



Two independently folding units of *Plasmodium* profilin suggest evolution via gene fusion

Saligram Prabhakar Bhargav¹ · Juha Vahokoski¹ · Juha Pekka Kallio^{3,4} · Andrew E. Torda⁵ · Petri Kursula^{1,2,6} · Inari Kursula^{1,3,4,6}

Received: 15 April 2015 / Revised: 13 May 2015 / Accepted: 18 May 2015
© Springer Basel 2015

Abstract Gene fusion is a common mechanism of protein evolution that has mainly been discussed in the context of multidomain or symmetric proteins. Less is known about fusion of ancestral genes to produce small single-domain proteins. Here, we show with a domain-swapped mutant *Plasmodium* profilin that this small, globular, apparently single-domain protein consists of two foldons. The separation of binding sites for different protein ligands in the two halves suggests evolution via an ancient gene fusion event, analogous to the formation of multidomain proteins. Finally, the two fragments can be assembled together after expression as two separate gene products. The possibility to engineer both domain-swapped dimers and half-profilins that can be assembled back to a full profilin

provides perspectives for engineering of novel protein folds, e.g., with different scaffolding functions.

Keywords Actin-binding protein · Crystal structure · β -Hairpin · Modular · Protein interaction · Sequence analysis · Small-angle X-ray scattering

Abbreviations

Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
CD	Circular dichroism
DTT	Dithiothreitol
SEC	Size-exclusion chromatography
SAXS	Small-angle X-ray scattering
SR	Synchrotron radiation
TEV	Tobacco etch virus
TIM	Triosephosphate isomerase

S. P. Bhargav and J. Vahokoski contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s00018-015-1932-0) contains supplementary material, which is available to authorized users.

✉ Inari Kursula
inari.kursula@oulu.fi

¹ Faculty of Biochemistry and Molecular Medicine, University of Oulu, P.O. Box 5400, 90014 Oulu, Finland

² Biocenter Oulu, University of Oulu, P.O. Box 5000, 90014 Oulu, Finland

³ Helmholtz Centre for Infection Research, Notkestrasse 85, 22607 Hamburg, Germany

⁴ German Electron Synchrotron (DESY), Notkestrasse 85, 22607 Hamburg, Germany

⁵ Centre for Bioinformatics, University of Hamburg, Bundesstrasse 43, 20146 Hamburg, Germany

⁶ Department of Biomedicine, University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway

Introduction

Proteins have evolved during ~3.8 billion years to become complex and versatile molecular machines responsible for most life-supporting functions. Despite glimpses into the formation and evolution of certain protein fold families, the complicated processes of acquiring new protein folds are poorly understood. Most certainly, present proteins have formed from small and simple ancestors [1], and modular evolution of proteins from small fragments has been suggested already in connection with the discovery of introns in eukaryotic genes [2, 3]. Multimerization provides several structural and functional advantages, such as increased stability, possibilities for fine-tuning active site architectures and allosteric regulation, formation of larger binding surfaces, and simplified construction of large interaction

networks [4–7]. Many of the reasons favoring oligomerization in the course of evolution are similarly applicable to both homo- and hetero-oligomers. However, assembly of complicated multimeric complexes is costly and error-prone. Thus, in many cases, oligomerization of functionally closely related entities has further led to gene duplication or fusion.

The formation of new proteins via gene duplication has been studied especially in the context of proteins of the triosephosphate isomerase (TIM) barrel superfamily. TIM barrel proteins consist of eight repeating α/β modules organized into a symmetric barrel and have likely formed via sequential gene duplication events [8–10]. Proteins catalyzing consecutive reactions in metabolic pathways can be products of gene duplication or fusion [11, 12]. In fact, fusion of non-identical genes is the most common evolutionary pathway for the formation of multidomain proteins [13–15]. It has been favored especially in the evolution of protein–protein interaction modules, providing the advantage of simplified assembly and topology of protein complexes [16]. As yet, there is little evidence for ancient gene fusion events in the case of small, asymmetric single-domain proteins.

Actin was long held to be a hallmark of eukaryotes, but upon the identification of bacterial actin homologs, it has emerged as a genuinely ancient protein [17, 18]. Actin has more interaction partners than any other eukaryotic protein, and its large filaments are a remarkable example of self-assembly involving a plethora of regulatory proteins [19]. These actin-binding proteins must have evolved hand in hand with the ability of actin to self-assemble [20], which has imposed the need for additional regulation. Accordingly, *Apicomplexa*, which have poorly polymerizing actins, harbor only few actin regulators [21].

Among the core set of actin-regulating proteins are profilins. They are small single-domain proteins that bind monomeric actin and proline-rich sequence motifs on opposite faces [22, 23]. Furthermore, profilin binding to protein ligands is regulated by membrane binding via polyphosphoinositides [24]. Most profilins are strictly monomeric, although tetramerization has been reported for at least human and plant profilins, and some functional relevance of these tetramers has been suggested [25, 26]. Profilins are evolutionarily widespread and present in not only all eukaryotes, but also in viruses and cyanobacteria, which probably acquired them via horizontal gene transfer from vertebrates [27, 28]. Despite the high degree of conservation at the level of 3D structure, the conservation of profilin primary sequences is remarkably low. This is expected for an ancient protein, whose structure is preserved for functional reasons and has had enough evolutionary time to explore a vast range of compatible sequences.

Apicomplexan parasites have a single profilin that has one of the most divergent structures within the protein family [29]. These organisms have presumably branched out from a common eukaryotic ancestor as early as the three kingdoms of animals, plants, and fungi [30, 31]. Here, we show that *Plasmodium falciparum* profilin consists of two independently folding units and suggest that evolution via fusion of two ancestral genes has led to the present profilin fold.

Materials and methods

Protein production and characterization

A synthetic, codon-optimized *P. falciparum* profilin gene (Mr. Gene) was used for the construction of three mutants, where residues 64–69 ($\Delta 6$), 62–71 ($\Delta 10$), and 59–75 ($\Delta 17$) were deleted. The Stratagene QuikChange Lightning kit was used for site-directed mutagenesis. The constructs were cloned into the pET-M11 vector, which contains an N-terminal 6xHis tag, followed by a tobacco etch virus (TEV) protease [53] cleavage site. For co-expression of the profilin halves, cDNA fragments encoding the N- and C-terminal halves of *P. falciparum* profilin [amino acids 1–67 (M₁A...NG₆₇) and 68–171 (T₆₈K...SQ₁₇₁), respectively] were cloned into the pETDuet vector (Novagen). The C-terminal fragment included an additional methionine as the first amino acid and a non-cleavable hexahistidine tag after the last residue Gln171.

Wild-type profilin was expressed and purified as described [54]. All mutants were expressed in BL21(DE3) RIPL cells, cell pellets resuspended in lysis buffer [10 mM 2-amino-2-hydroxymethyl-propane-1,3-diol-HCl (Tris-HCl; pH 7.5), 300 mM NaCl, 5–10 % (v/v) glycerol], and lysed by sonication. The clarified supernatant was passed through a HisTrap Ni-affinity column (GE Healthcare), which was washed with the lysis buffer with 500 mM NaCl alone and supplemented with 10 and 25 mM imidazole. The cell pellets containing the half-profilin fragments were lysed and washed in the lysis buffer containing 20 mM imidazole. Finally, the His-tagged proteins were eluted with 300 mM imidazole in the lysis buffer. The affinity-purified wild-type, $\Delta 6$, $\Delta 10$, and $\Delta 17$ proteins were subjected to TEV digestion during an overnight dialysis and then passed through a HisTrap column to remove the His-tagged TEV protease and uncleaved protein. The proteins were then concentrated, and final purification was performed using size-exclusion chromatography (SEC) on a Superdex 75 16/60 column (GE Healthcare) in 10 mM Tris-HCl (pH 7.5) and 50 mM NaCl. SEC for the profilin halves was performed in 10 mM Tris-HCl (pH 7.5) and 100 mM NaCl. Peak fractions were pooled together, concentrated, and stored on ice.

To assess the role of oxidation on the oligomeric state, the wild-type profilin, $\Delta 6$, and $\Delta 10$ were incubated in the presence or absence of 10 mM dithiothreitol (DTT) on ice for 2 h and subsequently analyzed by analytical SEC using a Superdex 75 10/300 GL column (GE Healthcare) in 10 mM Tris-HCl (pH 7.5), 50 mM NaCl with or without 10 mM DTT.

Circular dichroism spectroscopy

The wild-type profilin, $\Delta 6$, and $\Delta 10$ were dialyzed in 10 mM Tris-HCl (pH 7.5), 50 mM NaF and diluted to 1.04, 0.86, and 1.40 mg ml⁻¹, respectively. Synchrotron radiation (SR) circular dichroism (CD) spectra were recorded on the CD1 beamline at the ASTRID storage ring, ISA, Århus (Denmark) in the wavelength range of 180–280 nm in a 100- μ m path length quartz cuvette. The DichroWeb server [55] was used for secondary structure determination using the CDSSTR [56] and SELCON3 [57] algorithms and the SP175 reference set optimized for 190–240 nm [58]. The $\Delta 17$ mutant and the half-profilin fragments were dialyzed in 10 mM phosphate buffer (pH 7.5) and 10 mM phosphate buffer (pH 7.5) with 150 mM NaF, respectively, and diluted to 0.05 and 0.04 mg ml⁻¹, respectively. CD spectra were measured on an Applied Photophysics Chirascan spectropolarimeter for the final measurements in a 1-mm path length quartz cuvette.

Thermal denaturation curves were measured using an Applied Photophysics Chirascan Plus spectropolarimeter equipped with a thermal control unit (Quantum Northwest, TC125) and a direct temperature probe. Proteins were dialyzed into 10 mM Tris (pH 7.5), 10 mM NaF and diluted to 0.16 mg ml⁻¹ (wild type) and 0.20 mg ml⁻¹ ($\Delta 6$). CD spectra were recorded between 190 and 260 nm, using a quartz cuvette with a 0.5-mm path length, a temperature range of 20–80 °C, and a heating rate of 1 °C/min.

Small-angle X-ray scattering and light scattering

SAXS experiments were carried out on beamlines X33 at EMBL/DESY, Hamburg (Germany), and I911-4 at MAX-Lab, Lund (Sweden). The wild-type, $\Delta 6$, and $\Delta 10$ were concentrated to 1–5 mg ml⁻¹, in either 25 mM sodium phosphate (pH 7.5) or 10 mM Tris-HCl (pH 7.5), 50 mM NaCl with or without 2 mM DTT. Analysis of the data was carried out using the ATSAS package [59]. Ab initio models were built using GASBOR [60]. Coupled rigid body and ab initio modeling of the dimers was done using the available crystal structures and BUNCH [61].

Static light scattering to determine the exact molecular mass of the SEC peak containing both profilin fragments was measured using a mini-DAWN TREOS multi-angle static light scattering detector (Wyatt Technology, Europe),

coupled to a refractive index detector (Shodex), after separating the proteins over a Superdex S200 increase 10/300GL column (GE Healthcare), equilibrated with 10 mM Tris-HCl (pH 7.5) and 100 mM NaCl. The molecular mass was determined based on the measured light scattering and refractive index using the Astra v. 5.3.4 software (Wyatt Technology).

Crystallographic methods

The $\Delta 6$ mutant was crystallized at a concentration of 13 mg ml⁻¹ in 1.8 M ammonium sulfate, 0.1 M 2-(*N*-morpholino)ethanesulfonic acid (pH 6) at room temperature. A 3.3-Å data set was collected from a single crystal on a PILATUS 6 M detector at the EMBL-Hamburg beamline P13 at PETRA III/DESY. The data were processed and scaled using the XDS package [62] and XDSi [63]. The wild-type profilin monomer (PDB code 2JKF; [29]) was used as a model for molecular replacement in PHASER [64]. The structure was refined using

Table 1 Data collection and refinement statistics

Data collection ^a	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	84.1, 246.6, 256.8
Resolution (Å)	20.0–3.30 (3.39–3.30)
No. of unique reflections	154,448 (11,332)
Redundancy	3.5 (3.4)
$\langle I/\sigma(I) \rangle$	8.1 (0.7)
R_{meas}^b	0.142 (2.066)
Completeness (%)	99.6 (98.3)
$CC_{1/2}^c$	0.997 (0.365)
Refinement ^a	
Resolution (Å)	20.0–3.30 (3.34–3.30)
$R_{\text{work}}/R_{\text{free}}$	0.240/0.263 (0.389/0.453)
No. atoms	
Protein	20,236
Ligand/ion	40
Average <i>B</i> factors (Å ²)	
Protein	125
Ligand/ion	157
Rms deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.925
Ramachandran plot (%)	
Most favored regions	93.7
Outliers	0.6

^a Values in parentheses are for the highest resolution shell

^b R_{meas} is the redundancy-independent *R* factor [77, 78]

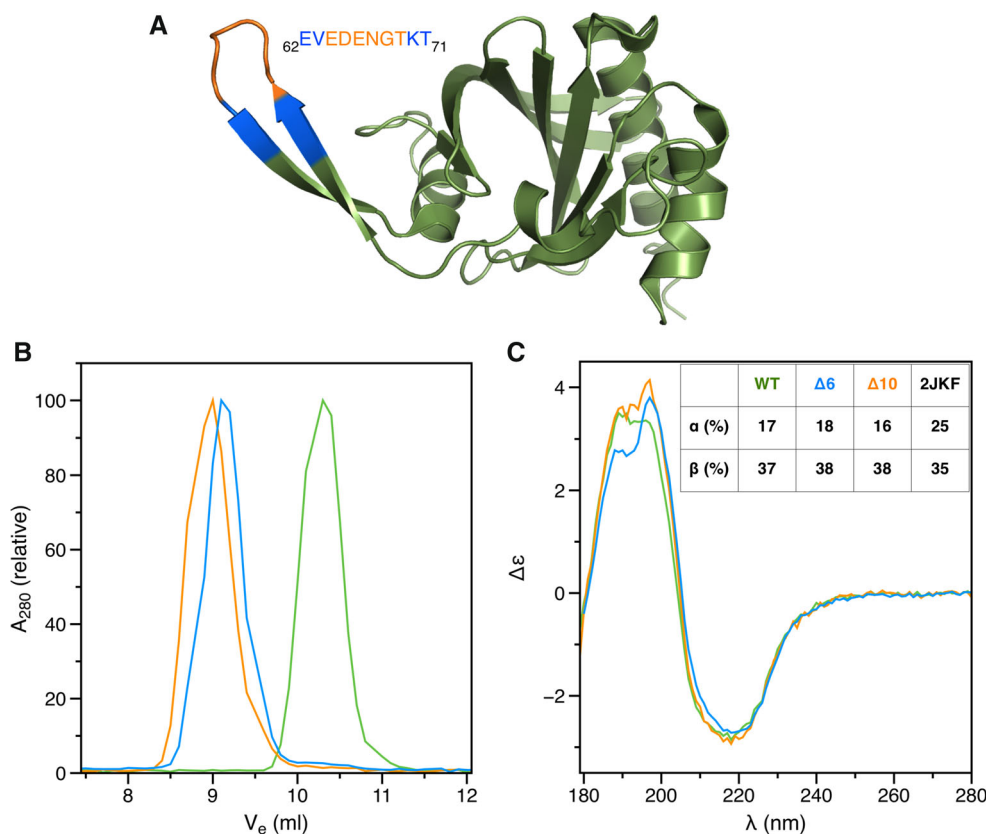
^c $CC_{1/2}$ is defined as the correlation coefficient between two random half-data sets [79]

phenix.refine [65] to final R/R_{free} factors of 0.240 and 0.263, respectively. The structure was validated using Molprobity [66]. Data collection and refinement statistics are shown in Table 1. All 3D structure diagrams were prepared using Pymol.

Sequence analyses

Sequence searches were made against the non-redundant database using PSI-BLAST with three iterations, accepting hits with an e -value $\leq 10^{-9}$, then one round with e -values $\leq 10^{-6}$, and a last round with e -values $\leq 10^{-9}$ [67]. The full-length sequence for each database hit was retrieved, and these were re-aligned using MAFFT in its most accurate mode with up to 200 iterations [68]. Structure-based sequence alignments to 2JKF were used as hard restraints for the sequences from PDB entries 1A0K [69], 1CQA [70], 1F2K, 1PRQ [71], 1YPR [72], 3D9Y [73], 3LEQ, 3NEC [74]. These seeds were removed before calculations of conservation. Sequence entropy S at each site in the alignment was based on only those homologs with an e value less than 10^{-9} and calculated from $S = -\sum_{i=1}^{20} p_i \log_{20} p_i$ where the summation runs over the 20 amino acid types and p_i is the frequency of amino acid type i at the position. SALAMI [75] was used for structure searches and structural alignments.

Fig. 1 Dimerization of profilin mutants. **a** Structure of the wild-type *P. falciparum* profilin (2JKF; [29]). Residues deleted in the different mutants are indicated with colors: orange $\Delta 6$; blue $\Delta 10$. **b** Size-exclusion chromatogram showing dimerization of the deletion mutants (green wild-type profilin; orange $\Delta 6$; blue $\Delta 10$). **c** SRCD spectra of the wild-type and mutant profilins (colors as above). Secondary structure contents calculated from the spectra using DichroWeb and as in the wild-type crystal structure (2JKF) are shown in the inset



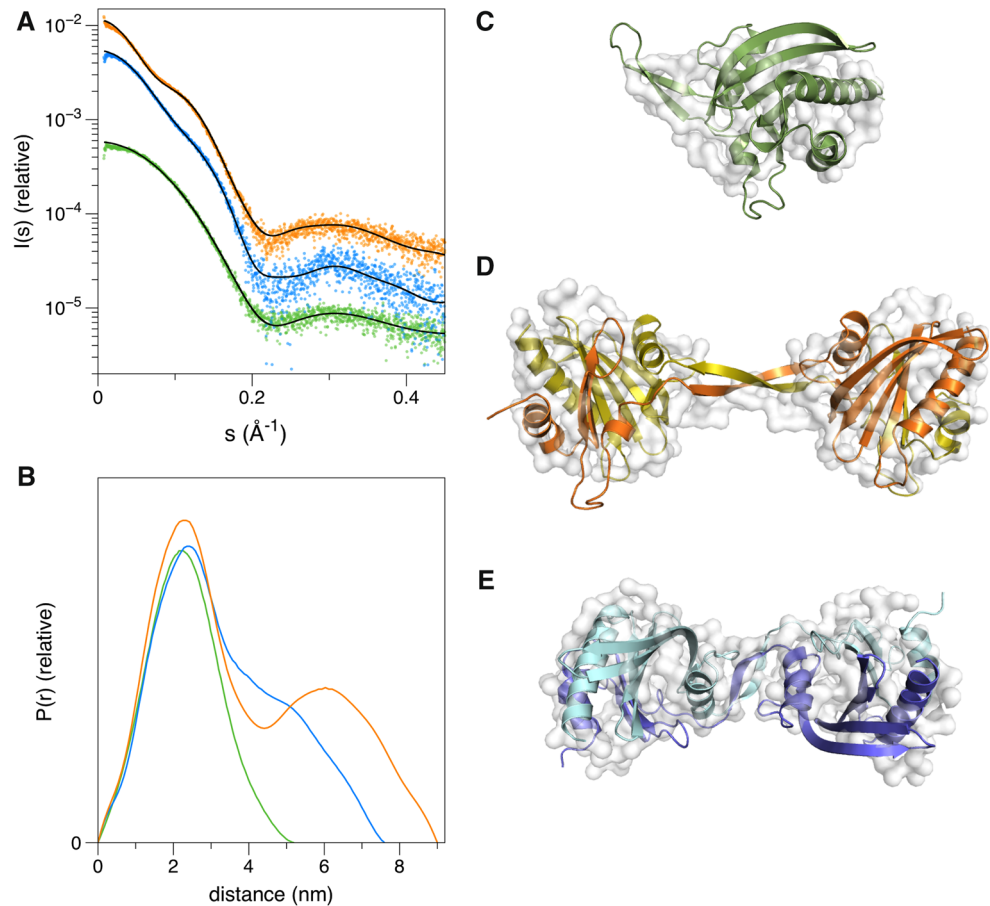
Results

Dimerization of mutant *Plasmodium* profilins

Despite low sequence similarity, *P. falciparum* profilin shares the overall fold with canonical profilins, containing a 7-stranded β sheet sandwiched between two α helices on each side (Fig. 1a). The most striking difference compared to higher eukaryotic profilins is a large β -hairpin extension, which we proposed participates in actin binding [29]. During the course of our work in characterizing the function of this motif, we created mutants that lack 6 ($\Delta 6$), 10 ($\Delta 10$), or 17 ($\Delta 17$) residues of the hairpin loop (Figs. 1a, S1A). Surprisingly, $\Delta 6$ and $\Delta 10$ form dimers, whereas $\Delta 17$ elutes in SEC as three peaks; one corresponding to a monomer, one to a dimer, and a third one in between the two (Figs. 1b, S1B). SRCD spectra indicated nearly identical secondary structure contents for the wild-type and the dimeric mutant proteins (Fig. 1c), while spectral features between 190 and 200 nm indicate minor differences between the structures. The middle peak of the $\Delta 17$ mutant is also folded with somewhat less β strands than the wild-type and the dimeric mutant proteins (Fig. S1C).

Small-angle X-ray scattering (SAXS) was further used to analyze the shape of the two dimers ($\Delta 6$ and $\Delta 10$) in solution (Fig. 2). Whereas the monomeric wild-type profilin has a globular shape similar to the crystal structure, $\Delta 6$

Fig. 2 Solution structures of the wild-type and mutant profilins. **a** SAXS curves and fits of the crystal structures (wild-type, $\Delta 6$)/model ($\Delta 10$) to the data. **b** Distance distribution curves. **c–e** SAXS models (white surface) calculated using GASBOR and fits of the crystal structures/model (colored cartoon representation) in them. Coloring in all panels follows that of Fig. 1



and $\Delta 10$ form elongated, dumbbell- or peanut-shaped structures that seem to have only a small contact interface between the monomers (Fig. 2). The $\Delta 6$ variant is more elongated than the $\Delta 10$ dimer, with a larger separation between the globular domains. Up to 10 mM DTT had no effect on the dimerization (Fig. S1D), suggesting that oligomerization does not occur via disulfide bridge formation due to exposure of the single buried cysteine residue (Fig. S1A).

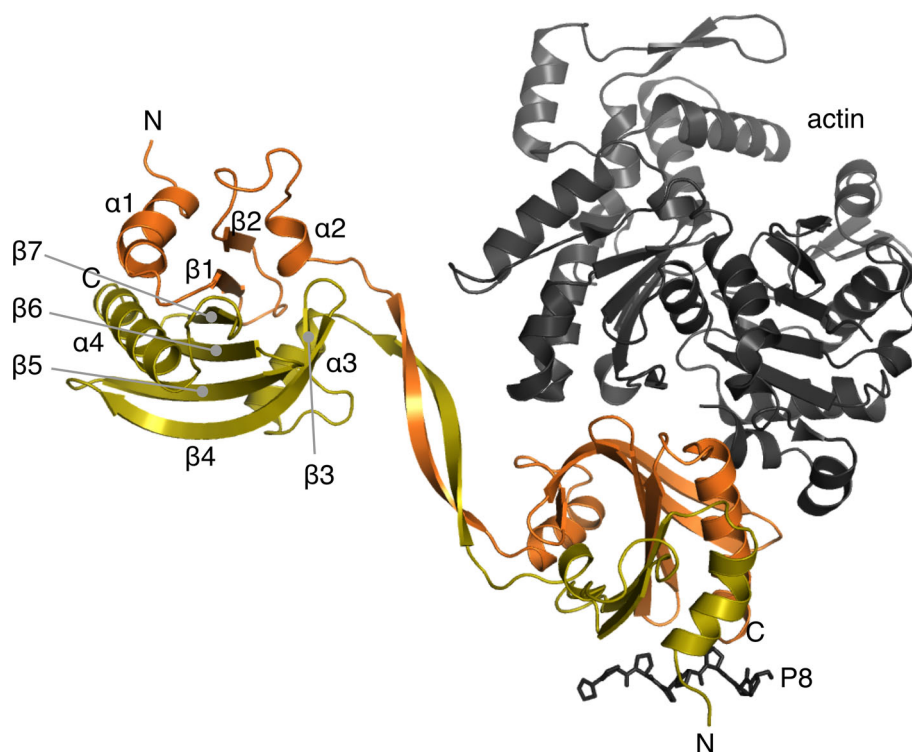
Dimerization occurs via domain swapping

Intrigued by the mutagenesis-induced homodimerization of this normally strictly monomeric protein and the peculiar shape of the dimers in solution, we determined the crystal structure of the $\Delta 6$ mutant. The protein crystallized with 8 dimers in the asymmetric unit. Dimerization occurs via domain swapping at the β -hairpin extension, such that the N-terminal half of one monomer combines with the C-terminal half of a second one (Fig. 3). Hence, the dimer is formed of two globular domains composed of two polypeptide chains, connected by a 2-stranded β sheet bridge. The crystal structure of the $\Delta 6$ mutant fits well to the SAXS data, explaining the tight, reducing agent

insensitive dimers with seemingly small inter-subunit contact areas (Fig. 2). Modeling of the $\Delta 10$ mutant based on the SAXS data indicates a similar arrangement, with a shorter linker between the globular domains. The structure of the globular domain formed through this domain swapping is, with an rmsd 0.8–1.0 Å, essentially indistinguishable from that of the wild-type monomeric profilin (Fig. 3).

The dimeric assembly highlights a feature that has passed unnoticed in previous comparisons of different profilin structures: *Plasmodium* profilin consists of two subdomains, of which the N-terminal one contains the proline-rich peptide-binding site and the C-terminal one the expected actin-binding site (Fig. 3). The central 7-stranded β sheet is assembled such that $\beta 7$ contacts $\beta 1$, and the N- and C-terminal subdomains both have a linear topology (Fig. 3). Such assembly suggests that this small, single-domain protein initially folds as two independent units that are only subsequently put together to form the full 3D fold. Thus, the intermediate form of the $\Delta 17$ mutant that elutes in SEC in between the monomer and dimer may represent an ‘open monomer’, which due to its elongated shape and longer hydrodynamic radius elutes earlier than the globular monomer of the same molecular weight (Fig. S1).

Fig. 3 Dimerization mode of the domain-swapped $\Delta 6$ mutant. The two chains are colored orange and yellow. The secondary structure elements of the core profilin fold are labeled in the upper left subunit. The N and C termini are labeled in both subunits. Actin and an octa-proline peptide are depicted on the lower right subunit as they are bound to a canonical profilin (PDB code: 2PAV [76]) and the wild-type *P. falciparum* profilin (PDB code: 2JKG [29]), respectively



Plasmodium profilin unfolds in multiple steps

We used CD spectroscopy to study thermal unfolding of the wild-type *P. falciparum* profilin and the $\Delta 6$ mutant (Fig. 4). Upon heating, the wild-type protein unfolds via a 3-step pathway with two small transitions at 29 and 40 °C, involving only minor changes in the secondary structure contents, and a final unfolding step at 58 °C. The dimeric mutant profilin has a lower overall thermal stability and shares the first transition at 29 °C with the wild-type protein. However, instead of the two latter steps, there is only one major unfolding step at 45 °C.

Sequence conservation of the half-profilins

Profilin is an abundant protein, and many eukaryotes have several profilin isoforms. Even with a conservative database search, thousands of related sequences, spanning the whole spectrum of eukaryotic organisms, are found. For conservation calculations, the 828 nearest homologs were used. Because of the large insertions in the apicomplexan sequences, we compared only residues at positions occupied in $\geq 95\%$ of the sequences (Fig. 5a). Overall, there is no large difference in the conservation of each of the subdomains. However, whereas a naïve BLAST search with the full-length *P. falciparum* profilin sequence immediately finds 76 related proteins, nearly all of them from apicomplexan parasites, and a search with the C-terminal

region (residues 75–173) gives almost the same results, the same search with the N-terminal 39 residues only identifies sequences from *Plasmodium* spp. and one sequence from *Sarcocystis neurona*, also an apicomplexan parasite. A search with the N-terminal part of a canonical profilin, e.g., human profilin 1, results in a wide range of other different profilins, and searching with the N terminus of *Toxoplasma* profilin also identifies a number of other apicomplexan profilins, excluding *Plasmodium* spp. Thus, based on the N terminus, there is a striking lack of sequence connection from *Plasmodium* profilin to other known proteins, even within the same phylum.

Exon/intron boundaries coincide with the profilin subdomains

The gene structure of several profilins has been analyzed before [27], but not in relation to the corresponding protein structures. Vertebrate profilins 1–3 have likely diverged from a common profilin 1-like ancestor from sea urchin [27]. Profilin 4 is an unconventional family member that is thought to be the most ancient of the present animal profilins and, therefore, closest to a common ancestor. Both human profilin 4 and *P. falciparum* profilin are encoded by four exons (Fig. 5b). Interestingly, the segment encoded by the second exon in *P. falciparum* profilin ends at the tip of the β -hairpin extension, separating the two subdomains. In human profilin 4, the first exon codes for the entire

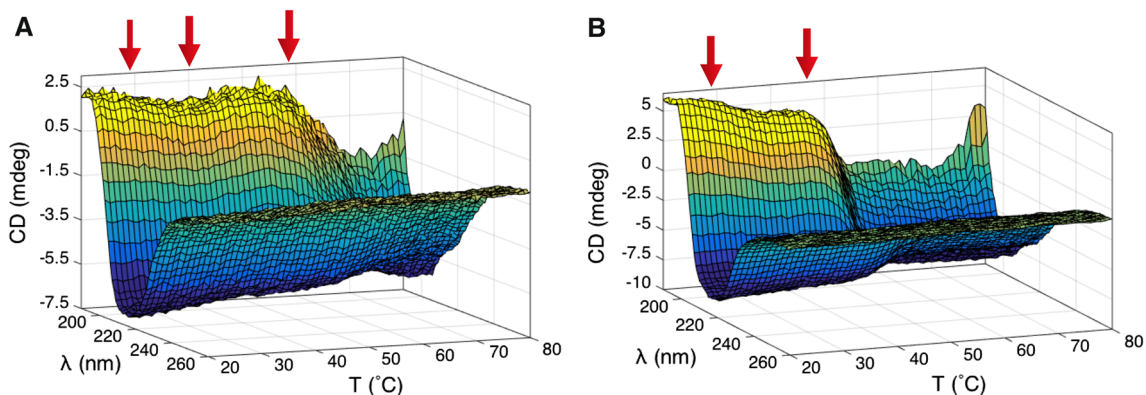


Fig. 4 Wild-type and mutant *Plasmodium* profilins unfold in multiple steps. **a** A thermal unfolding plot of the wild-type *Plasmodium* profilin shows three distinct transitions (indicated by red arrows) at

29, 40, and 58 °C. **b** The $\Delta 6$ mutant unfolds in two steps with transitions at 29 and 45 °C

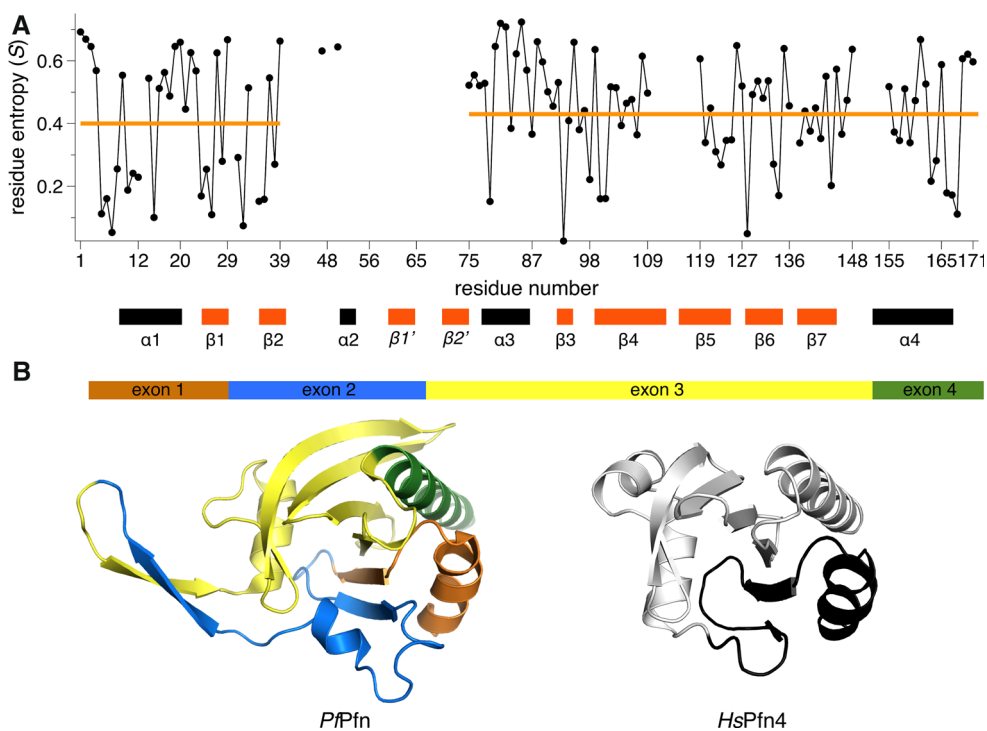


Fig. 5 Conservation and gene structure. **a** Conservation within 828 profilin sequences. *P. falciparum* profilin residue numbers are marked on the X axis. Residue entropy (S) on the Y axis (black dots) was only calculated at sites present in >95 % of the sequences. $S = 0$ means 100 % conservation; higher S means more variation. The dots are connected with a thin black line to guide the eye. The orange horizontal line shows the mean residue entropy (0.40 for the N-terminal and 0.43 for the C-terminal fragment). The secondary structure elements of *P. falciparum* profilin are depicted as orange and black bars below the residue numbers. **b** Gene structure of

Plasmodium profilin (*P/Pfn*) and human profilin 4 (*HsPfn4*) in relation to the protein structure. Exons 1–4 of the *Plasmodium* profilin gene are shown as bars with different colors, and their lengths are in relation to the protein sequence in panel **a**. The protein segments encoded by the exons are colored accordingly in the 3D structure presentation of the *Plasmodium* protein. In the human profilin 4 gene, exon 1 corresponds to exons 1 and 2 in the *Plasmodium* profilin gene and encodes the entire N-terminal segment (black), while the C-terminal fragment (light gray) is encoded by 3, instead of 2, exons

N-terminal fragment, ending exactly where the β -hairpin insertion in apicomplexan profilins would start. Most profilin genes contain fewer exons/introns, and have likely lost introns during evolution, as has been suggested for the gene

encoding profilin 3 in mammals and birds [27]. Taken further, the region encoded by exons 1 and 2 in *P. falciparum* profilin ($\alpha 1 \beta 1 / \beta 2 \alpha 2$) could be a result of duplication and permutation of an $\alpha \beta$ fragment.

Half-profilins can be co-expressed and assembled in vivo

To test whether the two subdomains could be assembled together when expressed as separate gene products, we cloned both fragments and co-expressed them in bacteria. The fragments were expressed in soluble form and eluted in SEC together as a symmetric peak corresponding to the elution volume of the wild-type profilin monomer (Fig. 6a). Static light scattering gave a mass of 23.3 ± 0.2 kDa for the peak, which is close to the calculated molecular mass (20.0 kDa) of a complex of the two fragments (Fig. 6b). Both fragments were also identified from the sample by mass spectrometry after SDS-PAGE. CD spectroscopy confirms that the two profilin fragments are folded with a slightly lower β -strand content than the wild-type protein (Fig. 6c). The reduced fraction of β strands probably reflects absent β -hairpin interactions due to the cutting of the protein at the tip of the *Apicomplexa*-specific insertion and this motif turning into mostly unstructured tails. Thus, the two separate half-profilin fragments, indeed, seem to assemble together into a dimer that resembles the normal profilin fold (Fig. 6d).

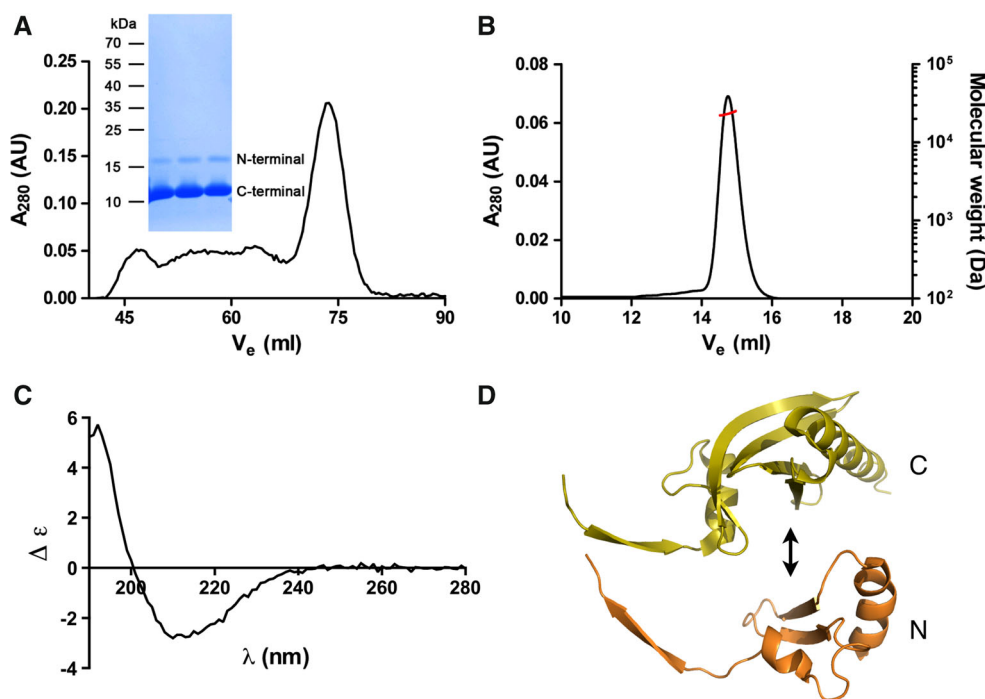


Fig. 6 Purification of half-profilins. **a** The co-expressed profilin fragments co-elute from SEC in a single symmetric peak, which corresponds to the full-length wild-type *Plasmodium* profilin. Both N- and C-terminal profilin fragments were identified by mass spectrometry from a Coomassie-stained SDS polyacrylamide gel (*inset*) of the SEC peak fraction. **b** Static light scattering gives a molecular weight of 23 kDa for the complex of the profilin fragments. The *left Y*

Discussion

Domain swapping, even when an artifact caused by extreme conditions or mutations, can tell one about the behavior of proteins in their natural state and environment, e.g., by revealing folding intermediates or independently folding units, also termed foldons [32–36]. Here, domain-swapped dimers indicate that the conserved, globular, seemingly single-domain profilin fold is actually composed of two foldons that are assembled into a complete profilin also when co-expressed as separate gene products. Simple topologies that enable fast and efficient folding and minimize the possibility of misfolding have been suggested to be a result of evolutionary optimization [37]. In profilins, the two β strands of the N-terminal and the five strands of the C-terminal subdomain have such simple topologies (Fig. 3) and seem to, indeed, fold separately before being assembled together. Unfolding happens in 2–3 steps. The acid-induced unfolding pathway of human platelet profilin also includes a stable intermediate that has been proposed to represent a physiological state relevant for the release of both actin and proline-rich ligands [38].

axis and the *black line* show the UV absorbance at 280 nm and the *right Y* axis the molecular weight (in Da) from the light scattering signal (*red line*). **c** CD spectrum of the complex of the two profilin fragments. **d** Putative assembly of the two profilin fragments. It seems likely that the long β extensions of each half (on the *left*) adopt a more disordered conformation

Proteins capable of domain swapping may be predisposed to evolving toward oligomers [39]. Conversely, we can ask: Are proteins that have evolved by fusion of two proteins likely to display domain swapping under certain conditions or due to small changes in the polypeptide chain? In this case, most likely the loss of a glycine at a tight turn leads to the observed domain swapping. The hairpin has to be flexible enough to allow for the chain to fold back during the assembly of the monomer. When this does not happen, the halves seek the closest partners from neighboring monomers. For example, the evolutionary loss of a conserved glycine residue in the PDZ domains of the giant scaffolding proteins periaxin and AHNAK2 is linked to the formation of highly intertwined, domain-swapped dimers [40]. In general, domain-swapped proteins tend to oligomerize, when hinge loops are shortened [7, 41]. On the other hand, making such connecting loops longer may generate monomers instead of domain-swapped dimers [7, 42].

Relating protein folding happening on a ms–s timescale to evolution taking place during billions of years prompted us to question the early origin of profilins. Do the folding intermediates represent smaller ancestral proteins that have been fused to form the currently widespread and structurally conserved eukaryotic profilin family? Proteins must have evolved from smaller, simple ancestral units. Have the two foldons seen in *Plasmodium* profilin once been present as two separate primitive proteins: one harboring an actin-binding site and a second one that could bind to other (proline-rich) regulatory partners? Assembly of large complexes with multiple pieces contains several error-prone steps. Therefore, permanent fusion of partners that work as common ‘hubs’ in multiprotein complexes has been favored during evolution [16]. Profilin can be seen as one such hub, connecting actin monomers to multiple regulatory proteins and networks, as well as to membranes. The actin-binding site in profilins involves residues from the C-terminal subdomain only. The N-terminal subdomain harbors most of the proline-rich peptide-binding residues. However, the C-terminal helix also participates in peptide binding in canonical profilins [22, 43]. This is not the case for *Plasmodium* profilin, which binds to octa-proline solely via its N-terminal subdomain [29]. The involvement of the C terminus in peptide binding may, thus, be a later adaptation to different regulatory proteins requiring higher affinity interfaces. Curiously, the supposedly ancestral mammalian profilin 4 binds neither actin nor proline-rich sequences. It is not clear whether it has never had these properties or whether they were lost at some point in time.

Profilins are remarkably widespread proteins, spanning all branches of the eukaryotic tree of life. In addition, considering close structural homologs, the profilin fold family would reach bacteria, with, e.g., the roadblock

proteins from *Streptomyces*, and cover a vast variety of tasks from scaffolding to enzymatic functions [44]. Therefore, if a fusion event is responsible for the modern profilin protein fold, it must have been a truly ancient event—or similar recombinations must have happened independently in different branches. Despite the high level of structural similarity, it seems from the sequence data that the N-terminal region of *Plasmodium* profilin may have an unusual evolutionary history that does not fit into a simple model of mutations with infrequent insertions. Rather, it has suffered large insertions and deletions.

While it has become clear that extensive building of new protein folds has occurred in Nature via duplication and recombination of simple, stable fragments, our ability to engineer functional protein folds de novo from small building blocks is still limited [45–48]. The feasibility of engineering novel proteins and activities by recombination of homologous fragments within the same fold family has been demonstrated [49, 50]. However, only the first steps have been taken in combining fragments from different, but still related, folds to create new functional proteins [51, 52]. Here, we demonstrate that also small, single-domain, asymmetric proteins can be split and engineered to adopt different topologies. This provides the possibility to create new scaffolding proteins and combine these with, e.g., enzymatic activities.

Acknowledgments We thank Moon Chatterjee and Esa-Pekka Kumpula for help with SRCD and SAXS data collection and the beamline staff at the CD1 beamline on ASTRID/ISA, Århus, the I911-4 beamline at MAX-Lab, Lund, and the EMBL-Hamburg beamlines X33 at DORIS and P13 at PETRA III, DESY, Hamburg, for excellent user support. The atomic coordinates and structure factor amplitudes have been submitted to the Protein Data Bank under the accession code 4D60. This work was supported by the Academy of Finland, the Sigrid Jusélius Foundation, the Emil Aaltonen Foundation, and the German Ministry of Education and Research.

References

1. Soding J, Lupas AN (2003) More than the sum of their parts: on the evolution of proteins from peptides. *BioEssays* 25:837–846
2. Gilbert W (1978) Why genes in pieces? *Nature* 271:501
3. Blake CCF (1978) Do genes-in-pieces imply proteins-in-pieces? *Nature* 273:267
4. Marianayagam NJ, Sunde M, Matthews JM (2004) The power of two: protein dimerization in biology. *Trends Biochem Sci* 29:618–625
5. Ispolatov I, Yuryev A, Mazo I, Maslov S (2005) Binding properties and evolution of homodimers in protein–protein interaction networks. *Nucleic Acids Res* 33:3629–3635
6. Jee J, Byeon IL, Louis JM, Gronenborn AM (2008) The point mutation A34F causes dimerization of GB1. *Proteins Struct Funct Bioinf* 71:1420–1431
7. Liu Y, Eisenberg D (2002) 3D domain swapping: as domains continue to swap. *Protein Sci* 11:1285–1299

8. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M (2000) Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* 289:1546–1550
9. Hocker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R (2001) Dissection of a (beta/alpha)₈-barrel enzyme into two folded halves. *Nat Struct Biol* 8:32–36
10. Richter M, Bosnali M, Carstensen L, Seitz T, Durchschlag H, Blanquart S, Merkl R, Sterner R (2010) Computational and experimental evidence for the evolution of a (beta/alpha)₈-barrel protein from an ancestral quarter-barrel stabilised by disulfide bonds. *J Mol Biol* 398:763–773
11. Horowitz NH (1945) On the evolution of biochemical syntheses. *Proc Natl Acad Sci USA* 31:153–157
12. List F, Sterner R, Wilmanns M (2011) Related (beta/alpha)₈-barrel proteins in histidine and tryptophan biosynthesis: a paradigm to study enzyme evolution. *ChemBioChem* 12:1487–1494
13. Buljan M, Frankish A, Bateman A (2010) Quantifying the mechanisms of domain gain in animal proteins. *Genome Biol* 11:R74
14. Marsh JA, Teichmann SA (2010) How do proteins gain new domains? *Genome Biol* 11:126
15. Pasek S, Risler JL, Brezellec P (2006) Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* 22:1418–1423
16. Marsh JA, Hernandez H, Hall Z, Ahnert SE, Perica T, Robinson CV, Teichmann SA (2013) Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* 153:461–470
17. Ettema TJ, Lindas AC, Bernander R (2011) An actin-based cytoskeleton in archaea. *Mol Microbiol* 80:1052–1061
18. van den Ent F, Amos LA, Lowe J (2001) Prokaryotic origin of the actin cytoskeleton. *Nature* 413:39–44
19. Dominguez R (2004) Actin-binding proteins—a unifying hypothesis. *Trends Biochem Sci* 29:572–578
20. Pollard TD (1984) Actin-binding protein evolution. *Nature* 312:403
21. Sattler JM, Ganter M, Hliscs M, Matuschewski K, Schuler H (2011) Actin regulation in the malaria parasite. *Eur J Cell Biol* 90:966–971
22. Metzler WJ, Bell AJ, Ernst E, Lavoie TB, Mueller L (1994) Identification of the poly-L-proline-binding site on human profilin. *J Biol Chem* 269:4620–4625
23. Schutt CE, Myslik JC, Rozycki MD, Goonesekere NC, Lindberg U (1993) The structure of crystalline profilin–beta-actin. *Nature* 365:810–816
24. Sohn RH, Chen J, Koblan KS, Bray PF, Goldschmidt-Clermont PJ (1995) Localization of a binding site for phosphatidylinositol 4,5-bisphosphate on human profilin. *J Biol Chem* 270:21114–21120
25. Babich M, Foti LRP, Sykaluk LL, Clark CR (1996) Profilin forms tetramers that bind to G-actin. *Biochem Biophys Res Commun* 218:125–131
26. Korupolu RV, Achary MS, Aneesa F, Sathish K, Wasia R, Sairam M, Nagarajaram HA, Singh SS (2009) Profilin oligomerization and its effect on poly (L-proline) binding and phosphorylation. *Int J Biol Macromol* 45:265–273
27. Polet D, Lambrechts A, Vandepoele K, Vandekerckhove J, Ampe C (2007) On the origin and evolution of vertebrate and viral profilins. *FEBS Lett* 581:211–217
28. Guljamow A, Jenke-Kodama H, Saumweber H, Quillardet P, Frangeul L, Castets AM, Bouchier C, de Tandeau Marsac N, Dittmann E (2007) Horizontal gene transfer of two cytoskeletal elements from a eukaryote to a cyanobacterium. *Curr Biol* 17:R757–R759
29. Kursula I, Kursula P, Ganter M, Panjekar S, Matuschewski K, Schuler H (2008) Structural basis for parasite-specific functions of the divergent profilin of *Plasmodium falciparum*. *Structure* 16:1638–1648
30. Douzery EJ, Snell EA, Bapteste E, Delsuc F, Philippe H (2004) The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci USA* 101:15386–15391
31. Escalante AA, Ayala FJ (1995) Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. *Proc Natl Acad Sci USA* 92:5793–5797
32. Fersht AR (1997) Nucleation mechanisms in protein folding. *Curr Opin Struct Biol* 7:3–9
33. Hayes MV, Sessions RB, Brady RL, Clarke AR (1999) Engineered assembly of intertwined oligomers of an immunoglobulin chain. *J Mol Biol* 285:1857–1867
34. Murray AJ, Head JG, Barker JJ, Brady RL (1998) Engineering an intertwined form of CD2 for stability and assembly. *Nat Struct Biol* 5:778–782
35. Newcomer ME (2002) Protein folding and three-dimensional domain swapping: a strained relationship? *Curr Opin Struct Biol* 12:48–53
36. Zegers I, Deswarte J, Wyns L (1999) Trimeric domain-swapped barnase. *Proc Natl Acad Sci USA* 96:818–822
37. Debes C, Wang M, Caetano-Anolles G, Grater F (2013) Evolutionary optimization of protein folding. *PLoS Comput Biol* 9:e1002861
38. McLachlan GD, Cahill SM, Girvin ME, Almo SC (2007) Acid-induced equilibrium folding intermediate of human platelet profilin. *Biochemistry* 46:6931–6943
39. Canals A, Pous J, Guasch A, Benito A, Ribo M, Vilanova M, Coll M (2001) The structure of an engineered domain-swapped ribonuclease dimer and its implications for the evolution of proteins toward oligomerization. *Structure* 9:967–976
40. Han H, Kursula P (2014) Periaxin and AHNAK nucleoprotein 2 form intertwined homodimers through domain swapping. *J Biol Chem* 289:14121–14131
41. Green SM, Gittis AG, Meeker AK, Lattman EE (1995) One-step evolution of a dimer from a monomeric protein. *Nat Struct Biol* 2:746–751
42. Albright RA, Mossing MC, Matthews BW (1996) High-resolution structure of an engineered Cro monomer shows changes in conformation relative to the native dimer. *Biochemistry* 35:735–742
43. Kursula P, Kursula I, Massimi M, Song YH, Downer J, Stanley WA, Witke W, Wilmanns M (2008) High-resolution structural analysis of mammalian profilin 2a complex formation with two physiological ligands: the formin homology 1 domain of mDial and the proline-rich domain of VASP. *J Mol Biol* 375:270–290
44. Aravind L, Mazumder R, Vasudevan S, Koonin EV (2002) Trends in protein evolution inferred from sequence and structure analysis. *Curr Opin Struct Biol* 12:392–399
45. Hocker B (2014) Design of proteins from smaller fragments—learning from evolution. *Curr Opin Struct Biol* 27C:56–62
46. Riechmann L, Winter G (2000) Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc Natl Acad Sci USA* 97:10068–10073
47. Riechmann L, Winter G (2006) Early protein evolution: building domains from ligand-binding polypeptide segments. *J Mol Biol* 363:460–468
48. Zhang D, Iyer LM, Burroughs AM, Aravind L (2014) Resilience of biochemical activity in protein domains in the face of structural divergence. *Curr Opin Struct Biol* 26:92–103
49. Claren J, Malisi C, Hocker B, Sterner R (2009) Establishing wild-type levels of catalytic activity on natural and artificial (beta/alpha)₈-barrel protein scaffolds. *Proc Natl Acad Sci USA* 106:3704–3709

50. Hocker B, Claren J, Sterner R (2004) Mimicking enzyme evolution by generating new (betaalpha)₈-barrels from (betaalpha)₄-half-barrels. *Proc Natl Acad Sci USA* 101:16448–16453
51. Eisenbeis S, Proffitt W, Coles M, Truffault V, Shanmugaratnam S, Meiler J, Hocker B (2012) Potential of fragment recombination for rational design of proteins. *J Am Chem Soc* 134:4019–4022
52. Shanmugaratnam S, Eisenbeis S, Hocker B (2012) A highly stable protein chimera built from fragments of different folds. *Protein Eng Des Sel* 25:699–703
53. van den Berg S, Lofdahl PA, Hard T, Berglund H (2006) Improved solubility of TEV protease by directed evolution. *J Biotechnol* 121:291–298
54. Ignatev A, Bhargav SP, Vahokoski J, Kursula P, Kursula I (2012) The lasso segment is required for functional dimerization of the Plasmodium formin 1 FH2 domain. *PLoS One* 7:e33586
55. Lobley A, Whitmore L, Wallace BA (2002) DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* 18:211–212
56. Compton LA, Johnson WCJ (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem* 155:155–167
57. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 287:252–260
58. Lees JG, Miles AJ, Wien F, Wallace BA (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* 22:1955–1962
59. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J Appl Crystallogr* 39:277–286
60. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80:2946–2953
61. Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89:1237–1250
62. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66:125–132
63. Kursula P (2004) XDSi: a graphical interface for the data processing program XDS. *J Appl Crystallogr* 37:347–348
64. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674
65. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221
66. Davis IW, Murray LW, Richardson JS, Richardson DC (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* 32:W615–W619
67. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
68. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
69. Thorn KS, Christensen HE, Shigeta R, Huddler D, Shalaby L, Lindberg U, Chua NH, Schutt CE (1997) The crystal structure of a major allergen from plants. *Structure* 5:19–32
70. Fedorov AA, Ball T, Mahoney NM, Valenta R, Almo SC (1997) The molecular basis for allergen cross-reactivity: crystal structure and IgE-epitope mapping of birch pollen profilin. *Structure* 5:33–45
71. Liu S, Fedorov AA, Pollard TD, Lattman EE, Almo SC, Magnus KA (1998) Crystal packing induces a conformational change in profilin-I from *Acanthamoeba castellanii*. *J Struct Biol* 123:22–29
72. Eads JC, Mahoney NM, Vorobiev S, Bresnick AR, Wen KK, Rubenstein PA, Haarer BK, Almo SC (1998) Structure determination and characterization of *Saccharomyces cerevisiae* profilin. *Biochemistry* 37:11171–11181
73. Ezezika OC, Younger NS, Lu J, Kaiser DA, Corbin ZA, Nolen BJ, Kovar DR, Pollard TD (2009) Incompatibility with formin Cdc12p prevents human profilin from substituting for fission yeast profilin: insights from crystal structures of fission yeast profilin. *J Biol Chem* 284:2088–2097
74. Kucera K, Koblansky AA, Saunders LP, Frederick KB, De La Cruz EM, Ghosh S, Modis Y (2010) Structure-based analysis of *Toxoplasma gondii* profilin: a parasite-specific motif is required for recognition by Toll-like receptor 11. *J Mol Biol* 403:616–629
75. Margraf T, Schenk G, Torda AE (2009) The SALAMI protein structure search server. *Nucleic Acids Res* 37:W480–W484
76. Ferron F, Rebowski G, Lee SH, Dominguez R (2007) Structural basis for the recruitment of profilin-actin complexes during filament elongation by Ena/VASP. *EMBO J* 26:4597–4606
77. Diederichs K, Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* 4:269–275
78. Weiss MS, Hilgenfeld R (1997) On the use of the merging R factor as a quality indicator for X-ray data. *J Appl Crystallogr* 30:203–205
79. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336:1030–1033