Structures of ubiquitin-like (Ubl) and Hsp90-like domains of sacsin provide insight into pathological mutations

Marie Ménade‡1, Guennadi Kozlov‡1, Jean-François Trempe1,2, Harshit Pande3, Solomon Shenker4, Sihara Wickremasinghe1, Xinlu Li1, Hamed Hojjat1, Marie-Josée Dicaire5, Bernard Brais5, Peter S. McPherson5, Michael J. H. Wong1, Jason C. Young‡, Xinlu Li‡, Marie-Josée Dicaire‡, and Kalle Gehring‡

From the ‡1Department of Biochemistry, McGill Centre for Structural Biology, McGill University, Montreal, Quebec H3G 0B1, Canada and ‡2Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada

Edited by George N. DeMartino

Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) is a neurodegenerative disease that is caused by mutations in the SACS gene. The product of this gene is a very large 520-kDa cytoplasmic protein, sacsin, with a ubiquitin-like (Ubl) domain at the N terminus followed by three large sacsin internal repeat (SIRPT) supradomains and C-terminal J and HEPN domains. The SIRPTs are predicted to contain Hsp90-like domains, suggesting a potential chaperone activity. In this work, we report the structures of the Hsp90-like Sr1 domain of SIRPT1 and the N-terminal Ubl domain determined at 1.55- and 2.1-Å resolutions, respectively. The Ubl domain crystallized as a swapped dimer that could be relevant in the context of full-length protein. The Sr1 domain displays the Bergerat protein fold with a characteristic nucleotide-binding pocket, although it binds nucleotides with very low affinity. The Sr1 structure reveals that ARSACS-causing missense mutations (R272H, R272C, and T201K) disrupt protein folding, most likely leading to sacsin degradation. This work lends structural support to the view of sacsin as a molecular chaperone and provides a framework for future studies of this protein.

Sacsin is a gigantic 520-kDa protein that is a product of the SACS gene, located on human chromosome 13 (13q12) (1, 2). Mutations in sacsin cause the neurodegenerative disease autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS)‡ first described in the Canadian province of Quebec in 1978 (3). Since then, cases have been reported worldwide with close to 200 mutations known (4, 5). In the Quebec population, ARSACS is characterized by an early-onset progressive cerebellar atrophy, peripheral neuropathy, and thickening of the retina; clinical symptoms in most cases consist of cerebellar ataxia, spasticity, distal amyotrophy, nystagmus, and dysarthria, although the identification of new mutations has broadened its clinical spectrum (6, 7). Sacsin is localized in the cytoplasm of neuronal cells in the soma, axons, and dendrites, and some is found in close proximity to mitochondria. Sacsin mRNA and its product are highly expressed in Purkinje cells (8). ARSACS patient fibroblasts and neurons with sacsin knockdown present a hyperfused mitochondrial network with some accumulation of mitochondria in the soma and proximal dendrites (9). Recent studies in patient fibroblasts demonstrate that sacsin knockout increases reactive oxygen species production (10, 11).

Little is known about the mechanism of sacsin function. A study of sacsin chaperone activity showed some protective function against polyglutamine-expanded toxicity of ataxin-1. Namely, when sacsin was knocked down in cells using siRNA, the toxicity of the spinocerebellar ataxia type-1 ataxin-1[Q82] mutant protein was enhanced, whereas there was no toxicity with the ataxin-1[Q30] protein that contains a normal number of glutamine residues (8). This suggests a possible role of sacsin in degradation of misfolded proteins. A firefly luciferase in vitro refolding experiment showed a weak activity with the first sacsin internal repeat, SIRPT1 (12).

Structurally, sacsin is a multidomain protein (Fig. 1A) with an N-terminal ubiquitin-like (Ubl) domain that was shown to bind to the proteasome via proteasomal 20S α subunit C8 (8). Sacsin has also been identified as a substrate of the E3 ubiquitin ligase Ube3A, the product of a gene associated with the neurological disorder Angelman syndrome (13). Partial sequence similarity of RAD23A(231–285), a binding partner of Ube3A, to residues repeat; HEPN, higher eukaryotes and prokaryotes nucleotide-binding; NBD, nucleotide-binding domain; GRP94, glucose-regulated protein 94; TRAP1, tumor necrosis factor receptor–associated protein 1; r.m.s.d., root mean square deviation; PDB, Protein Data Bank; 3-Met, F270M/K291M/ L318M; 4-Met, I113M/F270M/K291M/L308M; ITC, isothermal titration calorimetry; SAXS, small-angle X-ray scattering; Rg, radius of gyration; Bicene, N,N-bis(2-hydroxyethyl)glycine; CHESS, Cornell High-Energy Synchrotron Source.
Crystal structures of Ubl and Sr1 domains of human sacsin

3660–3735 of sacsin led to the suggestion that a putative xeroderma pigmentosum complementation group C domain exists in SIRPT3 (13, 14). Toward the C terminus, sacsin contains a J domain immediately followed by a HEPN (higher eukaryotes and prokaryotes nucleotide-binding) domain (8). The J domain was demonstrated to be functional using a complementation assay technique with Hsp70 chaperone (8). The sacsin J domain structure was previously determined by NMR (available as Protein Data Bank (PDB) code 1IUR). HEPN domains were originally identified in bacteria with a role in resistance to aminoglycosides antibiotics (15); it was subsequently recognized by bioinformatics analysis that the HEPN domain exists in humans exclusively in sacsin (16). A high-resolution crystal structure of the HEPN revealed that it forms a stable dimer that mediates sacsin dimerization and binds GTP and other nucleotides with low-micromolar affinity (17). More recently, a small-molecule screen was performed in the aim of identifying compounds binding HEPN with high affinity. This work showed preference of sacsin HEPN for compounds with a high number of phosphates (18).

Most of sacsin is composed of three large SIRPTs with each repeat divided into subrepeats, namely Sr1, Sr2, Sr3, and SrX (19). The second repeat does not contain SrX, making SIRPT2 smaller than SIRPT1 and SIRPT3. Within each of the SIRPTs, the first half of the Sr1 shows sequence homology to the nucleotide-binding domain (NBD) of Hsp90 (19).

Hsp90 is a superfamily of chaperones that includes cytoplasmic Hsp90α and Hsp90β, glucose-regulated protein 94 (GRP94) localized in the endoplasmic reticulum, and mitochondrial tumor necrosis factor receptor–associated protein 1 (TRAP1) (20, 21). Hsp90s comprise three functional domains starting with the NBD followed by a substrate-recognition domain and a dimerization domain at the C terminus (21). The NBDs adopt a Bergerat fold with a sandwich of four α-helices forming a nucleotide-binding pocket (22). ATP binding and hydrolysis are required for Hsp90 chaperone function (20).

The amino acid sequence of sacsin supports the hypothesis of a quality control function, but more studies are needed. Sr1 is a site for a number of disease-causing mutations: D168Y (23), T201K (24), R272C (25), R272H (19), and L308F (26). How these mutations contribute to ARSACS is unclear in the absence of structural information.

Here, we present atomic structures of two domains of human sacsin: the N-terminal Ubl domain and the Sr1 domain of SIRPT1. The Ubl crystallizes contain a swapped dimer arrangement of the protein domain, which is potentially relevant in the context of the full-length protein. The Sr1 displays the Bergerat protein fold with a characteristic nucleotide-binding pocket but binds nucleotides only weakly. Importantly, the structure demonstrates that the pathological missense mutations, R272H, R272C, and T201K, affect the protein fold and/or stability and most probably target sacsin for proteasomal degradation. The ARSACS D168Y mutation likely affects chaperone activity via interference with ATP binding.

Results

Structure of the Ubl domain of sacsin

To investigate sacsin function, we adopted a structure-directed approach. To elucidate the Ubl structure, we expressed sacsin (residues 2–85) in Escherichia coli, crystallized the purified protein, and obtained excellent X-ray diffraction. However, the structure could not be solved by molecular replacement with known ubiquitin or Ubl structures because sacsin Ubl has less than 30% sequence identity with ubiquitin. We sought to incorporate selenomethionine in the protein for experimental phasing, but the domain contains no methionines. We therefore made a series of mutants, based on sequence comparisons with other sacsin mammalian orthologs, to incorporate methionine: V9M, L26M, W29M, V34M, F39M, V46M, L58M, and L78M as well as a construct (residues 1–85) that includes the initiator methionine (M1). Of those mutants, only M1, L26M, F39M, and L78M were well expressed, and only M1 and L78M crystallized. Although the former did not yield any measurable anomalous scattering, the L78M mutant gave strong anomalous scattering and enabled the structure to be solved by multiwavelength anomalous dispersion. This mutant structure was subsequently used to determine the structure of native Ubl (Table 1).

The structure revealed the β-grasp fold characteristic of ubiquitin-like domains (Fig. 1B). Surprisingly, although the fold is identical to ubiquitin, the structure contains an unusual feature. The Ubl molecules form a dimer through swapping of their N-terminal β-strands. This dimeric arrangement is likely stabilized by an intramolecular disulfide bond between Cys17 and Cys20 (Fig. 1B). The dimerization caused extensive intermolecular contacts via β-sheet surfaces. This is unexpected because this surface typically contains a hydrophobic patch used for recognition of binding partners, including proteasome subunits. This hinted that the observed dimer could be a crystallization artifact; domain swapping often occurs unnaturally during crystallization of truncated proteins (27).

To experimentally determine the aggregation state of the Ubl domain in solution, we used multangle light scattering coupled with size-exclusion chromatography to measure the molecular mass of 10.5 kDa, which is in a good agreement with the expected mass of 9.9 kDa for monomeric Ubl (Fig. 1C). We repeated this experiment with a larger fragment of sacsin (residues 1–339) that includes the Ubl and the following Hsp90-like Sr1 domain. Again, the experimental molecular mass (35.5 kDa) correlated well with the calculated mass of the monomer (38.7 kDa). In conclusion, the isolated Ubl domain does not dimerize in solution, confirming that the swapped dimer observed in the Ubl crystal is likely a crystallization artifact.

When reconstructed from the swapped dimer, the structure of the sacsin Ubl monomer is very similar to ubiquitin with a root mean square deviation (r.m.s.d.) of 1.7 Å over 61 Ca atoms (Fig. 1C). Nevertheless, the surface composition is significantly different. Ubiquitin utilizes a highly conserved surface centered around Ile44 to recognize ubiquitin-binding proteins. In ubiquitin, the side chains of Ile44, Leu4, and Val48 form a hydrophobic patch, whereas Arg42 and His68 are responsible for charged...
intermolecular contacts (Fig. 1, D and E). In sacsin, only Arg50 is conserved. Ile44 is substituted by Trp52, which is hydrophobic but has a bulkier side chain, possibly affecting selection of binding partners. Two other hydrophobic residues found in ubiquitin (Val70 and Leu8) are replaced by Gln79 and Gly16 in sacsin. These differences result in surface-charge differences: the surface of ubiquitin has a much more pronounced positive charge (Fig. 1, D and E).

### High-resolution structure of the Sr1 domain of sacsin SIRPT1 repeat

Based on a bioinformatics analysis of sacsin (19), we designed several constructs of the sacsin Hsp90-like domain and screened them for crystallization. The first Sr1 domain (residues 89–336) (Fig. 2A) of the first SIRPT yielded well diffracting crystals. Because no protein structure with high enough sequence homology was available in the PDB, the sacsin Sr1 fragment was labeled with selenomethionine for experimental phasing of diffraction data. The crystallized construct naturally contains two methionine residues, Met148 and Met214, but the phasing of diffraction data. The crystallized construct naturally fragments was labeled with selenomethionine for experimental sequence homology was available in the PDB, the sacsin Sr1 repeat of the sacsin Hsp90-like domain and repeat (Table 1) was also crystallized and determined the structure of the homologous (91% identity) domain of sacsin from the green anole. This protein crystallized in yet another crystal form (space group C2). The crystals of WT human Sr1 contain three protein molecules in the asymmetric unit, whereas the other crystal forms have two molecules in the asymmetric unit. Having multiple crystal forms with each containing two or three molecules in the asymmetric unit is helpful for identifying regions of flexibility and bias resulting from crystal contacts.

The Sr1 structures show an α/β-sandwich where a central β-sheet is covered by α-helical bundles on both sides (Fig. 2B). The eight-stranded β-sheet is largely antiparallel except for strands β1 and β7, which are parallel to each other. The overall structure of the Sr1 domain is similar to a Bergerat fold (22), which is characteristic of nucleotide-binding domains involved in ATPase activity, in particular Hsp90 family proteins. A structure similarity search using the DALI database returned Hsp90 orthologs in the endoplasmic reticulum (GRP94), mitochondria (TRAP1), and cytoplasm (Hsp90) (Fig. S2). Comparison with these structures shows that similarity is mostly limited to the central β-sheet and the α-helices forming the nucleotide-binding site. Interestingly, the opposite side of the β-sheet is mostly solvent-accessible in the homologous structures, whereas it is covered by a four-helical bundle in sacsin.

Comparison of the multiple Sr1 molecules shows that the independently folded region extends from Pro94 to Ser335. Nevertheless, the majority of molecules (seven of nine) have the N terminus structured from Phe89, and in some, even the preceding linker is visible in electron density maps. This is because the N-terminal part is involved in crystal contacts largely driven by Phe89 inserting into a hydrophobic pocket proximal to the

### Table 1

| Data collection and refinement statistics | Ubl | Ubl (L78M) | Ubl (L78M) | Sr1 | Sr1 (3-Met) | Sr1 (4-Met) | Anole Sr1 |
|-----------------------------------------|-----|------------|------------|-----|-------------|-------------|-----------|
| Wavelength (Å)                          | 1.542 | 0.9787 (peak) | 0.9792 (inflection) | 0.977 | 0.977 | 0.977 | 0.977 |
| Space group C222, C222, C222, P1, P1, P43, C2 | 51.9, 62.6, 90.8 | 52.1, 61.5, 91.6 | 52.1, 61.5, 91.6 | 41.1, 41.8, 105.9 | 41.7, 41.4, 74.1 | 70.8, 70.8, 100.9 | 160.8, 38.2, 114.7 |
| Resolution (Å)                         | 50–2.10 | 45.8–2.40 | 36.6–2.90 | 50–1.55 | 50–1.90 | 50–1.80 | 50–1.85 |
| R_factor/Rfree (%)                      | 0.06 (0.29) | 0.061 (0.498) | 0.068 (0.707) | 0.05 (0.45) | 0.09 (0.46) | 0.10 (0.59) | 0.11 (0.57) |
| Completeness (%)                       | 23.8 (7.3) | 24.8 (5.5) | 23.9 (4.3) | 34.6 (2.9) | 43.5 (4.5) | 47.0 (6.4) | 20.4 (2.9) |
| Redundancy                             | 98.8 (82.1) | 98.9 (98.7) | 99.0 (98.6) | 96.8 (94.8) | 98.0 (96.9) | 100 (100) | 99.9 (100) |
| No. of reflections                     | 8,843 | 5,923 | 86,247 | 32,566 | 43,323 | 43,261 |
| No. of reflections/atom                | 0.187/0.247 | 0.218/0.284 | 0.182/0.214 | 0.187/0.226 | 0.198/0.230 | 0.196/0.235 |
| Protein                                | 1,229 | 1,063 | 5,622 | 3,709 | 3,950 | 3,721 |
| Water                                  | 59 | 19 | 390 | 167 | 203 | 253 |
| B-factors                              | 41.4 | 64.9 | 23.5 | 17.1 | 11.4 | 14.5 |
| Water r.m.s. deviations                | 45.2 | 57.5 | 39.2 | 37.3 | 25.6 | 33.5 |
| Bond lengths (Å)                       | 0.011 | 0.009 | 0.007 | 0.011 | 0.010 | 0.010 |
| Bond angles (°)                        | 1.27 | 1.1 | 1.16 | 1.37 | 1.34 | 1.28 |
| Ramachandran statistics (%)           | 97.3 | 96.0 | 97.3 | 97.0 | 97.1 | 97.1 |
| Additional allowed regions            | 2.7 | 4.0 | 2.7 | 3.0 | 2.9 | 2.9 |
| Deposition (PDB code)                  | 5VSX | 5VSZ | 5V44 | 5V45 | 5V46 | 5V47 |

* Highest-resolution shell is shown in parentheses.
This is a result of the crystal packing where conformation of this loop is stabilized via intermolecular crystal contacts. It is worth noting that, in the P4₃ crystals, this loop appears as a β-hairpin. This is in contrast to the helix-loop-helix arrangement observed in Hsp90 structures (Fig. S2). This helix-loop-helix is a flexible element from Hsp90, which closes over the bound ATP; it exposes hydrophobic residues to allow the dimerization crucial for the enzymatic activity of the protein (29, 30).

Given the sequence similarity of the solved fragment with Sr1 modules from SIRPT2 and SIRPT3 (~35% sequence identity), the structures of these modules are expected to be very similar (Fig. S4). Thus, our SIRPT1 Sr1 structures are good templates for interpreting structure–function studies of the other Sr1 modules.

**Sacsin Sr1 binds nucleotides with low affinity**

Because structurally related Hsp90-like domains bind ATP and possess ATPase activity, we sought to test nucleotide binding to SIRPT1 Sr1. We chose NMR because it is sensitive to even low-affinity binding, and Sr1 gives excellent NMR spectra. The ¹H–¹⁵N correlation spectrum of Sr1 shows good dispersion of signals characteristic of a well-folded protein. We first tested Sr1 for binding to ADP in the presence of Mg²⁺ as Hsp90 displays 4-fold higher affinity to ADP than to ATP (29 versus 132 μM) (31). If bound, ADP would cause shifts for a number of signals in the nucleotide-binding site. Somewhat unexpectedly, only high concentrations of ADP resulted in spectral changes, suggesting a $K_d$ of over 10 mM (Fig. 3A). Similar results were obtained in a titration with ATP (data not shown). This is consistent with no binding observed in isothermal titration calorimetry (ITC) experiments (Fig. S5). We sought to explain this marginal affinity for nucleotides by examining the nucleotide-bound structures of Hsp90-like domains.

The Bergerat fold is characterized by several conserved motifs involved in binding and hydrolysis of ATP (22). Structural comparison of Sr1 with nucleotide-bound TRAP1 shows that some features are conserved in sacsin (Fig. 3, B and C). Namely, motif I includes Glu116 that is expected to serve as a nucleophile, whereas Asn120 could help coordinate the Mg²⁺ ion involved in ATP binding. Thr269 in motif IV is part of a loop in Hsp90-like proteins that acts as a lid on the nucleotide-binding site and does not possess corresponding glycine residues. In the P4₃ crystal form of sacsin Sr1, this loop adopts a defined but irreg-
ular conformation, suggesting this stretch does not have propensity to form α-helices. Although the loop in Sr1 appears unlikely to support nucleotide binding, it is possible that another part of sacsin could function as an ATP-binding lid. All this points to insufficiency of Sr1 for ATPase activity. In agreement, steady-state ATPase assays did not show appreciable activity for the sacsin Sr1 domain (Fig. 3D).

Five ARSACS missense mutations (D168Y, T201K, R272H, R272C, and L308F) occur in SIRPT1 Sr1. We generated these

Role of pathological mutations

Five ARSACS missense mutations (D168Y, T201K, R272H, R272C, and L308F) occur in SIRPT1 Sr1. We generated these

Crystal structures of Ubl and Sr1 domains of human sacsin

Figure 2. Structure of sacsin Sr1 domain. A, domain architecture highlighting the SIRPT1 Sr1 domain. B, cartoon representation of human sacsin Sr1 domain (residues 89–336) colored from N terminus (blue) to C terminus (red). C, the N-terminal strands in the crystals are swapped. D, side chains of Phe89 and Gln91 are inserted into the surface of another Sr1 molecule.

Figure 3. Sacsin Sr1 binds nucleotides with low affinity. A, overlays of the 1H-15N correlation NMR spectra show shifts of some peaks upon addition of ADP at high concentrations. B and C, comparison of the Sr1 structure with nucleotide-bound TRAP1 shows conservation of many residues involved in nucleotide binding and hydrolysis. D, a steady-state ATPase activity assay shows low activity for the sacsin Sr1 domain alone and as a GST fusion. GST was used as a negative control, and Hsc70 (including sacsin DnaJ and C-Bag1) was used as a positive control.
mutants in the context of the isolated Sr1 domain (Fig. 4A). The D168Y and L308F mutations produced soluble protein in yields similar to the WT construct. Expression levels of the other three mutants were severely reduced (Fig. 4F). We used thermal shift assays (also called differential scanning fluorimetry) to measure the thermal stability of the D168Y and L308F mutants and compared them with the WT protein. The conservative L308F mutation slightly lowered the melting temperature compared with the WT protein (51 versus 52 °C), whereas the D168Y mutation (46 °C) significantly destabilized the protein (Fig. S6).

These results correlate well with the SRP3 structure. Thr201 is buried, and introducing a lysine side chain would disrupt the hydrophobic core of the protein (Fig. 4B). Arg272 is also buried; its guanidinium group makes hydrogen bonds to carboxyls of Tyr151 and Gly153 and to the side chain of Tyr157 (Fig. 4D). Mutating Arg272 to histidine or cysteine abrogates these contacts and would be expected to destabilize the protein. Asp168 and Leu308 are more intriguing as both residues are solvent-exposed, and their disease-related mutations do not cause structural changes. These residues could be involved in functional aspects of saccin. In particular, Asp168 is positioned in the putative nucleotide-binding site, and introduction of bulky tyrosine side chain could severely interfere with nucleotide binding (Fig. 4C). Thr201, Arg272, and Asp168 are invariantly conserved in all SIRPTs, highlighting their structural and functional importance (Fig. S4).

The function of Leu308 is less clear; it is located on a small hydrophobic patch on the domain surface and may be involved in binding substrates or another domain of saccin. Leu308 is substituted with isoleucine in SIRPT2 and methionine in SIRPT3, which leaves unanswered the question of why the L308F mutation causes ARSACS.

Conformation of Ubl-Sr1 in solution

Having solved the Ubl and Sr1 structures separately, we wondered whether they orient independently in solution or form a relatively rigid tandem. To address this, we performed small-angle X-ray scattering (SAXS) experiments with the saccin fragment containing both domains (residues 1–339). SAXS can be used to calculate the radius of gyration ($R_g$) and generate molecular shape reconstructions and analyze dynamics (32, 33). We collected scattering data on Ubl-Sr1 at concentrations up to 6 mg/ml (Fig. S7). The Ubl-Sr1 fragment showed good behavior with no apparent aggregation as attested by the linearity of the Guinier plot. The estimated molecular mass from the SAXS data is 44.5 kDa, which is consistent with the theoretical mass of 38.7 kDa. This is in agreement with multilangle light scattering data for the same construct.

We took advantage of the structure of individual domains to estimate the degree of flexibility using dynamic ensemble optimization (34). Briefly, a pool of 10,000 structures for the Ubl and Sr1 domains linked by a flexible tether was generated, and subsets from this pool were selected to fit the SAXS data. The distributions of $R_g$ from the fitted ensembles in two independent runs are narrower than that of the random pool (Fig. S7), suggesting that the two domains adopt a compact configuration. The statistical factor $R_flex$ of 67% is less than that for the random distribution (88%), and $R_g$ is slightly less than 1, suggesting that the system has limited flexibility. Starting from this observation, we used rigid-body modeling with CORAL to determine the global shape of the fragment (35). The structures of Ubl and Sr1 were used to define the boundaries of the rigid domains and length of the interdomain linker. Superposition of the 30 lowest-energy models using the Sr1 model revealed that 90% of them converged to one cluster (Figs. 5 and S7). Altogether, the cluster showed a good fit between the predicted and observed scattering (Fig. 5).

Although the calculations converged to a single orientation, we do not believe that the Ubl-Sr1 domains necessarily form a rigid pair in solution as NMR titrations of 15N-labeled Ubl domain with Sr1 did not detect an interaction (data not shown). However, transient interactions between Ubl and Sr1 are very likely considering the short length of the linker between the domains.

Discussion

In the 20 years since the discovery of saccin, progress in its structural characterization has been relatively limited. The full-length protein is very challenging to study due to its large size...
and low level of expression. At the same time, the success of the divide-and-conquer approach typically used by structural biologists has been restricted to only a few regions. The reasons include difficulties in defining the boundaries of the independently folded domains as well as the possible requirement for interactions between distant parts of sacsin for protein expression or stability. Before this study, the structures of only two domains have been determined: the HEPN domain (17) and DnaJ (PDB code 1IUR). Here, we double the number of known structures. Although the SIRPT1 Sr1 structure accounts for only 5.4% of amino acids in sacsin, the expected structural similarity to the Sr1 domain in SIRPT2 and SIRPT3 increases the coverage to ~16%. With the N-terminal Ubl domain and previous structures, we now have structural knowledge of close to 25% of this gigantic protein.

Sacsin is expected to be a dimer in solution due to dimerization of its C-terminal HEPN domain (17). There are still numerous uncharacterized regions in sacsin that could also dimerize. Although the isolated sacsin Ubl does not form a stable dimer in solution, it could form the swapped dimer observed in the protein crystals in the context of an already dimeric protein. Hsp90s are dimers, so it would not be surprising if the Sr1 domains were associated with as-yet-uncharacterized dimerization domains in sacsin.

The initial suggestion that sacsin is a putative chaperone came from identification of DnaJ and Hsp90-like sequences (2). This was followed by the demonstration of activity of the sacsin DnaJ domain in E. coli cells and ATPase and chaperone activity of the SIRPT1 fragment (12, 36). Although our work shows that an isolated Sr1 domain does not bind nucleotides with appreciable affinity, this is not inconsistent with Sr1 ATPase activity in the context of the larger SIRPT fragment or full-length sacsin. Similar uncertainties arose before the definitive demonstration of ATP binding by the Hsp90 NBD domain (31). Several facts argue for Sr1 ATPase activity. First, similarity between Sr1 and Hsp90 proteins is highest in the nucleotide-binding site of the NBD. Many of the residues involved in ATP binding and hydrolysis in Hsp90 proteins are conserved in all three Sr1 domains (Fig. S4). Sr1 domains have the highest occurrence of ARSACS-causing missense mutations among all the known sacsin domains, which underlines their functional importance (19). Second, the D168Y mutation decreased ATP hydrolysis by SIRPT1 (residues 1–1456) (36), which is consistent with a role in catalysis.

Given that SIRPT modules make up more than 80% of this large protein, it is likely that the SACS gene was formed via triplication of the SIRPT region. The Sr1 domains appear not to be functionally redundant considering that single mutations in any Sr1 lead to disease. The activity of all three modules is necessary to perform the protein function. There are six Sr1 modules in a sacsin dimer, possibly arranged as three pairs. To our knowledge, there have been no reports of dominant-negative alleles among ARSACS patients.

The importance of Ubl domain is less clear as no pathological missense mutations have been found in that domain. The HEPN domain would appear to be essential based on the pathology of a missense mutation (N4549D), but it is possible that the mutation has indirect effects on the protein expression or stability (37).

What is the function of yet uncharacterized sacsin domains? By analogy with Hsp90 proteins, the Sr2 modules, which directly follow Sr1s, could serve as substrate-binding domains. It is clear that more studies are needed to assign functions to the many parts of sacsin for which we have no structural information. Ultimately, we need to understand how the global structure of sacsin defines and affects its function. The structural characterization of larger multidomain fragments will be particularly important for understanding sacsin and identifying avenues for the treatment of ARSACS.

Experimental procedures

Cloning, protein expression, and purification

The Ubl (residues 2–85), Sr1 (residues 89–336), and Ubl-Sr1 (residues 1–339) domains were cloned by PCR using human cDNA as a template in pGEX-6P-1 (Amersham Biosciences) between BamHI and Xhol. The site-directed mutants of the Ubl and Sr1 domains were via standard site-directed mutagenesis using mismatched primers and PCR with TaqPlus Precision polymerase (Stratagene, La Jolla, CA) following the manufacturer’s protocol. Automated DNA sequencing was performed to verify all sequence modifications.

Proteins were expressed in a BL21 E. coli strain. The cells were grown at 37 °C in Luria broth to an optical density of 0.8, and expression was induced with 0.5 mM isopropyl-β-D-1-thio-galactopyranoside at 30 °C for 4 h or 18 °C overnight. After centrifuging the cells, the pellets were resuspended in phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 2 mM KH2PO4, pH 7.4) containing 1 mM phenylmethylsulfonyl fluoride, 0.1 mg/ml lysozyme, and 80 units of DNase and lysed by sonication. Cell debris was removed by centrifugation, and the GSH-S-transferase (GST) fusion protein was purified using GSH-agarose resin (Pierce). After eluting the proteins in PBS containing 20 mM GSH, the tag was removed by cleavage with the 3C rhinovirus protease. The proteins were additionally purified by size-exclusion chromatography on a Superdex75 gel filtration column (GE Healthcare) in Ubl buffer (10 mM HEPES, pH 7.0, 50 mM NaCl, 1 mM DTT) or Sr1 buffer (20 mM HEPES, pH 7.0, 100 mM NaCl, 3 mM DTT).
For \(^{15}\)N labeling, the cells were grown in M9 minimal medium supplemented with \([^{15}\text{N}]\)ammonium chloride as the sole source of nitrogen. For selenomethionine labeling, the plasmid was transformed into a methionine auxotroph DL41 (DE3) E. coli strain, and the cells were grown in LeMaster medium supplemented with selenomethionine. The expression and purification protocols were the same as for the native protein.

**Crystallography, data collection, and processing**

The Ubl domain (WT or SeMet-L78M) was crystallized using the hanging drop vapor diffusion method by equilibrating a 1.5–μl drop of protein (3.2 mg/ml) with 1.5 μl of reservoir solution containing 250 mM NaCl, 100 mM Bicine, pH 8.6, 1 mM lead acetate, and 10% glycerol. Ubl crystals grew within 16 h at 22 °C. For data collection, glycerol was added to a final concentration of 25% for cryoprotection.

For the Sr1 domain, initial conditions were identified by sitting drop vapor diffusion using the Classics II screen (Qiagen). The best crystals were obtained by equilibrating a 1.0-μl drop of protein (17 mg/ml) in 20 mM HEPES, pH 7.0, 100 mM NaCl, and 3 mM DTT with 1.0 μl of reservoir solution containing 0.1 M MES, pH 6.5, 28% (w/v) PEG 3350, and 0.2 mM ammonium acetate with the drop suspended over 1 ml of reservoir solution. Selenomethionine-labeled proteins were crystallized in similar conditions. Sr1 crystals grew in 3–6 days at 22 °C. For data collection, 20% glycerol was added for cryoprotection, and the crystals were flash cooled in an N\(_2\) cold stream.

The crystals were cryoprotected with 10–25% glycerol or ethylene glycol and flash-cooled in an N\(_2\) cold stream. X-ray diffraction data were collected using an ADSC Quantum-210 CCD detector (Area Detector Systems Corp.) on beamline F2 at CHESS. The SeMet-labeled crystals diffraction data were collected using a single-wavelength (0.977 Å) anomalous dispersion method with the program PHENIX (39). The initial model was built using PHENIX (39), improved using Coot (43), and further refined using Refmac5 (41). Data from the WT human and lizard Sr1 constructs were phased by molecular replacement using the 3-Met Sr1 structure as the search model. The initial model was adjusted and completed with Coot (43). The final models were improved by several cycles of refinement using Refmac5 (41). The refinement statistics are shown in Table 1.

The final models have no outliers in the Ramachandran plot. Coordinates of the WT Ubl domain, L78M Ubl mutant, WT Sr1, 3-Met and 4-Met Sr1 mutants, and the anole Sr1 domain have been deposited in the Protein Data Bank with the accession codes 5VSX, 5VSZ, 5V44, 5V45, 5V46, and 5V47, respectively.

**Multiwavelength light scattering measurements**

Size-exclusion chromatography and multiwavelength light scattering experiments were performed using an analytical Superose 6 column (GE Healthcare) equilibrated in 20 mM HEPES, pH 7.0, 100 mM NaCl, and 3 mM DTT with in-line static light scattering by absorbance and differential refractive index detectors (miniDAWN TREOS and Optilab-rEX, Wyatt Technology Corp.). The sacsin Ubl-Sr1 was loaded at 7 mg/ml and Ubl was loaded at 7.4 mg/ml with a sample size of 50 μl at 22 °C. Data were analyzed with the ASTRA software package (Wyatt Technology Corp.).

**NMR spectroscopy**

NMR titration experiments were performed at 298 K on a Bruker 600-MHz spectrometer. For titrations with nucleotides, NMR samples of the \(^{15}\)N-labeled Sr1 (residues 89–336) construct were prepared as 0.18–0.20 mM in 90% buffer A and 10% D\(_2\)O. \(^{15}\)N-H heteronuclear single quantum correlation spectroscopy (HSQC) titrations were performed by stepwise addition of ATP or ADP to a final nucleotide concentration of 100 mM. The NMR spectra were processed by NMRPipe (44) and analyzed using SPARKY (45).

**Isothermal titration calorimetry**

ITC experiments were performed on a MicroCal VP-ITC titration calorimeter (Malvern Instruments Ltd.). The sample cell was loaded with 30 μM Sr1 domain, and the syringe contained 0.3 mM nucleotide. All experiments were carried out at 293 K with 19 injections of 15 μl with stirring at 310 rpm. Results were analyzed using Origin software (MicroCal).

**Steady-state ATPase assays**

Steady-state ATPase experiments were based on published procedures (46). Reactions contained 4 μM Hsc70, 4 μM GST-sacsin DnaJ, and 4 μM C-terminal domain of Bag1 or 8 μM either GST-Sr1, Sr1, or GST supplemented with 2 mM ATP and 5 μCi/ml \([\alpha-\text{32P}]\)ATP (PerkinElmer Life Sciences). Reactions were assembled on ice and initiated by incubation at 30 °C.
Tm was used to determine the melting temperature (Tm). All measurements were done in triplicates. Data were run on the StepOne Plus quantitative real-time PCR system (Life Technologies). Samples were heated from 25 to 99 °C at a rate of 1 °C/min, and fluorescence signals were monitored by the StepOne Plus quantitative real-time PCR system (Life Technologies). All measurements were done in triplicates. Data were analyzed using Thermal Shift software (Life Technologies). The maximum change of fluorescence with respect to temperature was used to determine the melting temperature (Tm).

**SAXS**

Small-angle X-ray scattering data were collected on SIBYLS beamline (12.3.1) at the Advanced Light Source at Lawrence Berkeley National Laboratory. Scattering data were collected for 0.5–1 s at 20 °C at protein concentrations of 1, 2, and 3 mg/ml for the Ubl-Sr1 construct. Background scattering from the buffer (20 mM HEPES, pH 7.0, 100 mM NaCl, 2% glycerol, 3 mg/ml for the Ubl-Sr1 construct) was subtracted. The scattering data were acquired at the Macromolecular Diffraction (MacCHESS) facility at the Cornell High Energy Synchrotron Source (CHESS). CHESS is supported by the National Science Foundation (NSF) and NIGMS, National Institutes of Health (NIH) via NSF Award DMR-0225180, and the MacCHESS resource is supported by National Center for Research Resources (NCRR), NIH Award RR-01646. 

Crystal structures of Ubl and Sr1 domains of human sacsin

Aliquots were taken at various time points, inactivated with 40 mM EDTA, then loaded on polyethyleneimine cellulose TLC plates (Fisher), and developed in 0.5 M LiCl and 0.5 M formic acid. The plates were exposed on a phosphorimaging screen, and the ADP produced was quantified with a Typhoon scanner and ImageQuant software (GE Healthcare). Linear ATPase rates were calculated by regression as mol of ADP/(mol of enzyme × min).

**Thermal shift assays**

Each reaction contained a total 20 μl of solution with 20 μM WT or mutant Sr1 (residues 89–336) in buffer and 1× Protein Thermal Shift dye (Protein Thermal Shift Dye kit, Life Technologies). Samples were heated from 25 to 99 °C at a rate of 1 °C/min, and fluorescence signals were monitored by the StepOne Plus quantitative real-time PCR system (Life Technologies). All measurements were done in triplicates. Data were analyzed using Thermal Shift software (Life Technologies). The maximum change of fluorescence with respect to temperature was used to determine the melting temperature (Tm).

SAAS

Small-angle X-ray scattering data were collected on SIBYLS beamline (12.3.1) at the Advanced Light Source at Lawrence Berkeley National Laboratory. Scattering data were collected for 0.5–1 s at 20 °C at protein concentrations of 1, 2, and 3 mg/ml for the Ubl-Sr1 construct. Background scattering from the buffer (20 mM HEPES, pH 7.0, 100 mM NaCl, 2% glycerol, 3 mg/ml for the Ubl-Sr1 construct) was measured for 0.5–1 s. Scaling, buffer subtraction, and merging were performed using Primus, part of the ATSAS package (35). The pair-distance distribution was calculated using GNOM (47). Molecular mass was estimated using the Porod volume method (35).

For rigid-body modeling and dynamic ensemble analysis, the Ubl and Sr1 domains had to be modified. Briefly, a monomeric model of the sacsin Ubl (residues 6–79) was built by merging the first β-strand from chain A (residues 6–14) and the remainder of the structure from chain B (residues 20–79). The loop (residues 15–19) was built and regularized using Coot. The Sr1 was modified by building a loop between residues 174 and 190 for which no electron density was visible. The two domains were connected by a 14-residue flexible linker. EOM 2.0 was used to generate 10,000 structures using the two domains using the genetic algorithm for conformer selection. The genetic algorithm was performed 100 times twice to estimate the variability in the distribution of R2 values. The R2 flex of 67% is less than that for the random distribution (88%), and R2 root is slightly less than 1, suggesting that the system has limited flexibility (33). Rigid-body modeling of the SAXS data was performed using the software CORAL (35) using the same domain as described above. Thirty models with χ2 < 2.0 were generated. Ab initio modeling was performed using DAMMIF (48). Twenty models were generated and averaged using DAMAVER. The resulting coordinates were used to generate pseudodensities using Situs-pdb2vol (49) and contoured at the particle volume derived from the Porod invariant using UCSF-Chimera (40).
Crystal structures of Ubl and Sr1 domains of human sacsin

28. Sung, N., Lee, J., Kim, J. H., Chang, C., Tsai, F. T., and Lee, S. (2016) 2.4 Å resolution crystal structure of human TRAP1NM, the Hsp90 paralog in the mitochondrial matrix. *Acta Crystallogr. D Struct. Biol.* **72**, 904–911 CrossRef Medline

29. Ali, M. M., Roe, S. M., Vaughan, C. K., Meyer, P., Panaretou, B., Piper, P. W., Prodromou, C., and Pearl, L. H. (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 CrossRef Medline

30. Prodromou, C. (2012) The ‘active life’ of Hsp90 complexes. *Biochim. Biophys. Acta* **1823**, 614–623 CrossRef Medline

31. Prodromou, C., Roe, S. M., O’Brien, R., Ladbury, J. E., Piper, P. W., and Pearl, L. H. (1997) Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell* **90**, 65–75 CrossRef Medline

32. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 CrossRef Medline

33. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664 CrossRef Medline

34. Tria, G., Mertens, H. D., Kachala, M., and Svergun, D. I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *ILCifr* **2**, 207–217 CrossRef Medline

35. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kihkney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 CrossRef Medline

36. Anderson, J. F., Siller, E., and Barral, J. M. (2011) The neurodegenerative-disease-related protein sacsin is a molecular chaperone. *J. Mol. Biol.* **411**, 870–880 CrossRef Medline

37. Gavrilov, A., and de la Rosa, R. (2013) Solution scattering (SAXS) combined with crystallography and computational analysis of sacsin. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 224–226 CrossRef Medline

38. Richter, A. M., Ozgul, R. K., Poisson, V. C., and Topaloglu, H. (2004) Private SACS mutations in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) families from Turkey. *Neurogenetics* **5**, 165–170 CrossRef Medline

39. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 CrossRef Medline

40. 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. I., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 2126–2132 CrossRef Medline

41. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1621 CrossRef Medline

42. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 CrossRef Medline

43. Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003) Macromolecular solution scattering (SAXS) combined with crystallography and computational analysis of sacsin. *Acta Crystallogr. D Struct. Biol.* **59**, 904–911 CrossRef Medline

44. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for macromolecular structure refinement. *Coot* **6**, 277–293 Medline

45. Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 CrossRef Medline
Crystal structures of Ubl and Sr1 domains of human sacsin

46. Tzankov, S., Wong, M. J., Shi, K., Nassif, C., and Young, J. C. (2008) Functional divergence between co-chaperones of Hsc70. *J. Biol. Chem.* 283, 27100–27109 CrossRef Medline

47. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 25, 495–503 CrossRef

48. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* 42, 342–346 CrossRef Medline

49. Wriggers, W., and Chacón, P. (2001) Modeling tricks and fitting techniques for multiresolution structures. *Structure* 9, 779–788 CrossRef Medline