Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Protein engineering for crystallization of the GTPase Sar1 that regulates ER vesicle budding

Mingdong Huang,[a,b]† Jacques T. Weissman,[a] Chenqian Wang,[a] William E. Balch[a,b,c]* and Ian A. Wilson[b,d]

[a]Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA, [b]Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA, [c]The Institute of Childhood and Neglected Diseases, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA, and [d]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

† Current address: Division of Hemostasis and Thrombosis, Beth Israel Deaconess Medical Center, Harvard Medical School, RE-319, 330 Brookline Avenue, Boston, MA 02215, USA.

Correspondence e-mail: webalch@scripps.edu

Received 12 November 2001
Accepted 5 February 2002

Sar1 is an important and unique GTPase that regulates vesicle budding from the ER membrane. An effort to crystallize full-length hamster Sar1 was unsuccessful owing to the aggregation of Sar1 in solution as indicated by dynamic light-scattering measurements. It was presumed that a patch of hydrophobic residues in the N-terminal region of Sar1 was responsible for the aggregation. To attempt to improve protein crystallizability, the N-terminal residues of Sar1 were progressively truncated and the solution behavior of the resulting Sar1 variants was monitored by dynamic light scattering. Truncation of the first nine residues from the N-terminus led to a Sar1 variant that is monodisperse in solution. This well behaved Sar1 variant yielded crystals in just a few days that were ultimately refined to diffraction quality.

1. Introduction

GTPases are a family of molecular switch proteins that alternate between GTP- and GDP-bound forms and regulate a variety of cellular functions, including protein synthesis and translocation, cell proliferation and differentiation, and transmembrane signal sorting and amplification. Sar1 defines a unique family within the Ras superfamily of small GTPases and is significantly divergent from its closest evolutionary relative, ADP ribosylation factor 1 (ARF1), with which it has only 20–25% sequence identity. Sar1 and ARF1 are the pivotal regulatory GTPases involved in vesicle assembly and budding from the endoplasmic reticulum (ER) and Golgi, respectively. The crystal structure of ARF1 (Amor *et al.*, 1994; Greasley *et al.*, 1995) revealed major differences between ARF and other Ras-like GTPases that had important ramifications for understanding the function of ARF in membrane binding, nucleotide exchange and hydrolysis of nucleotides. Given that Sar1 lies on an isolated branch from all other members of the Ras superfamily in phylogenetic analyses, it is also expected to show distinct structural features that dictate its unique function in ER export.

Regions of key interest to the specific roles of Ras superfamily members are the N- or C-termini. They contain crucial functional regions that often undergo post-translational modifications essential for targeting the GTPase to the membrane to carry out its biological function. Most families of small GTPases, including Ras, Rho, Rac and Rab, have cysteine residue(s) at their C-termini and are modified with lipidic moieties such as farnesyl or geranylgeranyl. ARF proteins commonly have a myristoyl modification at their N-terminus. In contrast, Sar1 has no post-translational modifications at either the N- or C-terminus. Here, we describe the expression of the hamster Sar1 protein and our approach to solving the crystallization problem through protein engineering.

2. Experimental

2.1. Materials

PET11d vector was obtained from Novagen Inc. (Madison, Wisconsin, USA). Ni–NTA–agarose (NTA, nitrilotriacetic acid) beads were supplied by Qiagen (Valencia, CA, USA). The pCR2.1 TOPO vector was from Invitrogen (Carlsbad, California, USA). Crystallization screening kits were obtained from Hampton Research (Laguna Niguel, California, USA).

2.2. Plasmid construction

Sar1 variants lacking three, six or nine N-terminal residues (labeled Δ3-Sar1, Δ6-Sar1 and Δ9-Sar, respectively) were transferred into a pET11d vector containing a His_6 tag (Novagen, Madison, Wisconsin, USA) (Rowe & Balch, 1995).

2.3. Protein expression and purification of His_6-tagged Sar1 variants

Expression and purification of the His_6-tagged wild-type hamster Sar1 and its variants using an *Escherichia coli* expression system was modified from a previous protocol (Rowe & Balch, 1995). Changes include elimination of the inducer (which leads to extensive Sar1

aggregation) and altered fractionation conditions to increase the yield as described briefly below. *E. coli* BL21(DE3) cells carrying the Sar1 expression vector were cultured at 310 K in 1 l LB medium containing 100 µg ml^{-1} ampicillin overnight in the absence of the usual induction by isopropylthiogalactoside. The *E. coli* pellet collected from 500 ml of culture was resuspended in 25 ml lysis buffer containing 0.1 *M* NaCl, 50 m*M* Tris pH 8.0, 0.2 m*M* EDTA, 0.02 m*M* GDP, 1 m*M* PMSF and 10 m*M* β-mercaptoethanol. The cells were lysed through a sequential freeze–thaw procedure (3×) followed by a 30 min incubation at 277 K with the addition of lysozyme to a concentration of 1 mg ml^{-1}. The GDP and NaCl in the lysate were adjusted to concentrations of 0.1 m*M* and 0.3 *M*, respectively. $MgCl_2$ and DNAase were then added to 40 m*M* and 0.1%(*w*/*v*), respectively, and the lysate was incubated on ice for 30 min. After removing the cell debris from the lysate by centrifugation at 13 500 rev min^{-1} in a JA-20 Beckmann rotor for 20 min, the clear supernatant was incubated with 4 ml (bed volume) of nickel nitrilotriacetic acid (Ni–NTA) agarose on ice for 30 min. The Ni–NTA–agarose beads were then packed into a column and washed first with 150 ml washing buffer (50 m*M* HEPES pH 7.4, 0.3 *M* NaCl, 50 m*M* EGTA, 1 m*M* $MgCl_2$, 10 m*M* β-mercaptoethanol and 10 µ*M* GDP) and then with 150 ml washing buffer containing 20 m*M* imidazole. The protein was eluted with washing buffer containing 0.5 *M* imidazole. The fractions that contained Sar1 were pooled, frozen in liquid nitrogen and stored at 193 K for subsequent use. Typical expression yield of Sar1 after Ni–NTA–agarose purification was ∼200 mg per litre of cell culture.

2.4. Dynamic light-scattering measurements

Dynamic light scattering was performed with a DynaPro-801 molecular-sizing instrument (Protein Solutions, Inc., Charlottesville, VA, USA) equipped with a 20 µl microsampling cell at room temperature (295 K). All protein solutions were filtered through a 0.02 µm porosity membrane prior to loading the microsampling cell. 15–20 µl of protein solution at ∼1 mg ml^{-1} was used to measure the light-scattering signal. For each sample, at least ten light-scattering measurements were taken and the data were processed using *DynaLS* and *DYNAMICS* v4.0 software (Protein Solutions).

2.5. Crystallization of Sar1

For crystallization trials, Sar1 and its variant proteins were freshly thawed from 193 K, dialyzed into buffer consisting of 25 m*M* HEPES pH 7.4, 125 m*M* KOAc, 1 m*M* $MgCl_2$ for 3 h and concentrated to 12 mg ml^{-1}. A typical sitting-drop setup (McPherson, 1990) was used for crystallization trials at room temperature (295 K).

2.6. X-ray diffraction data collection

Diffraction data were collected either on an in-house 30 cm MAR image-plate detector mounted on a Siemens rotating-anode generator with a long Supper mirror and a 300 µm focal spot, operating at 50 mA and 80 kV, or on the 34.5 cm MAR image-plate detector at Stanford Synchrotron Radiation Laboratory (SSRL) at 1.08 Å wavelength, operating at 3 GeV and 30–100 mA. Crystals of Δ9-Sar1 were briefly soaked in a cryoprotecting buffer consisting of 30% polyethylene glycol (PEG) 4000, 0.1 *M* NaOAc pH 4.6, 0.2 *M* $(NH_4)_2SO_4$ and 25% glycerol and cryocooled in a gaseous liquid-nitrogen stream. The raw image data collected were processed and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

3. Results

3.1. Preliminary crystallization trials of full-length Sar1

Full-length hamster Sar1 was previously successfully expressed in *E. coli* (Rowe & Balch, 1995), but the yield was relatively low (1 mg per litre of cell culture). Using a modified protocol we were able to express and purify Sar1 in high yield (∼200 mg per litre of cell culture). The purified Sar1 had the expected molecular weight as judged by MALDI mass-spectrum analysis and its mobility in a reducing SDS–PAGE (data not shown). The protein can be concentrated to a very high concentration (60 mg ml^{-1}) without any noticeable protein precipitation. However, no crystals of full-length Sar1 were obtained following three months of effort and hundreds of crystallization trials.

3.2. Solution behavior of full-length and truncated Sar1 by dynamic light scattering

Dynamic light scattering measures the translational diffusion coefficient of the protein by monitoring the decay of an autocorrelation function with time. The hydrodynamic radius of the protein can then be calculated using the Stokes–Einstein equation. The molecular weight of the protein is estimated by fitting the measured hydrodynamic radius to a standard curve derived from 25 globular proteins. Despite the high solubility of full-length Sar1, dynamic light-scattering measurements showed that it was heavily aggregated, with an apparent molecular weight of >1447 kDa (Table 1). The addition of inorganic salts (sodium chloride, ammonium sulfate and magnesium chloride) at various concentrations or shifts of the solution pH values (from pH 4 to 9) did not decrease or inhibit aggregation, implying that the Sar1 aggregation is not through polar interactions. Thus, non-polar residues in Sar1 were apparently contributing to a significant degree of aggregation that could interfere with crystallization.

To provide alternative forms of Sar1 for crystallization, we systematically truncated the N-terminal residues given their highly hydrophobic sequence. Three Sar1 variant proteins were produced, with three, six and nine N-terminal residues deleted (denoted Δ3-Sar1, Δ6-Sar1 and Δ9-Sar1, respectively) from the full-length Sar1 sequence. Full-length Sar1 and its three N-terminal truncation variants had the same high solubility in aqueous buffer. Dynamic light-scattering measurements showed that one of the variants had a very different aggregation state in solution. While both the Δ3-Sar1 and Δ6-Sar1 variants were also highly aggregated in solution (comparable to the full-length Sar1; Table 1), the Δ9-Sar1 variant was well behaved and monodisperse with an apparent molecular weight of 43.6 kDa (Table 2), corresponding to a Δ9-Sar1 dimer.

3.3. Crystallization of Δ9-Sar1

The solubility profiles in crystallization trials for Δ3-Sar1 and Δ6-Sar1 proteins were similar to those of wild-type Sar1. No

Table 1
Molecular weights (kDa) of wild-type hamster Sar1 and its variants measured by dynamic light scattering.

		Bimodal cumulant analysis†	
Protein	Monomodal	First component	Second component
Full-length Sar1‡	1446.8	515.7 (81%)	3078.9 (19%)
Δ3-Sar1	7770.8	1548.6 (97%)	35804.8 (3%)
Δ6-Sar1	2384.6	207.5 (92%)	4173.7 (8%)
Δ9-Sar1	71.9	43.6 (99%)	1604.2 (1%)

† Bimodal cumulant analyses are based on the assumption of two components in the solution; values in parentheses are the percentages of each component. ‡ The theoretical molecular weight of wild-type Sar1 from its sequence is 22.4 kDa.

Table 2
X-ray data-collection statistics for Δ9-Sar1 crystal.

Data in parentheses are for the outer shell (1.70–1.78 Å).

Size of crystal (mm)	0.05 × 0.2 × 1
No. of crystals	1
Temperature (K)	100
Diffraction limit (Å)	1.7
No. of unique reflections	48000 (4269)
Total No. of observations	384830
Overall redundancy	8.0
Data completeness (%)	98.8 (88.8)
R_{sym}	0.050 (0.26)
$I/\sigma(I)$	24.5 (5.4)

crystals could be obtained for these two Sar1 variants. In contrast, the Δ9-Sar1 protein yielded two preliminary crystals in the first crystallization screen using Hampton sparse-matrix screening kit No. 1. The first crystal form (AS form) was obtained in a few days using ammonium sulfate (AS) as a precipitant in Hampton screening condition No. 39 (2% PEG 400, 2.0 *M* ammonium sulfate, 0.1 *M* sodium HEPES, pH 7.5). The second form (PEG form) was obtained in Hampton condition No. 20 (25% PEG 4000, 0.1 *M* sodium acetate, pH 4.6, 0.2 *M* ammonium sulfate). The second form (PEG form) was only obtained from Hampton condition No. 20. The morphology of the AS-form crystal appears rapidly as stacks of very fine needles, whereas the PEG-form crystal grew much more slowly and appeared at around one month. Efforts to improve the AS crystal form were terminated once the PEG-form crystals were found owing to their improved morphology (stacked thin plates) and larger crystals.

3.4. Optimization of crystallization conditions for Δ9-Sar1 crystals (PEG form)

When the Δ9-Sar1 protein was mixed with the precipitant solution (PEG 4000) during the crystallization setup, considerable precipitate appeared. The precipitate redissolved in a few days and the Δ9-Sar1 PEG form crystals grew over a period of several weeks. A new method was found that greatly accelerated crystallization from weeks to days. By simply shaking the crystallization tray in a shaker overnight at a rapid speed (150–200 rev min^{-1}) that did not spill the well solution, the precipitate dissolved quickly and crystals appeared after 3–4 d incubation in a 295 K incubator. They continued to grow to a maximal size of 0.05 × 0.2 × 1.0 mm in one to two weeks. This method allowed us to speed up the optimization process for generating suitable crystallization conditions. After subsequent fine grid crystallization screen, we found the optimal crystallization condition that yields large single crystals was 30% PEG 4000, 0.1 *M* sodium acetate pH 4.6 and 0.2 *M* ammonium sulfate (Fig. 1). The resulting Δ9-Sar1 crystals were quite thin (0.05 mm) and only diffracted to 2.5 Å using an in-house X-ray source, with a relatively high merging *R* value (0.176). However, they diffracted to 1.5 Å at SSRL beamline 7-1 and enabled a high-quality data set to be collected to 1.7 Å resolution (Table 2). The crystals belong to the monoclinic $P2_1$ space group, with unit-cell parameters a = 53.37, b = 61.69, c = 71.14 Å, β = 107.5°. Two Δ9-Sar1 molecules occupy the asymmetric unit as estimated from the Matthews coefficient of 2.53 Å^3 Da^{-1} (Matthews, 1968), with a corresponding solvent content of 48.9%.

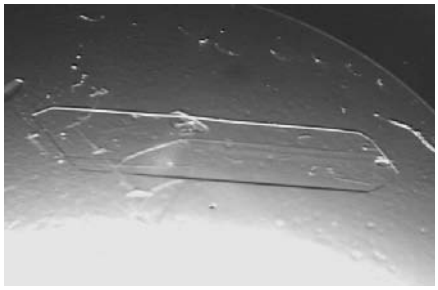

Figure 1
Single crystal of Δ9-Sar1 (PEG form) after optimization of crystallization conditions. The size of this crystal is approximately 0.05 × 0.2 × 1.0 mm.

4. Discussion

Sar1 is the prototypical member of an important GTPase family within the Ras superfamily that regulates vesicle budding from ER. Dynamic light scattering for full-length Sar1 suggested that despite its high solubility Sar1 forms soluble aggregates in solution and therefore explains the failure of our crystallization trials. This result further demonstrates the utility of dynamic light scattering as a tool to study solution behavior of proteins and their crystallizability (Ferre-D'Amare & Burley, 1994). The dynamic light-scattering measurements also suggested that the Sar1 aggregation was mediated primarily through hydrophobic interactions based on the inability to decrease aggregation with the addition of polar compounds. Interestingly, inspection of the hamster Sar1 sequence revealed a stretch of hydrophobic residues (MSFIFDWIYSGFSSVLQFLG-) at the N-terminus of Sar1. Since this was the only region in the sequence with extensive hydrophobicity, we assumed that residues within this region were responsible for the protein aggregation. To test this hypothesis, we gradually truncated residues from the N-terminus in order to monitor the solution behavior by dynamic light scattering. The removal of the first three (Met, Ser, Phe) or six (Met, Ser, Phe, Ile, Phe, Asp) residues did not significantly change the protein aggregation and crystals were not obtained from these Sar1 variants. However, removal of the first nine residues (Met-Ser-Phe-Ile-Phe-Asp-Trp-Ile-Tyr) led to a monodisperse protein solution with an apparent molecular weight of 43.6 kDa as measured by dynamic light scattering, which corresponds to a Δ9-Sar1 dimer. Preliminary crystals of this Sar1 variant were obtained quickly in the first crystallization tray. Moreover, we found we could accelerate the crystallization of Δ9-Sar1 crystal (PEG form) from weeks to days by shaking the crystallization tray overnight after initial setup. We speculate that this increase in crystallization rate may be related to the accelerated redissolution of the initial PEG/Δ9-Sar1 precipitate. Alternatively, it may be a consequence of accelerated crystal nucleation by rapidly bringing the precipitant concentration close to the metastable fluid–fluid critical point in the crystallization phase diagram (ten Wolde & Frenkel, 1997), since the vapor-diffusion equilibrium rate of PEG is much slower than salts such as ammonium sulfate (McPherson, 1990). Similar situations may apply to other crystallization studies.

In summary, we have found that the protein crystallization problem for Sar1 could be solved through a logical application of biophysical and molecular experiments, leading to the generation of large amounts of well behaved, compacted and monodisperse protein. Our protein-engineering approach further extends its utility as a tool for crystallization (Dale *et al.*, 1994; Scott *et al.*, 1998; for a review, see Price & Nagai, 1995).

This work was supported by NIH grants GM42336 (WEB) and GM 49497 (IAW). We thank Drs Robyn Steinfeld, Andreas Heines, Xiaoping Dao, Jeff Speir and Samantha Greasley for fruitful discussions and with help in data collection. We thank the staff of the SSRL beamline 7-1 for helpful support in synchrotron data collection.

References

Amor, J. C., Harrison, D. H., Kahn, R. A. & Ringe, D. (1994). *Nature (London)*, **372**, 704–708.
Dale, G. E., Broger, C., Langen, H., D'Arcy, A. & Stuber, D. (1994). *Protein Eng.* **7**, 933–999.
Ferre-D'Amare, A. R. & Burley, S. K. (1994). *Structure*, **2**, 357–359.

Greasley, S. E., Jhoti, H., Teahan, C., Solari, R., Fensome, A., Thomas, G. M., Cockcroft, S. & Bax, B. (1995). *Nature Struct. Biol.* **2**, 797–806.
McPherson, A. (1990). *Eur. J. Biochem.* **189**, 1–23.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Price, S. R. & Nagai, K. (1995). *Curr. Opin. Biotechnol.* **6**, 425–430.
Rowe, T. & Balch, W. E. (1995). *Methods Enzymol.* **257**, 49–53.
Scott, C. A., Garcia, K. C., Stura, E. A., Peterson, P. A., Wilson, I. A. & Teyton, L. (1998). *Protein Sci.* **7**, 413–418.
Wolde, P. R. ten & Frenkel, D. (1997). *Science*, **277**, 1975–1978.