The Sacsin Repeating Region (SRR): A Novel Hsp90-Related Supra-Domain Associated with Neurodegeneration

John F. Anderson¹, Efrain Siller¹ and Jose M. Barral¹,²* 

¹Department of Neuroscience and Cell Biology, The University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0620, USA 
²Sealy Center for Structural Biology and Molecular Biophysics, The University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0620, USA 

Received 17 February 2010; received in revised form 3 May 2010; accepted 11 May 2010 
Available online 19 May 2010 

Edited by M. Gottesman 

Keywords: protein supradomain; protein folding; ataxia; protein misfolding; molecular chaperone

Repeating elements within proteins provide versatile scaffolds for the generation of novel functions. Short tandem repeated motifs usually constitute an independent fold, such as the TPR and ankyrin repeats. A defining characteristic of these short motifs is their lack of occurrence in isolation. On the other hand, domains that often appear in repeats but that can also exist in isolation or within different 

protein contexts tend to be larger (>100 amino acid residues) and are generally functional by themselves. Examples include the thioredoxin domain and the immunoglobulin fold.

A common mechanism utilized throughout evolution for the generation of novel protein functions is domain recombination. There are certain combinations of domains, termed supra-domains, that are selected to be present in tandem, even though they are not necessarily part of a larger repeating unit. It is thought that a synergistic relationship that arises from the combination of partner domains provides a selective advantage to maintain them adjacent to one another. These supra-domains may move as a
We have found a new type of supra-domain in the protein sacsin, responsible for spastic ataxia of Charlevoix–Saguenay, an autosomal recessive human neurodegenerative disorder. The sacsin protein is peculiar in several respects. With a molecular mass of \( \sim 0.5 \) MDa, it is among the largest polypeptides known (4579 amino acid residues). Remarkably, the majority of the protein is encoded by a single, “giant exon” (\( \sim 11.5 \) kb), in fact the largest vertebrate exon identified to date. More N-terminal regions of the protein are encoded by traditional additional exons.

The original report on the sacsin protein described repeating regions, two of which contained similarity to the ATP-binding

**Fig. 1.** Phylogenetic distribution and domain architecture of the SRR domain. The approximate locations of SRR and additional domains identified by computer-based searches of the Conserved Domain Database (CDD) are indicated by colored boxes. True sacsin-like homologs are outlined by the gray rectangle (e.g., human sacsin). The J domain present in a plant SRR-containing protein is indicated by the red rectangle. A schematic representation of the composition of the SRR domain as an Hsp90-like region embedded within a region of novel sequence is provided in the top-right box, along with an approximate scale. Representative CDD abbreviations, names, and families for the additional domains identified are listed in the legend to the Supplementary Table.

**Fig. 2.** Multiple sequence alignment of representative SRR supra-domains. All SRRs identified were aligned using the E-INS-i strategy in MAFFT. Representative sequences were extracted from this alignment and edited using Jalview. The region of similarity to Hsp90 is indicated by an orange bar, while the novel sequence is marked with a gray bar, underneath the alignment. Residues known to coordinate nucleotide in yeast Hsp90 are marked with an “N”. Key residues in human sacsin are indicated (D168, R2703; see the text for discussion). A red arrow marks the position of a conserved arginine residue, which may participate in catalysis. The approximate location of predicted secondary structural elements is indicated by red cylinders (\( \alpha \)-helices) and green arrows (\( \beta \)-strands). Secondary-structure prediction was performed by submitting the alignment of all SRR domains to the Jpred server. The approximate location of known secondary-structure elements in yeast Hsp90 is indicated. Residues are colored according to the table provided at the lower right. Accession numbers for the sequences used are as follows: Hsapiens (Homo sapiens), NP_055178; Mmusculus (Mus musculus), Q9JLC8; Drerio (D. rerio), XP_001921714; Bfloridae2 (Branchiostoma floridae), XP_002599568; Tadhaerens (T. adhaerens), XP_002107908; Athaliana (Arabidopsis thaliana), NP_197702; Sclerotiorum (Sclerotinia sclerotiorum), XP_001590508; Ddiscoideum (Dictyostelium discoideum), XP_647015; Hwalsbyi (Haloquadratum walsbyi), YP_658972; Hmarisnigri (Methanoculleus marisnigri), YP_001047385; Bovatus (Bacteroides ovatus), ZP_02068428; Bcapillosus (Bacteroides capillosus), ZP_02038342; Afumigatus4 (Aspergillus fumigatus), EDP48569; Caprvum1 (Cryptosporidium parvum), XP_627068.
Characterization of a Sacsin Supra-Domain

Fig. 2 (legend on previous page)
Characterization of a Sacsin Supra-Domain
domain of Hsp90. The literature describes several other domains in human sacsin. These include ubiquitin, XPCB, J, and HEPN domains. The XPCB domain is a protein-interaction module that makes contacts via hydrophobic bonds. The J domain is present in co-chaperones involved in regulating the activity of the general chaperone Hsp70/DnaK while the HEPN domain is a putative nucleotide binding domain.

The presence of two regions of similarity to Hsp90 and the regions of self-homology identified in the literature led us to consider the nature of the repeating unit within sacsin. We performed a dot-plot analysis of the human sacsin protein against itself to define the borders of the repeating unit (data not shown). We identified the presence of three clear repeating homologous regions separated by longer regions of limited homology between the repeats (Fig. 1). There is weak self-similarity along the large inter-repeat tracts. However, these appear to be quite divergent and are clearly distinct from the three discrete repeats. For example, in human sacsin, 22% of the residues are identical across all three repeats, whereas an alignment between the three inter-repeat tracts reveals only ~4% identity. The identified repeats were unusually large, consisting of ~360 amino acid residues each.

Next, we investigated whether the two regions of similarity to Hsp90 and the regions of self-homology identified in the literature led us to consider the nature of the repeating unit within sacsin. We performed a dot-plot analysis of the human sacsin protein against itself to define the borders of the repeating unit (data not shown). We identified the presence of three clear repeating homologous regions separated by longer regions of limited homology between the repeats (Fig. 1). There is weak self-similarity along the large inter-repeat tracts. However, these appear to be quite divergent and are clearly distinct from the three discrete repeats. For example, in human sacsin, 22% of the residues are identical across all three repeats, whereas an alignment between the three inter-repeat tracts reveals only ~4% identity. The identified repeats were unusually large, consisting of ~360 amino acid residues each.

The PSI-BLAST search revealed that SRRs can be grouped into three SRR-containing proteins, which we defined as those that contain three SRR domains, one XPCB, one J, and one HEPN domain (Supplementary Table), and a broader group of “SRR-containing proteins”, which we defined as any protein containing at least one SRR. We have found that sacsin-like proteins are only present in animals, whereas SRR-containing proteins are widely represented across Eukarya, Bacteria, and Archaea. In addition to sacsin-like proteins, animals have several SRR-containing proteins with different identifiable domains, as do...
all other major kingdoms of eukaryotes (Fig. 1). Surprisingly, although this domain is very widespread, it is conspicuously absent from many organisms. For example, we could not identify any SRR domains in common model organisms, including *Escherichia coli*, *S. cerevisiae*, *Caenorhabditis elegans*, or *Drosophila melanogaster*.

A total of 261 unique SRR domains were identified out of the 181 SRR-containing proteins (Supplementary Table). As in human sacsin, these domains may occur in repeats of two to four copies per protein: 46 proteins contained two to four repeats. However, 135 of these proteins had a single SRR domain. Forty-eight of the proteins with single SRR domains were found to be in bacteria or archaea, and all archaeal SRR domains were found to be present in isolation (Fig. 1). The only animal example of a non-repeated SRR-containing animal gene we found was in *Trichoplax adhaerens*, perhaps the simplest free-living animal known.21 As a broad trend, more complicated organisms tended to have the SRR in a repeating fashion.

The strong tendency of SRR domains to occur in repeats along a single polypeptide chain prompted us to explore the significance and origins of this repetition. We constructed a phylogenetic tree of all SRR domains identified in animals (Fig. 3), which revealed a complicated past dominated by gene duplication and fusion events. The tree is readily divided into four groups. These correspond to groups A, B, C, and D. The latter includes genes from *Schistosoma mansoni*, *Hydra magnipapillata*, and *Danio rerio* that contain only one SRR domain, as well as four genes from *T. adhaerens* that contain only one SRR domain (blue boxes in Fig. 3). This group of genes in *T. adhaerens* appears to represent early and novel duplication events that never fused into a single polypeptide chain. These four genes are all relatively close together in the genome, within 65,000 base pairs.

Groups A, B, and C represent the origins of the three SRR domains present in that spatial order in most sacsin-like homologs. These appear to have triplicated and then fused into a single gene early on during evolution. However, several organisms provide examples of even more duplication events as well as convergent evolution. For example, zebrafish (*D. rerio*) contains two complete sacsin-like homologs. This is a rather simple case as it appears from the tree that a primitive fused gene

Fig. 4 (legend on next page)
Geldanamycin and radicicol were included at 60 μM in the green assay. Reactions were carried out with 1 μM protein (designated Spurpuratus_S_II in the tree). The sasxin-like homolog in this species appears not to have arisen as in most other animals. In this case, the SRR1 domain was not fused into the sasxin-like polypeptide and exists as its own protein (designated Spurpuratus_I in the tree). However, there are still three repeats in the S. purpuratus sasxin-like homolog. This likely occurred by duplication and fusion of two ancestral group B domains to become SRR1 and SRR2 of the final polypeptide. Notice that these two SRR domains arise from the same node in the tree, and thus, both the first and the second SRR in this gene fall into group B. This represents a likely example of convergent evolution, wherein this S. purpuratus homolog came to have a true sasxin homolog with three repeats via an alternate route.

Further examples of selective pressure to repeat the SRR domain are present in the SRR-containing proteins of the zebra finch Taeniopygia guttata. In this unusual case, there are two SRR-containing genes, but neither has three repeats. Rather, one gene contains two SRRs while the other one contains four. In order to attain a gene with four SRRs, this organism appears to have first followed the usual route to obtain a three repeat gene (with SRR1, SRR2, and SRR3 falling into Groups A, B, and C respectively) but subsequently duplicated the SRR3 and fused it at the C-terminus, to generate a gene with four SRR domains. Thus, both SRR3 and SRR4 are in Group C while SRR1 and SRR2 are in Groups A and B, respectively. This indicates a high level of selective pressure to repeat this domain. In plants, while there are no true sasxin-like proteins, a clear propensity for SRRs to occur as three tandem repeats was also observed (Fig. 1). We could not identify a common ancestral gene for plants and animals with three repeats and thus their three repeat proteins are likely another example of convergent evolution.

An additional example of convergent evolution in the plant SRR homologs involves the J domain. An SRR-containing protein in a variety of rice (Oryza officinalis) contains a J domain close to the C-terminus of the protein, in a configuration reminiscent of true sasxin-like homologs (outlined by a red box in Fig. 1). However, this protein does not contain any of the other domains present in sasxin-like proteins, and the sequence context around the J domain is clearly different from that found in the animal sasxin-like proteins. We did not identify any other SRR-containing proteins with J domains, and thus, it is likely that a

Fig. 4. Biochemical characterization of sasxin domains. A region of murine sasxin from the N-terminus to the beginning of SRR2 (residues 1-1456) was cloned into pVL1392 with an N-terminal His6 tag followed by a tobacco etch virus protease cleavage site. This was co-transfected with baculogold DNA (BD Biosciences) in Spodoptera frugiperda cells (sf9) to generate a recombinant baculovirus. A total of 2 × 10^7 sf9 cells are infected with a high titer viral stock, harvested after 72 h, and lysed by sonication. Recombinant protein was purified with a HisTrap Nickel column (GE Healthcare). Purity and identity of the protein were analyzed by SDS-PAGE, Western Blot (anti-His6 antibody, Millipore), and mass spectroscopy (UTMB Mass Spectrometry Core Facility). An oligonucleotide encompassing the mutated residue and convenient restriction sites was ordered from DNA2.0 to generate the mutant construct. The presence of the mutation was verified by DNA sequencing. (a) ATPase activity determinations. ATP hydrolysis was measured using the malachite green assay. Reactions were carried out with 1 μM protein, 1 mM ATP, 5 mM MgCl2, 30 mM Tris, pH 7.4, and 150 mM NaCl at 37 °C for 30 min. Increases in Abs600 over the first 5 min of the reactions were determined to be linear for obtaining v, and were correlated with standard curves of known concentrations of inorganic phosphate to determine kcat. Gelmanamycin and radiocil were included at 60 μM [or an equivalent amount (v/v) of dimethyl sulfoxide (DMSO) or ethanol (EtOH), respectively]. Values shown are background-subtracted and normalized to spontaneous ATPase hydrolysis. Reactions were performed in triplicate and are shown as mean±standard error of the mean. A t test was performed to determine the significance of the results. (b) Stimulation of Hsc70 ATPase activity by J domains. The J domain from mouse sasxin was cloned as a His6-tagged construct followed by a tobacco etch virus protease cleavage site. This was co-transfected with baculogold DNA (BD Biosciences) in Spodoptera frugiperda cells (sf9) to generate a recombinant baculovirus. A total of 2×10^7 sf9 cells are infected with a high titer viral stock, harvested after 72 h, and lysed by sonication. Recombinant protein was purified using a HisTrap Nickel column and Mono S cation exchange column (GE Healthcare). Hsc70 from bovine brain and recombinant human Hsp40 (Hdj-1) produced in E. coli were purified as previously described. Reactions were carried out with 1 μM protein, 1 mM ATP, 5 mM MgCl2, 30 mM Tris, pH 7.4, and 150 mM NaCl at 37 °C for 30 min. Increases in Abs600 over the first 5 min of the reactions were determined to be linear for obtaining v, and were correlated with standard curves of known concentrations of inorganic phosphate to determine kcat. Gelmanamycin and radiocil were included at 60 μM [or an equivalent amount (v/v) of dimethyl sulfoxide (DMSO) or ethanol (EtOH), respectively]. Values shown are background-subtracted and normalized to spontaneous ATPase hydrolysis. Reactions were performed in triplicate and are shown as mean±standard error of the mean. A t test was performed to determine the significance of the results. (b) Stimulation of Hsc70 ATPase activity by J domains. The J domain from mouse sasxin was cloned as a His6-tagged construct followed by a tobacco etch virus protease cleavage site. This was co-transfected with baculogold DNA (BD Biosciences) in Spodoptera frugiperda cells (sf9) to generate a recombinant baculovirus. A total of 2×10^7 sf9 cells are infected with a high titer viral stock, harvested after 72 h, and lysed by sonication. Recombinant protein was purified using a HisTrap Nickel column and Mono S cation exchange column (GE Healthcare). Hsc70 from bovine brain and recombinant human Hsp40 (Hdj-1) produced in E. coli were purified as previously described. Reactions were carried out with 1 μM protein, 1 mM ATP, 5 mM MgCl2, and 1 mM ATP. Reactions were performed for 10 min at 30 °C, in triplicate, background-subtracted, and normalized to Hsc70. Values shown are mean±standard error of the mean. Significance was determined using the t test. (c) Solubility analysis of wild-type and mutant sasxin SRRs. The fraction of soluble and insoluble recombinant protein for the N-terminal construct of mouse sasxin was determined by lysing cells with gentle sonication followed by centrifugation for 10 min at 20,000g. Similar amounts of Total (T), Super (S), or Pellet (P) fractions were analyzed by Western blot with an anti-His6 antibody (Millipore). (d) Emission spectra of bis-ANS binding to wild-type and mutant SRRs. Wild-type or mutant protein (100 nM) was combined with 3 μM bis-ANS in 30 mM Tris, pH 7.4, and 150 mM NaCl. The samples were excited at 295 nm and emission spectra were recorded from 305 to 405 nm. Spectra recorded with reactions containing buffer only were utilized for background subtraction.
convergent evolution strategy led to the placement of J domain C-terminally located to a region containing three SRRs in this case.

It is important to note that there is a strong tendency for the sequence C-terminal to the SRR domain to be devoid of identifiable domains for nearly 1000 residues or more. As mentioned earlier, there is weak self-similarity in these regions, but it appears to be rapidly diverging. We suggest that there may be a functional advantage for the SRR domains to be separated by a large region of sequence with few constraints. This tendency, combined with the occurrence of the SRR domain in repeats, suggests that there may be some higher-ordered structure produced by the repeated domains.

We next decided to characterize the biochemical properties of the domains described above. The N-terminal domain of Hsp90 is well known to bind and hydrolyze ATP. The importance of this activity is underscored by genetic experiments that demonstrate that ATP hydrolysis is indispensable for its chaperone activity in vivo. Mechanistically, ATP hydrolysis has been proposed to induce and/or regulate the conformation changes in this highly flexible molecule. Another important functional consequence of binding and hydrolysis of ATP is the formation of protein complexes. Indeed, several critical interactions of Hsp90, including those with the co-chaperones Stil and p23, are dependent on nucleotide hydrolysis. In the crystal structure of the yeast Hsp90 N-terminal domain, 17 residues are involved in nucleotide coordination. A multiple sequence alignment with all SRR domains (data not shown) revealed that 12 of these 17 residues are well conserved (in >70% of all sequences). Representative organisms taken from this alignment are shown in Fig. 2, with those residues identified from the crystal structure as putatively involved in coordinating nucleotide marked with an “N”.

In order to explore the biochemistry of the SRR domain, we generated a recombinant baculovirus for production in insect cells of the region of mouse sacsin from the N-terminus of the protein to the beginning of the second SRR domain. This protein was purified and its ATPase activity assayed by the malachite green method (Fig. 4). Wild-type SRR was found to hydrolyze ATP with a $k_{cat}$ of 2.5 min$^{-1}$, a specific activity similar to that of yeast Hsp90 ($k_{cat}$=1 min$^{-1}$). A useful property of several classes of Hsp90 proteins is the susceptibility of their ATPase activity to inhibition by certain antibiotics such as geldanamycin and radicicol. However, we found that our recombinant SRR is not inhibited by either of these compounds (Fig. 4). Similar insensitivity to geldanamycin has been reported for the cytosolic C. elegans Hsp90. Since SRRs are considerably more divergent from human Hsp90 than this C. elegans homolog (human sacsin SRR is only ~24% identical with the N-terminus of human Hsp90x1 while the C. elegans homolog is 74% identical), it is not entirely unexpected that SRRs are not sensitive to either of these drugs. These findings indicate that while these domains have retained their capacity to hydrolyze ATP, they constitute a class with sufficient structural divergence that it is no longer sensitive to these drugs, which may be of importance in the design of studies of sacsin function.

Next, we decided to assess whether the presence of a disease-causing mutation in this region (D168Y) had any measurable effects on the ability of this protein to hydrolyze ATP. Based on the alignment with yeast Hsp90, we identified the residue at the homologous position in the known crystal structure, which corresponds to K86. The presence of a charged side chain at this position is conserved in sacsin, albeit of opposite charge. The side chain of this residue appears to be solvent exposed and its available hydrogen bond donors and acceptors are not within allowable distance for standard hydrogen bonds. Thus, it is difficult to speculate on the structural-functional consequences resulting from the loss of charge in the human mutation. Nevertheless, we generated a construct identical with the above, except that the aspartic acid at position 168 was replaced by a tyrosine (D168Y), and examined the properties of the recombinant protein.

When overproduced in insect cells, the mutant protein remains soluble, indicating that the mutation does not result in global misfolding and aggregation (Fig. 4). Thus, we were able to purify this version and assessed its capacity to function as an ATPase. We found that the D168Y completely abrogated the ability of this protein to hydrolyze ATP (Fig. 4). This result serves as a control for the activity shown for the wild-type protein. Since the two proteins were expressed in the same cell type and purified by the same procedure, it is highly unlikely that the activity seen for the wild type is artifactual or the result of undetectable contamination by ATPases. Our results suggest that this mutation may not lead to global misfolding and/or degradation of the entire sacsin protein, but rather to loss of a biochemical function of one of its SRRs. Since the disease displays a clear recessive phenotype, the fact that the two other SRR domains are unable to confer function suggests that the activities of SRR domains are not simply redundant. Thus, it is likely that they function in a synergistic or cooperative fashion.

Since the mutant SRR is not globally misfolded, we decided to determine if we could detect more subtle conformation differences that could lead to a loss of nucleotide binding/hydrolysis. Thus, we performed fluorescence measurements of both wild-type and mutant versions of our construct. We began by probing the extent of surface hydrophobicity of both proteins by bis-1-anilinonaphthalene 8-sulfonate (ANS) fluorescence. Interestingly, there was a nearly 50% increase in bis-ANS emission with mutant sacsin than with wild type (Fig. 4). This may reflect the alteration of surface charge with the
homolog DnaJ in phenotypes of the genetic deletion of the Hsp40. Previous findings that it is capable of rescuing certain mutations have led to the occurrence of J domains in the same polypeptide beyond the biochemical activities demonstrated here for the individual components.

Acknowledgements

We thank E. R. Anderson, H. F. Epstein, and D. F. Boehnning for discussions and J. J. Chandler for technical assistance. This work was supported in part by Research Grant No. 5-FY07-641 from the March of Dimes Foundation and by an award from the Pew Scholars Program in the Biomedical Sciences to J.M.B.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.05.023

References


