Comparative Analysis and Functional Mapping of SACS Mutations Reveal Novel Insights into Sacsin Repeated Architecture

Alessandro Romano,1 Alessandra Tessa,2 Amilcare Barca,1 Fabiana Fattori,3 Maria Fulvia de Leva,4 Alessandra Terracciano,3 Carlo Storelli,1 Filippo Maria Santorelli,2,* and Tiziano Verri1,*

1Department of Biological and Environmental Sciences and Technologies, University of Salento, Lecce, Italy; 2Neurogenetics and Molecular Medicine, IRCCS Fondazione Stella Maris, Pisa, Italy; 3Neuromuscular Unit, IRCCS Bambino Gesù Hospital, Rome, Italy; 4Department of Neurology, Federico II University, Naples, Italy

Communicated by David N. Cooper

Received 28 May 2012; accepted revised manuscript 6 December 2012.

Published online 24 December 2012 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22269

ABSTRACT: Autosomal recessive spastic ataxia of Charlevoix–Saguenay (ARSACS) is a neurological disease with mutations in SACS, encoding sasin, a multidomain protein of 4,579 amino acids. The large size of SACS and its translated protein has hindered biochemical analysis of ARSACS, and how mutant sasins lead to disease remains largely unknown. Three repeated sequences, called sasin repeating region (SRR) supradomains, have been recognized, which contribute to sasin chaperone-like activity. We found that the three SRRs are much larger (≥1,100 residues) than previously described, and organized in discrete subrepeats. We named the large repeated regions Sacin Internal RePeaTs (SIRPT1, SIRPT2, and SIRPT3) and the subrepeats sr1, sr2, sr3, and srX. Comparative analysis of vertebrate sasins in combination with fine positional mapping of a set of human mutations revealed that sr1, sr2, sr3, and srX are functional. Notably, the position of the pathogenic mutations in sr1, sr2, sr3, and srX appeared to be related to the severity of the clinical phenotype, as assessed by defining a severity scoring system. Our results suggest that the relative position of mutations in subrepeats will variably influence sasin dysfunction. The characterization of the specific role of each repeated region will help in developing a comprehensive and integrated pathophysiological model of function for sasin.

Hum Mutat 34:525–537, 2013. © 2012 Wiley Periodicals, Inc.

KEY WORDS: autosomal recessive spastic ataxia of Charlevoix–Saguenay (ARSACS); comparative protein analysis; functional mapping of human mutations; neurodegeneration; protein domain architecture; repeated domains; SACS; sasin

Introduction

Autosomal recessive spastic ataxia of Charlevoix–Saguenay (ARSACS; MIM #270550) is an early-onset neurological disease presenting a founder effect in the Quebec regions of Charlevoix and Saguenay–Lac-St-Jean where the estimated carrier frequency is 1/22 [Bouchard et al., 1978; 1998]. The major clinical features of ARSACS include early-onset ataxia, later occurrence of spastic paraparesis, and brisk tendon reflexes, and an axonal sensory-motor peripheral neuropathy, with some instances of mental retardation or cognitive decline. Brain magnetic resonance imaging shows a distinct, tigroid appearance of the pons [Van Damme et al., 2009] and invariably an atrophied cerebellar vermis. Hypermyelination of the retinal nerve fibers [Bouchard et al., 1978, 1998] has long been considered a cardinal feature in Quebecois French–Canadian patients, and is not so obvious in cases from elsewhere [Crisciolo et al., 2004; Hara et al., 2005] or even absent. Several aspects including early appearance of abnormal pontocerebellar and retinal fibers seen at brain neuroimaging speak for a neurodevelopmental anomaly in ARSACS [Gazulla et al., 2012]. However, the progressive clinical course with involvement of the corticospinal tract and peripheral nerves in patients as well as studies in model mice questioned this hypothesis and suggested also the occurrence of a neurodegenerative process [Girard et al., 2012; Prodi et al., 2012].

The gene responsible for ARSACS (SACS) [Engert et al., 2000] is located on chromosome 13q12 and encodes sasin, a protein whose canonic variant is described as a polypeptide of 4,579 amino acids (GenBank acc. no. NP_055178.3). The enormous size of the SACS gene and translated protein has considerably hindered biochemical studies to date, and currently much more is known about the genetics of ARSACS than about the function of sasin in cells. Over the years, the number of ARSACS patients harboring mutations in the SACS gene has rapidly increased. They are distributed worldwide and are not limited to few ethnicities, and virtually any type of mutations has been discovered [Anheim et al., 2008].

How mutant sasin leads to neurodegeneration remains largely unknown. Earlier work had indicated that sasin might be involved in chaperone-mediated protein-folding activity [Engert et al., 2000] and play a role in regulating the Hsp70 chaperone machinery [Parfitt et al., 2009]. Recent biological and comparative genomic evidence suggested that sasin is organized in a repetitive supradomain structure of ~360 amino acids, named sasin repeating region (SRR)
[Anderson et al., 2010], which in turn might drive its function. Biochemical characterization demonstrated that such repetitive supradomain possesses ATPase activity, which appears to be a requirement for sacsin function, as a disease causing mutation leads to an alternate conformation incapable of hydrolyzing ATP [Anderson et al., 2010]. As well, this structure has been shown to enhance the refolding efficiency of a client protein, maintain it in soluble folding-competent states, and cooperate with members of the Hsp70 chaperone family to increase the yield of correctly folded client [Anderson et al., 2011]. Even more recently, sacsin has been shown to operate as a dimer and bind GTP at its C-terminus [Kozlov et al., 2011], with mutations in this region also resulting in loss of function. In addition, sacsin has been indicated as a potential substrate of the ubiquitin ligase Ube3A protein, which is responsible for Angelman syndrome (MIM #105830), a neurodevelopmental disorder with a motor component that shares same clinical aspects with ARSACS [Greer et al., 2010]. Such observations onto the function(s) of sacsin mainly arise from preliminary analysis on single putative domains that have been recognized along the sacsin sequence and are presently considered hallmarks of its structure. Finally, the generation of a sacsin knockout mouse is opening intriguing perspectives in the exploration of the pathophysiological basis of ARSACS, having shown that sacsin localizes to mitochondria and participates in regulation of mitochondrial dynamics via its interaction with dynamin-related protein 1 [Girard et al., 2012].

In the present work, we aimed at expanding our knowledge on the structure of sacsin. Three very large (≥1,100 amino acids) repeated regions were detected along the sacsin amino-acid sequence, each characterized by the occurrence of at least three subrepeats. A fourth subrepeat occurred in the first and third repeated region only. Such organization in domains is common to sacsin in all vertebrates including mammals, birds, reptiles, and fish. The comparative analysis of vertebrate sacsins architecture in combination with the fine positional mapping of a large set of disease causing mutations in human SACS well supported the concept of the functional nature of these novel domains. Furthermore, the location of a small selection of genetic variants detected in ARSACS was put in relation with the phenotype adopting a Spastic Ataxia (SPAX) rating system of clinical severity. Scoring mutations suggested original structure–function paradigms for sacsin, with hints on the relative relevance of novel and known domains in the activity of the protein.

Materials and Methods

Human SACS Gene, mRNA, and Protein Sequences and SNPs

The reference sequences for human (Homo sapiens) SACS gene (GenBank acc. no. NC_000013.10), mRNA (GenBank acc. no. NM_014363.4), and protein (GenBank acc. no. NP_055178.3) were as reported in Entrez Gene at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/gene). The human SACS gene SNPs mapped in this study (missense and nonsense mutations only) were from dbSNP at NCBI (http://www.ncbi.nlm.nih.gov/snp/) and from literature [Engert et al., 2006; Guernsey et al., 2010; Vermeul et al., 2009]. Throughout the manuscript, we systematically used names for both DNA and protein variations whenever appropriate, and adopted a mutation numbering system based on cDNA sequence as suggested by the internationally agreed mutation nomenclature (www.hgvs.org/).

Pattern and Profile Searches

Putative domains were defined using the pattern and profile searches tools included in the ExPaSy Proteomics Server (http://www.expasy.org/resources); in particular, the Simple Modular Architecture Research Tool 6 (SMART 6) (http://smart.embl-heidelberg.de/) [Letunic et al., 2009] and/or the ScanProsite tool (http://prosite.expasy.org/) [de Castro et al., 2006]. Internal repeats were detected by using the Prosopero program, as included in SMART 6. Default parameters were always used for analyses and only domains above threshold were represented. SIM, an alignment tool for analysis of local similarity in nucleotide and amino-acid sequences (http://web.expasy.org/sim/) [Huang and Miller, 1991], served to generate pairwise alignments of sacsin versus internal repeats using default parameters. The computed alignments were viewed using the graphical viewer program LALNVIEW (http://phil.univ-lyon1.fr/software/lalnview.html) [Duret et al., 1996]. Further details on the single computational tools and parameters used for analyses are reported in the legends to the figures as appropriate. Domains were drawn using the MyDomains image creator (http://prosite.expasy.org/mydomains).

Protein Sequence Alignments and Phylogenesis

On the basis of the genomic analysis detailed in the Supporting Information, the deduced protein sequences of orangutan, dog, horse, mouse, rat, chicken, zebra finch, anole lizard, fugu, tetraodon, stickleback, medaka, and zebrafish were obtained and used for alignments. Pairwise alignments of human versus the other vertebrate sacsin proteins were obtained by using SIM, as detailed above. Multiple sequence alignment of vertebrate sacsin proteins was obtained by using ClustalW2 using default parameters (http://www.ebi.ac.uk/Tools/clustalw2/index.html) [Larkin et al., 2007]. The phylogenetic reconstruction was generated by the neighbor-joining method [Saitou and Nei, 1987], as implemented in the Molecular Evolutionary Genetics Analysis 4 (MEGA4) software (http://www.megasoftware.net/) [Tamura et al., 2007].

Definition of a SPAX Scoring System (SPAX score) in ARSACS

Definition of a clinical score in ARSACS is lagging behind, although reliable and valid composite scores have been developed for the highly similar inherited ataxias [Truillas et al., 1997] and the hereditary spastic paraplegias [Schüle et al., 2006]. To define a posteriori a measure of disease severity in ARSACS and to correlate scores with type and location of mutations in sacsin, we put together a measure of severity in SPAX score that takes into account the ”core features” of ARSACS, including cerebellar ataxia, spastic paraplegia, and peripheral neuropathy. We are aware that SPAX scores are only an initial attempt to score disease severity, especially in the absence of functional tests, but the rating system has an intrinsic value in that it sums the gravity of the individual hallmarks of the disease through the use of validated scales. In particular, we used the parameters developed in the Scale for the Assessment and Rating of Ataxia [Schmutz-Hübsch et al., 2006] for cerebellar ataxia, the Spastic Paraplegia Rating Scale for motor symptoms and spasticity [Schüle et al., 2006], and the modified version of the Charcot–Marie–Tooth neuropathy score [Murphy et al., 2011] for peripheral neuropathy. In addition, cognitive impairment (0, absent to 3, if severe) and
ocular findings (from 0, normal to 4, maximal abnormality) were assessed. When visual abnormalities were detected only at optical coherence tomography, a unit was subtracted from the subscore. The several items of the scales were reviewed by two independent investigators blind to the genotype, duplicated items removed, data on single items averaged, and then corrected for disease duration whenever possible (or for averaged disease duration in a family). A grade of functional severity in ARSACS varying from 0 to 2 (maximal severity) was then calculated.

**Results**

**Identification of Novel Domains in Human SACS**

Along with the original description of human SACS [Engert et al., 2000], it was suggested that repeating regions, two of which containing the putative ATP-binding domain of Hsp90, might have occurred in the sacsin protein. At that time, human SACS was considered to consist of a single gigantic exon spanning 12,794 bp [Engert et al., 2000]. With the identification of nine (one noncoding and eight coding) additional exons upstream of this gigantic exon, the presence of conserved amino-acid sequences occurring in triplicate along the encoded protein started to be foreseen, and very recently the formal description of the SRR supradomain has been proposed [Anderson et al., 2010]. In this study, a systematic analysis of domains along the human sacsin amino-acid sequence was performed. In particular, besides the well-known ubiquitin-like (ubiquitin; PFAM acc. no. PF00240), DnaJ (DnaJ molecular chaperone homology domain; SMART acc. no. SM00271), and HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding domain; SMART acc. no. SM00748) domains (see Fig. 1B), two large Prospero repeats, corresponding to the 61–1,371 and 2,473–3,893 protein fragments of the human sacsin, were detected along the polypeptide chain (Fig. 1A).

Interestingly, both repeats shared similarity with a third homologous region in between them (along the 1,372–2,472 protein fragment), as detected by SIM analysis of sacsin protein versus each Prospero repeat (Fig. 1A). Similar results were also obtained by using the HHRepID program [Biegert and Soding, 2008] (data not shown). We named these three large homologous repeating regions Sacsin Internal RePeaTs (namely SIRPT1, SIRPT2, and SIRPT3; see Fig. 1). In spite of their low overall similarity (e.g., 16%–18% in human sacsin; with SIRPT1 vs. SIRPT2: 17%, SIRPT1 vs. SIRPT3: 16%, and SIRPT2 vs. SIRPT3: 18%), each SIRPT displayed, based on the degree of local similarity, at least three subrepeats (Fig. 1B) that were distanced by regions of extremely low similarity. We named these: subrepeat 1 (sr1), 2 (sr2), and 3 (sr3) (for position of the subrepeats along the protein, see Fig. 1B). Noteworthy, each sr1 contained a well-recognizable HATPase_c (Histidine kinase-like ATPases; SMART acc. no. SM00387) domain, which is adopted by the ATP-binding and catalytic domain of (among others) the members of the vast GHKL class of proteins (so-called after the founding members of the class: DNA Gyrase, Hsp90, bacterial histidine Kinases and MutL) [Dutta and Inouye, 2000] (Fig. 1B). Also, within the SIRPT architecture, sr1s and sr2s virtually corresponded to Anderson et al. SRR supradomain [Anderson et al., 2010] (Fig. 1B). On the other side, sr3 revealed no obvious relationships to any of the so far acknowledged domains included in databases. Besides the sr1, sr2, and sr3 domains described above, another repeated region could be identified in SIRPT1 and SIRPT3 in the area of very limited similarity. In fact, a long repeated region in SIRPT1 shared similarity with a homologous region in SIRPT3. We named this region srX (Fig. 1B). The srX domain had no obvious counterpart in the significantly shorter SIRPT2 (see Fig. 1A and B). Also, srX had no obvious similarity to any of the so far acknowledged domains in databases. Interestingly, in SIRPT3, the amino-acid sequence between srX and sr3 corresponded to a sacsin region previously reported to share limited homology with the Xeroderma Pigmentosum complementation group C binding (XPCB) domain of hHR23A [Kamionka and Feigon, 2004] and recently implicated in interactions with the ubiquitin ligase Ube3A [Greer et al., 2010] (Fig. 1B).

**Conservation of Sacsin Structural Organization among Vertebrates**

Comparative analysis of homologous proteins across phylogenetically distant species represents a powerful method for detecting conserved structural elements in proteins. Comparison of human sequences with sequences of other mammals, avians, reptiles, and teleost fish is valuable; in particular, teleosts offer maximal stringency for sequence comparisons among vertebrates. On this conceptual basis, we compared amino-acid sequence of sacsins from human with fish, having verified that: (1) genes encoding sacsin proteins are found in all vertebrate genomes sequenced so far, (2) sacsin proteins may have similar functional role(s) in all vertebrates, as supported by the evidence of similar expression patterns in mammals [Engert et al., 2000; Parfitt et al., 2009] and fish (such as zebrafish; see Supp. Fig. S1 and Supp. Table S1). In particular, as a result of a comprehensive gene analysis among vertebrates, sacsin proteins were deduced from human and other 13 vertebrate species, namely five mammals (orangutan, mouse, rat, horse, and dog), two birds (chicken and zebra finch), one reptile (anole lizard), and five fish (zebrafish, tetraodon, fugu, stickleback, and medaka). Then, the protein sequences were compared (for details, see Supp. Fig. S2), and the phylogenetic relationships among them are summarized in Figure 2A. With respect to the human protein, the other mammalian sacsins exhibited an overall degree of similarity (amino-acid identity) that varied from ~99% to ~93%, whereas the bird sacsins revealed an overall similarity of ~84% and the reptilian sacsin of 83% (for details, see Supp. Table S2). Fish proteins shared an overall degree of similarity with human sacsin that varied from ~70% to ~68% (Supp. Table S2). As expected, degrees of similarity locally varied along the protein sequence. The local degree of similarity is depicted in Figure 2B. In spite of these differences, all vertebrate sacsins conserved the same structural architecture as the human sacsin (see Supp. Fig. S3).

**Comparative Analysis of Vertebrate Protein Architecture and Positional Mapping of Human SACS Mutations Reveal the Functional Nature of the Sacsin Repeated Domains**

On the basis of our SIRPT-centered protein architecture and sequence similarity data, the following intra/intersequence alignment strategy was played out to identify and typify unique conserved elements in the repeated domains of the vertebrate sacsin proteins. Namely, the amino-acid sequences corresponding to SIRPT1-sr1, SIRPT2-sr1, and SIRPT3-sr1 from the human and other vertebrate sacsins were aligned against each other (see Supp. Fig. S4). The same procedure was applied to the sequences corresponding to SIRPT1-sr2, SIRPT2-sr2, and SIRPT3-sr2 (see Supp. Fig. S5), to SIRPT1-sr3, SIRPT2-sr3, and SIRPT3-sr3 (see Supp. Fig. S6), and to SIRPT1-srX and SIRPT3-srX (see Supp. Fig. S7) of the human and other vertebrate sacsins. Notably, in spite of the highly
selective alignment procedure, a number of amino-acid residues still kept appearing conserved in the same position of mate repeated domains.

If the sacsin repeated domains are functional, the amino acids that are found in these conserved positions should then be considered critical for sacsin function. Accordingly, in such repeated and conserved positions, one should expect to find more missense mutations associated with disease (missense pathogenic) than missense mutations not associated with disease (missense non-pathogenic) and/or nonsense (protein truncating) pathogenic mutations [Miller and Kumar, 2001; Miller et al., 2003]. To test this hypothesis, we collected missense (pathogenic and nonpathogenic) and nonsense mutations that have been reported to occur in human sacsin. In particular, Table 1 represents the recent update (January 2012) of all the acknowledged missense and nonsense mutations that are clearly pathogenic in ARSACS patients of different geographic origins (Supp. Appendix I lists pathogenic missense and nonsense mutations identified later than January 2012 and frameshift mutations not used in this study). Furthermore, Supp. Table S3 represents the list of all the missense mutations that have been described as SNPs in humans up to January 2012; for the most part, these mutations were recognized as undoubtedly nonpathogenic and were used for analysis (for details, see legend to Supp. Table S3) (Supp. Appendix II and Supp. Appendix III report a recent update of SNPs from dbSNP and NHLBI Exome Sequencing Project, respectively). Detailed positional information
Figure 2. Comparative analysis of vertebrate sacsin proteins. A: Unrooted phylogenetic tree depicting the evolutionary relationship of mammalian (orangutan, dog, horse, mouse, and rat), bird (chicken and zebra finch), reptilian (anole lizard), and fish (zebrafish, stickleback, medaka, fugu, and tetraodon) sacsin proteins. The unrooted tree was constructed using the neighbor-joining (NJ) method based on the alignment of the amino-acid sequences of the vertebrate sacsins. Bootstrap values (1,000 replicates) indicating the occurrence of nodes are reported above each branch in the figure. B: Schematic alignment of human versus the above listed vertebrate sacsin proteins. Pairwise sequence alignments and scores were generated using SIM. The computed alignments were visualized by LALNVIEW. Species are aligned according to their overall similarity (amino-acid identity) with respect to the human protein (from the highest to the lowest degree of overall similarity). Percent identity is reported in the figure. Different colors along the sequences are indicative of different degrees of similarity along the aligned sequences (black: 100%; white: nothing detected).
Table 1. Human SACS Gene (Missense, Nonsense, and Frameshift Leading to Immediate Premature Protein Truncation only) Pathogenic Mutations Mapped in this Study

| Mutation | Amino-acid change | Reference |
|----------|-------------------|-----------|
| **Origin** | **Exon** | **Original nucleotide position reported\(^a\)** | **Newly assigned nucleotide position in CDS\(^b\)** | **Original amino-acid position reported\(^c\)** | **Newly assigned amino-acid position in protein\(^d\)** |
| Synthetic construct | 3 | sdm | sdm | R50A | p.R50A | Parfitt et al., 2009 |
| Synthetic construct | 3 | sdm | sdm | R51A | p.R51A | Parfitt et al., 2009 |
| Synthetic construct | 4 | sdm | sdm | D60A | p.D60A | Parfitt et al., 2009 |
| The Netherlands | 7 | c.502G>T | c.502G>T | p.D168Y | p.D168Y | Vermeer et al. (2008, 2009) |
| Belgium | 7 | c.602C>A | c.602C>A | p.T201K | p.T201K | Bats et al. (2010) |
| Italy | 8 | c.815G>A | c.815G>A | p.L272H | p.L272H | In house database |
| Maritime Canada (Acadian descent) | 8 | c.826C>T | c.826C>T | p.R276C | p.R276C | Prodi et al. (2012) |
| Aragona-Spain/Croatia | 8 | c.922C>T | c.922C>T | p.L308F | p.L308F | Takado et al. (2007) |
| Japan | 8 | c.961C>T | c.961C>T | p.R321X | p.R321X | Vermeer et al. (2008, 2009) |
| Italy | 8 | c.1373C>T | c.1373C>T | p.T458I | p.T458I | In house database |
| The Netherlands | 8 | c.1420C>T | c.1420C>T | p.T474C | p.T474C | In house database |
| France | 8 | c.1475G>A | c.1475G>A | p.W492X | p.W492X | Vermeer et al. (2008, 2009) |
| Morocco | 8 | c.1667T>C | c.1667T>C | p.L556P | p.L556P | Bats et al. (2010) |
| Aragona-Spain | 8 | c.1894C>T | c.1894C>T | p.L632W | p.L632W | Gazulla et al. (2012) |
| The Netherlands | 9 | c.2182C>T | c.2182C>T | p.R742X | p.R742X | Terracciano et al. (2010) |
| Japan | 10 | g.2405T>C | c.2405T>C | p.L802P | p.L802P | Kamada et al. (2008) |
| Belgium | 10 | c.2971T>C | c.2971T>C | p.C991R | p.C991R | Bats et al. (2010) |
| Japan | 10 | 987T>C | c.3161T>C | p.F304S | p.F304S | Shimazaki et al. (2005) |
| Aragona-Spain | 10 | c.3198T>A | c.3198T>A | p.C1066Y | p.C1066Y | Gazulla et al. (2012) |
| Belgium | 10 | c.3491T>A | c.3491T>A | p.M1164K | p.M1164K | Ouyang et al. (2008) |
| Japan | 10 | 1858T>C | c.4035C>T | p.Q1345X | p.Q1345X | Okawa et al. (2006) |
| Italy | 10 | 3774C>T | c.4035C>T | p.Q1370X | p.Q1370X | Greco et al. (2004) |
| Turkey | 10 | g.2181T>C | c.4182T>C | p.C648R | p.C648R | Richter et al. (2004) |
| Italy | 10 | c.4198T>A | c.4198T>A | p.Y1400N | p.Y1400N | In house database |
| Italy | 10 | c.4567C>T | c.4567C>T | p.W1583R | p.W1583R | In house database |
| Serbia | 10 | c.4724G>C | c.4724G>C | p.R1575P | p.R1575P | Bats et al. (2010) |
| Spain | 10 | c.4784G>C | c.4784G>C | p.P1583R | p.P1583R | In house database |
| Belgium | 10 | c.4760A>G | c.4760A>G | p.H1587R | p.H1587R | Bats et al. (2010) |
| Algeria | 10 | c.4934G>A | c.4934G>A | p.E1654Q | p.E1654Q | In house database |
| The Netherlands (Turkish descent) | 10 | c.4957G>T | c.4957G>T | p.E1653X | p.E1653X | Vermeer et al. (2008, 2009) |
| The Netherlands | 10 | c.5125G>T | c.5125G>T | p.Q1709X | p.Q1709X | Vermeer et al. (2008, 2009) |
| The Netherlands | 10 | c.5143A>T | c.5143A>T | p.K1715X | p.K1715X | Vermeer et al. (2008, 2009) |
| Italy | 10 | c.5629C>T | c.5629C>T | p.R1877X | p.R1877X | Anezi et al. (2011) |
| Italy | 10 | c.5639C>T | c.5639C>T | p.T1880I | p.T1880I | In house database |
| Algeria | 10 | c.5719T>C | c.5719T>C | p.R1907X | p.R1907X | Prodi et al. (2012) |
| Tunisia | 10 | 3662T>C | c.5836T>C | p.W1946R | p.W1946R | El Euch-Fayache et al. (2003) |
| Japan | 10 | 6355C>T | c.7121T>C | p.L2374S | p.L2374S | Terracciano et al. (2009) |
| Algeria | 10 | 3649C>T | c.7121T>C | p.L2374S | p.L2374S | Terracciano et al. (2009) |
| Italy | 10 | c.6680T>C | c.7121T>C | p.L2374S | p.L2374S | Terracciano et al. (2009) |
| Belgium | 10 | c.7276C>T | c.7276C>T | p.R2426X | p.R2426X | Bats et al. (2010) |

(Continued)
Table 1. Continued

| Origin | Exon | Original nucleotide position reported<sup>a</sup> (see Reference, last column) | Newly assigned nucleotide position in CDS<sup>b</sup> | Original amino-acid position reported<sup>c</sup> (see Reference, last column) | Newly assigned amino-acid position in protein<sup>d</sup> | Reference |
|--------|------|-------------------------------------------------|------------------|----------------------------------------|-----------------------------------------------|----------|
| Quebec | 10   | g.5254T>C-T                                      | c.7504C>T        | R1752X                                 | p.R2502X                                       | Engert et al. (2000) |
| France | 10   | g.7673C>T                                        | c.7673C>T        | p.A2558V                                | p.A2558V                                        | Anheim et al. (2008) |
| Spain  | 10   | 7484C>T                                          | 8101C>T          | R2506C                                  | R2703C                                          | Criscuolo et al. (2005) |
| Peru   | 10   | c.8298_8299insTTT                                      | c.8298_8299insTTT | Y276X                                   | Y276X                                          | Baets et al. (2010) |
| Morocco| 10   | c.8393C>A                                        | c.8393C>A        | p.P798Q                                 | P2798Q                                         | Baets et al. (2010) |
| Aragona-Spain/Italy | 10   | c.8677A>T                                        | c.8677A>T        | R2893X                                  | R2893X                                          | In house database |
| Aragona-Spain | 10   | c.9760C>T                                        | c.9760C>T        | R3224X                                  | R3224X                                          | Gazulla et al. (2012) |
| Japan  | 10   | 7492T>C                                          | 9472T>C          | W2498R                                  | W2498R                                          | Ogawa et al. (2004) |
| Tunisia | 10   | c.10290C>G                                      | c.10290C>G       | Y3430X                                  | Y3430X                                          | Hmida Ben Brahim et al. (2011) |
| The Netherlands (English descent) | 10   | c.10442T>C                                      | c.10442T>C       | L3481P                                  | L3481P                                          | Vermeer et al. (2008, 2009) |
| The Netherlands | 10   | c.10906G>T                                       | c.10906G>T       | R3636X                                  | R3636X                                          | Vermeer et al. (2008, 2009) |
| Belgium | 10   | c.10907G>A                                      | c.10907G>A       | R3636Q                                  | R3636Q                                          | Baets et al. (2010) |
| Belgium | 10   | c.10934T>C                                      | c.10934T>C       | L364SP                                  | L364SP                                          | Baets et al. (2010) |
| Belgium | 10   | c.10954A>C                                      | c.10954A>C       | P3652T                                  | P3652T                                          | Baets et al. (2010) |
| Belgium | 10   | c.10957T>C                                      | c.10957T>C       | P36506                                  | P36506                                          | Breckpot et al. (2008) |
| Tunisia | 10   | c.10745C>T                                      | c.11185C>T       | Q3582X                                  | Q3582X                                          | Kamionka and Feigon (2004), Masciullo et al. (2008) |
| Tunisia | 10   | c.11374T>C                                      | c.11374T>C       | R3792X                                  | R3792X                                          | Bouhli et al. (2008) |
| Maritime Canada (Acadian descent) | 10   | c.11707T>C                                      | c.11707T>C       | R3903X                                  | R3903X                                          | Guernsey et al. (2010) |
| The Netherlands | 10   | c.12160T>C                                      | c.12160T>C       | Q4054X                                  | Q4054X                                          | Vermeer et al. (2008, 2009) |
| Tunisia | 10   | c.12426_12429insA                               | c.12426_12429insA | Y414X                                  | Y414X                                          | El-Euch-Fayad et al. (2003), Hmida Ben Brahim et al. (2011) |
| Italy  | 10   | 10723T>C                                         | 10723T>C         | R4325X                                  | R4325X                                          | Prodi et al. (2012) |
| The Netherlands | 10   | c.12992G>A                                      | c.12992G>A       | R4331Q                                  | R4331Q                                          | Vermeer et al. (2008, 2009) |
| Synthetic construct |  | sdm                                              | sdm              | R4331Q                                  | R4331Q                                          | Parfitt et al., 2009 |
| Italy  | 10   | c.12991C>T                                      | c.12991C>T       | R4331W                                  | R4331W                                          | Prodi et al. (2012) |
| Synthetic construct |  | sdm                                              | sdm              | R4331W                                  | R4331W                                          | Parfitt et al., 2009 |
| Belgium | 10   | c.13027G>A                                      | c.13027G>A       | H4337Q                                  | H4337Q                                          | Parfitt et al., 2009 |
| Belgium | 10   | c.13132T>C                                      | c.13132T>C       | E4343K                                  | E4343K                                          | Parfitt et al., 2009 |
| UK     | 10   | c.13237T>C                                      | c.13237T>C       | E4343K                                  | E4343K                                          | Baets et al. (2010) |
| Aragona-Spain | 10   | c.13405G>C                                      | c.13405G>C       | R4378X                                  | R4378X                                          | Anesi et al. (2011) |
| Belgium | 10   | c.13523A>C                                      | c.13523A>C       | Q4413X                                  | Q4413X                                          | Terracciano et al. (2010) |
| Turkey | 10   | g.13471A>G                                      | c.13645A>G       | N5499D                                  | N5499D                                          | Breckpot et al. (2008) |
| Synthetic construct |  | sdm                                              | sdm              | N5499D                                  | N5499D                                          | Kožlović et al. (2011) |

<sup>a</sup>Nucleotide positions and changes are indicated as reported in the original article and refer to different NCBI Reference Sequences (see Reference, last column). DNA mutation numbering system in use, based on cDNA sequence (with a “c.” symbol before the number) (www.ncbi.nlm.nih.gov), can be found for more recent mutations only. For older mutations, numbering system based on genomic sequences can be found (with a “g.” symbol before the number).

<sup>b</sup>On the basis of the following NCBI Reference Sequence: GenBank acc. no. NM_014363.4. DNA mutation numbering system in use is based on cDNA sequence (with a “c.” symbol before the number). Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, with the initiation codon being codon 1 (www.hgvs.org).

<sup>c</sup>Amino-acid positions and changes are indicated as reported in the original article and refer to different NCBI Reference Sequences (see Reference, last column). Amino-acid numbering system in use is based on protein sequence (with a “p.” symbol before the letter) (www.hgvs.org), can be found for more recent variants only.

<sup>d</sup>On the basis of the following NCBI Reference Sequence: GenBank acc. no. NM_014363.4. DNA mutation numbering system in use is based on cDNA sequence (with a “c.” symbol before the number). Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, with the initiation codon being codon 1 (www.hgvs.org).

<sup>e</sup>Double mutant alleles.

<sup>f</sup>Double mutant alleles.

<sup>g</sup>New variants not mapped in this study are reported in Supp. Appendix I.
Figure 3. Relative amount of conserved versus nonconserved missense mutations in SIRPT sr1, sr2, sr3, and srX domains. When mapped on our multiple alignments (see Supp. Figs. S4–S7), in each domain “conserved” (i.e., identical, conserved and semi-conserved, as assessed by ClustalW) pathogenic missense mutations were invariably over-represented with respect to missense nonpathogenic mutations (for details, see Supp. Table S5). Unclear mutations (i.e., variants not yet clearly associated with disease; for details, see Supp. Tables S3–S5) were omitted from the analysis.

and distribution of the mutations in the various domains along the human protein are summarized in Supp. Figure S2 and Supp. Table S4. All the mutations falling in positions within the sr1, sr2, sr3, and srX domains have been represented in Supp. Figures S4–S7. The relative amounts of missense pathogenic mutations, on one hand, and missense nonpathogenic mutations, on the other—expressed as percent of conserved vs. non-conserved mutations—are reported in Figure 3. As expected, with respect to the group of missense nonpathogenic mutations, missense pathogenic mutations were invariably over-represented in conserved positions in sr1, sr2, sr3, and srX (for details, see Supp. Table S5), thus suggesting that the four repeated domains of the SIRPT regions identified in this work (that include, with sr1 and sr2, and go beyond, with sr3 and srX, the SRR design; see Fig. 1B) [Anderson et al., 2010] do play a functional role in the sacsin protein.

The functional nature of sr1, sr2, sr3, and srX is also sustained by the observation that in human sacsin pathogenic missense mutations were found to be over-represented in these domains with respect to the regions between domains (interdomains), as qualitatively assessed by calculating the likelihood of occurrence of missense pathogenic mutations, that is, the ratio of the percentage mutations on a given region and the percentage of amino acids of the protein on the same region (for details, see Table 2). In particular, the calculated likelihood was 1.97, 1.42, 0.79, and 0.51 for sr1, sr2, srX, and sr3 domains, respectively, with respect to 0.35 for the interdomains.

Functional Relevance of sr1, sr2, srX, and sr3 in Sacsin Protein Based on Composite SPAX Scores Analysis

To investigate on the putative functional relevance of the various repeated domains that result from the proposed new sacsin architecture, we analyzed the clinical phenotype in patients selected for having, in a given domain (i.e., sr1, sr2, srX, and sr3), a missense pathogenic mutation (1) in homozygosis or (2) in heterozygosis with a frameshift mutation, a stop mutation, or a macrodeletion (see Table 3). It is reasonable to think that in an autosomal recessive disorder, such as ARSACS, frameshift mutations, stop mutations,
and macrodeletions can abolish sacsin function, although other mechanisms, such as dominant-negative effects, cannot totally be excluded until functional tests are performed. Under these conditions, we expect that differences in the clinical phenotypes observed in patients (1) are due to the effect(s) of the missense mutation on protein function and (2) provide (at least in part) information on the functional relevance of the protein domain where the missense mutation acts. In fact, although the nature of the substituted amino acid may contribute per se to the severity of the phenotype, it cannot be ignored that the effect of an amino-acid substitution depends on the protein domain where the substitution occurs. As a means to evaluate the pleomorphic clinical phenotype of AR-SACS, we defined a composite SPAX score, which takes into account the major core features (cognitive, cerebellar, spasticity, peripheral nerve, and retinopathy) that are part of the disease. This scoring system is largely based on validated rating scales for spasticity, peripheral neuropathy, and cerebellar function, corrected for disease duration and used to evaluate the severity of the clinical phenotype (see Table 3).

As a way to define the maximal severity of the disease, in our analysis, we initially calculated SPAX scores from patients in which both alleles were predicted to generate truncated sacsins (due to presence on both alleles of either a frameshift mutation or a stop mutation or a macrodeletion; for a description of various combinations of alleles, please see Supp. Table S6). As it results from the analysis of Figure 4, these patients formed a homogeneous group that ranked at the highest SPAX scores among those calculated in this study (for comparison, see also Table 3), with values varying from 1.48 to 1.84. Conversely, when SPAX scores were calculated from patients carrying a pathogenic missense mutation in sr1, sr2, srX, or sr3 (in homozygosity or heterozygosity with a frameshift mutation, a stop mutation or a macrodeletion, as described above), it was evident that the severity of the clinical phenotype largely varied (see Fig. 4 and Table 3) from values similar to those observed in patients carrying (a) truncated protein(s), for example, 1.69 for the c.1420C>T (p.R474C)/c.5719C>T (p.R1907X), which suggests nearly complete loss of function and (in at least two cases, see Fig. 4 and Table 3) from values similar to those observed in patients carrying a pathogenic missense mutation in sr1, sr2, srX, or sr3 (in homozygosity or heterozygosity with a frameshift mutation, a stop mutation or a macrodeletion), which suggests subsistence of partial or residual protein activity. Overall, the presence of the missense pathogenic mutations in sr1, sr2, srX, or sr3 established on average a set of phenotypes (i.e., SPAX scores) significantly less severe (i.e., lower) than those observed for mutations that generated truncated proteins (ANOVA; \( P < 0.0001 \)). We assumed that such a behavior correlated to the relevance that the domain in which the mutation falls had for sacsin activity. In particular, a trend to lower SPAX scores passing from sr1 and sr2 to srX and sr3 could be observed, suggesting (1) that alterations in srX and sr3 do cause less harmful, although measurable, effects on the function of the protein with respect to sr1 and sr2, and thus (2) that srX and sr3 play a "minority" role in the operational mechanism of the protein with respect to sr1 and sr2.

### Discussion

In this study, a systematic inspection of vertebrate sacsins has been carried out to identify repeated domains along the protein. By using a combination of standard databank consulting tools and bioinformatics methods, three large (\( \geq 1,100 \) amino acids) repeated regions have been identified. Such internal repeats, named SIRPT1, SIRPT2, and SIRPT3, cover \(~84\)% of the protein sequence, and each contains three subrepeats, named sr1, sr2, and sr3, with sr1 and sr2 falling into Anderson et al. SRR supradomain [Anderson et al., 2010]. In addition, a fourth subrepeat, named srX, occurs in the first and the third internal repeat only, in a region between sr2 and sr3. Our SIRPT-based architectural structure is invariably conserved in all vertebrate sacsins. This is not unexpected, as vertebrate sacsins share a high degree of similarity at both global and local level (this study), and most probably exert similar functional roles, a notion also supported by the observation that similar expression patterns can be found in both mammals [Engert et al., 2000; Parfitt et al., 2009] and fish (this study).

All the different subrepeats identified within the SIRPT architecture most likely represent regions involved in sacsin function. To answer this question, we have developed a strategy that combines very stringent alignments of the vertebrate sacsin domains with positional mapping of the human SACS mutations (for details, see Results). As a matter of fact, at least two pieces of evidence come out from our analyses indicating that the different subrepeats identified do represent functional regions. First, in sr1, sr2, srX, and sr3, missense pathogenic mutations are invariably over-represented in conserved positions with respect to missense non-pathogenic mutations. Second, missense pathogenic mutations are over-represented in sr1, sr2, sr3, and srX with respect to the regions between domains [Miller et al., 2003], this scheme being fully applicable also to the well-known DnaJ and HEPN domains. All together, these findings indicate that there is a strong tendency in the sacsin protein to gather the missense mutations associated with disease within the newly identified or the already known domains.

Sacsins is considered to operate in a chaperone-like manner, but very limited information is available on its activity, mainly due to the technical difficulties of managing with such an unusually long protein by means of standard biochemical, cellular, or molecular biology assays [Anderson et al., 2010; Kozlov et al., 2011; Parfitt et al., 2009]. Under such circumstances, achieving information on the functional role(s) of our sr1, sr2, srX, and sr3 domains represents a difficult task. In the effort to obtain new hints on the impact of the newly identified domains in the activity of the protein, we have developed a procedure that allows evaluation of the functional relevance of the domains by measuring the severity of the clinical phenotype, quantified in terms of SPAX score, in patients selected for carrying missense pathogenic mutations in sr1, sr2, srX, and sr3 in homozygosity or heterozygosity with a null allele (for details,

### Table 2. Percentage of the Amino Acids in a Given Region (% protein), Percentage of the Mutations in the Same Region (% mutation) and Likelihood of a Mutation Occurring in the Region (% mutation/% protein), Calculated as the Ratio of the Percentage Mutations on a Given Region and the Percentage of Amino Acids of the Protein on the Region

| Region          | Whole sacsin | Ubiquitin-like | sr1 | sr2 | sr3 | srX | XPCB | DnaJ | HEPN | Interdomains |
|-----------------|--------------|----------------|-----|-----|-----|-----|------|------|------|--------------|
| Protein (fragment) length (aa) | 4,579        | 72             | 817 | 436 | 481 | 1098 | 76   | 60   | 117 | 1,422 |
| % Protein       | 100          | 1.57           | 17.84 | 9.52 | 10.51 | 23.98 | 1.66 | 1.31 | 2.56 | 31.05 |
| Missense mutations | 37          | 0              | 13  | 5   | 2   | 7   | 0    | 3    | 3   | 4     |
| % Mutation      | 100          | 0              | 35.13 | 13.51 | 5.41 | 18.92 | 0.00 | 8.11 | 8.11 | 10.81 |
| Likelihood (% mutation/% protein) | 1.00      | 0.00           | 1.97 | 1.42 | 0.51 | 0.79 | 0.00 | 6.19 | 3.17 | 0.35 |

aa, amino acids.
Table 3. Composite SPAX Score Assigned to Selected ARSACS Patients

| SACS mutation | Sacsin internal repeat (SIRPT) subregion (sr) or interdomain where the mutation is located | Duration (yrs) | Onset | Cognitive | Cerebellar | Spasticity | Peripheral neuropathy | Retinal | Composite SPAX score |
|---------------|-----------------------------------------------------------------------------------------|---------------|-------|-----------|-----------|-----------|----------------------|--------|---------------------|
| c.4182T>C (p.C1398R) | SIRPT1–sr3 and SIRP2–sr1 | 16 | 3 | 1 | 1 | 1 | 1 | 0.88 |
| c.4198T>A (p.Y1400N) | SIRPT1–sr3 and SIRP2–sr1 | 11 | 2 | 0 | 1 | 2 | 2 | 1 | 0.55 |
| c.4198T>A (p.Y1400N) | SIRPT1–sr3 and SIRP2–sr1 | 15 | 3 | 1 | 3 | 1 | 3 | 3 | 1.24 |
| c.4182T>C (p.C1398R) | SIRPT1–sr3 and SIRP2–sr1 | 15 | 1 | 1 | 3 | 1 | 3 | 3 | 1.34 |
| c.814C>T (p.R272C) | SIRPT1–sr3 and SIRP2–sr1 | 24 | 3 | 1 | 2 | 2 | 3 | 0 | 1.10 |
| c.5639C>T (p.T1880I) | SIRPT1–sr3 and SIRP2–sr1 | 24 | 3 | 1 | 2 | 2 | 3 | 3 | 1.30 |
| c.1894C>T (p.R632W) | SIRPT1–sr3 and SIRP2–sr1 | 17 | 2 | 1 | 3 | 2 | 1 | 0 | 0.67 |
| c.9742T>C (p.W3248R) | SIRPT1–sr3 and SIRP2–sr1 | 22 | 2 | 2 | 1 | 2 | 1 | 3 | 0.84 |
| c.2405T>C (p.L802F) | SIRPT1–sr3 and SIRP2–sr1 | 36 | 3 | 3 | 0 | 2 | 0 | 0 | 0.81 |
| c.31405G>C (p.A4469P) | SIRPT1–sr3 and SIRP2–sr1 | 17 | 2 | 1 | 2 | 0 | 3 | 0 | 0.87 |
| c.1894C>T (p.R632W) | SIRPT1–sr3 and SIRP2–sr1 | 26 | 1 | 2 | 1 | 1 | 2 | 0 | 0.73 |
| c.9742T>C (p.W3248R) | SIRPT1–sr3 and SIRP2–sr1 | 24 | 2 | 0 | 3 | 2 | 1 | 2 | 0.70 |
| c.1894C>T (p.R632W) | SIRPT1–sr3 and SIRP2–sr1 | 13 | 3 | 1 | 3 | 2 | 2 | 1 | 1.69 |
| c.2405T>C (p.L802F) | SIRPT1–sr3 and SIRP2–sr1 | 34 | 2 | 0 | 3 | 3 | 1 | 1 | 1.50 |
| c.31405G>C (p.A4469P) | SIRPT1–sr3 and SIRP2–sr1 | 17 | 2 | 1 | 1 | 2 | 1 | 0 | 0.67 |
| c.1894C>T (p.R632W) | SIRPT1–sr3 and SIRP2–sr1 | 17 | 2 | 1 | 1 | 2 | 1 | 0 | 0.67 |
| c.9742T>C (p.W3248R) | SIRPT1–sr3 and SIRP2–sr1 | 22 | 2 | 2 | 1 | 2 | 1 | 3 | 0.84 |
| c.2405T>C (p.L802F) | SIRPT1–sr3 and SIRP2–sr1 | 17 | 1 | 2 | 3 | 1 | 2 | 0 | 0.88 |
| c.31405G>C (p.A4469P) | SIRPT1–sr3 and SIRP2–sr1 | 40 | 2 | 0 | 1 | 2 | 1 | 0 | 0.76 |
| c.1894C>T (p.R632W) | SIRPT1–sr3 and SIRP2–sr1 | 36 | 3 | 3 | 0 | 2 | 0 | 0 | 0.81 |
| c.9742T>C (p.W3248R) | SIRPT1–sr3 and SIRP2–sr1 | 26 | 1 | 2 | 1 | 2 | 0 | 3 | 0.83 |
| c.2405T>C (p.L802F) | SIRPT1–sr3 and SIRP2–sr1 | 15 | 2 | 1 | 2 | 0 | 3 | 3 | 0.87 |
| c.31405G>C (p.A4469P) | SIRPT1–sr3 and SIRP2–sr1 | 26 | 1 | 1 | 2 | 2 | 1 | 0 | 0.73 |
| c.1894C>T (p.R632W) | SIRPT1–sr3 and SIRP2–sr1 | 17 | 2 | 1 | 1 | 2 | 1 | 0 | 0.69 |
| c.9742T>C (p.W3248R) | SIRPT1–sr3 and SIRP2–sr1 | 24 | 2 | 2 | 1 | 2 | 1 | 0 | 0.66 |

del, base deletion (microdeletion); Δ, deletion (macrodeletion); fs, frameshift; ins, insertion; yrs, years.

See also Table 1.

Numbering based on the following NCBI Reference Sequences: GenBank acc. no. NM_014363.4, for nucleotide, and GenBank acc. n. NP_055178.3, for protein. DNA mutation numbering system in use is based on cDNA sequence (with a “c.” symbol before the number). Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, with the initiation codon being codon 1. Amino-acid change numbering system in use is based on protein sequence (with a “p.” symbol before the letter) (www.hgvs.org). Original nucleotide and amino-acid positions and changes can be found in the original articles (see Reference). (Continued)
Individual Items to Score Disease Severity

Table 3. Continued

| Score | Onset   | Cognitive            | Cerebellar<sup>a</sup> | Spasticity<sup>b</sup> | Peripheral neuropathy<sup>c</sup> | Retinal                           |
|-------|---------|----------------------|------------------------|------------------------|----------------------------------|----------------------------------|
| 0     | Adult   | Absent               | Absent                 | Absent                 | Absent                           | Absent                           |
| 1     | Juvenile| Mild decline         | Mild                   | Mild                   | Mild                             | No functional impairment but aware of worsened acuities |
| 2     | Teen    | IQ lower than peers  | Moderate               | Moderate               | Moderate                         | Reduced night vision             |
| 3     | Early-onset | Marked mental retardation | Severe               | Severe                 | Severe                           | Abnormal fundoscopy or ERG       |

IQ, intelligence quotient; ERG, electroretinogram.

<sup>a</sup>On the basis of SARA: Scale for the Assessment and Rating of Ataxia and IACRS (Inherited Ataxia Clinical Rating Scale).

<sup>b</sup>On the basis of SPRS: Spastic Paraplegia Rating Scale.

<sup>c</sup>On the basis of CMT (Charcot–Marie–Tooth) neuropathy score (second version).

Our results extend and refine the current knowledge on the organization of some sacsin domains. In particular, the sr1 and sr2 domains identified in this work substantially form the SRR supradomain recently defined by others [Anderson et al., 2010]. This supradomain is composed of an N-terminal portion (~160 residues), which is homologous to the HATPase<sub>c</sub> domain of Hsp90, and a C-terminal portion (~200 residues), which consists of a novel sequence invariably connected to the HATPase<sub>c</sub> domain [Anderson et al., 2010]. Our bioinformatics approach divides this SRR supradomain in two well-defined repeated domains, that is, sr1 and sr2, which are separated by an evident nonrepeated linker segment. This organization is coherent with a system that works as an Hsp90-like protein. In fact, in Hsp90-type chaperones, the ATP binding domain is connected to the middle domain via a divergent linker region. In particular, in our sacsin organization, sr1 represents the ATP binding domain and sr2 the middle domain. Notably, in Hsp90 the middle domain invariably contains an arginine residue accepting phosphate after ATP hydrolysis [Pearl and Prodromou, 2006]. This phosphoacceptor arginine, already observed by Anderson et al. (2010) as invariably conserved in each C-terminal region of their SRR supradomains, does occur in each sr2 domain. Interestingly, our study clearly demonstrates the crucial role of this arginine in the operational mechanism of sacsin. In fact, a mutation occurring on one of such conserved arginines, namely c.1420C>T (p.R474C) in SIRPT1-sr2, associates to one of the highest SPAX scores (1.69) found in this survey.
In this article, we report for the first time the occurrence of two novel repeated domains, namely srX and sr3, downstream of the Hsp90-type regions discussed above. Such domains share no similarity to any domains reported so far in databanks, and no obvious role can be assigned to them. However, in the context of an Hsp90-like scheme of function, srX and/or sr3, located near the sr1/sr2 “biochemical clamp” that allows ATP binding and hydrolysis, may participate (via dimerization, client binding, cochaperone interaction, regulation, etc.) to sacsin chaperone activity. In this respect, there has been recent demonstration that a large sacsin region (RegA), virtually corresponding to our SIRPI, do exhibit a chaperone-like activity that can be detected in vitro by standard biochemical approaches [Anderson et al., 2011]. Such protein module is composed of the Hsp90-like region and of a large undefined downstream region. However, our study identifies srX and sr3 as functional elements in that large undefined region and likely involved in the chaperone activity of the whole module.

In conclusion, we used a functional comparative genomics approach that combines bioinformatics sequence examination tools to mapping and phenotypical analysis of human mutations, to provide novel information on the organization in repeated domains of sacsin. In particular, our results establish that large portions of the protein can be arranged in a few and well-defined repeated domains. The demonstration of the functional nature of sr1, sr2, srX, and sr3 suggests that these regions contribute to the activity of the protein. Further studies are needed to define the specific role(s) of such domains, in the perspective of developing a comprehensive and integrated model of function for sacsin in the context of cell pathology. In a larger perspective, our approach that combines comparative analysis of vertebrate protein sequences/architecture, positional mapping of human mutations, and severity of clinical phenotype can be tentatively applied in the biomedical field to shed light on the functional nature of other proteins associated to disease but of yet unknown function.

Acknowledgment
Disclosure statement: The authors declare no conflict of interest.

References
Anderson JF, Siller E, Barral JM. 2010. The sacsin repeating region (SRR): a novel Hsp90-like domain encoding an 11.5-kb ORF. Nat Genet 24:120–125.

Anderson JF, Siller E, Barral JM. 2010. The sacsin region (RegA), virtually corresponding to our SIRPI, do exhibit a chaperone-like activity that can be detected in vitro by standard biochemical approaches [Anderson et al., 2011]. Such protein module is composed of the Hsp90-like region and of a large undefined downstream region. However, our study identifies srX and sr3 as functional elements in that large undefined region and likely involved in the chaperone activity of the whole module.

In conclusion, we used a functional comparative genomics approach that combines bioinformatics sequence examination tools to mapping and phenotypical analysis of human mutations, to provide novel information on the organization in repeated domains of sacsin. In particularly, our results establish that large portions of the protein can be arranged in a few and well-defined repeated domains. The demonstration of the functional nature of sr1, sr2, srX, and sr3 suggests that these regions contribute to the activity of the protein. Further studies are needed to define the specific role(s) of such domains, in the perspective of developing a comprehensive and integrated model of function for sacsin in the context of cell pathology. In a larger perspective, our approach that combines comparative analysis of vertebrate protein sequences/architecture, positional mapping of human mutations, and severity of clinical phenotype can be tentatively applied in the biomedical field to shed light on the functional nature of other proteins associated to disease but of yet unknown function.

Acknowledgment
Disclosure statement: The authors declare no conflict of interest.

References
Anderson JF, Siller E, Barral JM. 2010. The sacsin repeating region (SRR): a novel Hsp90-related supra-domain associated with neurodegeneration. J Mol Biol 400:665–674.

Anderson JF, Siller E, Barral JM. 2011. The neurodegenerative-disease-related protein 536
HUMAN MUTATION, Vol. 34, No. 3, 525–537, 2013

References
Anderson JF, Siller E, Barral JM. 2010. The sacsin repeating region (SRR): a novel Hsp90-like domain encoding an 11.5-kb ORF. Nat Genet 24:120–125.

Anderson JF, Siller E, Barral JM. 2010. The sacsin region (RegA), virtually corresponding to our SIRPI, do exhibit a chaperone-like activity that can be detected in vitro by standard biochemical approaches [Anderson et al., 2011]. Such protein module is composed of the Hsp90-like region and of a large undefined downstream region. However, our study identifies srX and sr3 as functional elements in that large undefined region and likely involved in the chaperone activity of the whole module.

In conclusion, we used a functional comparative genomics approach that combines bioinformatics sequence examination tools to mapping and phenotypical analysis of human mutations, to provide novel information on the organization in repeated domains of sacsin. In particularly, our results establish that large portions of the protein can be arranged in a few and well-defined repeated domains. The demonstration of the functional nature of sr1, sr2, srX, and sr3 suggests that these regions contribute to the activity of the protein. Further studies are needed to define the specific role(s) of such domains, in the perspective of developing a comprehensive and integrated model of function for sacsin in the context of cell pathology. In a larger perspective, our approach that combines comparative analysis of vertebrate protein sequences/architecture, positional mapping of human mutations, and severity of clinical phenotype can be tentatively applied in the biomedical field to shed light on the functional nature of other proteins associated to disease but of yet unknown function.

Acknowledgment
Disclosure statement: The authors declare no conflict of interest.

References
Anderson JF, Siller E, Barral JM. 2010. The sacsin repeating region (SRR): a novel Hsp90-related supra-domain associated with neurodegeneration. J Mol Biol 400:665–674.

Anderson JF, Siller E, Barral JM. 2011. The neurodegenerative-disease-related protein 536
HUMAN MUTATION, Vol. 34, No. 3, 525–537, 2013

References
Anderson JF, Siller E, Barral JM. 2010. The sacsin repeating region (SRR): a novel Hsp90-like domain encoding an 11.5-kb ORF. Nat Genet 24:120–125.

Anderson JF, Siller E, Barral JM. 2010. The sacsin region (RegA), virtually corresponding to our SIRPI, do exhibit a chaperone-like activity that can be detected in vitro by standard biochemical approaches [Anderson et al., 2011]. Such protein module is composed of the Hsp90-like region and of a large undefined downstream region. However, our study identifies srX and sr3 as functional elements in that large undefined region and likely involved in the chaperone activity of the whole module.

In conclusion, we used a functional comparative genomics approach that combines bioinformatics sequence examination tools to mapping and phenotypical analysis of human mutations, to provide novel information on the organization in repeated domains of sacsin. In particularly, our results establish that large portions of the protein can be arranged in a few and well-defined repeated domains. The demonstration of the functional nature of sr1, sr2, srX, and sr3 suggests that these regions contribute to the activity of the protein. Further studies are needed to define the specific role(s) of such domains, in the perspective of developing a comprehensive and integrated model of function for sacsin in the context of cell pathology. In a larger perspective, our approach that combines comparative analysis of vertebrate protein sequences/architecture, positional mapping of human mutations, and severity of clinical phenotype can be tentatively applied in the biomedical field to shed light on the functional nature of other proteins associated to disease but of yet unknown function.

Acknowledgment
Disclosure statement: The authors declare no conflict of interest.

References
Anderson JF, Siller E, Barral JM. 2010. The sacsin repeating region (SRR): a novel Hsp90-related supra-domain associated with neurodegeneration. J Mol Biol 400:665–674.

Anderson JF, Siller E, Barral JM. 2011. The neurodegenerative-disease-related protein 536
HUMAN MUTATION, Vol. 34, No. 3, 525–537, 2013

References
Anderson JF, Siller E, Barral JM. 2010. The sacsin repeating region (SRR): a novel Hsp90-like domain encoding an 11.5-kb ORF. Nat Genet 24:120–125.

Anderson JF, Siller E, Barral JM. 2010. The sacsin region (RegA), virtually corresponding to our SIRPI, do exhibit a chaperone-like activity that can be detected in vitro by standard biochemical approaches [Anderson et al., 2011]. Such protein module is composed of the Hsp90-like region and of a large undefined downstream region. However, our study identifies srX and sr3 as functional elements in that large undefined region and likely involved in the chaperone activity of the whole module.

In conclusion, we used a functional comparative genomics approach that combines bioinformatics sequence examination tools to mapping and phenotypical analysis of human mutations, to provide novel information on the organization in repeated domains of sacsin. In particularly, our results establish that large portions of the protein can be arranged in a few and well-defined repeated domains. The demonstration of the functional nature of sr1, sr2, srX, and sr3 suggests that these regions contribute to the activity of the protein. Further studies are needed to define the specific role(s) of such domains, in the perspective of developing a comprehensive and integrated model of function for sacsin in the context of cell pathology. In a larger perspective, our approach that combines comparative analysis of vertebrate protein sequences/architecture, positional mapping of human mutations, and severity of clinical phenotype can be tentatively applied in the biomedical field to shed light on the functional nature of other proteins associated to disease but of yet unknown function.

Acknowledgment
Disclosure statement: The authors declare no conflict of interest.

References
Anderson JF, Siller E, Barral JM. 2010. The sacsin repeating region (SRR): a novel Hsp90-related supra-domain associated with neurodegeneration. J Mol Biol 400:665–674.

Anderson JF, Siller E, Barral JM. 2011. The neurodegenerative-disease-related protein 536
HUMAN MUTATION, Vol. 34, No. 3, 525–537, 2013

References
Anderson JF, Siller E, Barral JM. 2010. The sacsin repeating region (SRR): a novel Hsp90-like domain encoding an 11.5-kb ORF. Nat Genet 24:120–125.
Masciullo M, Modoni A, Fattori F, Santoro M, Denora PS, Tonali P, Santorelli FM, Silvestri G. 2008. A novel mutation in the SACS gene associated with a complicated form of spastic ataxia. J Neurol 255:1429–1431.

Miller MP, Kumar S. 2001. Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet 10:2319–2328.

Miller MP, Parkes JD, Rissing SW, Kumar S. 2003. Quantifying the intragenic distribution of human disease mutations. Ann Hum Genet 67:567–579.

Murphy SM, Herrmann DN, McDermott MP, Scherer SS, Shy ME, Reilly MM, Parryson D. 2011. Reliability of the CMT neuropathy score (second version) in Charcot-Marie-Tooth disease. J Peripher Nerv Syst 16:191–198.

Ogawa T, Takiyama Y, Sakoe K, Mori K, Namekawa M, Shimazaki H, Nakano I, Nishizawa M. 2004. Identification of a SACS gene missense mutation in ARSACS. Neurology 62:107–109.

Okawa S, Sugawara M, Watanabe S, Imota T, Toyoshima I. 2006. A novel sacsin mutation in a Japanese woman showing clinical uniformity of autosomal recessive spastic ataxia of Charlevoix-Saguenay. J Neurol Neurosurg Psychiatry 77:280–282.

Ouyang Y, Segers K, Bouquiaux O, Wang FC, Janin N, Andris C, Shimazaki H, Sakoe K, Nakano I, Takiyama Y. 2008. Novel SACS mutation in a Belgian family with spastic ataxia of Charlevoix-Saguenay. J Neurol Neurosurg Psychiatry 77:280–282.

Parfitt DA, Michael GJ, Vermeulen EG, Prodromou NV, Webb TR, Gallo JM, Chertham ME, Nicoll WS, Blatch GL, Chapple JP. 2009. The ataxia protein sacsin is a functional co-chaperone that protects against polyglutamine-expanded ataxin-1. Hum Mol Genet 18:1556–1565.

Pearl LH, Prodromou C. 2006. Structure and mechanism of the Hsp90 molecular chaperone machinery. Annu Rev Biochem 75:271–294.

Prodi E, Grisoli M, Panzeri M, Minati L, Fattori F, Erbetta A, Uziel G, D’Arrigo S, Tessa A, Ciano C, Santorelli FM, Savoiardo M, et al. 2012. Supratentorial and pontine MRI abnormalities characterize recessive spastic ataxia of Charlevoix-Saguenay. J Neurol Neurosurg Psychiatry 83:150–158.

Richter AM, Ozgul RK, Poisson VC, Topaloglu H. 2004. Private SACS mutations in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) families from Turkey. Neurogenetics 5:165–170.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Schmitz-Hubsch T, du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, Giunti P, Globovs C, Infantti J, Kang JS, Kremer B, Mariotti C, et al. 2006. Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology 66:1717–1720.

Schüle R, Holland-Letz T, Klümpe S, Kassubek J, Kleopstock T, Mall V, Otto S, Winner B, Schöls L. 2006. The Spastic Paraplegia Rating Scale (SPRS): a reliable and valid measure of disease severity. Neurology 67:430–434.

Shimazaki H, Takiyama Y, Sakoe K, Ando Y, Nakano I. 2005. A phenotype without spasticity in sacsin-related ataxia. Neurology 64:2129–2131.

Takado Y, Hara K, Shimohata T, Tokiguchi S, Onodera O, Nishizawa M. 2007. New mutation in the non-gigantic exon of SACS in Japanese siblings. Mov Disord 22:748–749.

Takiyama Y. 2006. Autosomal recessive spastic ataxia of Charlevoix-Saguenay. Neuropathology 26:368–375.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599.

Terraciano A, Casali C, Greco GS, Orteschi D, Di Giandomenico S, Seminara L, Di Fabio R, Carrozzo R, Simonati A, Stevanin G, Zollino M, Santorelli FM. 2009. An inherited large-scale rearrangement in SACS associated with spastic ataxia and hearing loss. Neurogenetics 10:151–155.

Terracciano A, Foulds NC, Ditchfield A, Bunyan DJ, Crolla JA, Huang S, Santorelli FM, Hammann SR. 2010. Pseudodominant inheritance of spastic ataxia of Charlevoix-Saguenay. Neurology 74:1152–1154.

Trouillas P, Takayanagi T, Hallett M, Currier RD, Subramony SH, Wessel K, Breyer A, Diener HC, Massaquoi S, Gomez CM, Coutinho P, Ben Hamida M, et al. 1997. International cooperative ataxia rating scale for pharmacological assessment of the cerebellar syndrome. The Ataxia Neuropathology Committee of the World Federation of Neurology. J Neurol Sci 145:205–211.

Van Damme P, Demaereel P, Spileers W, Robberecht W. 2009. Autosomal recessive spastic ataxia of Charlevoix-Saguenay. Neurology 72:1790.

Vermeer S, Meijer RP, Hoiste TG, Bodmer D, Bosgoed EA, Cremers FP, Kremer BH, Knors NV, Scheffer H. 2009. Design and validation of a conformation sensitive capillary electrophoresis-based mutation scanning system and automated data analysis of the more than 15 kbp-spanning coding sequence of the SACS gene. J Mol Diagn 11:514–523.

Vermeer S, Meijer RP, Pijl BJ, Timmermans J, Cruysberg JR, Bos MM, Schelhaas HJ, van de Warrenburg BP, Knors NV, Scheffer H, Kremer B. 2008. ARSACS in the Dutch population: a frequent cause of early-onset cerebellar ataxia. Neurogenetics 9:207–214.

Yamamoto Y, Nakamori M, Konaka K, Nagano S, Shimazaki H, Takiyama Y, Sakoda S. 2006. Sacsin-related ataxia caused by the novel nonsense mutation Arg4325X. J Neurol 253:1372–1373.