005). (*b*) Side chains of residues that are dissimilar between the whale shark and human proteins are represented as spheres on each subdomain of the whale shark CYRI-B structure. Non-conservative differences between the two proteins are shown in bold in (*a*) and (*b*). (*c*) Views of the electrostatic surface of whale shark (left) and human (*Phyre* homology model; right) CYRI-B proteins. A 360° rotation tour is presented in Supplementary Movie S3.

Figure 4
Comparison of whale shark CYRI-A and CYRI-B. (*a*) Radical substitutions between CYRI-A and CYRI-B as indicated by *Clustal Omega* (Sievers *et al.*, 2011) are shown as blue spheres on the CYRI-B structure. The corresponding residues in CYRI-B are indicated. (*b*) Electrostatic surfaces of CYRI-A (*Phyre* homology model) and CYRI-B (crystal structure). (*c*) Sequence alignment of multiple CYRI-A and CYRI-B proteins highlighting residues 73 and 109 (*R. typus* CYRI-B numbering). Abbreviations are as follows: If, *Ictalurus furcatus*; Rt, *R. typus*; Xl, *Xenopus laevis*; Ch, *Crotalus horridus*; Hs, *H. sapiens*; Fc, *Felis catus*; Ms, *Mus musculus*. The full sequence alignment is provided in Supplementary Fig. S3(*a*).

3.3. Similarity to CYFIP1 from the WAVE regulatory complex (WRC)

The ubiquitous CYFIP protein is a component of the WAVE regulatory complex (WRC). Two CYFIP isoforms exist, CYFIP1 and CYFIP2, which both contain a DUF1394 domain (residues 59–301; human numbering) that is also shared by CYRI proteins. Despite a relatively low sequence identity of 21% between human CYRI-B and the human CYFIP1 DUF1394 domain, the two regions are predicted to be structurally homologous (Yuki *et al.*, 2019). Comparison of the CYFIP1 DUF1394 domain (PDB entry 3p8c) with our structure of CYRI-B confirms this prediction (Fig. 5*a*). The two proteins are strikingly similar and can be superposed with an r.m.s.d. of 3.4 Å over 240 C$^{\alpha}$ atoms and of between 1.8 and 2.4 Å for the individual N-terminal, medial and C-terminal subdomains (Fig. 5*b*).

Interestingly, the disordered loop (residues 169–181) connecting α7 and α8 is six residues longer than the corresponding region of CYFIP1 (residues 195–201). This loop constitutes one of the less conserved regions in the CYRI-A and CYRI-B proteins (Supplementary Fig. S3*a*). Residues 171–175, which are missing in our structure, could not be built based on the CYFIP1 structure as the flanking regions of the loop fold slightly differently in the two proteins (Supplementary Figs. S4*a* and S4*b*). Thus, helix α7 of CYRI-B is one turn longer than the corresponding helix in CYFIP1. The

supplementary turn allows an interaction between Asn74 and the side chain of Gln318 belonging to a symmetry mate. Similarly, helix α8 of CYRI-B terminates one turn earlier, with the loop unfolding towards the back of the structure, where Asn177 interacts with the main chain of Asp290 belonging to a different symmetry mate. Moreover, the position of the loop in CYFIP1 is not compatible with the packing of the CYRI-B crystal (Supplementary Fig. S4*c*). Indeed, the α4 apex of a symmetry mate would clash with an extra α-turn of α8, whereas the packing of the WRC crystal (PDB entry 3p8c) allows both conformations of this linker region.

We examined the conservation of the spatially proximal cysteines Cys231 and Cys253 (see Section 3.1) in CYRI proteins and their CYFIP homologues. These cysteines are conserved in both the CYRI-A and CYRI-B isoforms, but in the CYFIP homologues, only the equivalent of Cys231 is conserved (Supplementary Fig. S2*b*). Examination of CYFIP protein sequences and the CYFIP1 structure (PDB entry 3p8c; Chen *et al.*, 2010) does not reveal a plausible partner for Cys231. The significance of these conserved cysteines for the biological function of CYRI and CYFIP remains to be clarified.

The comparison of the two protein structures also highlights differences in the C-terminal subdomain. In CYRI-B, a loop connects α13 and α14, whereas in the equivalent region of CYFIP1, an 11-residue β-hairpin is present (Fig. 5*a*). This
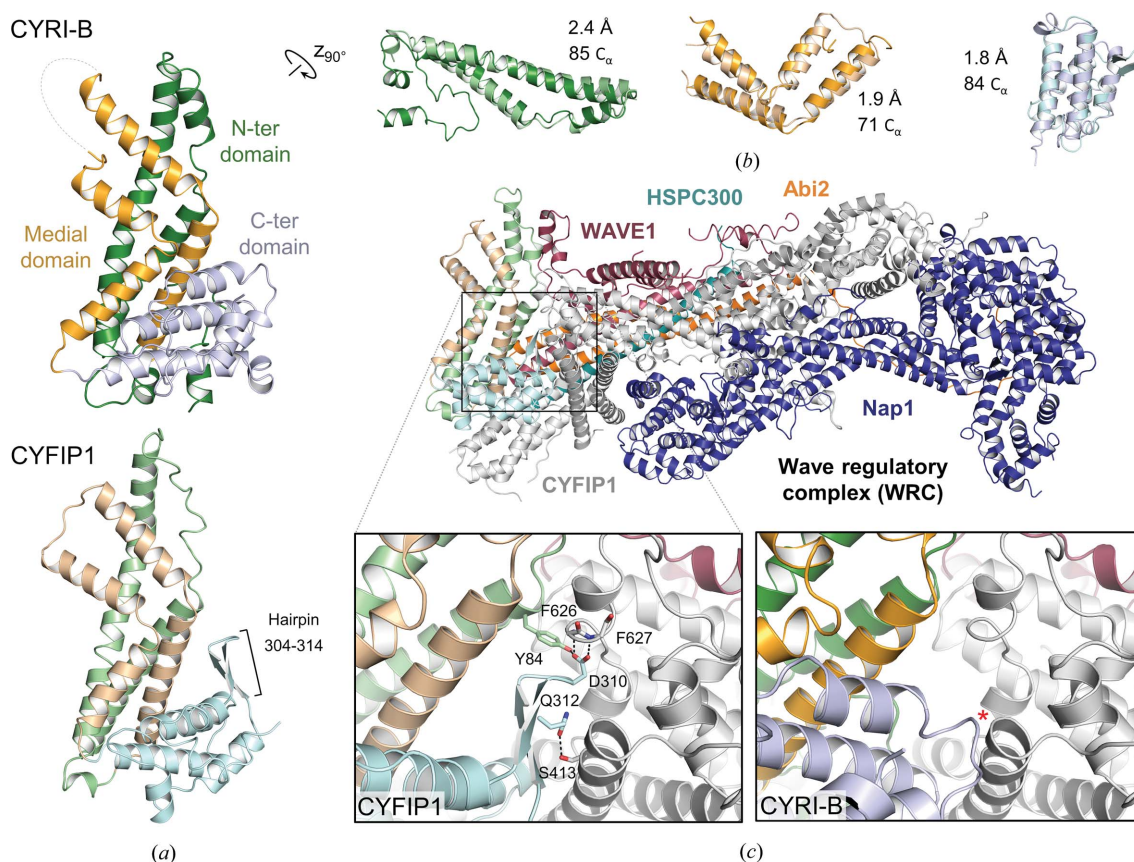
Figure 5
CYRI-B and CYFIP1 present a high degree of structural similarity. (*a*) Side-by-side view of the CYRI-B (left) and CYFIP1 DUF1394 domain (PDB entry 3p8c; right) structures. The colour code of CYRI-B is as in Fig. 1. The presence of an extra hairpin in CYFIP1 is indicated. (*b*) Secondary-structure alignment of the N-terminal (green), medial (yellow) and C-terminal (blue) subdomains of the two proteins. The r.m.s.d. value and number of C^{α} atoms are shown beside each alignment (from *SUPERPOSE*; Krissinel & Henrick, 2004). (*c*) Top, overall structure of the WAVE regulatory complex (WRC; PDB entry 3p8c). The DUF1394 domain of CYFIP1 is coloured as in (*a*). Bottom left, close-up view of the 304–314 hairpin of CYFIP1 showing residues involved in hydrogen-bond interactions with the rest of the protein (grey). Bottom right, superposition of CYRI-B on the DUF1394 domain of CYFIP1 showing the absence of the 304–314 hairpin. An asterisk indicates a steric clash.

hairpin mediates both intradomain and interdomain interactions in CYFIP1 (Fig. 5*c*, bottom left). Thus, the side chain of Asp310, which is located at the apex of the hairpin, forms a hydrogen bond to the Tyr84 hydroxyl of the N-terminal subdomain and the main-chain amides of Phe626 and Phe627 from the FragX-IP domain (residues 389–1222 of CYFIP1). Interaction with the FragX-IP domain is further stabilized by a hydrogen bond between Gln312 and the side chain of Ser413.

CYFIP1 has been shown to bind Rac1 (Chen *et al.*, 2010), and mutations that impair binding form a patch on CYFIP1 (Fig. 6*a*). Two of these mutations, C179R and R190D, are positioned on the DUF1394 domain of CYFIP1. Similarly, Shang and coworkers depicted the interaction of CYRI-B with Rac1 using co-immunoprecipitation and GST pull-down assays (Shang *et al.*, 2018). The association was reinforced with active forms of Rac1, *i.e.* the nucleotide-bound form and the constitutively activated G12V mutant. They subsequently identified a series of mutations which abolished the function of CYRI-B *in vivo*.

We mapped these mutations onto our structure of CYRI-B (Figs. 6*a* and 6*b*). None of the residues are involved in intra-

protein interactions, which is also the case for the two CYFIP1 mutants described to diminish the interaction between a modified version of the WRC and Rac1 (Fig. 6*b*; Chen *et al.*, 2010).

In both CYRI-B and CYFIP1, helix α7, which comprises the majority of the residues that are predicted to constitute the Rac1-binding region, is stabilized by interactions with the antiparallel helix α8 and through a conserved aspartate that interacts with the loop joining helices α2 and α3 of the N-terminal subdomain (Fig. 6*b*, left). Thus, in CYRI-B, Asp155 forms hydrogen bonds to the hydroxyl group of Tyr56 and the backbone amide group of Ala59, while the equivalent CYFIP1 residue, Asp184, forms a hydrogen-bonding network with Trp86, Cys89 and Ser90 (Fig. 6*b*, right).

The highly conserved Arg161 of CYRI-B adopts an identical orientation to the equivalent residue in CYFIP1 (Arg190). This arginine is believed to play a central role in the association with Rac1 (Chen *et al.*, 2010; Fort *et al.*, 2018; Shang *et al.*, 2018). The side chain of a second conserved arginine (Arg160) in CYRI proteins forms extensive hydrogen bonds to the side-chain carbonyl groups of Asn181 and Asn185 and to the hydroxyl group of Ser188, whereas the
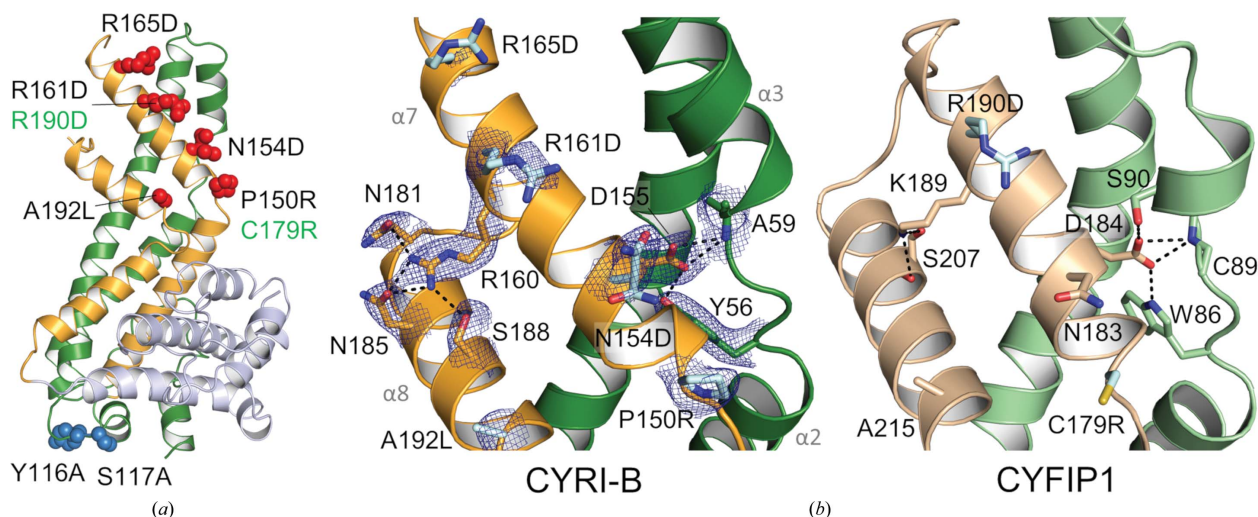
Figure 6
Predicted association of Rac1 with CYRI-B and CYFIP1. (*a*) Site-directed mutagenesis data mapped onto the *R. typus* CYRI-B crystal structure. Mutations in CYRI-B (black labels) that impair (red spheres) or do not affect (blue spheres) its function *in vivo* (Shang *et al.*, 2018) are shown. Green labels indicate mutations in the CYFIP1 DUF1394 domain that perturb the association of the WRC complex with Rac1 (Chen *et al.*, 2010). (*b*) Close-up views of the predicted Rac1-interacting regions of CYRI-B (left) and CYFIP1 (right), coloured as in Fig. 5. Residues leading to effective mutations in (*a*) are represented as cyan sticks with the corresponding mutated residue indicated. Electron densities of CYRI-B side chains are shown as a blue mesh contoured at a σ level of 1.

equivalent residue in CYFIP1 (Lys189) interacts with the carbonyl and hydroxyl groups of Ser207. Interestingly, mutation of the conserved Arg160 to aspartate abolishes the interaction of CYRI-B with the activated Q61L Rac1 variant (Fort *et al.*, 2018). It is possible that this is owing to a local disruption of the secondary structure as Arg160 stabilizes the interaction of α7 and α8. Alternatively, the presence of two consecutive positively charged amino acids may be crucial for the recruitment of Rac1, in which case Arg160 may alter its conformation to interact with specific residues of Rac1.

4. Conclusion

In summary, we have revealed the structure of CYRI-B, also called FAM49B, the only described negative regulator of cellular actin assembly mediated by the WAVE regulatory complex (WRC; Fort *et al.*, 2018; Whitelaw *et al.*, 2019). Our structure, the first to be reported of any member of the CYRI family, reveals a protein that is entirely composed of α-helices and organized into three distinct subdomains. CYRI-B exhibits significant structural homology to CYFIP1, a component of the WRC that interacts with the small GTPase Rac1 to promote actin polymerization (Bompard & Caron, 2004; Pollitt & Insall, 2009). The similarity in structure suggests that CYRI-B interacts with Rac1 in the same manner as CYFIP1 and therefore regulates actin-dependent processes by competition with the WRC for Rac1 binding. Residues previously described as vital for CYRI-B function (Shang *et al.*, 2018) cluster on a conserved helix which is likely to be central to Rac1 interaction. This study provides the structural basis to understand the function of CYRI proteins in a variety of fundamental cellular processes.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* D**66**, 213–221.
Alekhina, O., Burstein, E. & Billadeau, D. D. (2017). *J. Cell Sci.* **130**, 2235–2241.
Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A. & Yu, Y.-K. (2005). *FEBS J.* **272**, 5101–5109.
Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Berg, S. van den, Löfdahl, P. Å., Härd, T. & Berglund, H. (2006). *J. Biotechnol.* **121**, 291–298.
Bompard, G. & Caron, E. (2004). *J. Cell Biol.* **166**, 957–962.
Buracco, S., Claydon, S. & Insall, R. (2019). *F1000Research.* **8**, F1000 Faculty Rev-1977.
Campellone, K. G. & Welch, M. D. (2010). *Nat. Rev. Mol. Cell Biol.* **11**, 237–251.
Chattaragada, M. S., Riganti, C., Sassoe, M., Principe, M., Santamorena, M. M., Roux, C., Curcio, C., Evangelista, A., Allavena, P.,

Salvia, R., Rusev, B., Scarpa, A., Cappello, P. & Novelli, F. (2018). *Oncogene*, **37**, 697–709.

Chen, B., Chou, H.-T., Brautigam, C. A., Xing, W., Yang, S., Henry, L., Doolittle, L. K., Walz, T. & Rosen, M. K. (2017). *eLife*, **6**, e29795.

Chen, B., Padrick, S. B., Henry, L. & Rosen, M. K. (2014). *Methods Enzymol.* **540**, 55–72.

Chen, Z., Borek, D., Padrick, S. B., Gomez, T. S., Metlagel, Z., Ismail, A. M., Umetani, J., Billadeau, D. D., Otwinowski, Z. & Rosen, M. K. (2010). *Nature*, **468**, 533–538.

Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.

Cowtan, K. (2010). *Acta Cryst.* D**66**, 470–478.

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G. & Baker, N. A. (2007). *Nucleic Acids Res.* **35**, W522–W525.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). *Nucleic Acids Res.* **32**, W665–W667.

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.

Fort, L., Batista, J. M., Thomason, P. A., Spence, H. J., Whitelaw, J. A., Tweedy, L., Greaves, J., Martin, K. J., Anderson, K. I., Brown, P., Lilla, S., Neilson, M. P., Tafelmeyer, P., Zanivan, S., Ismail, S., Bryant, D. M., Tomkinson, N. C. O., Chamberlain, L. H., Mastick, G. S., Insall, R. H. & Machesky, L. M. (2018). *Nat. Cell Biol.* **20**, 1159–1171.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. & Smith, H. O. (2009). *Nat. Methods*, **6**, 343–345.

Grosse-Kunstleve, R. W. & Adams, P. D. (2003). *Acta Cryst.* D**59**, 1966–1973.

Ismail, A. M., Padrick, S. B., Chen, B., Umetani, J. & Rosen, M. K. (2009). *Nat. Struct. Mol. Biol.* **16**, 561–563.

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.-W., Holst, M. J., McCammon, J. A. & Baker, N. A. (2018). *Protein Sci.* **27**, 112–128.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Kurisu, S. & Takenawa, T. (2009). *Genome Biol.* **10**, 226.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.

Long, Y., Marian, T. A. & Wei, Z. (2019). *Biochem. Biophys. Res. Commun.* **513**, 1027–1034.

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

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Miki, H., Suetsugu, S. & Takenawa, T. (1998). *EMBO J.* **17**, 6932–6941.

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.

Pollard, T. D. & Cooper, J. A. (2009). *Science*, **326**, 1208–1212.

Pollitt, A. Y. & Insall, R. H. (2009). *J. Cell Sci.* **122**, 2575–2578.

Rottner, K., Faix, J., Bogdan, S., Linder, S. & Kerkhoff, E. (2017). *J. Cell Sci.* **130**, 3427–3435.

Rottner, K. & Schaks, M. (2019). *Curr. Opin. Cell Biol.* **56**, 53–63.

Rotty, J. D., Wu, C. & Bear, J. E. (2013). *Nat. Rev. Mol. Cell Biol.* **14**, 7–12.

Schaks, M., Singh, S. P., Kage, F., Thomason, P., Klünemann, T., Steffen, A., Blankenfeldt, W., Stradal, T. E., Insall, R. H. & Rottner, K. (2018). *Curr. Biol.* **28**, 3674–3684.

Shang, W., Jiang, Y., Boettcher, M., Ding, K., Mollenauer, M., Liu, Z., Wen, X., Liu, C., Hao, P., Zhao, S., McManus, M. T., Wei, L., Weiss, A. & Wang, H. (2018). *Proc. Natl Acad. Sci. USA*, **115**, E4051–E4060.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Mol. Syst. Biol.* **7**, 539.

Steffen, A., Ladwein, M., Dimchev, G. A., Hein, A., Schwenkmezger, L., Arens, S., Ladwein, K. I., Holleboom, J. M., Schur, F., Small, J. V., Schwarz, J., Gerhard, R., Faix, J., Stradal, T. E. B., Brakebusch, C. & Rottner, K. (2013). *J. Cell Sci.* **126**, 4572–4588.

Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.

Whitelaw, J. A., Lilla, S., Paul, N. R., Fort, L., Zanivan, S. & Machesky, L. M. (2019). *Commun. Integr. Biol.* **12**, 112–118.

Yuki, K. E., Marei, H., Fiskin, E., Eva, M. M., Gopal, A. A., Schwartzentruber, J. A., Majewski, J., Cellier, M., Mandl, J. N., Vidal, S. M., Malo, D. & Dikic, I. (2019). *Nat. Microbiol.* **4**, 1516–1531.