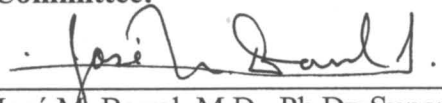


Copyright
by
John Francis Anderson
2011

The Dissertation Committee for John Francis Anderson Certifies that this is the approved
version of the following dissertation:

THE ROLE OF SACSIN AS A MOLECULAR CHAPERONE

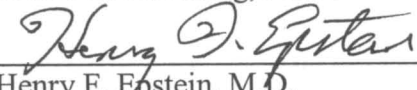
Committee:



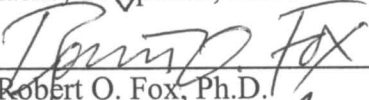
José M. Barral, M.D., Ph.D., Supervisor



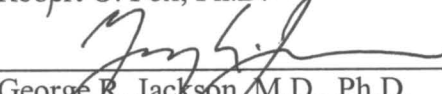
Darren F. Boehning, Ph.D.



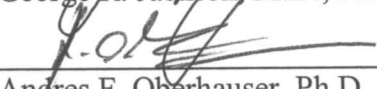
Henry F. Epstein, M.D.



Robert O. Fox, Ph.D.



George R. Jackson, M.D., Ph.D.



Andres F. Oberhauser, Ph.D.



Dean, Graduate School

THE ROLE OF SACSIN AS A MOLECULAR CHAPERONE

by

John Francis Anderson, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas Medical Branch

May 2011

Dedication

To my wife, Erika R. Anderson.

Acknowledgements

I am indebted to my mentor, Dr. José M. Barral for his investment in my education. Dr. Barral taught me that curiosity drives scientific investigation and rigorous experiments derive new knowledge. This combination of values is rare to find in a mentor and I am extremely grateful for my experience in his lab. I would like to acknowledge Jason J. Chandler and Paige Spencer for technical assistance. I am especially grateful to Dr. Efrain Siller for daily assistance and discussions on all aspects of this project. I would like to thank Dr. Christian Kaiser providing me a practical education on the design and performance biochemical experiments. I acknowledge Dr. Bernard Brais for generously sharing his time and resources; he provided the rare opportunity to visit an ARSACS patient in her home as well as supplied us with *SACS* knockout mouse brains. My committee members, Dr. Henry F. Epstein, Dr. Andres F. Oberhauser, Dr. George R Jackson, and Dr. Robert O. Fox and Dr. Darren F. Boehning provided invaluable guidance throughout this work for which I am very grateful. I would like to especially thank Dr. Darren F. Boehning and Dr. Henry F. Epstein for being involved in the details of my project from the beginning of my education at UTMB, through sharing reagents and expertise as well as serving on my committee. Financial support for this study was supplied by the March of Dimes Foundation, the Pew Foundation and the Jean C. and William D. Willis foundation. I am especially grateful to Dr. William D. Willis and others who contributed to the Jean C. and William D. Willis foundation.

THE ROLE OF SACSIN AS A MOLECULAR CHAPERONE

Publication No. _____

John Francis Anderson, Ph.D.

The University of Texas Medical Branch, 2011

Supervisor: José M. Barral

The central pathological finding in neurodegenerative disease is loss of neurons. In many disorders, this loss of neurons appears to be related to protein misfolding, comprising a large public health burden. For example, the two most common neurodegenerative diseases, Alzheimer's and Parkinson's diseases, possess protein misfolding as a core component of their pathology. In order to properly fold, many proteins require the assistance of molecular chaperones. While substantial gains in our knowledge of the function of general chaperones have been made in the last two decades, the role of molecular chaperones in brain-specific processes is not clearly defined. In this study, we examined the function of the protein sasin, mutated in autosomal recessive spastic ataxia of Charlevoix-Saguenay. Pathologically, these patients demonstrate loss of neurons in the cerebellum and cervical spinal cord along with inclusions reminiscent of misfolded proteins in the remaining neurons. Sasin contains regions of similarity to both molecular chaperones and co-chaperones. Thus, this disease may represent a useful model to study the role of molecular chaperones in neurodegenerative disease. We

performed bioinformatics and biochemical investigations to determine the function of salsin. We found that this protein contains a novel supra-domain present three times. We show that this domain is both ATPase active and contains molecular chaperone activity. Additionally, we determined that a region of salsin possesses co-chaperone activity. Thus, salsin is a novel molecular chaperone, with built-in co-chaperone modules, directly involved in a neurodegenerative disease.

Table of Contents

List of Tables	xi
List of Figures	xii
Abbreviations	xv
SECTION 1: INTRODUCTION	1
Chapter 1: Protein Biogenesis and Folding	1
A. Protein Biogenesis and Stability	1
B. Protein Folding and Neurodegeneration	18
C. The Role of Molecular Chaperones in Protein Folding	24
Chapter 2: Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay	50
A. Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay	50
B. The Sacsin Protein	61
SECTION 2: RESULTS	71
Chapter 3: Sacsin Contains a Novel Supra-Domain with Similarity to a Region of Hsp90	71
A. There Is a Repeating Supra-Domain in Sacsin with Similarity to Hsp90: SRR ..	72
B. The Phylogeny of SRR Domains Reveals Multiple Instances of Convergent Evolution	83
C. Conclusions	89
Chapter 4: The SRR Domain Is ATPase Active	91
A. Production of a Soluble Fragment of Sacsin	91
B. Purification of a Soluble Fragment of Sacsin	98
C. Biochemical Properties of Sacsin	105
Chapter 5: Sacsin Is Both a Chaperone and a Co-Chaperone	112
A. Sacsin Is a Chaperone	112
B. Sacsin Is a Co-chaperone	127
C. Conclusions	131

Chapter 6: Potential Client Proteins of Sacsin.....	133
A. Sacsin Interacts with MAP1A.....	133
B. Analysis of Sacsin Knockout Mouse Brain.....	138
C. Conclusions	142
SECTION 3: DISCUSSION	143
Chapter 7: The SRR Supra-Domain	143
Chapter 8: The Protein Folding Activities of Sacsin	149
Chapter 9: Conclusions and Perspectives	154
Conclusions.....	154
Perspectives.....	154
SECTION 4: MATERIALS AND METHODS.....	156
Chapter 10: Methods.....	156
A. Bioinformatics Analysis.....	156
B. Buffer Compositions	156
C. Cloning.....	159
D. Purification of Proteins	160
E. Cell Culture	165
F. SDS-PAGE	165
G. Western Blots.....	166
H. Fluorescence Spectroscopy.....	166
I. Immunoprecipitations.....	167
J. Luciferase Refoldings.....	167
K. Dot Blot Analysis.....	168
L. Immunofluorescence	168
M. ATPase Assays	169
N. Methanol/Chloroform Precipitation.....	169

SECTION 5: APPENDICES	170
Bibliography	183
Vita	208

List of Tables

Table 1.1:	Disorders due to tRNA charging mutations.	8
Table 1.2:	Ribosomopathies and anomalies of protein elongation rates.	12
Table 1.3:	Chaperonopathies.	13
Table 1.4:	Disorders associated with abnormal protein degradation.	15
Table 2.1:	General signs and symptoms of ARSACS.	53
Table 2.2:	Clinical features of ARSACS patients from various countries.	57
Table A.1:	Mutations identified in human sarsin.	170
Table A.2:	Characteristics of SRR containing proteins.	181
Table A.3:	Nomenclature for other domains in SRR containing proteins.	182

List of Figures

Figure 1.1: Overview of protein biogenesis and stability.....	3
Figure 1.2: The folding landscape.	26
Figure 1.3: Chaperone-assisted protein folding.....	32
Figure 1.4: The Hsp70/Hsp40 chaperone cycle.	40
Figure 1.5: Structural flexibility of Hsp90.	46
Figure 1.6: The ATPase cycle of Hsp90.	48
Figure 2.1: Geography of Charlevoix and Saguenay regions of Quebec.	51
Figure 2.2: Ataxia of fine motor movements in an ARSACS patient.	54
Figure 2.3: Images of the fundus in an ARSACS patient.....	56
Figure 2.4: MRI study of an ARSACS patient.....	59
Figure 2.5: Genomic map of the <i>SACS</i> gene.	62
Figure 2.6: Location of human mutations in saccin.	62
Figure 2.7: Known and hypothesized structural and functional data for saccin.....	65
Figure 3.1: Dot-plot analysis of saccin sequence.	72
Figure 3.2: Multiple sequence alignment of human SRR domains.....	73
Figure 3.3: Repeat location and length in human saccin.	74
Figure 3.4: Similarity between Hsp90 and the repeat in human saccin.....	76
Figure 3.5: Multiple sequence alignment of representative SRR supra-domains.	78
Figure 3.5: Multiple sequence alignment of representative SRR supra-domains.	79
Figure 3.6: Topology of conservation from yeast Hsp90 to human SRR domains.....	82
Figure 3.7: Phylogenetic distribution and domain architecture of SRR domains.	84
Figure 3.8: Phylogenetic relationships among animal SRR supra-domains.	88
Figure 4.1: Analysis of the solubility of saccin fragments.	92

Figure 4.2: Functional regions of sacsins.....	96
Figure 4.3: Solubility analysis of wild-type and mutant RegA.....	97
Figure 4.4: IMAC purification of RegA.....	99
Figure 4.5: RegA saturates nickel resin at low concentrations.	101
Figure 4.6: Chromatogram of RegA purification.	103
Figure 4.7: Gel filtration calibration.....	105
Figure 4.8: Size Exclusion Chromatography of RegA.....	106
Figure 4.9: RegA of sacsins is ATPase active.	108
Figure 4.10: Spectroscopic effects of a human mutation in RegA.....	110
Figure 5.1: Scheme for chaperone-assisted refolding of FLuc.....	113
Figure 5.2 Chaperone-assisted refolding of Fluc.	115
Figure 5.3: RegA of sacsins is a chaperone.	116
Figure 5.4: Quantification of the chaperone activity of RegA.	117
Figure 5.5: RegA of sacsins maintains FLuc soluble.....	120
Figure 5.6: Scheme for holdase experiments.	121
Figure 5.7: RegA maintains FLuc in a folding-competent state.	123
Figure 5.8: RegA does not prevent the thermal inactivation of FLuc.	124
Figure 5.9: RegA has a high affinity for client protein.....	126
Figure 5.10: RegA does not depend on DnaJ to cooperate with DnaK.....	127
Figure 5.11: Similarity between the J domain of sacsins and other J domains.....	129
Figure 5.12: The J domain of sacsins is functional.....	130
Figure 6.1: The J domain construct of sacsins pulls down MAP1A.....	134
Figure 6.2: Co-immunoprecipitation of sacsins and MAP1A.....	135
Figure 6.3: Map1a and sacsins co-localize in the cerebellum.....	136

Figure 6.4: RegA of sacsini does not interact with MAP1A, as assessed by a pull-down assay.	137
Figure 6.5: Protocol for fractionation of sacsini knockout mouse brain.	139
Figure 6.6: SDS-PAGE of sacsini knockout mouse brain fractions.	140

Abbreviations

A β ₄₂	amyloid- β -42 peptide
AD	Alzheimer's disease
APP	amyloid precursor protein
ARSACS	autosomal recessive spastic ataxia of Charlevoix-Saguenay
ATP	adenosine triphosphate
BSA	bovine serum albumin
bvRegA	baculovirus Region A
CAT	computerized axial tomography
CD	C-terminal domain
CDD	conserved domain database
CHIP	carboxyl terminus of hsp70-interacting protein
CMT	charcot-marie-tooth disease
CMT2D	CMT subtype 2D
CNS	central nervous system
DBA	diamond blackfan anemia
DTT	dithiothreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
ER	endoplasmic reticulum
FLuc	firefly luciferase
GARS	glycyl-tRNA synthetase

GdmCl	guanidinium chloride
GFP	green fluorescent protein
GTP	guanosine triphosphate
HD	huntington's disease
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEPN	higher eukaryotes and prokaryotes nucleotide binding domain
hHR23A	human homolog of RAD23 A
HOP	Hsp70/Hs90 organizing protein
HPD	histidine-proline-aspartic acid
HRP	horseradish peroxidase
Hsp	heat shock protein
Hsp40	heat shock protein 40
Hsp70	heat shock protein 70
Hsp90	heat shock protein 90
Hsp110	heat shock protein 110
IMAC	immobilized metal affinity chromatography
IQ	intelligence quotient
JBS	johanson-blizzard syndrome
KCl	potassium chloride
kDa	kilodalton
K/E	DnaK/GrpE
K/J/E	DnaK/DnaJ/GrpE
KNTase	kanamycin nucleotidyl transferase
KOAc	potassium acetate

KOH	potassium hydroxide
LB	lewy body
LGMD2C	limb girdle muscular dystrophy type 2c
LTP	long term potentiation
M	molar
MAP	microtubule associated protein
MAP1A	microtubule associated protein 1A
MD	middle domain
MDR1	multidrug resistance protein 1 gene
mg	milligram
MgOAc	magnesium acetate
ml	milliliter
mM	millimolar
MOPS	3-(N-morpholino)propanesulfonic acid
MRI	magnetic resonance imaging
mRNA	messenger RNA
NaCl	sodium chloride
NBD	nucleotide binding domain
ND	N-terminal domain
NEF	nucleotide exchange factor
NGF	nerve growth factor
Ni-NTA	nickel-nitriloacetic
nM	nanomolar
NMDA	<i>N</i> -methyl-D-aspartate

NR	non-redundant
ORF	open reading frame
PBS	phosphate buffered saline
PD	parkinson's disease
PDB	protein databank
PGP	p-glycoprotein
PS1	presinilin-1
PS2	presinilin-2
PSI-BLAST	position specific iterated basic local alignment search tool
RegA	Region A
RNA	ribonucleic acid
rRNA	ribosomal RNA
RPS19	ribosomal protein S19
<i>SACS</i>	spastic ataxia of Charlevoix-Saguenay (gene)
SAXS	small angle x-ray scattering
SBD	substrate binding domain
SCA-1	spinocerebellar ataxia type-1
SDS	sodium dodecyl sulfate
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
SRR	sacsin repeating region
TEV	tobacco etch virus
TPR	tetratricopeptide repeat
TRiC	TCP-1 ring complex

Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
Ubl	ubiquitin-like domain
UBR1	ubiquitin protein ligase E3 component n-recognin 1
μg	microgram
μl	microliter
μM	micromolar
UPS	ubiquitin proteasome system
USA	United States of America
XPCB	xeroderma pigmentosum C-binding domain

SECTION 1: INTRODUCTION

Chapter 1: Protein Biogenesis and Folding¹

Much of the biochemical work performed by the cell is accomplished by proteins. Proper protein function requires the coordination of numerous processes, from protein biogenesis to stability/degradation of the final product. Numerous human disorders have been identified as directly involved in these processes. In particular, many neurodegenerative disorders possess protein misfolding as a component of their pathology. These diseases provide a substantial public health burden to our society. A complex set of proteins referred to as molecular chaperones are present in the cytosol to assist in proper protein folding and mitigate the problem of protein misfolding.

A. PROTEIN BIOGENESIS AND STABILITY

Although an increasing role for RNA as a central regulatory molecule in the cell has recently become evident in addition to its essential template (mRNA), adaptor (tRNA) and catalytic (rRNA) functions underlying gene expression, most of the biochemical work of the cell is accomplished by proteins. The complexities of the folds conferred to proteins by their constituent amino acids allow proteins to carry out very diverse functions, including chemical catalysis (enzymes), exquisitely specific molecular recognition (antibodies, adaptor proteins, ligand binding proteins), formation of robust structural frameworks (cytoskeletal proteins), and transport of specific molecules (channels, pores). When coupled to each other, these functions underlie virtually all

¹ Portions of this chapter are reproduced from Protein Pept Lett. 2011 Feb;18(2):110-21. Disorders of protein biogenesis and stability. Anderson JF, Siller E, Barral JM. Copyright 2011, Bentham Science Publishers Ltd.

living processes, including acquisition and use of nutrients (transporters, metabolic enzymes), cell movement and division (cell adhesion and motor proteins), cell signaling (channels, kinases, phosphatases, small molecule hydrolyzing proteins, scaffolding proteins), communication with the environment (receptor proteins), and maintenance, transmission and expression of genetic information (nucleic acid polymerases, surveillance and repair proteins, transcription and translation factors, ribosomal proteins, molecular chaperones and proteases). Harmonious execution of these functions is essential for cellular homeostasis and requires that proteins *appear at* and *disappear from* specific subcellular locations at precisely regulated rates. The principles behind the biogenesis and stability of the proteome are essentially conserved throughout evolution and have been referred to as “protein quality control” mechanisms (Bukau et al., 2006; Kubota, 2009). Thus, it would be expected that errors in these essential process would not be compatible with life. However, it has become recently evident that there exist mutations in components of almost every step of protein production and degradation that do not result in lethality, but rather lead to a variety of human syndromes. The following discussion will outline the general principles of protein biogenesis and stability (Figure 1.1) followed by an overview of human disease-causing mutations in the proteins that comprise these processes (Tables 1.1 – 1.4).

Overview of Protein Biogenesis and Stability

Aminoacylation of tRNAs

In order for the ribosome to perform its mRNA-directed polypeptide synthesis function, every amino acid must be previously attached to its corresponding tRNA (Figure 1.1). This step, known as “tRNA charging” serves two critical purposes. First, chemical activation of the carboxyl group of every amino acid provides the chemical energy utilized subsequently for peptide bond formation. Second, the interaction of the anticodon loop of the amino acid-bearing tRNA with the appropriate codon along the mRNA allows correct amino acid placement along the nascent polypeptide chain. tRNA charging is an ATP-dependent reaction carried out in the cytosol by a class of enzymes known as aminoacyl-tRNA synthetases. Since the ribosome does not check the identity of

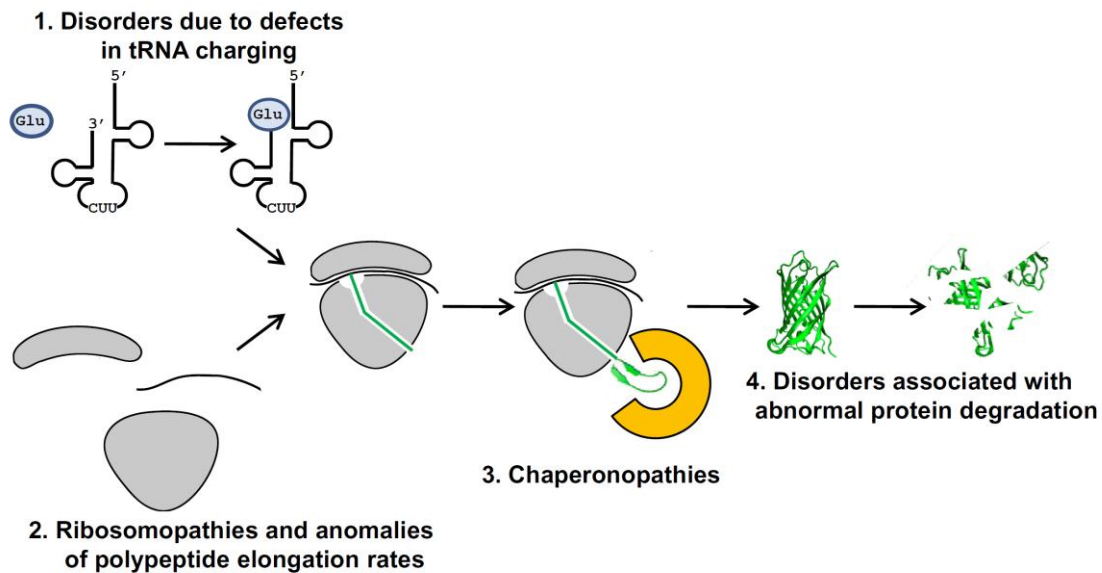


Figure 1.1: Overview of protein biogenesis and stability.

The general processes required for proper protein biogenesis and stability are presented graphically. See text for discussion. Reproduced from Protein Pept Lett. 2011 Feb;18(2):110-21. Disorders of protein biogenesis and stability. Anderson JF, Siller E, Barral JM. Copyright 2011, Bentham Science Publishers Ltd.

the amino acid attached to a tRNA during polypeptide elongation, the accuracy of these enzymes is critical to ensure fidelity during translation. Such proofreading is derived from the binding energy between the enzyme and its substrates during two successive binding and catalysis events, and has been referred to as a “double sieve” mechanism (Fukai et al., 2000; Nureki et al., 1998).

Messenger RNA-Directed Polypeptide Synthesis

Once amino acids are correctly attached to their respective tRNAs, the ribosome can decode the information contained along mRNA to generate polypeptide chains (Figure 1.1). In broad terms, this process is carried out in three successive stages: initiation, elongation and termination. An initiation complex is formed when the large ribosomal subunit binds to a pre-formed complex consisting of the small ribosomal subunit, the mRNA and the initiating methionyl-tRNA. This process requires GTP and is facilitated by cytosolic proteins known as initiation factors. During elongation, amino acids are covalently attached to the carboxyl terminus of the nascent polypeptide chain (which consists of solely of methionine at the very beginning) as directed by the anticodon:codon interactions, through a series of events currently understood in atomic detail (Ramakrishnan, 2010; Steitz, 2010; Yonath, 2010) that require cytosolic proteins known as elongation factors. Ribosomal movement along the mRNA is facilitated by the energy derived from tRNA binding as well as GTP hydrolysis carried out by elongation factors. Once the ribosome encounters a termination codon, polypeptide synthesis stops and the newly made polypeptide is released with the aid of cytosolic proteins known as release factors.

Protein Folding and Posttranslational Modifications

Newly synthesized proteins must fold into a precise three-dimensional structure in order to be functional (Figure 1.1 and discussed below). As part of the folding process, certain enzymes catalyze the rearrangement of certain special covalent bonds within some protein molecules. These include the peptidyl-prolyl cis/trans isomerases (Schmid, 2001) and protein disulfide isomerases (Gleiter and Bardwell, 2008; Wilkinson and Gilbert, 2004). While undergoing folding (or shortly afterwards) many newly synthesized proteins undergo a variety of additional enzyme-mediated posttranslational modifications, including proteolytic removal of signal sequences, addition of various chemical groups (such as phosphoryl, acetyl, methyl and carboxyl groups), attachment of sugars and incorporation of prosthetic groups (Walsh et al., 2005). It is also known that for many proteins whose function requires assembly into larger quaternary structural complexes, folding and posttranslational modifications are coupled to or dependent upon incorporation into such complexes (Dulyaninova et al., 2005; Hulmes, 2002; Kim et al., 2008).

Protein Stability and Degradation

Once a protein is synthesized and folded, it will perform its biological function for a certain length of time, until it is ultimately degraded (Figure 1.1). The abundance of a particular protein will depend on the rates of production and turn over. Recent large-scale analyses have demonstrated that in mammalian cells, although protein half-lives can vary greatly, there is a bimodal distribution of protein stabilities centered at ~0.5 and 2 hours (Yen et al., 2008). The process of protein degradation is typically an active and controlled process that allows protein function to be turned off abruptly and irreversibly and plays a

critical role in many cellular process, including cell cycle progression and signal transduction cascades (Konstantinova et al., 2008).

Several protein degradation systems are employed by the cell. These include the ubiquitin/proteasome system (UPS) (Jung et al., 2009), lysosomal degradation (Ciechanover, 2005; Cuervo and Dice, 1998) and other specialized proteases (Mogk et al., 2008). A prominent fraction of the regulated protein degradation is carried out by the UPS. The proteasome is a hollow cylindrical particle that passes unfolded polypeptides through its core, which contains many protease active sites that carry out peptide bond hydrolysis (Jung et al., 2009). Importantly, there are regulatory subunits on the ends of the proteasome that assist in recognizing proteins marked for degradation as well as threading them through the protease. Proteins that need to be degraded are typically marked by covalent attachment of a string of ubiquitin proteins through isopeptide bonds at lysine residues (Pickart and Eddins, 2004). Therefore, the regulatory component of this system is the ubiquitination of proteins. Ubiquitination is carried out by a series of proteins known as E1, E2 and E3 enzymes (Tanaka et al., 1998). Ubiquitin is first covalently attached to an E1 ubiquitin activating enzyme. The activated ubiquitin is then transferred to an E2 ubiquitin conjugating enzyme. Next, the E2 ubiquitin complex is brought into close proximity to a protein that is ready to be degraded by a E3 ubiquitin ligase allowing the transfer of ubiquitin to the substrate protein. The later steps usually impart specificity to the process. Thus, there are many more E3 than E2 enzymes and many more E2 than E1 enzymes, as specificity increases during the process.

Human Disorders Related to Protein Biogenesis and Stability

Disorders Due to Defects in tRNA Charging

In order for peptide bond formation to occur, aminoacyl tRNAs must be present at the appropriate location and concentration during translation. In eukaryotes, most protein synthesis is carried out by ribosomes present in the cytoplasm (and the cytoplasmic face of the rough endoplasmic reticulum). However, mitochondrial ribosomes present in the matrix, translate the limited number of mRNAs encoded in the mitochondrial genome. As mentioned above, the process of charging tRNA with amino acids is catalyzed by aminoacyl tRNA synthetases, both in the cytoplasm and in the mitochondrial matrix. Several human diseases have been determined to be due to mutations in genes encoding some of these enzymes, including glycyl-tRNA synthetase (GARS), which functions both in the cytosol and in mitochondria (Antonellis et al., 2006), cytoplasmic tyrosyl- and alanyl-tRNA synthetases and mitochondrial aspartyl- and arginyl-tRNA synthetases (Table 1.1). Remarkably, the great majority of diseases related to tRNA charging thus far identified result in abnormalities that affect primarily the nervous system. A particularly informative example of a possible explanation behind this phenomenon is provided by the syndrome that results from mutations in GARS: autosomal dominant Charcot-Marie-Tooth Disease type 2D (Skre, 1974).

Charcot-Marie-Tooth (CMT) is primarily a disease of peripheral axons with a heterogeneous clinical presentation (Skre, 1974). A complex classification scheme categorizes the disease into types and subtypes based on inheritance, symptoms and molecular genetics. CMT type 2 is an autosomal dominant disease that primarily exhibits foot deformities and ataxia with relatively slow progression (Gemignani and Marbini, 2001). The patients typically exhibit symptoms starting in the second decade of life.

Subtype 2D (CMT2D) exhibits weakness and muscle atrophy that is more severe in the hands than the feet, but is typically a milder disease (Ionasescu et al., 1996). Pathologically, this disease demonstrates axonal neuropathy with axonal atrophy and regeneration both occurring (Gemignani and Marbini, 2001).

In 2003, CMT2D was determined to be due to mutations in the glycyl-tRNA synthetase gene (GARS) (Antonellis et al., 2003). Remarkably, *in vitro* characterization of human disease causing mutations in GARS did not reveal a direct correlation between disease and presence or absence of aminoacylation activity. In other words, certain mutated proteins were competent to perform their expected activity while other mutants that lead to a similar disease phenotype displayed defective biochemical activity. Further work revealed a characteristic distribution of this enzyme in neurons: it was detected in granules in the cell body and axons (Antonellis et al., 2006). This and other studies revealed that mutant versions of GARS did not localize properly to neurites or axons (Antonellis et al., 2006; Nangle et al., 2007; Xie et al., 2007). Thus, the disease phenotype cannot be directly attributed to impaired aminoacylation, but is rather explained by a lack of appropriate subcellular localization in susceptible tissues. This

Table 1.1: Disorders due to tRNA charging mutations.

<i>Disease</i>	<i>Protein (gene)</i>	<i>System Affected</i>	<i>Function</i>	<i>Reference</i>
Charcot-Marie-Tooth disease Type 2D Distal Spinal Muscular Atrophy Type V	glycyl-tRNA synthetase (GARS)	Nervous System	tRNA charging	(Antonellis et al., 2003)
Dominant Intermediate Charcot-Marie-Tooth Neuropathy	tyrosyl-tRNA synthetase, cytoplasmic (YARS)	Nervous System	tRNA charging	(Jordanova et al., 2006)
Leukoencephalopathy	aspartyl-tRNA synthetase, mitochondrial (DARS2)	Nervous System	tRNA charging	(Scheper et al., 2007)
Infantile Encephalopathy Associated with Pontocerebellar Hypoplasia	probable arginyl-tRNA synthetase, mitochondrial (RARS2)	Nervous System	tRNA charging	(Edvardson et al., 2007)
Axonal Charcot-Marie-Tooth Disease (CMT2)	alanyl-tRNA synthetase, cytoplasmic (AARS)	Nervous System	tRNA charging	(Latour et al., 2009)

may also explain why an enzyme that is presumably essential in all cells for protein translation to occur only produces disease phenotypes in particular cellular populations. It appears that peripheral axons retain a requirement for properly charged amino acids that cannot be overcome by the machinery present in the cell body, which could be due to the limitations in diffusion in such long cellular processes. Thus, CMT2D constitutes an example in which defective biological function is not necessarily due to decreased enzymatic activity, but rather to inappropriate spatial distribution.

Ribosomopathies and Anomalies of Protein Elongation Rates

The ribosome is the cellular particle responsible for the decoding of the genetic information present in mRNA and peptide bond formation, processes absolutely fundamental to living systems. Surprisingly, mutations in components of the ribosome have been identified that do not result in lethality, but rather lead to several human diseases. These disorders have been collectively referred to as “ribosomopathies” (Narla and Ebert, 2010). For example, Diamond-Blackfan anemia (DBA) was determined to be due to mutations in ribosomal protein S19 (RPS19) in about 25% of the genotyped cases (Draptchinskaya et al., 1999). During the last decade, several additional proteins of both the small and large ribosomal subunits have been implicated in other forms of this disease (Table 1.2). Interestingly, different ribosome-associated proteins have also been implicated in other human disorders that affect multiple systems (Table 1.2).

DBA comprises various developmental defects as well as severe hypoplastic anemia presenting in childhood. Congenital anomalies include thumb and hand malformations, cleft palate and other facial anomalies, as well as urogenital abnormalities (Janov et al., 1996). The cellular defect in DBA patients is largely limited to the erythrocytic lineage. There is typically a macrocytic anemia with decreased reticulocytes.

Other lineages, such as platelets and leukocytes, are typically normal (Willig et al., 2000). The anemia in DBA is often responsive to corticosteroid treatment for unknown reasons. Other treatment options are transfusions and bone marrow transplant (Willig et al., 2000).

In order to understand the molecular basis for this disease, investigations have been directed at understanding the function of RPS19. Studies in yeast revealed that RPS19 is involved in ribosome biogenesis. Specifically, it participates in the cleavage of 90S pre-rRNA into pre-40S and pre-60S particles (Leger-Silvestre et al., 2005). Similar results were obtained in mammalian cell lines, indicating the RPS19 is an assembly factor involved in producing 40S ribosome subunits (Angelini et al., 2007). The precise reason behind the erythropoietic phenotype in DBA patients is still not understood, but may be related to a reduction in translational capacity in red blood cell precursors due to an overall reduction in ribosome numbers in these cells. This hypothesis assumes that red blood cells and their precursors are particularly sensitive to translational capacity, which is supported by studies that reveal an overall decrease in ribosomes as well as other ribosome proteins in blood samples from these patients (Badhai et al., 2009; Robledo et al., 2008). The molecular etiology of the phenotype in DBA is likely more complex than afforded by this explanation and may involve other non-ribosomal functions of RPS19. Indeed, extra-ribosomal functions for RPS19 have been identified, such as interaction with macrophage migration inhibitory factor (Filip et al., 2009) and regulation of p53 function (Danilova et al., 2008; McGowan et al., 2008).

Once an active ribosome has been engaged in translation, variations in the sequence of the mRNA can result in the production of a defective protein product. Obvious examples arise from mutations that exert their effect by altering the amino acid sequence and thus the structure and function of the protein. However, there are a few

examples of mutations that do not directly alter the primary sequence of the protein but nevertheless alter the biological activity of the protein product. There are several explanations for how this may occur, including alteration of mRNA stability and transport to the correct subcellular location. However, it has become recently evident that variations in rates of synthesis of identical polypeptides can lead to decreased protein function, due to defects in protein folding (Siller et al., 2010; Zhang et al., 2009). One particularly well studied example arises from synonymous polymorphisms in the multidrug resistance protein 1 gene (MDR1), whose protein product is known as P-glycoprotein (PGP) (Ambudkar et al., 2003) (Table 1.2).

MDR1 is not directly involved in a defined clinical syndrome, but polymorphisms in this gene have been determined to affect how individuals respond to toxic chemicals and therapeutic agents. MDR1 functions as a small molecule efflux transporter. Specifically, it limits absorption of compounds from the gastrointestinal tract and promotes the excretion of compounds in the liver and kidney and it prevents the access of compounds to the central nervous system by efflux through the blood brain barrier (Kim et al., 1998; Pauli-Magnus and Kroetz, 2004). MDR1 displays broad specificity, so many drugs are affected by its activity, particularly chemotherapeutic agents used in cancer treatment (Seelig et al., 2000). Thus, overexpression of MDR1 in cancer cells can portend a poor prognosis, since these cells become more effective at extruding the chemotherapy agents used in treatment (Ng et al., 2000). The expression and activity of MDR1 have been shown to be affected by single nucleotide polymorphisms (SNPs) and thus there is great interest in identifying which patients may be more prone to the effects of drug interactions as well as more resistant to cancer therapy drugs (Chen et al., 1997; Kioka et al., 1989).

In a study of SNPs in MDR1, a synonymous polymorphism in exon 12 was identified (C3435T) (Hoffmeyer et al., 2000). Surprisingly, this SNP was determined to be biologically relevant as it altered both the expression and the activity of the protein. Further investigations were undertaken to determine how a “silent” mutation could lead to an alteration in activity (Kimchi-Sarfaty et al., 2007). It was determined that this SNP leads to altered protein conformation and activity in the presence of equivalent levels of mRNA relative to the wild-type gene. This effect becomes more pronounced when PGP is overproduced and may be related to the depletion of certain tRNA species. The SNP changes the codon from a common to rare one. This is hypothesized to lead to pausing or slowing of translation, particularly when the cognate tRNA is depleted (*e.g.*, under conditions of high PGP expression). Indeed, the authors demonstrated that altering the codon to a rarer one leads to even larger alterations in protein activity (Kimchi-Sarfaty et

Table 1.2: Ribosomopathies and anomalies of protein elongation rates.

<i>Disease</i>	<i>Protein (gene)</i>	<i>System Affected</i>	<i>Function</i>	<i>Reference</i>
Diamond-Blackfan Anemia	40S ribosomal protein S19 (RPS19) 40S ribosomal protein S24 (RPS24) 40S ribosomal protein S17 (RPS17) 60S ribosomal protein L35a (RPL35A) 60S ribosomal protein L5 (RPL5) 60S ribosomal protein L11 (RPL11) 40S ribosomal protein S7 (RPS7) 40S ribosomal protein S10 (RPS10) 40S ribosomal protein S26 (RPS26)	Hematological System	Ribosomal Protein	(Draptchinskaia et al., 1999) (Gazda et al., 2006) (Cmejla et al., 2007) (Farrar et al., 2008) (Doherty et al., 2010)
Shwachman-Diamond Syndrome	ribosome maturation protein SBDS (SBDS)	Multi-system	Ribosomal Maturation Protein	(Boocock et al., 2003)
5q- syndrome	40S ribosomal protein S14 (RPS14)	Hematological System/Cancer	Ribosomal Protein	(Ebert et al., 2008)
X-linked Dyskeratosis Congenita	H/ACA ribonucleoprotein complex subunit 4 (DKC1)	Dermatological	rRNA Processing	(Heiss et al., 1998)
Pharmacological Metabolism Polymorphisms	multidrug resistance protein 1 (ABCB1)	Drug Metabolism	Efflux Transporter	(Hoffmeyer et al., 2000; Kimchi-Sarfaty et al., 2007)
X-Linked Spinal Muscular Atrophy	ubiquitin-like modifier activating enzyme 1 (UBE1)	Nervous System	Ubiquitin Activating Enzyme (E1)	(Ramser et al., 2008)

al., 2007). Thus, it appears that a change from the optimum codon usage in PGP to one that leads to a change in elongation rate leads to the production of a protein with an altered conformation (perhaps partially misfolded) that is not as competent to perform its function.

Chaperonopathies

As described above, many proteins require the assistance of molecular chaperones to fold with high efficiency in the cytosol. Human disease-causing mutations have been identified in several chaperones or proteins with clear similarities to known chaperones

Table 1.3: Chaperonopathies.

<i>Disease</i>	<i>Protein (gene)</i>	<i>System Affected</i>	<i>Function</i>	<i>Reference</i>
Hereditary Spastic Paraplegia 13 MitCHAP-60 disease	60 kDa heat shock protein, mitochondrial (HSPD1)	Nervous System	Chaperonin	(Hansen et al., 2002; Magen et al., 2008)
desmin-related myopathy Congenital Cataract	alpha-crystallin B chain (CRYAB) alpha-crystallin B chain (CRYAB) alpha-crystallin A chain (CRYAA)	Musculoskeletal System/Nervous System	Chaperone Chaperone	(Vicart et al., 1998) (Litt et al., 1998)
Kenny-Caffey Syndrome Sanjad-Sakati Syndrome	tubulin-specific chaperone E (TBCE)	Multi-System Nervous System	Chaperone	(Sanjad et al., 1991) (Diaz et al., 1999; Parvari et al., 2002) (Diaz et al., 1999)
Bardet-Biedl Syndrome Type 6 McKusick-Kaufman Syndrome	McKusick-Kaufman/Bardet- Biedl syndromes putative chaperonin (MKKS)	Multi-System Nervous System	Putative Chaperone	(Katsanis et al., 2000; Slavotinek et al., 2000) (Stone et al., 2000)
Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay	Sacsin (SACS)	Nervous System	Putative Chaperone	(Engert et al., 2000)
Axonal Charcot-Marie-Tooth Disease Distal Hereditary Motor Neuropathy	heat shock 27kDa protein 2 (HSPB2)	Nervous System	Chaperone	(Evgrafov et al., 2004)
Axonal Charcot-Marie-Tooth Disease Type 2L Distal Hereditary Motor Neuropathy Type II	heat shock 22kDa protein 8 (HSPB8)	Nervous System	Chaperone	(Tang et al., 2005) (Irobi et al., 2004)
Mutilating Sensory Neuropathy with Spastic Paraplegia	T-complex protein 1 subunit epsilon (CCT5)	Nervous System	Chaperonin	(Bouhouche et al., 2006)
Early-onset Torsion Dystonia	Torsin-1A (TOR1A)	Nervous System	Chaperone	(Burdette et al., 2010; Ozeliuss et al., 1997)

(Barral et al., 2004), and they have been referred to as “chaperonopathies” (Macario and Conway de Macario, 2007) (Table 1.3). This class includes mutations in member of several classes of general chaperones (*e.g.*, the cylindrical chaperonins, small heat shock proteins and Hsp90-related members) as well as more specialized co-chaperones (*e.g.*, DnaJ-like and tubulin specific chaperones). An illustrative example of this class, which contains regions with similarities to both general chaperones and co-chaperones is the putative chaperone sarsin, mutated in Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay (ARSACS) (Bouchard et al., 1978; Engert et al., 2000). An expanded discussion on ARSACS and sarsin will be covered in Chapter 2.

Disorders Associated with Abnormal Protein Degradation

In recent years, a considerable number of diseases have been identified that alter the function of the degradation machinery of the cell, including components that mark proteins for degradation (*e.g.*, E1, E2 and E3 ubiquitin ligases) and actual components with proteolytic activities (*e.g.*, ubiquitin hydrolases and AAA proteases) (Table 1.4). These findings highlight the notion that, in addition to their proper biogenesis, precise temporal and spatial degradation of proteins is critical for cellular homeostasis. For example, the tumor suppressor p53 is marked for degradation by the ubiquitin ligase mdm2 (Clegg et al., 2008). p53 is synthesized during mitosis to halt cell division and participate in “quality-control” of the genome as it replicates. Once this process is complete, mdm2 must mark p53 for degradation in order for mitosis and cell division to continue. Therefore, cell division requires a form of regulated degradation to properly proceed. Alternatively, over-expression of mdm2 will lead to excessive degradation of p53 and cell division may occur without proper quality-control of DNA replication, a situation that is often observed in human malignancies (Levav-Cohen et al., 2005). The

E3 ubiquitin ligase UBR1 is the gene mutated in Johanson-Blizzard Syndrome (JBS) (Zenker et al., 2005) and provides an instructive example of how a specific mutation affects cellular homeostasis through perturbation of an entire degradatory mechanism: the

Table 1.4: Disorders associated with abnormal protein degradation.

<i>Disease</i>	<i>Protein (gene)</i>	<i>System Affected</i>	<i>Function</i>	<i>Reference</i>
Lafora Disease	NHL repeat-containing protein 1 (NHLRC1)	Nervous System	E3 Ubiquitin Ligase	(Chan et al., 2003; Gentry et al., 2005)
Parkinson's Disease	E3 ubiquitin-protein ligase parkin (PARK2) ubiquitin carboxyl-terminal hydrolase isozyme L1 (UCHL1)	Nervous System	E3 Ubiquitin Ligase Ubiquitin Hydrolase	(Kitada et al., 1998; Shimura et al., 2000) (Leroy et al., 1998)
Familial Breast Cancer	breast cancer type 1 susceptibility protein (BRCA1)	Cancer	E3 Ubiquitin Ligase	(Hashizume et al., 2001)
Von Hippel-Lindau Disease	von Hippel-Lindau disease tumor suppressor (VHL)	Renal System	E3 Ubiquitin Ligase complex	(Iwai et al., 1999)
Angelman Syndrome	ubiquitin-protein ligase E3A (UBE3A)	Multi-System Nervous System	E3 Ubiquitin Ligase	(Kishino et al., 1997)
Multisystem Autoimmune Disease	E3 ubiquitin-protein ligase Itchy homolog (ITCH)	Immune System	E3 Ubiquitin Ligase	(Lohr et al., 2010)
Limb Girdle Muscular Dystrophy Type 2H Sarcotubular Myopathy Bardet-Biedl Syndrome Type 11	E3 ubiquitin-protein ligase TRIM32 (TRIM32)	Musculoskeletal System	E3 Ubiquitin Ligase	(Frosk et al., 2002; Locke et al., 2009) (Schoser et al., 2005) (Chiang et al., 2006)
Johanson-Blizzard Syndrome	E3 ubiquitin-protein ligase UBR1 (UBR1)	Muti-System	E3 Ubiquitin Ligase	(Zenker et al., 2005)
3-M Syndrome	cullin-7 (CUL7)	Multi-System	E3 Ubiquitin Ligase complex	(Huber et al., 2005)
Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy	autoimmune regulator (AIRE)	Multi-System	E3 Ubiquitin Ligase	(Nagamine et al., 1997; Uchida et al., 2004)
Fanconi Anemia Group L	E3 ubiquitin-protein ligase FANCL (FANCL)	Hematological System	E3 Ubiquitin Ligase	(Meetei et al., 2003)
Opitz Syndrome	midline 1 (MID1)	Developmental Defects	E3 Ubiquitin Ligase	(Quaderi et al., 1997; Trockenbacher et al., 2001)
X-Linked Mental Retardation	ubiquitin-conjugating enzyme E2 A (UBE2A)	Nervous System	E2 Ubiquitin Conjugating Enzyme	(Nascimento et al., 2006)
X-Linked Spinal Muscular Atrophy	ubiquitin-like modifier activating enzyme 1 (UBE1)	Nervous System	Ubiquitin Activating Enzyme (E1)	(Ramser et al., 2008)
Spinocerebellar Ataxia Type 28	AFG3-like protein 2 (AFG3L2)	Nervous System	AAA Protease	(Di Bella et al., 2010)
Hereditary Spastic Paraplegia Type 7	Paraplegin (SPG7)	Nervous System	AAA Protease	(Casari et al., 1998)
Hereditary Spastic Paraplegia Type 4	Spastin (SPAST)	Nervous System	AAA Protease	(Evans et al., 2005; Hazan et al., 1999)

N-end rule pathway (Bachmair et al., 1986).

JBS was identified as a clinical entity in 1971 (Johanson and Blizzard, 1971) and is associated with a wide range of congenital malformations and deficiencies, including hypoplastic alae nasi (outer portion of the nose), pancreatic insufficiency, short stature, mental retardation, and dental abnormalities (Hurst and Baraitser, 1989). The diagnosis is usually made by recognition of a characteristic facial morphology. The pancreatic deficiency is responsible for much of the mortality, with malabsorption leading to failure to thrive and hypoproteinemia and further subsequent complications. Children can survive infancy but generally require long-term medical care (Hurst and Baraitser, 1989). Currently, the treatment is aimed at symptom relief with enzyme replacement therapy and surgical correction of congenital abnormalities as the main therapeutic options (Rezaei et al., 2011).

The N-end rule pathway states that the half-life of a protein is related to the identity of its N-terminal residue (Bachmair et al., 1986). This pathway is based on the recognition of destabilizing N-terminal residues (N-degrons) and subsequent ligation of a polyubiquitin chain to internal lysine residue(s) (Bachmair and Varshavsky, 1989), which leads to targeting to and proteolytic degradation by the proteasome. At least one recognition component of this pathway in humans was determined to be accomplished by UBR1 (Kwon et al., 1998). This protein is an E3 ubiquitin ligase as well as a recognition component of the system and is thus responsible for identifying proteins that are marked to have short half-lives (due to a destabilizing N-terminal residue). UBR1 performs its recognition function through high affinity direct interactions with the N-degron (Xia et al., 2008). Since all proteins are required to include an N-terminal methionine due to the nature of the start codon, this system implies that machinery exists in the cell to remove

and/or modify this residue. Indeed this has been found to be the case (reviewed in (Tasaki and Kwon, 2007)). Interestingly, the role of UBR1 in protein quality control has been expanded to include functions independent of the N-end rule pathway (Nillegoda et al., 2010). Similar to other ubiquitin ligases, such as parkin and CHIP, UBR1 has been shown to be capable of providing a link between protein folding (and misfolding) and degradation in the yeast model system. In particular, UBR1 was shown to specifically recognize unfolded luciferase and mark it for degradation during cytosolic translation (Nillegoda et al., 2010).

Conclusions

Protein biogenesis is an integral aspect of cellular function. Thus, it is somewhat surprising to realize the breadth of diseases that can be caused due to mutations in components of these pathways. One may expect that mutations affecting such basic processes as ribosome biogenesis and tRNA charging would result in lethality of the organism. The fact that these mutations are compatible with life suggests that these systems contain certain elements of redundancy and tissue specificity. This is particularly evident in the tissue-restricted patterns observed for some of these diseases. For example, DBA and CMT2D both involve processes that are globally employed (ribosome biogenesis and tRNA charging, respectively), but the disease is primarily manifested in the hematological and peripheral nervous system, respectively. Currently, there is no convincing explanation for this selectivity, but it may be related to tissue specific requirements for bursts of protein synthesis. For example, production of erythrocytes requires the production of large quantities of hemoglobin. A subtle reduction in the capacity for translation may not be so acutely detrimental in other cell types that do not require a consistent production of high levels of protein for function during a particular

developmental stage. Also, in the peripheral nervous system, there is a requirement for the movement of nutrients and macromolecules over very long distances relative to the molecules involved (*e.g.*, from the cell body to the ends of neuritic processes that may be up to a meter away). Therefore, mutations that do not necessarily abrogate the activity of a particular enzyme (such as a tRNA synthetase) but disrupt its capacity to properly migrate to a certain sub-cellular location will affect certain cells (such as neurons) more prominently than others. In addition to the importance of adequate cellular location for protein production highlighted by protein biogenesis diseases, the multiple human disorders that arise from abnormalities in protein degradation molecules emphasize the critical role of precise temporal termination of protein function in the maintenance of cellular homeostasis.

B. PROTEIN FOLDING AND NEURODEGENERATION

While all of the components of protein biogenesis are critical for cellular function, the public health burden of protein misfolding disorders is particularly high (Dobson, 1999). These diseases can be due to mutations in molecular chaperones as described above, due to mutations in the misfolded protein itself (*e.g.* superoxide dismutase in amyotrophic lateral sclerosis), or due to age related damage to the proteome (Shaw and Valentine, 2007). It is interesting to note that protein misfolding disorders often strike the central nervous system (CNS) (Agorogiannis et al., 2004). The explanation for why neurons are particularly sensitive to protein misfolding is not clear.

Neurodegenerative Disorders

Alzheimer's disease (AD) provides an example of a protein misfolding disease with substantial public health burden. While the precise pathogenesis of this disease is unclear, all hypotheses point to a misfolded protein as the toxic agent. AD is the most

prevalent neurodegenerative disorder (Muchowski, 2002). The burden of AD in the United States of America in the year 2000 was 4.5 million people; it is estimated that by 2050, almost three-fold more people will be affected (Hebert et al., 2003). This disease results in dementia, robbing the patients of their cognition and memory while providing substantial emotional and financial burdens for their caretakers. While difficult to estimate, the cost of AD to the American economy is clearly in the billions of dollars (Bloom et al., 2003). The major pathological finding in AD is neurodegeneration: neuronal loss in the cortex leading to gross atrophy of the brain. This process can occur in a spontaneous age-dependent fashion or rarely *via* mutations in the amyloid precursor protein (APP), or the enzymes that process it, presenilin-1 and presenilin-2 (PS1 and PS2 respectively) (Selkoe, 2001). The hallmark pathological features of AD include extracellular amyloid plaques and intraneuronal neurofibrillary tangles (Agorogiannis et al., 2004). These species may contain many components, but the plaques are largely composed of misfolded amyloid- β -42 peptide ($A\beta_{42}$), a proteolytic product of APP; and the tangles are composed of aggregated microtubule-associated protein tau. These features have historically been considered to be toxic and result in neuron death. Interestingly, the quantity of plaques is not directly correlated to the degree of dementia (Hardy and Selkoe, 2002). This and other observations have led to the hypothesis that misfolded but soluble low-order oligomers of $A\beta_{42}$ may be the toxic species (Sakono and Zako, 2010).

Parkinson's disease (PD) is the second most common neurodegenerative disease and also includes protein misfolding as a core component of its pathology. The prevalence of PD increases with age, from 1% of people age 65 and older to 4% of people age 85 and older (Bekris et al., 2010). This disorder is classically described by

four characteristics: (1) tremor at rest, (2) rigidity, (3) akinesia or bradykinesia, and (4) postural instability (Jankovic, 2008). In addition to these core characteristics, there may also be dysfunction of the autonomic nervous system, cognitive disturbances, and sleep problems. The cause of PD is usually multifactorial and cannot be ascribed to a specific single gene; however there are many clear monogenic forms of this disease. Mutations in over ten genes result in this disease and many more appear to be linked to the development of PD (Bekris et al., 2010). These genes do not appear to comprise a single pathway that leads to disease; however, several are involved in the UPS in some manner. It is important to note that the vast majority of cases of PD are sporadic. Pathologically, the disease results from loss of dopaminergic cells in the substantia nigra in the brainstem. The remaining cells contain Lewy bodies (LBs): intracytoplasmic inclusions composed principally of misfolded α -synuclein (Love, 2005). The pathogenesis of LBs and their role in the disease are not clearly defined, but protein misfolding is a core component of this pathological structure.

Huntington's disease (HD) is associated with protein misfolding but is distinct from AD and PD, which can occur sporadically without mutations in a specific gene, in that it invariably occurs due to autosomal dominant mutations in the huntingtin gene. This disease is composed of the triad of movement disorder (chorea), cognitive decline and behavioral changes (Cardoso, 2009). HD is progressive and fatal. This rare disorder (4-10 people per 100,000) exemplifies a distinct pathogenic process from AD and PD: a trinucleotide repeat expansion (CAG) (Ross and Tabrizi, 2011). These repeat expansions in the huntingtin gene result in an increased number of glutamine residues in the protein product resulting in the disease phenotype. Interestingly, the age of onset is directly related to the length of the repeat (Ross and Tabrizi, 2011). Pathologically, there is

striking degeneration of the striatum and lesser degrees of degeneration throughout the brain. The physiological function of huntingtin is not entirely understood, but current models envision the polyglutamine expanded protein assuming a misfolded conformation which is toxic (Ross and Tabrizi, 2011). Consistent with this hypothesis, intranuclear inclusions of misfolded and aggregated huntingtin are observed in HD.

Mechanisms of Protein Misfolding in Neurodegeneration

The public health importance of protein misfolding diseases has led to substantial research into their pathogenesis. Currently, there are at least two mechanisms identified by which a misfolded protein can result in neurodegeneration: a gain of function or a loss of function of the misfolded species (Winklhofer et al., 2008). A gain of function is demonstrated by the oligomers of misfolded A β ₄₂ in the amyloid pathway in AD; whereas a loss of function is demonstrated by the deficiency of parkin activity in certain familial PD cases.

Amyloid is a term for a conformation seen in certain protein aggregates. This term was coined by Virchow who noticed lesions in the brain that stained strongly with iodine, in a manner similar to starch; thus, he believed that these lesions were composed of carbohydrates (Dumoulin and Bader, 2006). It was determined in 1859 that these lesions were proteinaceous, yet the term amyloid has remained. Amyloid is defined histologically as a plaque which exhibits birefringence under polarized light when stained with Congo Red (Frid et al., 2007). Many proteins possess the capacity to form amyloid under the appropriate conditions (Dobson, 1999). When this occurs in the human body, it often results in disease. The common denominator required for the formation of amyloid is the conversion of a native structure to a misfolded β -sheet rich conformation which is prone to forming intermolecular aggregates with other misfolded molecules (Dobson, 2004).

The end result is the production of amyloid fibrils seen grossly as plaques. Typically, this process is thought to occur through successive stages of aggregation, starting at the formation of a misfolded β -sheet rich structure, the generation of soluble oligomers which then combine to form small pre-fibrillar aggregates followed by assembly into protofibrils and finally the mature fibril (Caughey and Lansbury, 2003). The formation of amyloid is a kinetically slow process which results in an extraordinarily stable conformational species which is no longer competent to perform a biological function, similar to disordered aggregates. These fibrils are distinct from disordered aggregates in that they tend to have a regular repeating structure that is consistent across amyloids formed from many different types of proteins.

While the formation of amyloid plaques is a striking gross phenomenon in AD, it is not thought that these are the toxic species in this disease. Rather the soluble low-order aggregates that are precursors to the fibrils that comprise the plaques are thought to be responsible for the toxic effects. It appears that dimers, trimers and other oligomeric species of $A\beta_{42}$ are responsible for the neuronal dysfunction and death (Crews and Masliah, 2010; Sakono and Zako, 2010). Indeed, there is a hypothesis that the amyloid plaques are actually protective, since they sequester the soluble oligomers into an insoluble and thus biologically inactive pool. Numerous mechanisms of toxicity have been suggested/described for the oligomers. Binding of oligomers to the nerve growth factor (NGF) receptor and *N*-methyl-D-aspartate (NMDA) receptor along with toxic downstream effects have been demonstrated (De Felice et al., 2007; Yamamoto et al., 2007). Several studies have shown that oligomers can disrupt long-term potentiation (LTP) (Selkoe, 2008; Viola et al., 2008). It has even been suggested that oligomeric assemblies can form membrane pores leading to dysregulation of the membrane potential

(Kawahara and Kuroda, 2000). Finally, a recent report suggested that amyloid aggregates in general have a propensity to sequester many proteins in the cell which may also contribute to their toxicity (Olzscha et al., 2011). In each of these scenarios, the misfolded oligomeric assembly has gained a new function that is toxic to the cell.

A second mechanism by which protein misfolding can result in neurodegeneration is through loss of function of the misfolded protein (Winklhofer et al., 2008). When the protein is misfolded, it is functionally absent and there is a deficiency of its activity in the cell (Welch, 2004). The cell often degrades these misfolded proteins. This is analogous to a loss of function mutation, except that the protein is misfolded, and may not necessarily be mutated. However, many autosomal recessive diseases result at least in part from mutation-induced misfolding and subsequent degradation of the protein involved; for example, this phenomenon is known to occur in cystic fibrosis (Cheung and Deber, 2008). Thus, mutations that lead to misfolding and spontaneous age-dependent protein misfolding can both lead to disease through a loss of function mechanism.

An example of loss of function due to misfolding in neurodegeneration is provided by certain types of PD. Some cases of autosomal recessive PD are due to mutations in the gene encoding parkin (Kitada et al., 1998). The autosomal recessive inheritance of this form of PD indicates that it may be due to a loss of function of the involved gene. Importantly, it appears that mutations in parkin largely lead to misfolding and aggregation of the protein (Winklhofer et al., 2008). Parkin, an E3 ubiquitin ligase, is involved in conjugating ubiquitin to proteins destined for degradation by the proteasome (Table 1.4). Thus, it performs a quality-control/regulatory function in the cell. There are numerous substrates that parkin can direct to the proteasome. In a parkin knockout model of PD, at least two substrates were observed to accumulate in the substantia nigra,

indicating that the regulation of their degradation was perturbed possibly leading to pathogenic effects (Ko et al., 2006; Ko et al., 2005). However, parkin may also be involved in cellular-protective roles that are regulated by ubiquitination but not directly tied to the proteasome (Henn et al., 2007). Regardless, a loss of function of parkin leads to either accumulation of substrates or a loss of its protective activities. Importantly, a misfolding/damage-related loss of function of parkin in an age-dependent manner has been implicated in *sporadic* PD as well (Wang et al., 2005). Indeed, the aggregation of parkin seen in familial PD due to mutations in parkin is also seen in sporadic PD which possess an intact genome (LaVoie et al., 2005). Additionally, parkin appears to be particularly susceptible to stress-related misfolding (Wong et al., 2007).

The cell utilizes molecular chaperones to mitigate protein misfolding. Thus, much interest has been given to the involvement of chaperones in the pathogenesis of neurodegenerative diseases. Age and stress-dependent alterations in expression and accumulation of chaperones have been proposed to be important mediators of the sporadic late-onset forms of several neurodegenerative diseases (Voisine et al., 2010). The aggregates/inclusions observed in these diseases often co-localize with molecular chaperones such as Hsp70 (Muchowski and Wacker, 2005). Indeed, several studies have demonstrated that certain chaperones can reduce the toxic effects of misfolded proteins when co-expressed with toxic mutant proteins in AD, PD and HD models (Muchowski, 2002; Muchowski and Wacker, 2005).

C. THE ROLE OF MOLECULAR CHAPERONES IN PROTEIN FOLDING

Protein folding is the process that results in the transformation of the one-dimensional information contained in the primary sequence into the three-dimensional structure required for protein function. All of the information required for folding is

contained in the primary sequence (Anfinsen and Scheraga, 1975); however, efficient folding *in vivo* often requires the assistance of molecular chaperones (Young et al., 2004). An understanding of protein folding and the mechanisms of chaperone function are critical in determining how protein misfolding diseases develop.

Principles of Protein Folding and Chaperone Mechanism

Protein Folding in Vitro

The importance of folding for proper protein function has inspired much research into how proteins attain their native state. A cohesive mechanism for protein folding has been elusive for many years. Cyrus Levinthal described a *gedanken* experiment in 1968 which implied that if a protein sampled all the conformations available then it would take longer than the age of the universe to find the native state, however, proteins are known to fold *in vitro* on the order of seconds or less; this discrepancy is known as Levinthal's paradox (Nolting and Agard, 2008). Thus, random sampling of all conformations is not a realistic means of folding and there must be constraints/principles that define a rapid pathway of folding to the native state. In recent years experimental and computational tools have become available to probe the events that occur during the first fractions of a second along the path to the native state allowing the development of folding models (Bartlett and Radford, 2009).

The energy landscape model has emerged as the prominent paradigm for protein folding (Figure 1.2) (Bryngelson et al., 1995; Onuchic et al., 1997). Folding is envisioned as occurring on an energy landscape shaped as a funnel which points towards a unique native state, that represents a thermodynamic minimum for the molecule (Figure 1.2a). The energy landscape is composed of the terms entropy and enthalpy which can be defined as the degree of disorder in a system and the heat content in a system,

respectively. This model implies that the unfolded state has the most entropy and most enthalpy and the largest conformational ensemble; entropy and enthalpy decrease as more ordered structures are formed. Finally, an ordered native state with the lowest enthalpy is attained. The reduction in enthalpy outweighs the reduction in entropy leading to the lowest overall free energy in the native state. Additionally, the formation of aggregates can be plotted on the same axes, but are difficult to include in the energy funnel since these have an extremely low enthalpy (*e.g.* they may be considered practically

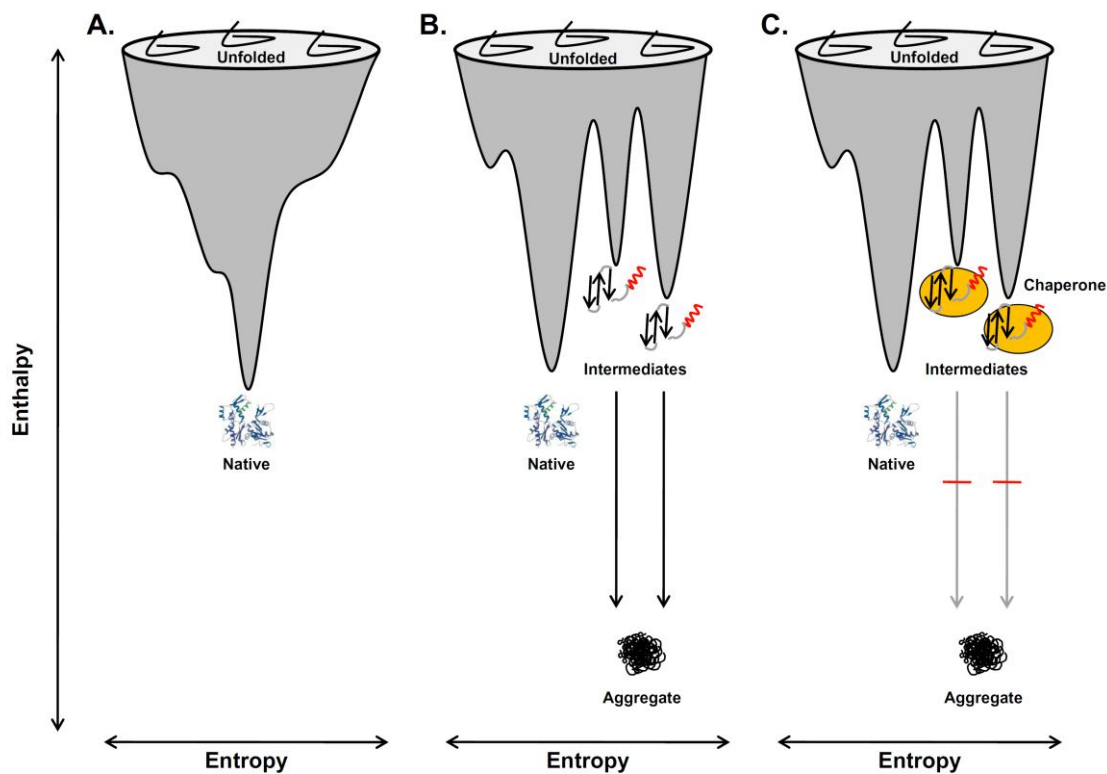


Figure 1.2: The folding landscape.

Schematic representation of protein folding according to the energy-landscape funnel model. The conformational space a protein is free to explore is represented by an energy funnel; enthalpy decreases from top to bottom, entropy is lowest in the middle and increases towards either side of the funnel. **(a)** Schematic of a smooth folding landscape with no kinetic traps. **(b)** Schematic of a rough folding landscape with multiple intermediates that are at risk for aggregation. **(c)** Schematic of a rough folding landscape demonstrating a chaperone's effect on at-risk intermediates.

irreversible at biological time scales and temperatures) but often require an additional input to the system (*e.g.* thermal, redox) in order to be populated due to a large kinetic barrier to their formation.

The degree of smoothness or ruggedness in the landscape is variable among proteins (Kapon et al., 2008). A smooth landscape would be found in small proteins that fold without long-lived intermediates. A rugged landscape implies that kinetically-trapped intermediates are formed during folding (Figure 1.2b). These intermediates possess a kinetically-stable energetic minimum that must be overcome to reach the native state (Bartlett and Radford, 2009). These structures may be either on or off pathway based on the degree of native or non-native like contacts present. On pathway intermediates are populated during a folding process that results in acquisition of the native state, while off pathway intermediates typically do not progress to the native state. Debate exists as to whether the folding process traverses intermediates through parallel pathways or through a preferred pathway (Hartl and Hayer-Hartl, 2009). As folding proceeds, more stable intermediates are populated but fewer of these exist. Interestingly, even the denatured state has a relatively limited ensemble of conformations with distinctive structural characteristics that may even recapitulate some native-like interactions (Kristjansdottir et al., 2005; Lindorff-Larsen et al., 2004).

As mentioned above, many of the intermediates in the folding pathway are thought to possess native contacts and are on pathway to the native state. This and other observations have led to the nucleation-condensation model of protein folding (Fersht, 2000; Karplus and Weaver, 1976). In this model, native elements of secondary structure and perhaps even tertiary structure are formed very early in folding and may even be present in the denatured state ensemble. These native-like regions are stable enough to

maintain their structure and act as nucleation centers for folding. Further elements of less stable secondary and tertiary structure develop (“condense”) around these. This effectively constricts the number of conformations that the protein can sample as it folds, which was the major problem identified in Levinthal’s paradox.

Consistent with this model is the realization that across many proteins with varying types of folds, there is a strong correlation between folding rate and “contact-order”, the distance between two residues in the primary sequence that make a contact in the native state (Plaxco et al., 1998). In simple terms, this means that proteins that fold from nucleation centers composed of residues separated by longer distances in the primary sequence take longer to fold. Thus, the formation of the nucleation center is rate limiting for folding and the rate of formation of the nucleation center is directly correlated to the physical distance between the residues that compose it. This provides a substrate for natural selection to operate on, as adjusting the contact order of folding nuclei could alter the kinetics of folding.

In contrast, non-native-like energetically minimal structures may also form during folding. This has been termed frustration: multiple states of similar and minimal enthalpy in a protein preventing the formation of one discrete native structure. Frustration would lead to multiple folds that are energetically similar to the native state but of a different structures. These would compete with the native state in the folding process, decreasing efficiency. In general, proteins are considered to be minimally frustrated: there is one unique native state that is the energetic minimum (Bryngelson and Wolynes, 1987). Thus, evolution selects for not only functional proteins, but efficiently folding proteins.

However, local areas of frustration are occasionally identified in the native state. These areas of frustration are typically found in two situations. (1) They are put to

biological use in molecular motors that traverse multiple conformations. (2) They represent a compromise with function in that the residues that contribute to frustration are required for function; consistent with this hypothesis is the observation that frustrated residues often segregate with functional residues (Ferreiro et al., 2007). For example, an active site in an enzyme may require certain residues in a precise orientation for function; however, this constraint on conformation may increase the free energy of the native state to the point that other non-functional structures are also populated. Thus the opposing evolutionary forces to fold efficiently yet retain function meet at these physical locations.

In addition to the intra-molecular principles of protein folding discussed above (*e.g.* contact order, frustration) proteins must also manage inter-molecular interactions. The major inter-molecular force encountered during folding which may impede efficient folding is the hydrophobic interaction. Extensive research has revealed that a fundamental principle underlying protein folding is the burial of hydrophobic amino acid residues. In the native state most of the hydrophobic residues are excluded from the solvent and buried in the core of the protein. This burial begins very early in the folding process, simultaneous to or even more rapid than the acquisition of secondary structure (Camilloni et al., 2008; Sadqi et al., 2003). The hydrophobic collapse may be a driving force in the formation of nucleation centers or an unrelated consequence of the exclusion of water (Camilloni et al., 2008). In fact, the contribution of the polypeptide backbone has been proposed to be the driving force behind the folding process (Street et al., 2006). Regardless, the speed of the hydrophobic collapse emphasizes the power of this interaction and the necessity to properly manage hydrophobic residues during the folding process. Indeed, transient exposure of hydrophobic residues in unfolded and partially folded intermediates has the potential to result in inter-molecular aggregation,

particularly if the intermediates are relatively stable and long-lived (Figure 1.2b). Thus, a major selective pressure encountered in the evolution of protein folding pathways is the management of hydrophobic residues. It is likely obligatory for certain proteins to populate an intermediate state with exposed hydrophobics in order to fold to a functional native state. In these cases, the cell utilizes molecular chaperones to prevent aggregation and allow efficient folding. A general principle of chaperone function is binding to these hydrophobics and preventing their inappropriate association followed by releasing them and allowing the protein to fold (Figure 1.2c). Chaperone action can also be coordinated with the proteolytic machinery of the cell to allow “triaging” of folding/misfolded proteins.

Protein Folding in Vivo

The majority of our understanding of protein folding has resulted from *in vitro* refolding studies. These systems are valuable due to their simplicity and the plethora of tools that can probe structure at extremely short time-scales (Bartlett and Radford, 2009). However, the setting under which proteins fold in the cell, *de novo* folding, is remarkably different. As the ribosome synthesizes polypeptides, they emerge from the ribosome and fold for the first time: a fundamentally different paradigm than refolding experiments performed *in vitro*. Protein folding in the cell must obey the same laws of physics, thus it is possible that the same biophysical principles extracted from *in vitro* refolding experiments likely apply to *de novo* folding; however ribosomal synthesis in the context of the cytoplasm raises unique challenges to the nascent polypeptides.

Refolding studies *in vitro* are typically carried out under very dilute and homogenous conditions (*e.g.* 0.01 – 0.1 mg/ml) in order to reduce aggregation during refolding. Since aggregation is a concentration dependent phenomenon, it can be

prevented by simple dilution in many cases (Haber and Anfinsen, 1961; Kiefhaber et al., 1991); additionally the exclusion of other macromolecular co-factors also prevents the possibility of aggregation between different proteins. However, the cellular environment is much less amenable to protein folding. The overall concentration of macromolecules in the cytosol is approximately 300-400 mg/ml (Ellis and Minton, 2003). This results in a large excluded volume effect raising the effective concentration of unfolded proteins during synthesis, increasing the possibility of aggregation. Additionally, numerous ribosomes often translate along the same mRNA simultaneously in a structure known as the polysome (Warner et al., 1962). It has often been suggested that the formation of the polysome results in an elevated local concentration of nascent polypeptide chains, increasing the chance of aggregation. However, recent evidence indicates that the polysome possesses a rational organization that orients nascent chains away from one another, minimizing the risk of inter-chain contact (Brandt et al., 2010; Brandt et al., 2009). Even so, for larger proteins, the advantage of optimized orientation is eventually overcome by the increased diffusional capacity of the growing ribosome-tethered chain (*i.e.* longer nascent chains may be able to diffuse far enough to interact with other ribosome associated nascent chains).

In vitro protein refolding experiments allow the protein to refold in the context of the entire polypeptide chain. In contrast, in the cell, the polypeptide is synthesized in a vectorial manner; the mRNA is “read” by the decoding center of the ribosome in a 5’ to 3’ direction and the peptidyl-transferase center catalyzes peptide bond formation in an N-terminal to C-terminal manner. Thus, N-terminal regions are generated before C-terminal regions. Additionally, the rate of protein synthesis is often significantly slower (seconds to minutes) than the observed rates of refolding upon dilution from denaturant

(nanoseconds to seconds) (Chang et al., 2005). As the N-terminal end of the polypeptide emerges from the ribosomal exit tunnel, it can begin folding before the C-terminal portions have even been synthesized (Figure 1.3a). This raises several significant issues in *de novo* protein folding that are not encountered during *in vitro* protein folding.

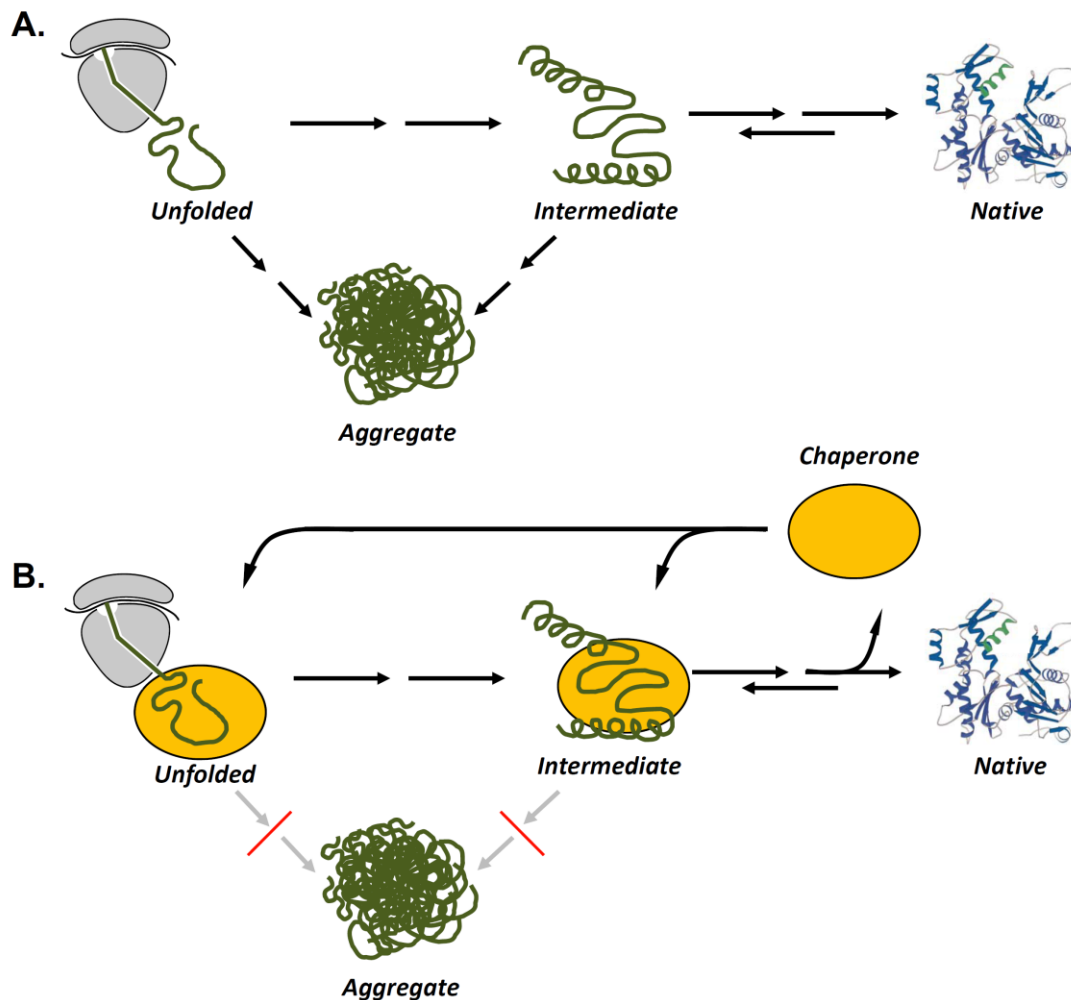


Figure 1.3: Chaperone-assisted protein folding.

(a) Schematic of *de novo* protein folding. Unfolded and intermediately folded proteins expose hydrophobic residues that are at risk for aggregation, a process that competes with the formation of the native state. **(b)** Schematic of the effect of chaperones on *de novo* protein folding. Chaperones transiently bind to hydrophobic residues in unfolded and partially folded proteins, preventing aggregation and allow efficient folding to the native state.

The vectorial nature of protein synthesis places a physical constraint on the conformational space the protein can explore as it folds. This reduction in available conformations partially mitigates the problem of Levinthal's paradox. However, if N-terminal residues interact with C-terminal residues in the native structure, then folding must be delayed until the C-terminal portions of the protein emerge from the ribosomal exit tunnel. This period of waiting in a partially folded state can present the problem of aggregation for the newly synthesized protein, as it must wait in an intermediately folded state that is not optimized for management of hydrophobic residues. These intermediates may possess exposed hydrophobic residues that have the potential to aggregate while the rest of the protein is being synthesized (Figure 1.3a). Similarly, if a folding nucleation center is composed of residues separated by large distances in the primary sequence (high contact order), then a delay in folding is necessitated to allow the formation of this center; this again requires a period of waiting in a partially folded state which may be susceptible to aggregation.

The ability to restrict the folding space of the nascent chain as well as the necessity for folding to be matched to the sequence available requires the rate of translation elongation to be coupled to folding. Ample evidence has demonstrated that the speed by which ribosomes move along mRNA is not uniform but is rather influenced by local stable mRNA structure and/or the presence of particular kinds of codons (Komar, 2009; Sorensen and Pedersen, 1991; Tsai et al., 2008). Since translation is spatially and temporally coupled to protein folding, it has long been hypothesized that subtle variations in the manner by which mRNAs are translated could have significant impact on the folding of the encoded polypeptides (*e.g.* (Purvis et al., 1987)). Recent experimental evidence suggests that this is indeed the case (Siller et al., 2010; Zhang et al., 2009).

Similarly, several studies have found significant correlations between the so-called *codon adaptation index* (Sharp and Li, 1987) along an mRNA and the structural features of the encoded polypeptide. However, it has become increasingly clear that the correlation between codon frequency and translation speed is not absolute and necessitates refinement (Kudla et al., 2009; Saunders and Deane, 2010).

Thus, there is a critical role for translation rates to be matched to the folding regime employed by a particular protein. For example, some proteins may begin to fold while they are being translated (co-translational folding) while others may not fold until translation has been completed (post-translation folding) (Agashe et al., 2004; Kaiser et al., 2006; Netzer and Hartl, 1997). The folding pathway of co-translationally folding proteins presumably evolved under the spatial constraints imposed by the vectorial nature of protein synthesis. Thus a change in elongation rate at a sensitive location along a polypeptide may expose the nascent polypeptide to a novel folding environment (*i.e.*, additional or restricted sequence along the C-terminus of the polypeptide) and alter the folding efficiency by allowing inappropriate intra-molecular contacts or preventing appropriate intra-molecular contacts. Recently, experimental evidence has demonstrated that variations in translation rates either due to ribosomal mutations (Siller et al., 2010) or to alterations in tRNA concentration (Zhang et al., 2009) have significant effects on protein folding of certain proteins. Several factors may contribute to translation rate and thus may have a general effect on *de novo* folding efficiency, including the codon composition of the message (Sorensen and Pedersen, 1991; Zhang et al., 2009), the concentration of cognate amino-acyl tRNAs (Varenne et al., 1984; Zhang et al., 2009), and intrinsic properties of the ribosome (Chang et al., 2005).

Chaperone-Assisted Protein Folding

The potential folding impediments that proteins encounter when synthesized in the cell are mitigated by a set of molecules known as molecular chaperones (Figures 1.2c and 1.3b). These are a diverse set of proteins present in all kingdoms of life that share the common function of enhancing the folding efficiency of other proteins known as the “clients” of the chaperone. . Over the last twenty years, substantial gains have been made in our knowledge of chaperone function (Bukau et al., 2006; Frydman, 2001; Hartl and Hayer-Hartl, 2002; Hartl and Hayer-Hartl, 2009; Walter and Buchner, 2002; Young et al., 2004). Several general principles of chaperone-assisted protein folding have emerged from these studies, while these may not be universally true, these represent the current major paradigms in chaperone biology:

- (1) Chaperones bind promiscuously to hydrophobic regions of unfolded or partially-folded intermediates, preventing inappropriate interactions and subsequent aggregation (Figure 1.2c, Figure 1.3b) (Hartl and Hayer-Hartl, 2002).
- (2) Chaperones do not provide structural information to the client protein (Anfinsen and Scheraga, 1975).
- (3) Chaperones are not a component of the final structure of the client protein (Figure 1.3b) (Walter and Buchner, 2002).
- (4) Chaperones are not folding catalysts; with few exceptions, they do not accelerate folding rates, they usually delay them. Consistent with this notion, they operate in a supra-stoichiometric fashion with respect to their clients (Kiefhaber et al., 1991).

- (5) Chaperones typically operate as a network, not individually; a client usually traverses multiple chaperones sequentially on its way to the native state (Young et al., 2004).

A general paradigm for chaperone-assisted *de novo* folding can be described as follows. As translation proceeds, the polypeptide emerges from the ribosomal exit tunnel in a non-folded state, except for perhaps the formation of some α -helical structures (Bhushan et al., 2010). Upon leaving the steric constraints presented by the ribosomal exit tunnel, folding is allowed to begin. As folding continues, intermediate structure(s) may be populated along the way to the final structure. The unfolded and intermediately folded structures often have exposed hydrophobic patches that will be buried in the core of the native structure. These patches are prone to inappropriately interacting with other folding proteins, leading to irreversible protein aggregation (Figure 1.3). The exposure of hydrophobic patches constitutes a signal for chaperone recruitment. The chaperone binds to these in a relatively non sequence-specific fashion. While a chaperone is bound, folding and (intra- and inter-molecular) misfolding are prevented. The chaperone then releases the client into solution, where it can attempt to fold (Figure 1.2c and 1.3b). If the protein folds successfully, it does not re-bind to the chaperone, whereas failure usually results in iterative cycles of chaperone binding and release. During this chaperone-assisted process of folding, translation continues to proceed (relatively slowly compared to folding), allowing additional regions of the polypeptide to be synthesized and exposed to the cytosol. Thus, by maintaining the more N-terminal regions of a polypeptide free from aggregation and in a folding competent state, chaperones are able to overcome the mismatch between folding rates and elongation rates. The release of the client protein is a

critical component of chaperone action; yet the dynamics of this activity remain largely unknown and are an exciting area of study.

Chaperone action also extends beyond nascent chain interactions. Client proteins typically “flux” through chaperone networks. Consistent with this notion, certain chaperones classes (*e.g.* Hsp90) are selected to interact with nearly folded clients and help to “polish” the final structure (Jakob et al., 1995). Additionally, other chaperone families, known as chaperonins, (*e.g.* TRiC) act as molecular cages and encapsulate the entire polypeptide; these chaperones generally prefer translation to be complete to function (McCallum et al., 2000).

Chaperones also have a role in repairing “damaged” proteins (Landsverk et al., 2007; Richter et al., 2010). Protein damage can be defined as the misfolding and aggregation that occurs to a previously folded protein due to stochastic processes or an external stressor (*e.g.* heat); indeed a disaggregase function has been ascribed to certain classes of co-chaperones in conjunction with the DnaK chaperone family in bacteria. Historically, chaperones were discovered as a set of genes that are up-regulated during heat shock, hence the term “heat shock protein” (Hsp) for many of the chaperones (Ashburner and Bonner, 1979; McKenzie et al., 1975). The stressors encountered by a cell extend beyond heat and include altered redox conditions, heavy metal toxicity, ethanol and others (Yura et al., 1984). These stressors induce proteins to partially unfold and become at risk for aggregation due to the exposure of hydrophobic residues. The regulation of the heat shock response in eukaryotes has been described by the chaperone-titration model (Richter et al., 2010; Voellmy and Boellmann, 2007). A simplified form of this model states that the transcription factor Hsf1 is bound by chaperones and kept in a monomeric state. When there is an excess of unfolded proteins, the chaperones are

titrated away from Hsf1 to assist in protein folding activities. The free Hsf1 molecule then trimerizes into its active form and translocates to the nucleus where it acts as a transcription factor to up-regulate transcription of the heat shock genes. Thus, protein unfolding is the signal for induction of the response and leads to the transcription/translation of numerous genes, including chaperones, which are involved in mitigating and repairing the damage. A similar pathway is thought to operate in bacteria involving a signaling molecule known as σ_{32} (Arsene et al., 2000).

The flux of client proteins through chaperone networks involves coordinated communication between members of different chaperone families. For example, certain newly synthesized proteins in the eukaryotic cytosol have been proposed to traverse a chaperone network that begins with the chaperone/co-chaperone pair Hsp70/Hsp40 during or immediately after translation, followed by a transition to the chaperone Hsp90 and/or the chaperonin TRiC (Young et al., 2004). Other substrates, such as actin, join the network by binding to prefoldin after translation and are then passed off to TRiC (Vainberg et al., 1998). This cooperation between chaperones often involves a physical adaptor. For example, the co-chaperone Hop can simultaneously bind Hsp70 and Hsp90 facilitating client transfer (Scheufler et al., 2000).

There is a broad array of co-chaperones that perform regulatory activities in addition to the adaptor activities described above. For example, Hsp90 is regulated by many co-chaperones including p23, CDC37, and FKBP52 (Grammatikakis et al., 1999; Johnson and Toft, 1994). These co-chaperones may alter chaperone function or affect client protein function. For example, the co-chaperone Aha1 accelerates the ATPase activity of Hsp90 while FKBP52 possesses peptidyl-prolyl cis/trans isomerase activity allowing it to assist in structural transitions in the client protein (Panaretou et al., 2002;

Peattie et al., 1992). An important connection is also made to the proteasome to allow the degradation of terminally misfolded polypeptides by the ubiquitin ligase CHIP which can directly bind to Hsp70 and initiate the formation of poly-ubiquitin chains on the client (Ballinger et al., 1999; Connell et al., 2001).

Certain proteins appear to have a preference/requirement to follow a specific chaperone pathway while other proteins may only need to associate with one chaperone in the network to fold to the native state. Still other proteins require interactions with specific chaperones to fold properly (*e.g.* desmin and α B crystallin) (Vicart et al., 1998). The following sections will describe two chaperone systems in detail as models for chaperone function. A third major chaperone system conserved across evolution which will not be discussed in detail are the chaperonins. These assist in protein folding by encapsulating their clients and removing them from the general cytosol, a phenomenon which has been described as infinite dilution.

The Hsp70/Hsp40 Chaperone System

One of the earliest identified chaperones is the Hsp70/Hsp40 system (known as DnaK/DnaJ in bacteria) (Bardwell and Craig, 1984; Craig et al., 1979; McKenzie et al., 1975; Saito and Uchida, 1977). This chaperone system is integral to *de novo* folding and protein disaggregation (in conjunction with Clp family members) during stress; additionally, it is involved in mitochondrial import, endoplasmic reticulum (ER) import, protein trafficking, and endocytosis (Mayer and Bukau, 2005). Hsp70 is a general chaperone and Hsp40 is a co-chaperone. These, along with a nucleotide exchange factor, comprise the core components of a chaperone system that operates *via* iterative cycles of client protein binding and release regulated by ATP hydrolysis (Szabo et al., 1994).

Hsp70 is composed of an N-terminal nucleotide binding domain (NBD) structurally related to actin (Flaherty et al., 1991), followed by a substrate binding domain (SBD) composed of a beta-sandwich floor and an alpha helical lid (Figure 1.4a) (Jiang et al., 2005; Zhu et al., 1996). The lid can open or close over the SBD in a motion regulated by ATP binding and hydrolysis. This is thought to be accomplished by an allosteric mechanism (Mapa et al., 2010). In the ADP bound state, the NBD and SBD move relatively independently of one another within a defined realm and the lid is closed

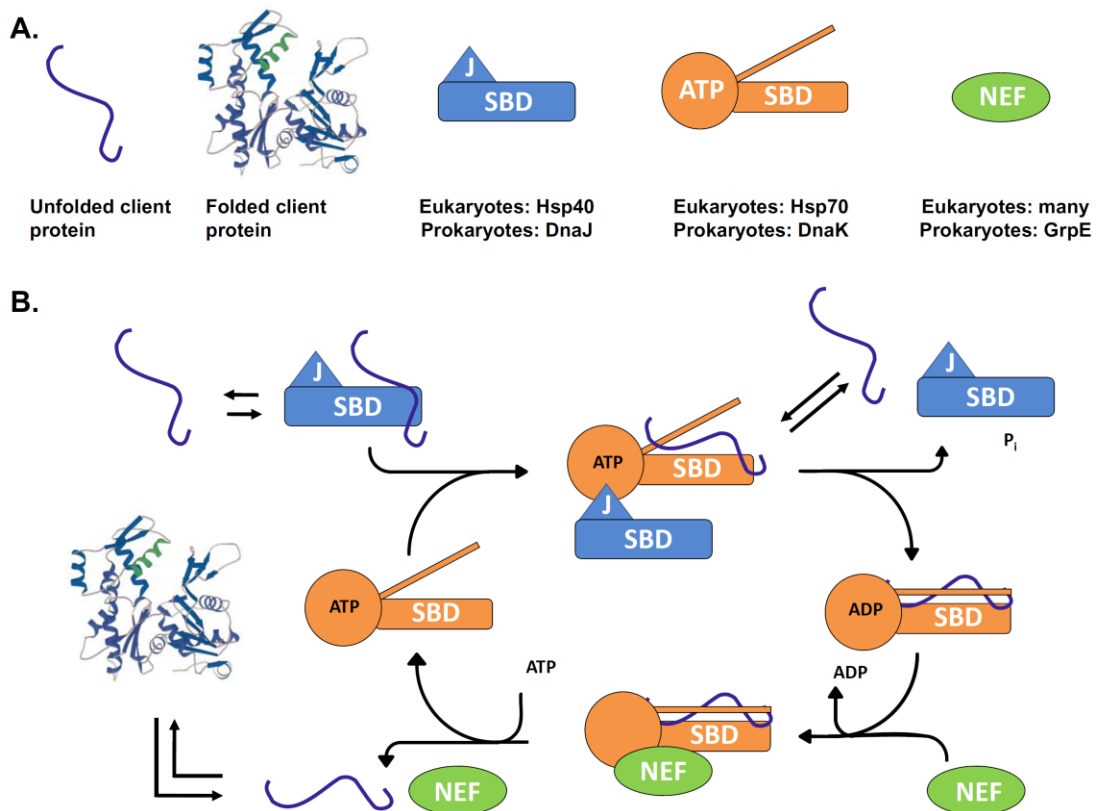


Figure 1.4: The Hsp70/Hsp40 chaperone cycle.

(a) The components of the cycle and the domain structure of Hsp70 and Hsp40. Hsp70 contains an N-terminal ATP binding domain and a C-terminal SBD; Hsp40 contains an N-terminal J domain and a C-terminal substrate binding domain (SBD); nucleotide exchange factor (NEF) domain structure is variable. **(b)** Model for the ATP-dependent chaperone cycle of Hsp70 as regulated by the co-chaperone Hsp40, a NEF and ATP. See text for a detailed explanation.

(Bertelsen et al., 2009). Upon ATP binding it is thought that a buried surface is revealed on the NBD and it interacts with the SBD inducing an allosteric conformation change in the SBD and lid opening.

The substrate specificity of Hsp70 has been investigated revealing that, consistent with the paradigm of chaperone function, it has a preference for stretches of hydrophobic amino acids. In particular, regions of five hydrophobic amino acids were found to bind to the SBD while positively charged residues were enriched immediately outside the cleft (Rudiger et al., 1997a). Others found that in addition to a preference for peptides enriched for hydrophobic residues, negatively charged residues are poor binders (Gragerov et al., 1994). The signature motif that Hsp70 has a high affinity for is found every 36 residues on average and these are usually buried in the native structure (Rudiger et al., 1997b). A structure of the SBD bound to a client peptide revealed that hydrogen bonds are formed along the backbone of the peptide in order for the SBD to recognize the client in an extended conformation (Zhu et al., 1996). Additionally, hydrophobic residues constitute the floor of the client binding cleft creating a Van-der-Waals surface that is primed to bind the hydrophobic residues in the client (Zhu et al., 1996). Indeed the hydrophobic cleft is thought to possess significant flexibility and the client is thought to bind in an induced-fit manner (Mayer et al., 2000; Pellecchia et al., 2000; Wang et al., 1998). Thus, the non-specific/promiscuous nature of chaperone client interactions is exemplified by the backbone hydrogen bonds and flexible hydrophobic cleft in Hsp70.

Hsp40 is composed of an N-terminal J domain and a C-terminal SBD (Figure 1.4a) (Kelley, 1998). The J domain is comprised of a helix-loop-helix with an invariant histidine-proline-aspartic acid (HPD) motif that interacts with Hsp70 and accelerates its ATPase activity (Mayer et al., 2000). The substrate binding domain (SBD) is able to

rapidly bind unfolded proteins, preventing their aggregation and then delivers these client to the chaperone Hsp70 for efficient and rapid folding (Langer et al., 1992). Hsp40 exists as a dimer and utilizes a clamp-like structure at the SBD interface to bind clients (Li et al., 2003). There are numerous J domain containing protein families in the cell. These have been divided into several classes (Cheetham and Caplan, 1998). For example, class I proteins are full-length homologs of the canonical *E. coli* DnaJ, whereas class III proteins only contain a J domain and are typically involved in specialized functions such as clathrin uncoating or protein import into mitochondria or the ER.

The function of NEFs in the Hsp70 reaction cycle is to induce the release of ADP from the SBD and the binding of ATP (Szabo et al., 1994). In bacteria the NEF GrpE is required for an efficient folding cycle due to the relatively slow release of ADP from Hsp70 (Langer et al., 1992). In eukaryotes, ADP release is more rapid and an NEF is not strictly required and may or may not be present; indeed, Hsp70/Hsp40 can perform chaperone activity *in vitro* without an NEF, whereas DnaK/DnaJ cannot function *in vitro* without GrpE (Brehmer et al., 2001). There are several classes of NEFs in eukaryotes which share a common function but are not phylogenetically related, the BAG domain family, the Hsp110/Grp107 family and the HspBP1/Fes1p family (Hohfeld and Jentsch, 1997; Kabani et al., 2002; Steel et al., 2004). Interestingly, Hsp110 is a homolog of Hsp70 that has been shown to possess an affinity for unfolded proteins but cannot functionally replace Hsp70 (Easton et al., 2000); this protein appears to have been optimized to enhance nucleotide exchange.

These components participate together in a chaperone cycle as demonstrated in Figure 1.4b (Hartl and Hayer-Hartl, 2002). In the ATP-bound state, Hsp70 assumes an open conformation for the client protein that has relatively low affinity but rapid binding

and release rates. Hsp40 binds to unfolded polypeptides through its C-terminal domain and delivers these to Hsp70. Through its J domain, Hsp40 functions to enhance the hydrolysis rate of Hsp70. Upon ATP hydrolysis, the conformation of Hsp70 is altered and it assumes a high affinity for client protein with slow rates of binding and release. During this phase of the cycle, the client protein is protected from inappropriately contacting other partially folded molecules and the consequent aggregation that may result. In order for folding to proceed, Hsp70 must release the client protein. This phase of the cycle is regulated by a nucleotide exchange factor which accelerates the exchange of ADP for ATP. As ADP is released and ATP is rebound, Hsp70 again assumes an open conformation with a rapid release rate for the client protein. As the client is released, it is free to fold on its own. Hsp40 is involved in this process in two ways: it delivers unfolded client proteins through its C-terminal domain and then stimulates Hsp70 to hydrolyze ATP through its J domain.

The Hsp90 Chaperone Machine

Hsp90 is an abundant molecular chaperone required for viability in eukaryotic cells (Borkovich et al., 1989). It constitutes approximately 1% of cellular protein in unstressed yeast cells and can increase several-fold upon heat stress (Pearl and Prodromou, 2001). The client proteins that interact with Hsp90 have been more easily defined than those of other chaperones because they are closer to their native state and form more stable complexes with the chaperone and are thus more amenable to isolation (Pearl and Prodromou, 2006). These clients represent a diverse assortment of kinases, hormone receptors, and other proteins. Since many cell proliferation signal molecules require Hsp90 for activation, this protein has been an area of intensive investigation for anti-neoplastic drugs (Holzbeierlein et al., 2010).

Hsp90 is present in bacteria (where it is known as HtpG) and eukaryotes but has not been identified in archaea (Pearl and Prodromou, 2006). There are several isoforms of Hsp90 in eukaryotic cells including inducible and constitutive cytosolic forms, ER-(Grp94) and mitochondrial-(Trap1) specific genes. All Hsp90 homologs are composed of three domains, an N-terminal domain (ND), a middle domain (MD) and a C-terminal domain (CD) (Figures 1.5 and 1.6) (Pearl and Prodromou, 2006; Wandinger et al., 2008). Hsp90 is constitutively dimerized through the CD with a dissociation constant of ~60 nM (Richter et al., 2001). Additionally, in eukaryotic Hsp90s there is a linker composed of numerous charged residues connecting the ND and MD which is not present in bacterial genes (Figure 1.6) (Scheibel et al., 1999).

The role of Hsp90 *in vivo* to enhance the maturation of steroid hormone receptors was one of the initial functions ascribed to Hsp90 (Pelham, 1986; Picard, 2006). This was hypothesized to be a molecular chaperone function and Hsp90 has since been confirmed to exert chaperone activity *in vitro*. Hsp90 interacts with incompletely folded polypeptides and maintains them in a folding competent state for subsequent action by other chaperone systems *in vitro*. Specifically, Hsp90 has been demonstrated to exert this activity towards β -galactosidase and luciferase (Freeman and Morimoto, 1996; Yonehara et al., 1996). In contrast to the Hsp70/Hsp40 system, Hsp90 has not been demonstrated to be effective in refolding model clients out of denaturant and requires other chaperone systems to “renature” model proteins (Freeman and Morimoto, 1996). However, Hsp90 is effective at preventing the thermal aggregation of model proteins such as citrate synthase *in vitro* (Jakob et al., 1995; Wiech et al., 1992).

Hsp90 does not appear to have a preference for extended regions of polypeptides like Hsp70. In contrast, it interacts with near-native proteins and assists to “polish” their

final structure or keep them stable until they are bound to ligands or other partner proteins (Mayer, 2010; Wandinger et al., 2008). The mechanism by which Hsp90 accomplishes this feat is not clear. In fact, the client binding site has not been definitively established for this molecule. Indeed, all the domains of Hsp90 have been demonstrated to interact with client protein and possess chaperone activity (Scheibel et al., 1998; Young et al., 1997). For example, electron microscopy investigations demonstrated a client asymmetrically bound to both the ND and MD of one monomer (Vaughan et al., 2006). It is commonly suggested that client protein binds in the V-shaped interface of the dimer and chaperone activity is accomplished via a molecular clamp mechanism (Figure 1.5) (Mayer, 2010; Stirling et al., 2006).

Initial investigations into Hsp90 function revealed conflicting results regarding its ATPase activity. Careful investigations revealed that there was little to no biochemical evidence of nucleotide binding or hydrolysis in yeast or human Hsp90 molecules (Jakob et al., 1996; Scheibel et al., 1997). This finding was consistent with the realization that the chaperone activity of Hsp90 *in vitro* did not depend on ATP (Jakob et al., 1995; Johnson et al., 2000; Wiech et al., 1992). However, others performed rigorous investigations into the putative ATPase activity of Hsp90 and were able to detect binding and hydrolysis (Grenert et al., 1999; Grenert et al., 1997; Panaretou et al., 1998; Sullivan et al., 1997). Conclusive proof that Hsp90 is an ATPase was realized when a crystal structure of the ND was determined in complex with ATP (Prodromou et al., 1997). The importance of ATP in Hsp90 function is demonstrated by ATPase deficient mutants that cannot perform their function *in vivo* as well as the effectiveness of ansamycin antibiotics, such as geldanamycin, which competitively inhibit ATP binding and

effectively disrupt the ATPase activity of Hsp90 as well as its *in vivo* function (Grenert et al., 1997; Obermann et al., 1998; Stebbins et al., 1997).

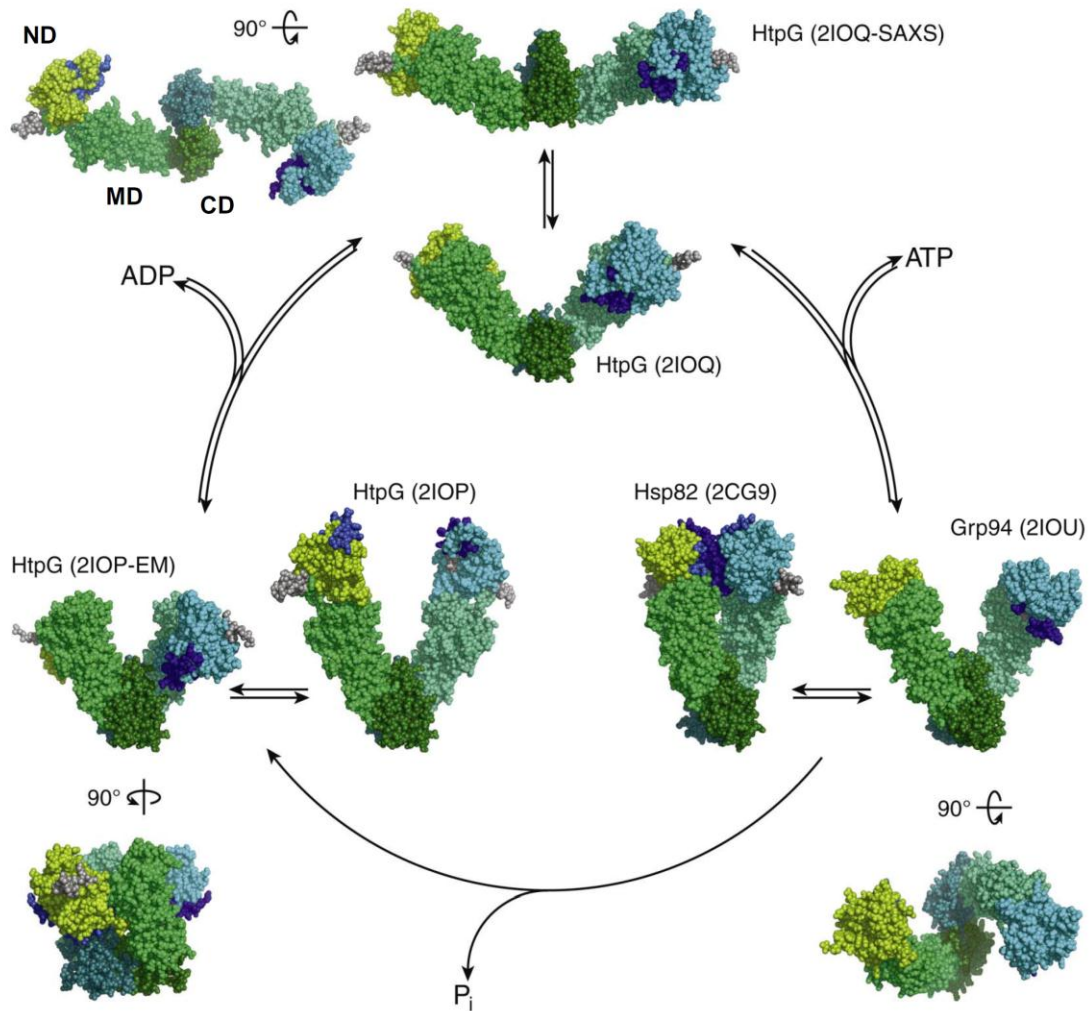


Figure 1.5: Structural flexibility of Hsp90.

The experimentally determined Hsp90 dimer in various nucleotide bound states is presented in surface representation. The ND, MD and CD are labeled in the top-left representation. The charged linker is shaded gray. The ND dimer interface is shaded dark blue. The PDB identifier is shown along with the specific isoforms of Hsp90 and the method of determination of the structure (SAXS is small angle x-ray scattering; EM is electron microscopy; unlabeled is X-ray crystallography). HtpG is the bacterial Hsp90. Hsp82 is the yeast cytosolic Hsp90. Grp94 is the eukaryotic ER Hsp90. Reprinted from Molecular Cell, 39/3, Matthias P. Mayer, Gymnastics of Molecular Chaperones, 321-331, Copyright (2010), with permission from Elsevier.

Structural studies have demonstrated that the nucleotide sits in the binding pocket in an unusual “kinked” conformation (Ali et al., 2006). The ribose and adenine groups are situated at the floor of the pocket and the phosphate groups emerge up and out towards the solvent. Hsp90 is classified as a split ATPase in that residues from the MD are required for catalytic activity. An arginine in the MD contacts the gamma phosphate of ATP and is essential for hydrolysis (Ali et al., 2006; Meyer et al., 2003). These unusual characteristics of the Hsp90 ATP-binding fold along with its slow catalytic rate are probably the reason for the difficulty in firmly establishing that this molecule is indeed an ATPase. The nucleotide binding region of Hsp90 is in a specialized class of structurally related folds found in DNA topoisomerases, DNA mismatch repair proteins, and histidine kinases. This fold is referred to as the Bergerat fold and the family is known as the GHKL family (Dutta and Inouye, 2000).

Hsp90 possesses extraordinary flexibility (Figure 1.5). Several full-length crystallographic structures, electron microscopy structures, and small angle X-ray scattering (SAXS) correlates of structure for this molecule demonstrate the dimer situated in vastly different conformations (Ali et al., 2006; Dollins et al., 2007; Shiau et al., 2006). The variety of conformations that Hsp90 can assume are demonstrated in Figure 1.5; note that Hsp90 is constitutively dimerized *via* the CD, however it can transiently dimerize through the ND as well. Both the open and closed conformations also show extensive flexibility in both *en bloc* motions and secondary structural transitions including lateral, longitudinal and twisting motions (Graf et al., 2009; Phillips et al., 2007). These conformational changes appear to be preferentially populated in different nucleotide states but are not stringently coupled to ATP hydrolysis (Figure 1.5) (Hessling et al.,

2009; Mickler et al., 2009). For example, multiple conformations of Hsp90 have been observed each for the apo state, ATP bound, and ADP bound state (Figure 1.5).

Single molecule fluorescence resonance energy transfer (FRET)-based kinetic analysis has allowed the development of a model that incorporates the structural transitions with the nucleotide state (Figure 1.6) (Hessling et al., 2009). Consistent with the structural data demonstrating various conformations in each nucleotide state, the authors found that the presence of nucleotide changed the preference of Hsp90 for certain conformations but it could still sample each conformation through stochastic thermal fluctuations (Mickler et al., 2009). The cycle is traversed as follows. The Hsp90 dimer in the ATP-free state is in an open conformation that is competent to bind ATP (1). Upon

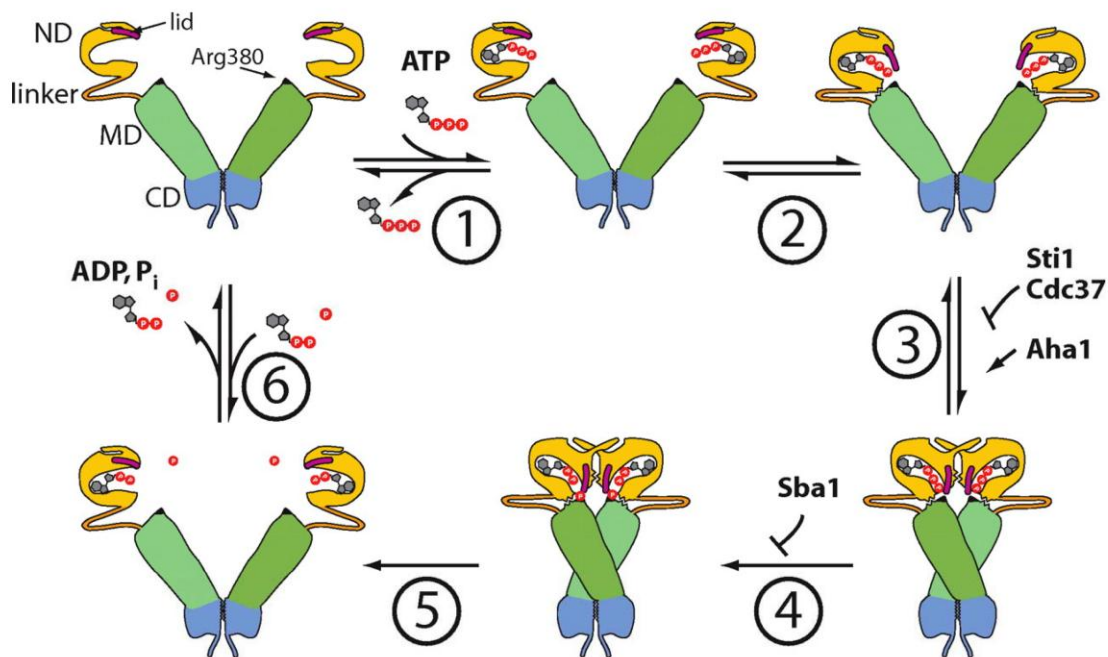


Figure 1.6: The ATPase cycle of Hsp90.

The Hsp90 dimer is represented with the ND, MD, CD, linker and lid shaded as demonstrated. This research was originally published in The Journal of Biological Chemistry. Wandinger SK, Richter, K, Buchner J. The Hsp90 Chaperone Machinery. *J Biol Chem.* 2008; 283:18473-7. © the American Society for Biochemistry and Molecular Biology.

ATP binding, a lid segment of the ND closes over ATP and the ND interacts with the MD (2). The dimer then forms a second dimerization interface involving strand swapping between the two NDs (3). The full split ATPase is now formed and ATP is committed to hydrolysis (4). Once ATP is hydrolyzed, the ND dimer separates and the open conformation is reassumed (5). Finally, ADP is released and the cycle is ready to begin again (6) (Figure 1.6).

In addition to nucleotide regulation, these states are also bound by the plethora of co-chaperones that Hsp90 interacts with. These cofactors stabilize the various states Hsp90 can sample, modulating the conformation cycle of the chaperone. For example, Hop (known as Sti1 in yeast) interacts with Hsp90 in an open conformation and prevents the formation of the closed complex which is necessary for nucleotide hydrolysis (Figure 1.6, step (3)) (Hessling et al., 2009). Thus, Hop prevents ATP hydrolysis while not affecting binding and is a non-competitive inhibitor (decreases the K_{cat} while not affect the K_m .) (Richter et al., 2003). As mentioned earlier, Hop also interacts with Hsp70, thus it is suggested that Hop brings these two chaperones together and locks Hsp90 in an open conformation that is competent to bind client, streamlining the client protein transition process (Scheufler et al., 2000; Wandinger et al., 2008). Another important co-factor of Hsp90 is UNC45, this protein directly interacts with Hsp90 and can act as a chaperone for myosin as well as the progesterone receptor (Barral et al., 2002; Chadli et al., 2006). In contrast to Hsp70 which is regulated by only a few co-chaperone molecules, there are dozens of partner proteins that bind to and regulate Hsp90 function. These molecules are involved in activities ranging from delivering client protein (Hop), activating ATPase activity (Aha1), inhibiting ATPase activity (Hop, CDC37), and catalyzing peptidyl-prolyl isomerization reactions (FKBP52).

Chapter 2: Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay

The role of molecular chaperones in protein misfolding diseases is demonstrated by the neurodegenerative disease, Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay (ARSACS). This disorder is due to loss-of-function mutations in a gene which encodes a protein with homology to both molecular chaperones and co-chaperones: sarsin.

A. AUTOSOMAL RECESSIVE SPASTIC ATAXIA OF CHARLEVOIX-SAGUENAY

ARSACS is emerging as world-wide etiology of hereditary spastic ataxia. This neurodegenerative disorder is caused mutations in a gene which encodes the protein sarsin. Sarsin possesses regions of similarity to both molecular chaperones and co-chaperones. Thus, ARSACS is an example of a disorder with a direct pathological link between a putative molecular chaperone and neurodegenerative disease. An understanding of sarsin function may shed light on how efficient protein folding is accomplished in the central nervous system (CNS).

A Historical Perspective

In the mid-1970s, neurologists in Canada became aware of a distinct ataxia common in the Charlevoix-Saguenay region of Quebec (Figure 2.1). This ataxia was prevalent in a genetically isolated population and was inherited in a strictly autosomal recessive fashion. Previously, these patients had been diagnosed with Friedreich's ataxia, spastic paraplegia or a "*forme de passage*" between the two diseases; however the neurologists realized they had a distinct phenotype. Based on the unique symptoms and the inheritance pattern, a new syndrome was defined for these patients: Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay (Bouchard et al., 1978).

The Charlevoix county of Quebec was populated by about 40 families from Quebec City between 1665 and 1725 (Bouchard et al., 1978). This region is approximately 60 miles northeast of Quebec City and is on the shores of the St. Lawrence River (Figure 2.1). This cohort became genetically isolated due to the mountainous terrain and lack of accessible roads and there was little contact with the communities outside of their region. They populated the region relatively rapidly and by 1825 required additional arable land to sustain the needs of the population. Thus, a group of families left for the Saguenay region northwest of Charlevoix and towards Lac-Saint-Jean between 1838 and 1855 (Figure 2.1). Within a few years (~1850) these families comprised 80% of the population of about 5,000 people who lived in Saguenay. By the late 1900s the population of the Charlevoix-Saguenay region had grown to over 300,000. Based on

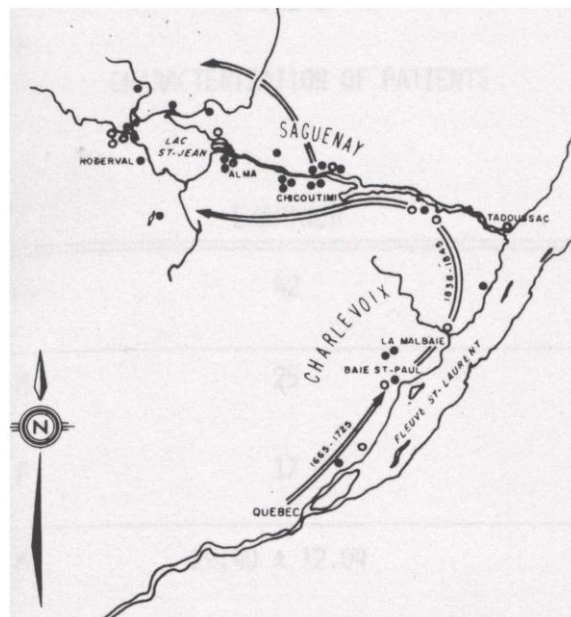


Figure 2.1: Geography of Charlevoix and Saguenay regions of Quebec.

A population from Canada migrated to Baie St-Paul in Charlevoix from 1665-1725. Another migration to the Saguenay and Lac St-Jean regions occurred from 1838-1855. Fleuve St-Laurent is the St. Lawrence River. Reproduced with permission from The Canadian Journal of Neurological Sciences, J.P. Bouchard, A. Barbeau, R. Bouchard and R.W. Bouchard, 1978, 5(1), 61-69.

genealogical records it is thought that the majority of this population are descendants of the original colonists with relatively little immigration to the region (Richter et al., 1993).

Thus, there is a significant founder effect in the Charlevoix-Saguenay region. The population has grown to nearly 500,000 and most of these appear to be descendent from 40 families. In fact, it has been suggested that the disease can be traced to a single couple living in Quebec City in the mid 1600s (Bouchard et al., 1978); however, it is difficult to validate this prospect. Regardless, a small founder group populated this region and the population is relatively genetically homogenous leading to a high incidence of ARSACS. There is not substantial consanguinity in the population, but all of the inhabitants are distantly related to one another (Richter et al., 1993). This has led to a carrier frequency for ARSACS that was estimated to be at 1/22 in the mid-1900s (De Braekeleer et al., 1993). Based on the above statistics, it can be estimated that the disease prevalence is at least 0.1%; it may be much higher since not all patients are included in the literature. Interestingly, ARSACS is not the only genetic disease that is prevalent in this isolated population. Familial tyrosinemia and familial agenesis of the corpus callosum with sensorimotor neuropathy are also prevalent in Charlevoix-Saguenay, consistent with the complications of a genetically homogenous population (Laberge, 1969; Larbrisseau et al., 1984).

Due to the high prevalence of ARSACS and a genetically homogenous population, linkage mapping was pursued to identify the locus of the gene involved (Engert et al., 2000; Engert et al., 1999; Richter et al., 1999). In 2000, these efforts were successful leading to the identification of the *SACS* gene (discussed below) (Engert et al., 2000). Since the molecular identification, patients worldwide with similar symptoms were tested for this disease and indeed it was shown to be a prominent global cause of

ataxia (Gomez, 2004). Currently, patients have been identified in Tunisia (Mrissa et al., 2000), Japan (Ogawa et al., 2004), Italy (Criscuolo et al., 2004), Turkey (Richter et al., 2004), Spain (Criscuolo et al., 2005), Belgium (Breckpot et al., 2008), The Netherlands (Vermeer et al., 2008), France (Anheim et al., 2009), Morocco (Baets et al., 2010), Serbia (Baets et al., 2010), Hungary (Baets et al., 2010) Algeria (H'Mida-Ben Brahim et al., 2011) and Germany (Gerwig et al., 2010). Importantly, one Italian patient inherited a mutation from a father from The United States of America (USA) (Anesi et al., 2011). Thus, at least one carrier of ARSACS has been identified in the USA and it is reasonable to suggest that ARSACS patients are present in this country but do not yet have a molecular diagnosis or have not been published.

Signs and Symptoms of ARSACS

The initial report defining ARSACS as a clinical entity described a progressive spastic ataxia that strikes in childhood (Bouchard et al., 1978). The term ataxia refers to a gross lack of coordination of body movements often manifested as unsteadiness or

Onset

Unsteadiness at gait initiation

Progressive signs

Mostly spastic ataxia of the four limbs

Slurred and dysrhythmic speech

Discrete to severe distal amyotrophy

Absent ankle jerks after 25 years of age

Early non-progressive signs

Increased deep tendon reflexes

Bilateral abnormal plantar response

Marked saccadic alteration of ocular pursuit

At funduscopy: prominent myelinated fibers radiating from the disc and embedding retinal vessels

Table 2.1: General signs and symptoms of ARSACS.

Reprinted from Neuromuscular Disorders, 8/7, Bouchard JP, Richter A, Mathieu J, Brunet D, Hudson TJ, Morgan K, Melançon SB., Autosomal recessive spastic ataxia of Charlevoix-Saguenay, 475, Copyright (1998), with permission from Elsevier.

staggering. The core finding of the disease is ataxia that is noted at around the age of gait initiation and is manifested as clumsiness and frequent falls. However, the disease is progressive and becomes most apparent in the late teens to early twenties (Richter et al., 1993). As disease progression occurs, spastic ataxia of the limbs worsens and the speech becomes abnormal. The muscles begin to atrophy and the ankle jerk reflex will disappear. Early in the disease, several non-progressive signs are apparent; the deep tendon reflexes are hyperactive (although progression of the disease results in a loss of the ankle jerk reflex), the plantar response is abnormal bilaterally (Babinski's sign is present), and there is saccadic alteration in smooth ocular pursuit. The main features are listed in Table 2.1.

Ataxia of fine motor skills is more pronounced than truncal ataxia. For example,

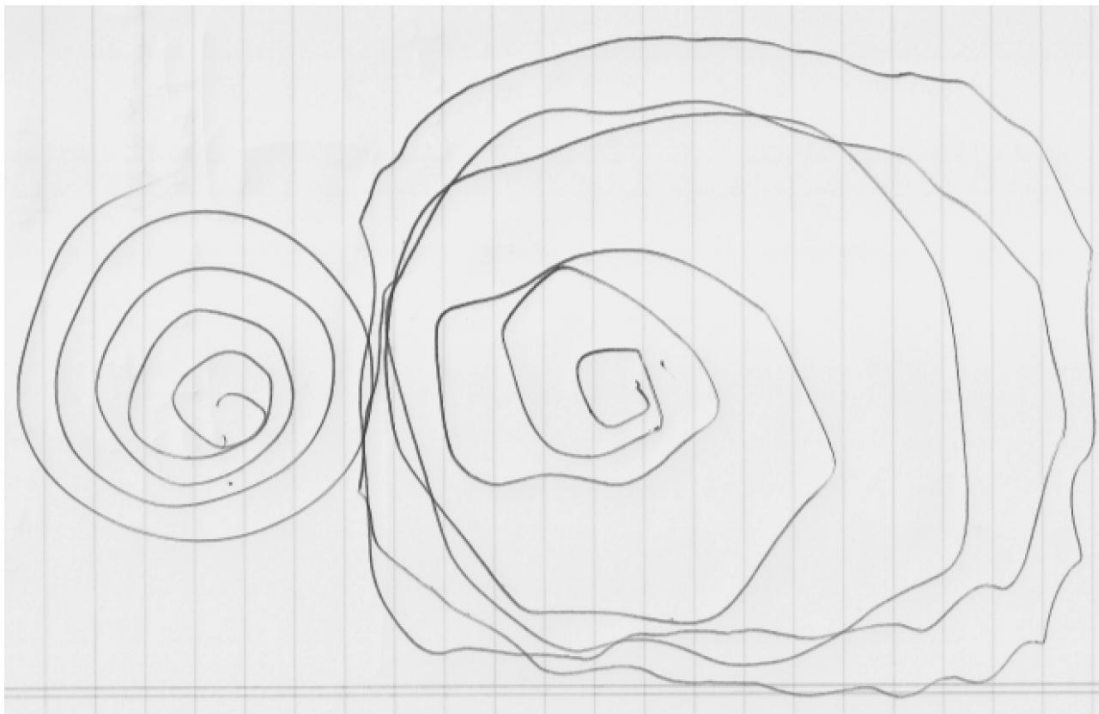


Figure 2.2: Ataxia of fine motor movements in an ARSACS patient.

A spiral shape was drawn on a piece of paper (left) and the patient (54 years old) was asked to copy it (right). Note the over-correction type mistakes made during the drawing. The image is from a personal visit to a patient with Dr. Bernard Brais, Montreal, Canada, 2010.

the patient can remain seated without exhibiting any signs of ataxia, but cannot perform rapid fine or course alternating movements. The fine motor deficit is demonstrated by asking the patient to copy a spiral figure on a piece of paper (Figure 2.2). The patient demonstrates a tendency to make an error and then overcorrect, indicative of a cerebellar defect. Similarly, early in the disease, the patients will display a staggering gait as they start to fall and then overcorrect and fall in the other direction. Additionally, the gait has been described as jerky, and with a scissoring characteristic (Dupre et al., 2006).

In addition to the ataxia, the patients are spastic and amyotrophic. Peripheral denervation occurs early in the disease and leads to both increased muscle tone and loss of muscle mass. The spasticity and amyotrophy are particularly evident in the hands and feet; by the age of 20 their combined effect leads to deformities of these extremities (Richter et al., 1993). This is a particularly debilitating phenomenon which often requires surgery for functional improvement (Bouchard and Langlois, 1999). As described above, the patients are hyper-reflexic early in the disease and demonstrate clonus in the ankles and knees; however, as the disease progresses the ankle jerk reflex disappears entirely. This phenomenon is due to denervation of the musculature of the lower extremity. Muscle weakness becomes more prominent as the disease progresses and appears to be due to denervation and amyotrophy rather than intrinsic muscle pathology. The patients will require a wheelchair for locomotion eventually. In one case-series, the mean age at becoming wheelchair-bound was 41 years old (Richter et al., 1993).

There is also a sensory deficit consistent with abnormalities of information conduction along the posterior columns (Bouchard et al., 1978). Cutaneous sensation is typically intact, but vibration sense is impaired. This occurs in an age-dependent and anatomic-dependent fashion. The children typically have normal vibratory sensation that

decreases as they age. Additionally, the loss of vibratory sense begins in the toes and progresses up to the level of the ankle. There is altered proprioception as demonstrated by the reduced ability to detect toe positioning. The anatomic location and age-dependent nature of these abnormalities is consistent with the loss of the ankle jerk reflex later in disease.

The neurological deficits are also noted in the control of voice and eye movement (Bouchard et al., 1978). The patients are dysarthric and speak with a slurred voice. This finding becomes apparent in childhood and progresses into adulthood. However, the patients are able to communicate verbally throughout life. The patients exhibit nystagmus upon both horizontal and vertical gaze (Dionne et al., 1979). A saccadic defect is grossly apparent during ocular pursuit movements. These findings appear in children and do not seem to progress with age (Dionne et al., 1979). Fundusoscopic examination of the retina

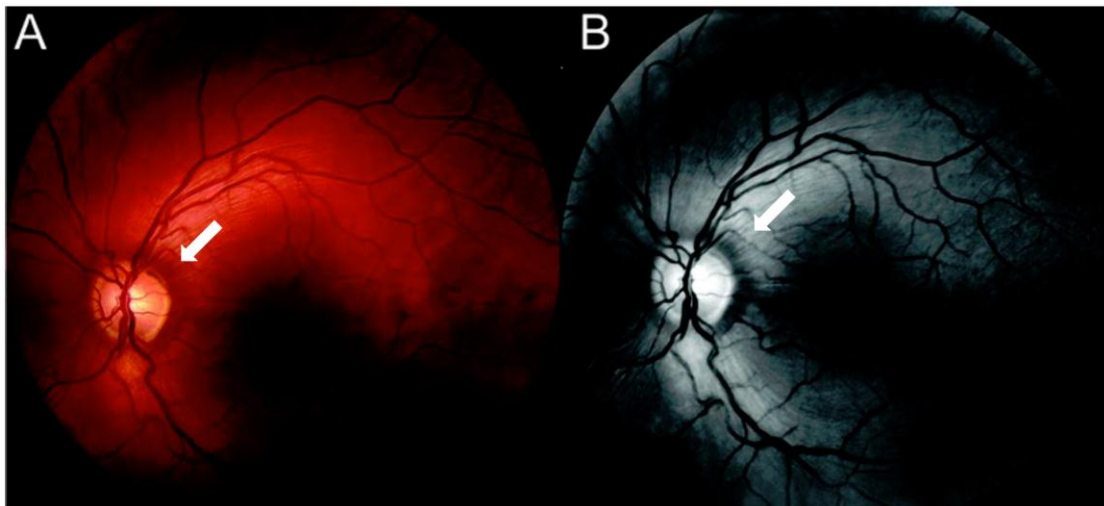


Figure 2.3: Images of the fundus in an ARSACS patient.

(a) Color and (b) black and white pictures of the fundus in an ARSACS patient demonstrating hyper-myelinated retinal fibers emanating from the optic disc (arrows). Reproduced with permission from: Philip Van Damme, Philippe Demaerel, Werner Spileers, Wim Robberecht, *Autosomal recessive spastic ataxia of Charlevoix-Saguenay*, Neurology, volume 72, issue 20, page 1790. © by AAN Enterprises.

may reveal increased presence of nerve fibers, often described as hyper-myelinated retinal nerve fibers (Figure 2.3). The consequences of this finding are not clear; however, most patients report visual defects that are not correctable by refraction. Indeed, their visual acuity is similar to the general population (Bouchard et al., 1978). However, many complain of a “film” over their vision, yet the lens is clear. This appears to be due to an intrinsic retinopathy that has not been investigated.

There is a relatively homogenous clinical picture from patients possessing diverse mutations throughout the world. In 2007, Takiyama and colleagues tabulated the signs and symptoms of all the case identified to date (Table 2.2) (Takiyama, 2007). This analysis revealed a remarkably homogenous clinical picture with many of the variations occurring in the progressive signs which may become more apparent in the patients as they age. In all ARSACS patients identified, ataxia is an invariant finding. The most significant variation is the degree of spasticity, which is a core component of the Quebec

	Italy ^a	Japan ^b	Quebec ^c	Spain ^d	Tunisia ^e	Turkey ^f
Number of families	3	9	24	1	4	4
Number of patients	4	13	42	2	18	7
Age at examination						
Range (years)	21–43	25–57	9–52	26–37	16–55	11–20
Mean (years)	33.5	38.7	30.5	31.5	32.6	14.9
Age at onset						
Range (years)	2–2	3–9	1–1.5	infancy	1–14	1.5–3.5
Mean (years)	2.0	5.4	?	infancy	4.5	2.4
Ataxia (%)	100	100	100	100	100	100
Dysarthria	100	100	100	100	77.8	100
Nystagmus	66.7?	92.3	100	100	100?	71.4
Distal amyotrophy	66.7?	84.6	?	100?	66.7	71.4
Babinski sign	100?	92.3	100	100	?	100
Hyperreflexia	100	69.2	100	100	100	100
Spasticity	100	76.9	100	100	100?	100
Pes cavus	100?	84.6	85.7	100	55.6?	?
Retinal striations	0?	38.5	100	0	11.1	71.4

Table 2.2: Clinical features of ARSACS patients from various countries. With kind permission from Springer Science+Business Media: The Cerebellum, Saccinopathies: saccin-related ataxia, 6, 2007, 354, Yoshihisa Takiyama, Table 1.

patients. Almost all patients with ARSACS are spastic, yet a few non-spastic patients were identified in Japan (Shimazaki et al., 2007; Shimazaki et al., 2005). However, these patients may have been spastic at some point in their disease and it decreased rather than progressed. Mental deficiency is also a component of a minority of the patients, but is not a standard finding for this disease (Yamamoto et al., 2005). Retinal striations appear to be the most variable feature in ARSACS patients. This may be related to variations in the retinal pathology or to different subjective thresholds for a clinician to define the retina as abnormal.

This disease is compatible with life well into adulthood. In one case series of 34 patients the mean age at death was 51 years old (Bouchard et al., 1998). Some patients did survive into their 70's. However by this age, the patients were bed-ridden and typically succumbed to infection.

Diagnostic Studies and Pathology

Except in a few case reports, ARSACS patients do not demonstrate significant mental deficiencies. Slight deficits in non-verbal Intelligence Quotient (IQ) testing may be present, but these could be related to a physical inability to perform certain tasks due to the ataxia. In agreement with this, verbal IQ is found to be in the normal range in most patients (Bouchard et al., 1978). There may be a trend towards early-onset dementia (Bernard Brais, personal communication), but this finding has not been systematically studied.

Electromyography confirms denervation of distal muscles. This sign appears early in the disease and is progressive (Bouchard et al., 1979a). As the disease progresses, motor nerve conduction studies will demonstrate reduced conduction velocities of evoked action potentials. In many patients it is not possible to evoke a motor action potential

after the patient reaches 40 years old; consistently, the patients often become wheel-chair dependent at this age. Sensory nerve conduction studies reveal absent potentials in agreement with the abolished vibratory sensation (Bouchard et al., 1979a). Non-specific generalized abnormalities are often found on electroencephalograms as well as an increased prevalence of epilepsy (Bouchard et al., 1979b).

Sectional imaging of the brain using Magnetic Resonance Imaging (MRI) or Computerized Axial Tomography scans (CAT) demonstrate neurodegeneration (Richter et al., 1993). The cerebellum and cervical spinal cord are grossly atrophied and there is concomitant enlargement of the cisterns. The vermis of the cerebellum demonstrates the

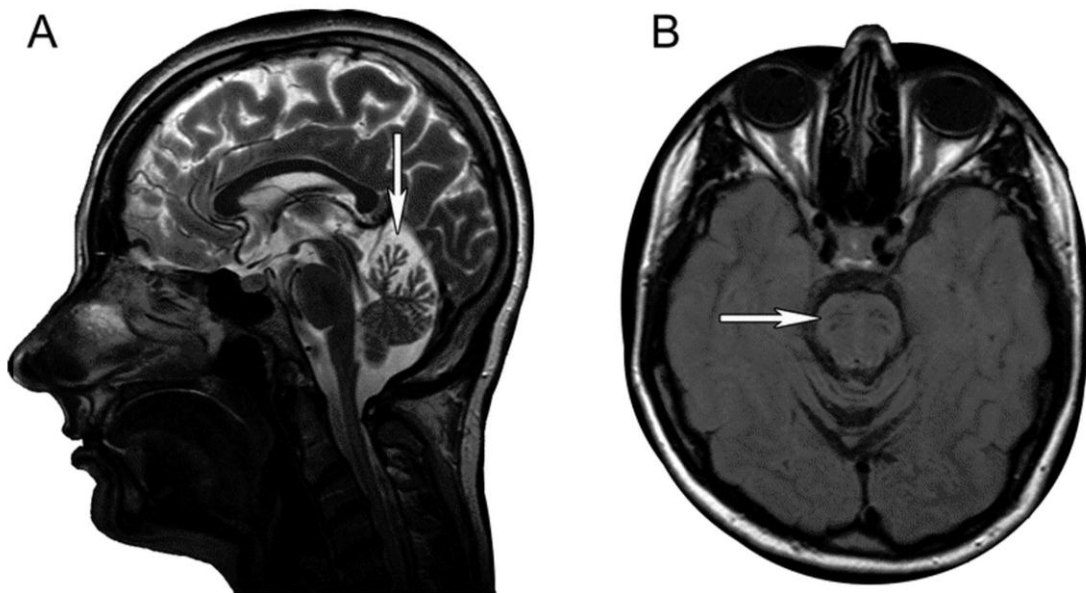


Figure 2.4: MRI study of an ARSACS patient.

(a) Sagittal section through the brain of an ARSACS patient on a T2-weighted MRI study. The arrow denotes the atrophy of the vermis. (b) Transverse section of the same patient. The arrow denotes the linear hypointensities in the Pons. Reproduced with permission from: Philip Van Damme, Philippe Demaerel, Werner Spileers, Wim Robberecht, *Autosomal recessive spastic ataxia of Charlevoix-Saguenay*, *Neurology*, volume 72, issue 20, page 1790, 2009. © by AAN Enterprises.

most pronounced atrophy with relative sparing of the inferior structures in the cerebellum (Figure 2.4a). Additional signs of atrophy extend to the thoracic segments of the spinal cord. The cerebrum does not display any significant focal abnormalities of the white matter although there is some non-specific global cerebral atrophy in later life (Martin et al., 2007). A characteristic finding in MRI studies is linear hypointensities in the Pons (Figure 2.4b). This is a symmetric finding that is localized to the anatomic site of the corticospinal tracts, and is thought to be due to disease involvement in the pontocerebellar fiber and/or pontine nuclei (Martin et al., 2007).

Since the gene was identified it is now possible to molecularly diagnose the patients (Engert et al., 2000). Efforts have begun to develop an inexpensive and efficient genetic test (Mercier et al., 2001; Vermeer et al., 2009), however these methods are not in widespread use due to availability and technical drawbacks. Currently, definitive diagnosis requires specialized testing in a research lab to directly sequence the gene.

There is limited data in the literature on the histological pathology in the central nervous system. Autopsy findings on a 21-year old patient whose death was not directly caused by the disease demonstrated findings consistent with imaging studies (Richter et al., 1993). There was significant atrophy of the cerebellar vermis, most pronounced in the superior and anterior regions. The atrophy appeared to be caused by an absence of Purkinje cells confirming that the disease strikes neurons rather than supporting cells. Throughout the cerebellar cortex, there was thinning of the molecular and granular layers with decreased cell numbers. The spinal cord also demonstrated pathology with a loss of myelin stained structures in the lateral corticospinal tracts, consistent with the atrophy observed on imaging studies. This finding was less pronounced in the more distal regions of the spinal cord, but in these regions loss of staining of the spinocerebellar tract became

more prominent. Thus, there is loss of structures in pathways leading both to the cerebellum and from the cerebellum. Interestingly, the presence of “dense, lipofuscin-like granules” was demonstrated in one patient (Richter, 1996). This finding prompted the authors to suggest that this disease may be due to lysosomal storage defects, however no alterations in known lysosomal enzymes were found.

B. THE SACSIN PROTEIN

The *SACS* gene was identified through linkage analysis and its protein product was termed sacin. Many mutations have been identified throughout this gene. Sacin is a multi-domain protein of unknown function. Cellular and biochemical studies are only now beginning to shed light on its biology.

Gene Mapping and Mutations

The homogenous clinical phenotype along with the genetically isolated/homogenous population in Charlevoix-Saguenay made ARSACS an ideal disorder for linkage analysis to identify the locus (Bouchard et al., 1978). Since ARSACS shares certain features with Friedreich’s ataxia, the investigators first confirmed that the same gene was not responsible for these diseases. Friedreich’s ataxia maps to a locus on chromosome 9; whereas it was definitively determined that ARSACS was not linked to this chromosome; thus, these diseases were determined to be caused by mutations in distinct genes (Richter et al., 1993). Subsequent analysis revealed that the ARSACS locus was linked to chromosome 13q (Bouchard et al., 1998). High resolution genetic mapping led to the determination that chromosome 13q11 was the locus of the ARSACS gene (Engert et al., 1999; Richter et al., 1999). Several genes in this region were directly sequenced and compared to unaffected controls to identify the responsible gene. This resulted in the discovery of two mutations in a single large open reading frame (ORF): a

deletion, 80454delT, and a nonsense mutation, C79114T that were determined to be responsible for ARSACS in this population (Table A.1) (Engert et al., 2000).

The initial identification of the gene described it as a single large exon. The entire gene was thought to be coded by a single 11.5 kb exon, the largest vertebrate exon known (Engert et al., 2000). While the majority of the coding sequence is contained in this exon, it has since been determined that eight small exons are present in the 5' direction (Figure 2.5) (Ouyang et al., 2006). A total of 74 mutations world-wide have been identified in the *SACS* gene (as of April 2011). These include 29 missense mutations, 20 deletions leading to frameshift, 4 insertions leading to frameshift, 18 nonsense mutations, 1 codon deletion and 2 genomic deletions of the entire gene (Figure 2.6 and Table A.1). The clinical phenotype shows some variability across these mutations, but there is not a consistent link between any one phenotypic parameter and a specific mutation. Rather, it appears that environmental and other genetic factors may be important in the specific pattern of disease expression in a particular patient.

Interestingly, two large genomic deletions have been identified in ARSACS

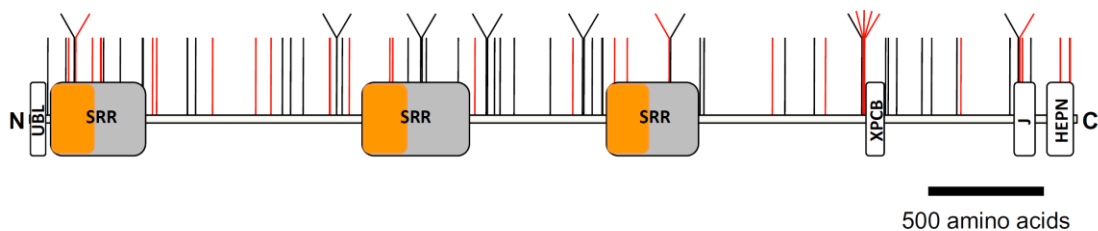


Figure 2.6: Location of human mutations in saccin.

Cartoon of the saccin protein domain structure with the locations of human disease-causing mutations marked by red or black lines (red lines are missense mutations, black lines are other types of mutations). Closely overlapping mutations are identified by a bent line. Abbreviations are as follows: Ubl, ubiquitin-like domain; SRR, saccin repeating region; XPCB, xeroderma pigmentosum c-binding domain; J, J domain; HEPN, higher eukaryotes and prokaryotes nucleotide binding domain. Detailed description of these domains is given in the main text and in Chapter 3. The orange region in the SRR domain represents similarity to the molecular chaperone Hsp90.

patients. An Italian patient was identified with a frameshift deletion mutant in the *SACS* gene (Terracciano et al., 2009). The *SACS* gene in both parents was sequenced and it was found that the mother did not possess this mutation. Careful analysis revealed that in addition to the mutation described above, this patient also possessed a large genomic deletion on the other chromosome. This deletion encompassed six genes including *SACS*. Another patient was identified in the same cohort who possessed a similar deletion. Interestingly, these patients exhibited hearing loss which has not been identified in other ARSACS patients. Likewise, a patient was identified in Belgium who possessed a similar genomic deletion on one chromosome and a hemizygous mutation in *SACS* on the other chromosome (Breckpot et al., 2008). This deletion encompassed the same six genes as the Italian patients, but was a *de novo* mutation in this patient and may only be somatic (*i.e.* neither of his parents had a deletion). Furthermore, a distinct chromosomal deletion was detected in a Canadian patient who lived outside the Charlevoix-Saguenay region (McMillan et al., 2009). This patient exhibited signs of both Limb Girdle Muscular Dystrophy type 2C (LGMD2C) and ARSACS. The gene responsible for LGMD2C is *SGCG* and is directly adjacent to the *SACS* gene on chromosome 13. Genetic analysis revealed a homozygous deletion of both genes. The parents were both heterozygous for the same deletion. This patient had a classical clinical presentation of both LGMD2C and ARSACS, additionally she did not have hearing loss as in the other cases of ARSACS due to chromosomal deletion, indicating that this feature is caused by deletion of the larger region of the chromosome and not by deletion of *SACS* alone.

An additional genetic mechanism of acquiring ARSACS was demonstrated by an Italian patient (Anesi et al., 2011). This patient appeared to possess a homozygous mutation in *SACS* but the mutation was only identified in the father (who was from the

USA). Additional chromosomal analysis revealed that he possessed two copies of his father's chromosome 13. Thus he acquired ARSACS through uniparental disomy (UPD).

All of the genetic evidence demonstrates that ARSACS is due to loss of function of the saccin protein. The most significant evidence is that deletion of the entire gene gives rise to the same phenotype as all of the other types of mutations. These chromosomal deletions are clearly loss of function, since the entire gene is absent from the chromosome and there is no gene from which to make a protein product. Thus the phenotype is not due to gain of function of the mutated protein or toxic aggregates of the mutated protein.

Domain Architecture of the Saccin Protein

The initial report identifying the saccin protein, along with subsequent studies, revealed the presence of multiple domains in this large protein (Figure 2.7). These include an Ubiquitin-like domain (Ubl) at the extreme N-terminus of the protein, followed by three saccin repeating regions (SRR) (described in Chapter 3). The C-terminal third of the protein is composed of a Xeroderma Pigmentosum C-Binding domain (XPCB), J domain (J), and Higher Eukaryotes and Prokaryotes Nucleotide binding domain (HEPN).

Ubiquitin-like domains are regions of similarity to the entire ubiquitin protein contained within a larger polypeptide (Buchberger, 2002). In multi-domain proteins, these typically function as modules for interaction with the proteasome (Hartmann-Petersen and Gordon, 2004). Based on sequence homology, saccin was suggested to possess a Ubl domain (Figure 2.7) (Parfitt et al., 2009). The 124 N-terminal residues of saccin, which contain this Ubl domain, was found to co-immunoprecipitate with a subunit of the 20S proteasome (Parfitt et al., 2009). Thus, saccin is capable of interacting with the

proteasome; the significance of this interaction for saccsin function is not currently known. Interestingly, saccsin was identified as a target for ubiquitination itself. The E3 ubiquitin ligase Ube3A, involved in Angelman syndrome, ubiquitinates saccsin in neuronal cells (Greer et al., 2010). Since Angelman syndrome possesses a jerky movement abnormality, the authors suggested that dysregulation of saccsin levels may underlie the motor deficits seen in this disorder.

Saccsin possesses three regions of similarity to the N-terminal region of Hsp90. As described in Chapter 1, this region of Hsp90 binds and hydrolyzes ATP. Additionally, it

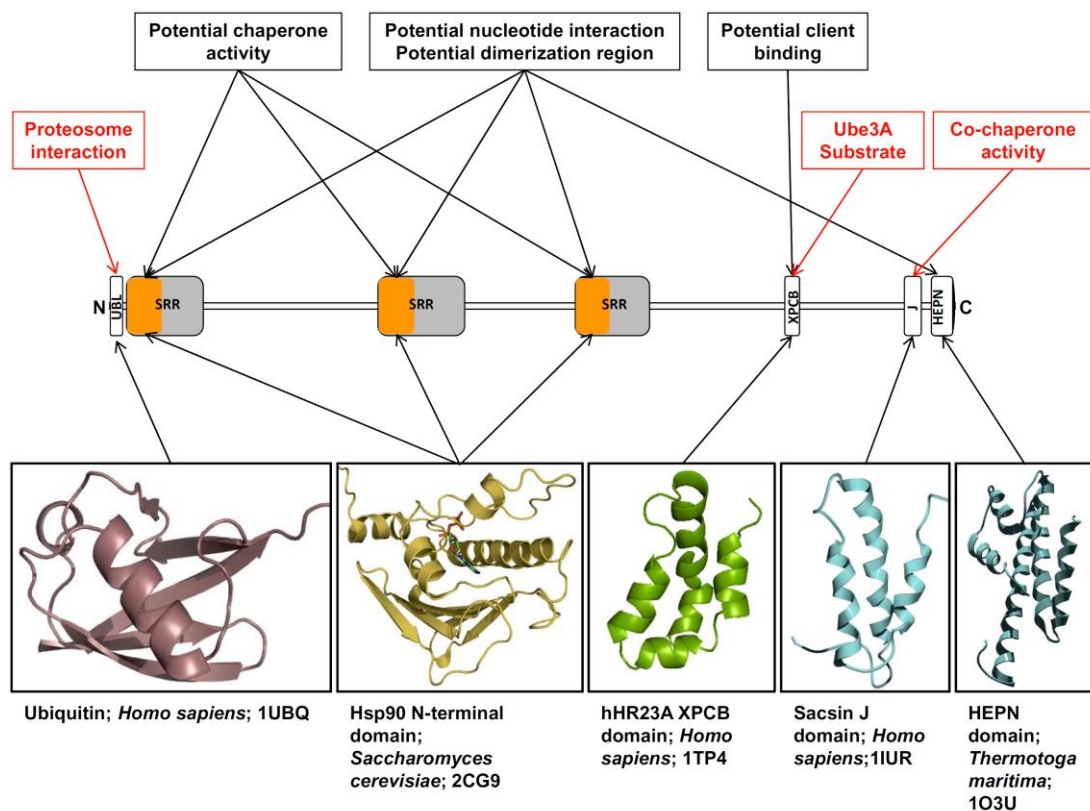


Figure 2.7: Known and hypothesized structural and functional data for saccsin. The potential functions and putative domains involved are marked in black squares with black arrows. Known functions of saccsin are marked in red squares with red arrows. The structures of domains similar to saccsin are shown with the species and PDB ID.

has been shown to possess chaperone activity independent of the rest of the molecule. A detailed discussion of the role of the Hsp90-like regions in saccin function will be given in Chapter 3.

Saccin was determined to possess a region of similarity to the XPCB domain of hHR23A, a homolog of Rad23 (Figure 2.7) (Kamionka and Feigon, 2004). Rad23 is a component of the nucleotide excision repair pathway. It contains an Ubl domain followed by the XPCB domain and two ubiquitin-associated domains (UBA). The XPCB domain is also found in the co-chaperone Hop/Sti1 where it is known as a Sti1 domain. As described in Chapter 1, Hop is a co-chaperone that links the Hsp70 and Hsp90 chaperone machinery. In Rad23, XPCB is a protein interaction module for the XPC protein (Masutani et al., 1994). The interaction between XPCB and its partner protein, XPC, is resistant to salt, and was thus suggested to occur through hydrophobic interactions (Masutani et al., 1994). Structural studies of the XPCB domain revealed extensive hydrophobic surface regions that are likely the protein-interaction regions (Kamionka and Feigon, 2004; Kim et al., 2005). XPC is an unstable and poorly soluble protein that requires interaction with the XPCB domain for stabilization (Ng et al., 2003). The hydrophobic interaction surface along with the stabilization function are similar to a chaperone-client interaction, thus this domain may be a client protein interaction motif.

The initial report identifying saccin noted the presence of a region of similarity to the J domain of the co-chaperone Hsp40/DnaJ (Engert et al., 2000). This domain was described in Chapter 1 as important for stimulating the ATPase activity of Hsp70. Parfitt and colleagues determined that the saccin J domain has *bona fide* co-chaperone function in an *in vivo* complementation assay (Parfitt et al., 2009). Indeed, mutation of the histidine to glutamine in the critical HPD motif resulted in a loss of complementation

confirming the result was specific to saccin. Thus, saccin is capable of acting as a co-chaperone. Interestingly, a human disease causing mutation in the J domain (R4331Q) did not disrupt the J domain function in this assay.

The final domain identified in saccin is the HEPN domain (Figure 2.7). This unusual domain was identified only in bacteria, archaea and vertebrates and not in “lower” eukaryotes (Grynberg et al., 2003). In bacteria and archaea, these domains are part of proteins that also contain a nucleotidyl transferase (NT) domain or are positioned adjacent to a gene encoding an NT in the genome. HEPN was only identified in saccin homologs which were present in vertebrates and not in less complicated eukaryotes. The structure of a HEPN domain from *Thermotoga maritima* was solved and revealed that this protein is likely a dimer and is composed of five alpha-helices (Erlandsen et al., 2004). Database searches revealed structural homology to the C-terminal domain of kanamycin nucleotidyl transferase (KNTase), a nucleotide-dependent inactivator of the antibiotic kanamycin (Erlandsen et al., 2004). Interestingly, the N-terminal portion of KNTase is similar to the NT that is usually found in close relation to HEPN domains. Importantly, both the N-terminal and C-terminal domains in KNTase are involved in binding nucleotide (Pedersen et al., 1995). Thus, HEPN domains appear to be half of a functional nucleotide binding domain. The role of this domain in saccin function has not been investigated.

The presence of regions of similarity to a chaperone (Hsp90) and co-chaperones (Hsp40 and Sti1) reveals a potential role for saccin in chaperone-assisted protein folding activities. The presence of a functional Ubl domain also links saccin function to the protein degradation machinery in the cell. Thus, saccin has been suggested to be involved in protein quality control processes in the cell. There are four regions of potential

nucleotide binding in saccin (three Hsp90-like regions and the HEPN domain, Figure 2.7), indicating that saccin function is regulated by nucleotide binding and hydrolysis possibly at several levels. Additionally, these four regions are implicated in protein dimerization, transiently in Hsp90 and constitutively in HEPN domains (Figure 2.7).

Localization and Functions of Saccin

The cellular and sub-cellular localization of saccin has been investigated. Initial studies utilized northern blot analysis to determine which tissues expressed saccin mRNA (Engert et al., 2000). The message was identified in fibroblasts, brain, and skeletal muscle. *In situ* hybridization studies in human, monkey and rat brain revealed the message was present in all areas examined (*e.g.* cerebral cortex, cerebellum and hippocampus). Another group undertook a detailed analysis of saccin localization in the brain at the protein level utilizing polyclonal antibodies (Parfitt et al., 2009). This investigation revealed a largely neuronal expression pattern throughout the cerebrum, cerebellum and brainstem. Neurons that did not demonstrate at least some staining for saccin could not be identified. Strong staining was observed in the Purkinje cells in the cerebellum, particularly the molecular layer.

The subcellular localization of saccin in neurons was also investigated (Parfitt et al., 2009). Saccin was found in the soma and throughout the axons and dendrites in a predominantly cytoplasm distribution. The staining was mostly diffuse with small regions of punctuate staining. Additionally, a small amount of nuclear staining was observed. A fraction of saccin staining appeared to associate with the cytoplasmic face of mitochondria.

Parfitt and colleagues decided to investigate the effect of saccin on the overexpression of an ataxia-related protein, ataxin-1 (Parfitt et al., 2009). When

containing a polyglutamine expansion, this protein leads to the neurodegenerative disease spinocerebellar ataxia type 1 (SCA-1), an autosomal dominant cerebellar ataxia. In cellular models, overexpression of polyglutamine expanded ataxin-1 leads to aggregation and cell death. When overexpressed in a neuroblastoma cell line (SH-SY5Y) endogenous salsin was found to co-localize with intranuclear inclusions of mutant ataxin-1. Additionally, upon siRNA mediated knockdown of salsin, polyglutamine expanded ataxin-1 appeared to be even more toxic. Thus, salsin was concluded to exert a protective role *in vivo*. The authors hypothesized that this protection may be accomplished by enhancing the degradation of misfolded proteins.

Recently, a yeast-two hybrid analysis was utilized to develop an interaction network for known ataxia-related proteins (Lim et al., 2006). Along with many other ataxia-related proteins, a C-terminal portion of salsin was utilized as bait in this screen. Four interacting proteins were identified for this portion of salsin. The cytoskeletal protein α -actinin-4 (Otey and Carpen, 2004), the transcriptional regulator Thg-1pit (Canterini et al., 2005), the synaptic protein PICK1 (Xu and Xia, 2006), and ZNRD1, potentially involved in gene regulation (Hong et al., 2005), were found to be associated with salsin. A sample validation group showed that 83% of the hits could be independently confirmed, thus it is possible that some of these are not true hits; for example, Thg-1pit and ZNRD1 are both involved in gene regulation so these hits may be due to auto-activation. Analysis of the interaction network revealed that proteins involved in neurodegeneration of Purkinje cells are highly interconnected. Thus, a few common pathways may be involved in the neurodegeneration observed in this diverse set of ataxias.

Two reports identified sacsin as regulated by various cellular stressors. Jin and colleagues performed quantitative mass spectroscopy experiments to identify proteins whose expression was altered in a dopaminergic cell line (MES cells) upon treatment with rotenone, an inhibitor of the mitochondrial complex 1 (Jin et al., 2007). Rotenone has a broad range of effects, many of which mimic the cellular pathology seen in Parkinson's disease. In this study a portion of sacsin was confirmed to be mitochondrial by subcellular fractionation, while the majority was cytoplasmic, consistent with the study by Parfitt and colleagues (Parfitt et al., 2009). Additionally, sacsin protein levels were observed to decrease greater than two-fold upon treatment with the inhibitor. Rise and colleagues studied the effect of nodavirus infection on gene expression in Atlantic cod (Rise et al., 2010). Infection with this virus results in nervous system necrosis. *SACS* mRNA was found to be increased greater than tenfold in fish either infected with the virus or injected with a double stranded RNA mimic of viral infection (Rise et al., 2010). The biological significance of these findings is not clear, but it appears that sacsin levels are regulated or affected by various nervous system stressors.

SECTION 2: RESULTS

Chapter 3: Sacsin Contains a Novel Supra-Domain with Similarity to a Region of Hsp90²

In order to further understand the role of chaperone-assisted protein folding in the CNS during health and disease, we studied the function of sacsins as a potential model chaperone directly implicated in a neurodegenerative disease. Our initial evaluation of the function of sacsins involved a detailed study of its primary sequence utilizing bioinformatics tools. This analysis led to the description of a new type of large supra-domain present three times in sacsins, which we named the sacsins repeating region (SRR). Protein supra-domains are defined as recurring arrangements of two or three domains present adjacent to each other along a polypeptide chain. Such combinations have novel functions beyond those of the individual partner domains that compose them, which can exist in isolation (Vogel et al., 2004). We found that this supra-domain has a broad phylogenetic distribution, with examples in archaea, bacteria and eukarya. There is strong selective pressure for this arrangement to occur as part of very large repeated regions as well as *in cis* with a J domain. We identified multiple strategies of convergent evolution to attain such configurations.

² Portions of this chapter are reprinted from The Journal of Molecular Biology, 400 / 4, John F. Anderson, Efrain Siller, Jose M. Barral, The Sacsins Repeating Region (SRR): A Novel Hsp90-Related Supra-Domain Associated with Neurodegeneration, 665-674, Copyright (2010), with permission from Elsevier.

A. THERE IS A REPEATING SUPRA-DOMAIN IN SACSIN WITH SIMILARITY TO HSP90: SRR

The initial report identifying the *SACS* gene described two regions of similarity to Hsp90 and three regions of self-homology (Engert et al., 2000). This led us to consider the nature of the repeating unit within sacsins. Therefore, we decided to define the borders of this repeat and determine its relationship to Hsp90 utilizing bioinformatics tools. Residues with sequence similarity were mapped on the known Hsp90 structure to investigate the degree of similarity between sacsins and Hsp90 from a structural perspective.

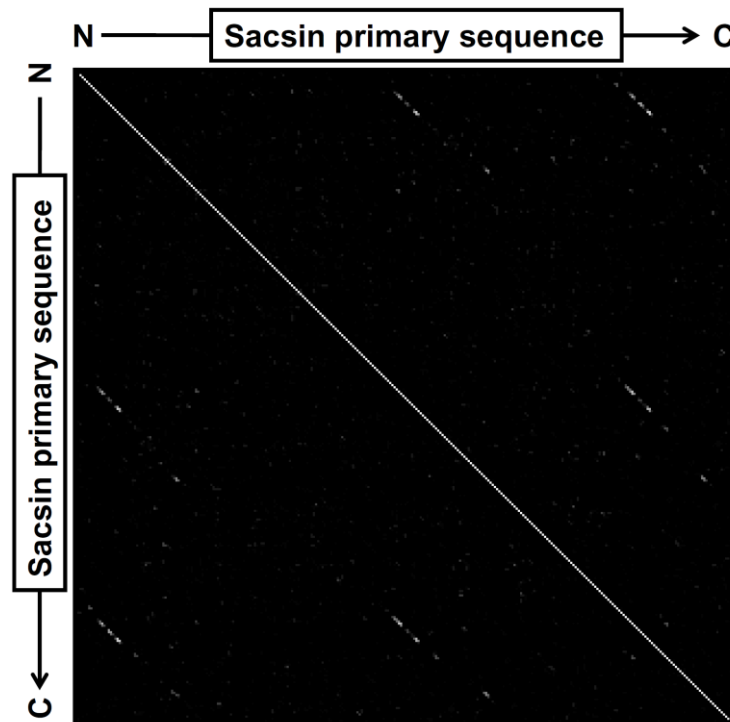


Figure 3.1: Dot-plot analysis of sacsins sequence.

The N-terminal two-thirds of sacsins were plotted on each axis and the figure was generated utilizing the dotlet algorithm (Junier and Pagni, 2000). The white diagonal line represents the identity between the two axes. The three discrete dashed white lines represent repeating regions of similarity in sacsins.

Identification of a Discrete Repeat Present Three Times in Sacsin

In order to investigate the repeat in detail, we needed to first determine its borders. The use of global multiple sequence alignment algorithms to identify the borders of similar regions requires an element of bias since the ends of the sequences are forced into alignment. Likewise, without *a priori* knowledge of the location of the borders, a local multiple sequence alignment can be misleading as well. Thus, in order to identify the borders of the repeat without bias we utilized a dot-plot analysis (Junier and Pagni, 2000).

The N-terminal two-thirds of sacsins, which contains all of the repeating elements identified in the literature, were plotted on both axes of a dot matrix utilizing the dotlet algorithm (Junier and Pagni, 2000). This program utilizes an amino acid similarity matrix to generate a graphical representation of the similarity between the sequences on the X and Y axes. A dot is placed at the intersection of similar residues. Adjusting the window size and signal-to-noise ratio allows the generation of a plot with sufficient contrast to identify the borders of similar regions. This analysis revealed the presence of three clear repeating regions separated by longer regions of little or no self-similarity between the repeats (Figure 3.1). Note the solid diagonal line stretching from the top left of the plot to the bottom right, which represents the identity of the sequences on each axis. The three

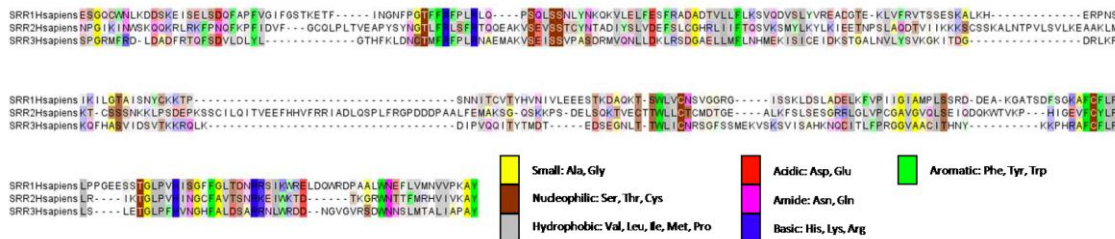


Figure 3.2: Multiple sequence alignment of human SRR domains. Multiple sequence alignment of the three SRR domains in human sacsins. Similar residues are colored according to the scheme at the bottom right.

dashed regions on each side of this solid line represent the repeating regions. Interestingly, each repeat appeared to be composed of a portion of high self-similarity on the N-terminal portion, followed by a stretch of reduced self-similarity and a clear ending point with high self-similarity.

We cross-referenced the plot with the sequence from human sacsin and were able to unambiguously identify a glycine followed by glutamine at the N-terminus of each repeat and an alanine followed by a tyrosine at the C-terminus. We defined these residues as the border of the repeat and utilized these entire regions for more detailed global alignment analysis of their similarities. A multiple sequence alignment of the three repeats from human sacsin is shown in Figure 3.2. The alignment is colored based on amino acid similarity. Consistent with the dot-plot analysis, the alignment demonstrates substantial self-similarity along the N-terminal portion and the C-terminal portion, with

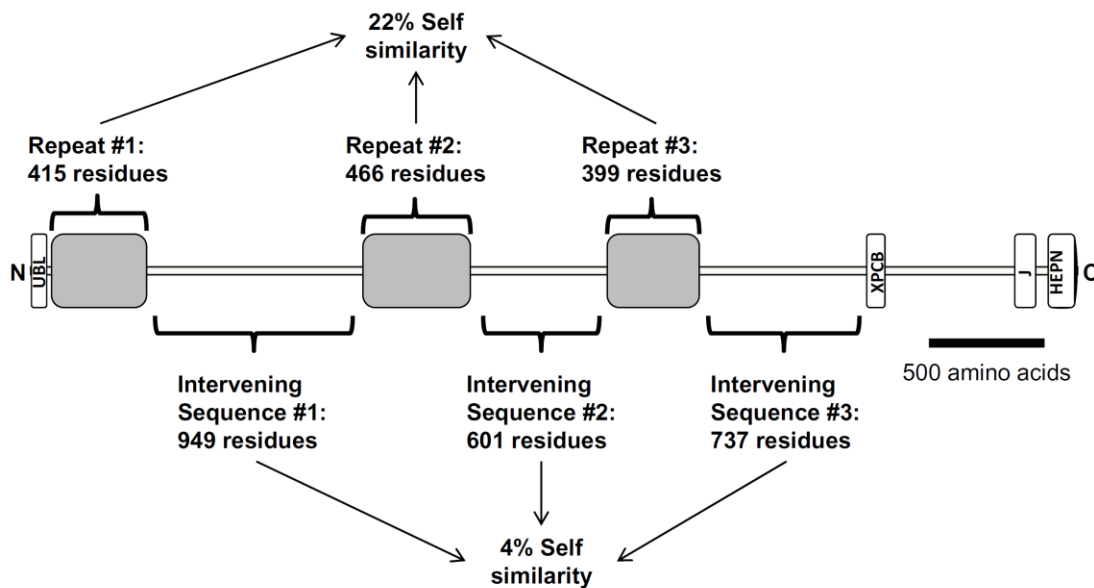


Figure 3.3: Repeat location and length in human sacsin.

The location of the repeating unit in human sacsin is marked with a gray box. The length of the repeats and the degree of self-similarity are noted. The length of the intervening sequence between repeats and the degree of self-similarity are noted.

more divergent regions in the middle of repeat. We were intrigued to find that these repeats were unusually large: in human saccin the average length was 427 residues, while the intervening sequence between the repeats was 762 residues (Figure 3.3).

Next, we examined the long intervening sequences for any additional self-similarity or identifiable domains. Interestingly, we did find weak self-similarity along these inter-repeat tracts (Intervening Sequence; Figure 3.3). However, these appear to be extremely divergent and are clearly distinct from the three discrete repeats. For example, in human saccin, an alignment of all three repeats contains 22% identity (Figures 3.2 and 3.3); whereas an alignment between all three inter-repeat tracts contains only 4% identity (Figure 3.3). This is clearly below the threshold from which one could state that a similar structure is likely formed based on sequence similarity (Abagyan and Batalov, 1997). Interestingly, the intervening sequence did not contain any identifiable domains based on extensive searches of the non-redundant (NR) sequence database and the conserved domain database (CDD). Intrinsic disorder may be a component of these long regions of intervening sequence, but this possibility has not been explored.

Large bloc gene duplication and fusion (*e.g.* of each repeat plus the ~700 residues after the repeat) is a plausible explanation for the evolution of the repeating regions in saccin. This hypothesis assumes that the three discrete regions identified in the dot plot (Figure 3.1) appear to have experienced selective pressure to retain a specific degree of conservation, while the intervening sequence did not experience such stringent constraints on its sequence. The total amino acid content of the intervening sequences is 2,287 residues. This is a substantial amount of protein to be present without identifiable domains or putative functions. The observation that these were not shortened or removed during evolution raises the possibility that a specific function is imparted by these

divergent regions. However, the lack of sequence similarity among the intervening sequences argues that the determinants for function are contained in the discrete repeating regions and it is merely the presence of a long bloc of protein separating these repeats domains that is required. The putative function of the intervening sequence is not apparent from bioinformatics analysis.

Similarity of the repeat to Hsp90

Since multiple regions of similarity to Hsp90 along the saccin sequence were described in the literature (Engert et al., 2000), we assumed these would be related to the

A.

```

Hsp90:      1  yknkeiflrelshnsdalldkiryksldpdpqletdpdlfilitkpeqkileirsgigmkkaelnnlgtiakegkafmealsagadvsmigqsgvg
Saccin repeat #3: 1  ypskkmilkelionadga-kateicfyfppchpvririfddwapiqgpalcvynagpfteddvrofnnlkgkthgpnpy-----ktgqyrgg
Hsp90:      101  gvalfladrvqvllkanddeqyvwssag-----gactvtldvne-----grgtllifll
Saccin repeat #3: 88  gnpvntcopsfllgnd--ilclfparyagatsispggmfrldadartqfsylylglthfndctsfsp

```

B.

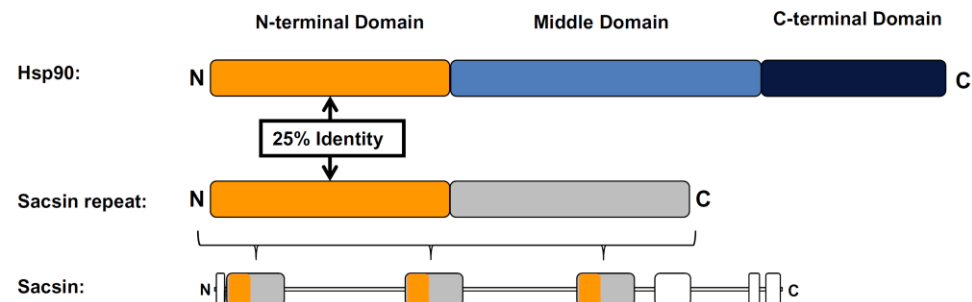


Figure 3.4: Similarity between Hsp90 and the repeat in human saccin.

(a) Pairwise sequence alignment between the cytosolic yeast Hsp90 (Hsp82) and the third repeat in human saccin. Identical residues are marked with a black square, similar with a gray one. Residues involved in nucleotide coordination in the crystal structure of yeast Hsp90 are marked with an asterisk. (b) Cartoon depiction of the regions of similarity between the repeats in human saccin and yeast Hsp90. The domains of Hsp90 are marked in different colors and labeled. The region of similarity between the repeat and Hsp90 is denoted in orange. Twenty-five percent of the amino acid residues in the N-terminal domain of yeast Hsp90 are identical to the third repeat in human saccin. The location and orientation of the repeat in full length saccin is shown on the bottom.

repeating regions we had identified. Therefore, we investigated the similarity between Hsp90 and the repeats. We performed a pair-wise alignment with the third repeat in human saccin and the *Saccharomyces cerevisiae* Hsp90 protein, whose structure is known in atomic detail (Figure 3.4a). The alignment revealed that a region of homology (~160 residues) to the entire ATP-binding domain of Hsp90 was *embedded within* the N-terminal half of each of the larger repeating regions (orange boxes in Figure 3.4b). Interestingly, there was no further similarity to the other domains in Hsp90. This observation was surprising since Hsp90 is constitutively dimerized *via* its C-terminal domain and requires an arginine residue in the middle domain to catalyze ATP hydrolysis. It is difficult to envision in what capacity an isolated N-terminal domain from Hsp90 can function without the other portions of this molecule.

In order to assign a putative function to the C-terminal region of the repeat (gray region in Figure 3.4b), we performed position specific iterated basic local alignment search tool (PSI-BLAST) searches with these regions from human saccin. Surprisingly, every homolog identified by the PSI-BLAST search contained a region of similarity to the ATP-binding domain of Hsp90 intimately connected to its N-terminus. In other words, we could not identify any sequences similar to the C-terminal portion of the repeat (~200 residues) that were not associated with a region of similarity to Hsp90 (gray boxes in Figure 3.3b). Therefore, we determined that this combination comprises a novel supra-domain which we termed a saccin repeat region (SRR).

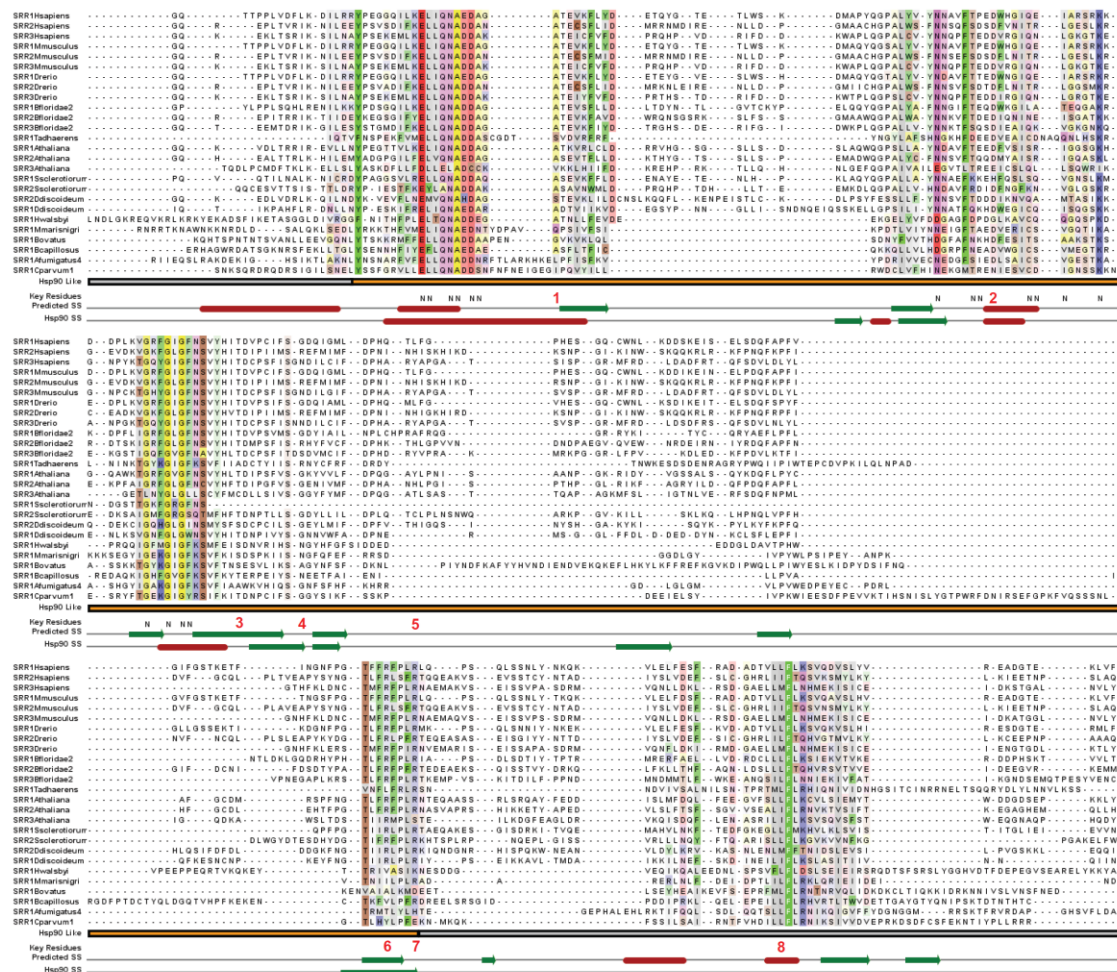


Figure 3.5: Multiple sequence alignment of representative SRR supra-domains. The region of similarity to Hsp90 is indicated by an orange bar, the novel sequence is marked with a gray bar. Residues known to coordinate nucleotide in yeast Hsp90 are marked with an “N”. Key residues in human saccin are indicated. The approximate location of predicted secondary structural elements is indicated. The approximate location of known secondary structure elements in yeast Hsp90 are indicated (Ali et al., 2006). Residues are colored according to the table provided at the lower right. Residues mutated in ARSACS are marked with a red number; the exact mutations are: 1, A2258V; 2, D168Y; 3, T201K; 4, R1575P; 5, H1587R; 6, R272C; 7, R2703C; 8, L308F; 9, P2798Q. See Table A.1 for details on these mutations. *Continued on following page.*



Figure 3.5: Multiple sequence alignment of representative SRR supra-domains. *Continued from previous page.* Accession numbers are: Hsapiens (*Homo sapiens*), NP_055178. Mmusculus (*Mus musculus*), Q9JLC8. Drerio (*Danio rerio*), XP_001921714. Bfloridae2 (*Branchiostoma floridae*), XP_002599568. Tadhaerens (*Trichoplax adhaerens*), XP_002107908. Athaliana (*Arabidopsis thaliana*), NP_197702. Ssclerotiorum (*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 fumigates*), EDP48569. Caprvum1 (*Cryptosporidium parvum*), XP_627068. Reprinted from The Journal of Molecular Biology, 400 / 4, John F. Anderson, Efrain Siller, Jose M. Barral, The Sacsin Repeating Region (SRR): A Novel Hsp90-Related Supra-Domain Associated with Neurodegeneration, 665-674, Copyright (2010), with permission from Elsevier.

We next performed secondary structure prediction on the entire SRR sequence to shed light on the putative structure of this supra-domain utilizing the Jpred algorithm (Cole et al., 2008). All SRR domains identified in the PSI-BLAST were aligned and submitted to the JPred server. The predicted secondary structure elements are shown in Figure 3.5 along with an alignment of several representative sequences which were extracted from the master alignment. The location of known secondary structure elements of the ATP-binding domain of yeast Hsp90 are also shown for reference (Ali et al., 2006). When the prediction was compared to these known secondary structure elements, most (6 out of 9) were predicted to be of the same nature (α -helix or β -strand) and are present in the same location (Figure 3.5). These results are in good agreement with the known accuracy of this algorithm (Cole et al., 2008). This analysis provides evidence that the same secondary structure elements are being formed in SRR domains and Hsp90 and thus these may form similar structures. The novel C-terminal region of the SRR domain was predicted to consist of nine β -strands and three α -helices. These predicted structural elements fall into the most highly conserved regions and many of these conserved residues are hydrophobic (highlighted in yellow, green or gray colors in Figure 3.5).

In the crystal structure of the yeast Hsp90 N-terminal domain, 17 residues are involved in nucleotide coordination (Ali et al., 2006). We located these residues in our representative multiple sequence alignment (Figure 3.5) and marked them with an “N”. We consulted the master multiple sequence alignment of all SRR domains identified in the PSI-BLAST (represented in Figure 3.5), and found that 12 of these 17 residues are well conserved (identical in >70% of all sequences) (Ali et al., 2006). The strong conservation of predicted secondary structure along with the residues involved in nucleotide coordination across such diverse organisms suggests a likely role for SRR

domains in binding and hydrolyzing ATP. Importantly, two of the non-conserved residues (a glycine and a methionine) only make backbone van der Waals contacts to the nucleotide, thus the identity of the amino acid at these positions may not be critical.

In eukaryotic Hsp90 family members, the nucleotide binding domain is linked to the middle domain *via* a divergent and highly charged linker region which is essential for ATP hydrolysis (Scheibel et al., 1999). However, no such highly charged sequences were present within SRR domains. A similar arrangement is observed in bacterial Hsp90 homologs which are also missing the charged linker. As described earlier, Hsp90 is a split ATPase and nucleotide hydrolysis necessitates a contact between the gamma phosphate of the nucleotide and an arginine residue in the middle domain (Arg380 in yeast Hsp90) (Ali et al., 2006). Therefore, in order for SRR domains to be ATPase active, the C-terminal portion would likely require a basic residue to perform the catalysis. We examined this region of the SRR domain for conserved basic residues and were able to identify an arginine close to the C-terminus of the SRR which was absolutely conserved in the SRR sequences analyzed (Figure 3.5, red arrow). We propose that this residue is involved in ATP hydrolysis.

In order to visualize the conserved residues between yeast Hsp90 and human sarsin SRR domains we took advantage of the coordinates for the yeast Hsp90 structure (Ali et al., 2006). The nucleotide pocket in Hsp90 is composed of two helices that the ribose and base wrap around and a loop that forms a lid over the triphosphate (Figure 3.6). We extracted this critical region of the structure and labeled the residues involved in nucleotide coordination that were conserved between SRR domains and yeast Hsp90 in black. This analysis revealed conservation on the side of the helices that face the nucleotide and a lack of conservation on the opposite side. Thus, these helices are

experiencing highly localized selective pressure, wherein the critical residues for coordinating nucleotide are on one face of the helix and are conserved, while the residues not involved are on the other and are not conserved. The pattern of conservation provides strong evidence that these domains have retained their capacity to bind ATP.

Significantly, nine missense mutations in the three SRR domains have been identified that lead to the ARSACS disease phenotype. In the first SRR domain (SRRa), these include an aspartic acid to tyrosine mutation (D168Y), a threonine to lysine mutations (T201K), an arginine to cysteine mutation (R272C) in the region similar to Hsp90 and a leucine to phenylalanine mutation (L308F) in the C-terminal portion. In the second SRR (SRRb), these include an arginine to proline mutation (R1575P) and a histidine to arginine mutation (H1587R), both in the region similar to Hsp90. In the third SRR domain (SRRc) these include an alanine to valine mutation (A2558V) and an arginine to cysteine mutation (R2703C) in the region similar to Hsp90 and a proline to glutamine mutation (P2798Q) in the C-terminal portion. These mutations are denoted in

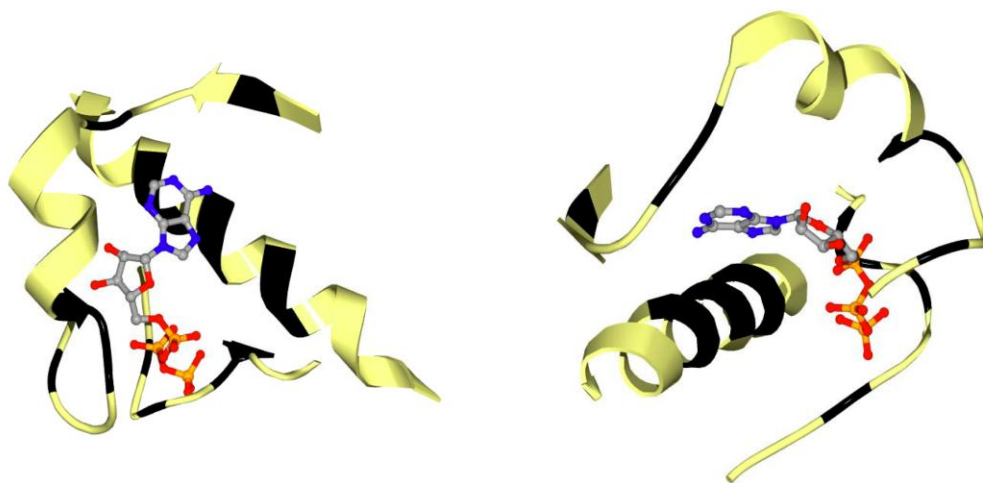


Figure 3.6: Topology of conservation from yeast Hsp90 to human SRR domains. Ribbon diagram of the nucleotide binding pocket of yeast Hsp90 with residues conserved in human sarsin SRR domains colored black (*i.e.* residues marked with an asterisk in Figure 2.4a). ATP is shown in ball and stick representation.

Figure 3.5 and are detailed in Table A.1.

B. THE PHYLOGENY OF SRR DOMAINS REVEALS MULTIPLE INSTANCES OF CONVERGENT EVOLUTION

In order to further understand the biology of SRR domains, we decided to study the origins of these complex arrangements. We identified all the SRR domain-containing proteins in the non-redundant (NR) sequence database. These proteins were analyzed to determine which other domains they might contain and their evolutionary relationships.

SRR Domains Are Present in All Kingdoms of Life

The SRR domain-containing proteins identified in the PSI-BLAST were manually curated and the other domains contained in these proteins were determined by searching the conserved domain database (CDD). This analysis revealed that SRRs can be present in a wide variety of protein contexts (Figure 3.7). Thus we decided to differentiate between "sacsin-like" proteins, which we defined as those that contain three SRR domains, one XPCB, one J, and one HEPN domains (Table A.2) and SRR-containing proteins, which we defined as any protein containing at least one SRR. We found that sacsin-like proteins are only present in animals, whereas SRR-containing proteins are widely represented across eukaryotes, bacteria and archaea. In addition to sacsin-like proteins, animals have several SRR-containing proteins with additional identifiable domains, as do all other major kingdoms of eukaryotes (Figure 3.7). Surprisingly, although this domain is very widespread, it is conspicuously absent from many organisms. For example, we could not identify any SRR domains in several model organisms, including *E. coli*, *S. cerevisiae*, *C. elegans* or *D. melanogaster*.

We identified 181 SRR domain-containing proteins with a total of 261 unique SRR domains (Table A.2). As in human sacsin, these domains may occur in repeats of 2 to 4 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

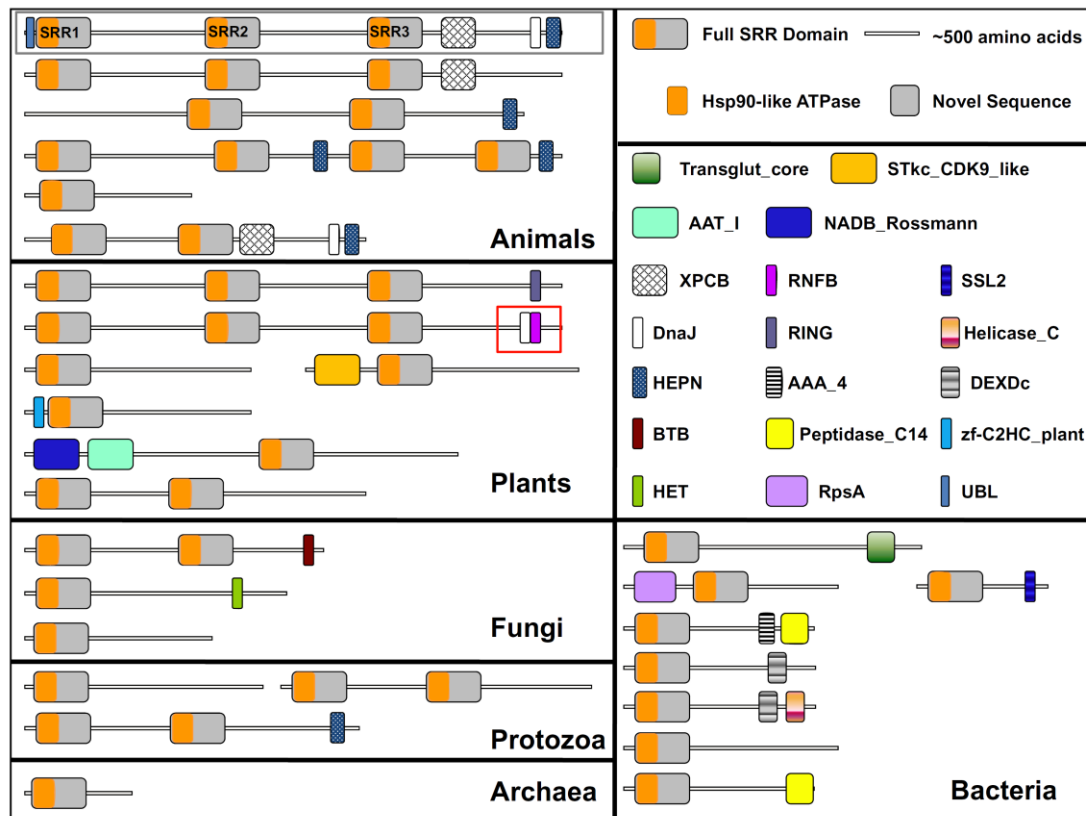


Figure 3.7: Phylogenetic distribution and domain architecture of SRR domains. The approximate location of SRR and additional domains identified by computer-based searches of the Conserved Domain Database (CDD) are indicated by colored boxes. True sacsin-like homologs are indicated by the gray outline (*e.g.* human sacsin). The J domain present in a plant SRR-containing protein is indicated by the red outline. 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. Representative CDD abbreviations, names and families for the additional domains identified are listed Table A.3. Reprinted from The Journal of Molecular Biology, 400 / 4, John F. Anderson, Efrain Siller, Jose M. Barral, The Sacsin Repeating Region (SRR): A Novel Hsp90-Related Supra-Domain Associated with Neurodegeneration, 665-674, Copyright (2010), with permission from Elsevier.

are in bacteria or archaea and all archaeal SRR domains are present in isolation (Figure 3.7). The only example of a non-repeated SRR-containing animal gene we found was in *Trichoplax adhaerens*, perhaps the simplest free living animal known (Srivastava et al., 2008). As a broad trend, more complicated organisms tend to have the SRR in a repeating fashion.

Interestingly, we identified a HEPN domain in a protozoan SRR-containing protein. This protein contains two SRR domains and a single HEPN domain and is found in the organism *Polysphondylium pallidum*. This is a slime mold found throughout the world (Kawakami and Hagiwara, 2008). This domain had originally been termed HEPN due its presence only in higher eukaryotes and prokaryotes (Grynberg et al., 2003). However, its phylogenetic distribution is apparently wider than originally suspected.

Convergent Evolution of a Repeating Arrangement of SRR Domains

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 (Figure 3.8), 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 and C as well as group D. The latter includes genes from *S. mansoni*, *H. magnipapillata* and *D. rerio* that contain only one SRR domain, as well as four genes from *T. adhaerens* that contain only one SRR domain (blue boxes in Figure 3.8).

The homolog in *T. adhaerens* was intriguing, since it possesses four genes with SRR domains, yet none of these possess SRR domains in a repeating arrangement. The SRR domains in each of these four genes are more closely related to one another than they are to any other SRR domains in the tree. We suggest that they arose from a

common ancestor prior to the development of animal proteins that possess repeating SRR domains. These genes would then represent early and distinct duplication events which never fused into a single polypeptide chain. Consistent with this hypothesis, we identified the physical location of these four genes in the genome and found that they are all relatively near to one another, within 65,000 base pairs.

Groups A, B and C represent the origins of the three SRR domains present in that order in most sacsins-like homologs. These appear to have triplicated and then fused into a single gene early on during evolution. For example, in human sacsins (gray box in Figure 3.8), there is a single SRR domain-containing gene with three repeats. The first SRR domain found in the sequence of this gene (SRR1) is found in group A, the second SRR domain (SRR2) is found in group B, and the third SRR domain (SRR3) is found in group C. The same pattern is observed for most other animal homologs that contain SRR domains. However, several organisms provide examples of even more duplication events as well as convergent evolution.

For example, zebrafish (*D. rerio*) contains two complete sacsins-like homologs. This is a rather simple case, as it appears from the tree that a primitive fused gene with three SRR domains was simply duplicated early on, as opposed to three novel duplication events that occurred independently and then fused (black boxes in Figure 3.8). This is indicated by the fact that the order of repeats in both genes fall into the appropriate group in the tree (*e.g.* both SRR1 are in group A, both SRR2 are in group B, etc.). Interestingly, there is another gene in *D. rerio* that contains a single SRR domain (Drerio_III). This domain falls into group D and is highly dissimilar to the other SRR domains found in the Drerio_I and Drerio_II genes. The significance of this finding is not clear, but it may represent an aborted attempt at gene duplication that was not fused into the standard

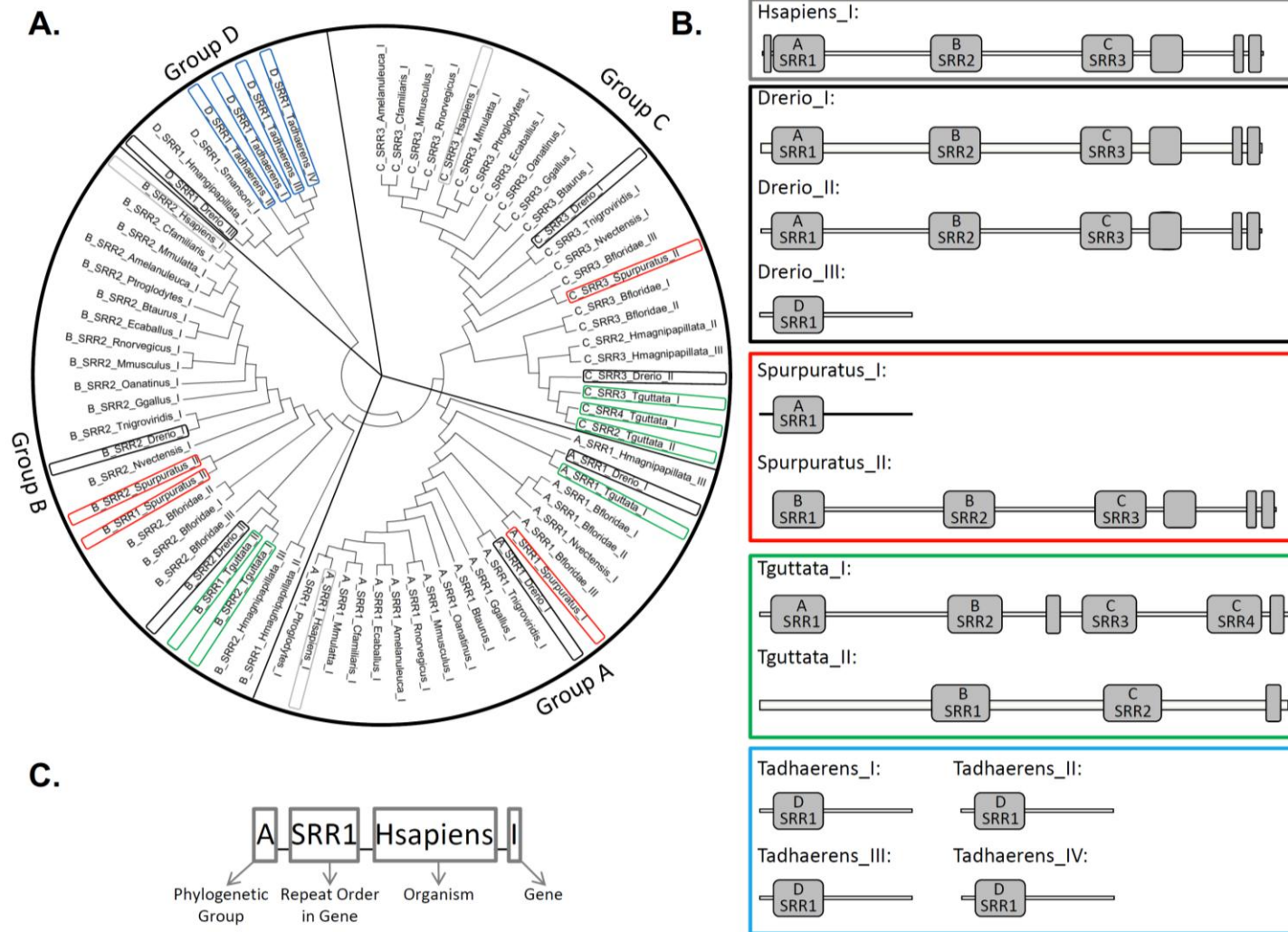


Figure 3.8: Phylogenetic relationships among animal SRR supra-domains. *Legend on the following page.*

Figure 3.8: Phylogenetic relationships among animal SRR supra-domains.

Figure on previous page. Phylogenetic tree construction was conducted *via* the neighbor-joining method in MEGA4 (Tamura et al., 2007) and is presented as the bootstrap consensus tree from 1000 replicates. Branch length was set to allow visual comparison of relationships. Particular examples discussed in the main text are marked in colored boxes. Repeat order is represented by the number immediately following the letters SRR (SRR1 for the first repeat in a protein, SRR2 for the second, *etc.*) See Table A.2 for full species names and accession numbers. Reprinted from The Journal of Molecular Biology, 400 / 4, John F. Anderson, Efrain Siller, Jose M. Barral, The Sacsin Repeating Region (SRR): A Novel Hsp90-Related Supra-Domain Associated with Neurodegeneration, 665-674, Copyright (2010), with permission from Elsevier.

three-repeat polypeptide and has since diverged into its own protein.

A more interesting example is provided by one of the sea urchin (*Strongylocentrotus purpuratus*) SRR-containing proteins (designated Spurpuratus_II in the tree). The appearance of this saccin-like homolog in this species appears to have followed a different route than in most other animals. In this organism, the SRR1 domain was not fused into the saccin-like polypeptide and exists as its own protein (designated Spurpuratus_I in the tree). However, there are still three repeats in the *S. purpuratus* saccin-like homolog (Spurpuratus_II). This likely occurred by duplication and fusion of two SRR2-like domains (belonging to group B) to become SRR1 and SRR2 of the final polypeptide (to yield a B-B-C rather than the standard A-B-C arrangement). 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 saccin homolog with three repeats *via* an alternate route than the other saccin homologs.

Further examples of selective pressure to repeat the SRR domain are present in the SRR-containing proteins of the Zebra Finch, *Taeniopygia guttata* (a species of bird). 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 (A-B-C-C arrangement). Thus, both SRR3 and SRR4 are in the 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 sacsin-like proteins, a clear propensity for SRRs to occur as three tandem repeats was also observed (Figure 3.7). We could not identify a common ancestral gene for plants and animals with three repeats, thus their three repeat proteins are likely another example of convergent evolution.

Convergent Evolution of SRR Domains in a Polypeptide with J Domains

An additional example of convergent evolution is found in one of the plant homologs and 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 the sacsin-like homologs (red box in Figure 3.7). However, this protein does not contain any of the other domains present in sacsin-like proteins, and the sequence context around the J domain is clearly different from that found in the animal sacsin-like proteins. We did not identify any other SRR-containing proteins with J domains, and thus it is likely that a convergent evolution strategy led to the placement of a J domain C-terminal to a region containing three SRRs.

C. CONCLUSIONS

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 (Vogel et al., 2004). 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 phylogenetic unit, but their constituent domains may also exist in isolation.

Through our analysis, we identified a new supra-domain: the SRR domain. These are large modules (~380 residues) in which one of the component partners (~200 residues) appears to be incapable of existing in a context other than immediately adjacent to the well-characterized Hsp90-like ATPase domain. Thus, our sequence analysis suggests that SRRs constitute a novel class of supra-domains. Previously, supra-domains have been identified as domains which *can exist in isolation* but may also co-evolve as a single phylogenetic unit (Vogel et al., 2004). However, although SRRs consist of two domains, only one appears to be able to exist in isolation: the nucleotide-binding domain of Hsp90, which can be present as an ATP-binding fold in many proteins and in many contexts (*e.g.* DNA topoisomerases and histidine kinases) (Dutta and Inouye, 2000). The additional component of each SRR consists of novel sequence which appears to be incapable of existing in isolation, but is rather intimately connected to the Hsp90 ATPase domain as a single phylogenetic element.

This domain has a strong tendency to occur in repetition. We identified multiple convergent evolution strategies as well as several unique duplication events that resulted in this arrangement. Additional evidence of convergent evolution is indicated by the separate addition of a J domain in a plant homolog. The significance of the repeating arrangement of SRR domains or their co-evolution with J domains is not immediately clear.

Chapter 4: The SRR Domain Is ATPase Active³

In order to biochemically characterize SRR domains, we cloned and purified a soluble fragment of murine saccin which contained the first SRR domain. Biochemical characterization demonstrated that this domain possesses ATPase activity. The activity appears to be a requirement for saccin function, as a disease-causing mutation leads to an alternate conformation completely incapable of hydrolyzing ATP.

A. PRODUCTION OF A SOLUBLE FRAGMENT OF SACSIN

The initial biochemical investigation of saccin function involved the putative ATPase activity of the Hsp90-like regions. We identified a human disease causing mutation in the ATPase like region of the third SRR in human saccin in a patient from Spain (R2703C) (Criscuolo et al., 2005). We located this mutation in our alignment along the third human SRR domain and yeast Hsp90 and found that this arginine corresponds to a lysine (Figure 3.5, mutation number seven). Thus, a basic residue is conserved at this position. We examined the crystal structure of yeast Hsp90 to determine its three-dimensional location. We observed that the lysine in the yeast structure forms a salt bridge with an aspartic acid at the ends of two β -strands that are situated at the core of a β -sheet. This β -sheet comprises the floor of the ATP binding pocket. We found that the aspartic acid is conserved in SRR domains, and thus we hypothesized that this salt bridge may be conserved as well. If this suggestion is correct, then the mutation of an arginine to cysteine would disrupt the positive charge required for the formation of the salt-bridge.

³Portions of this chapter are reprinted from The Journal of Molecular Biology, 400 / 4, John F. Anderson, Efrain Siller, Jose M. Barral, The Saccin Repeating Region (SRR): A Novel Hsp90-Related Supra-Domain Associated with Neurodegeneration, 665-674, Copyright (2010), with permission from Elsevier.

This would destabilize the nucleotide binding pocket, and likely result in the loss of the structural determinants required for ATP binding and hydrolysis. Thus we reasoned that there was some evidence that the third SRR domain is capable of binding ATP since it appears to possess some of the sequence determinants required for binding in yeast Hsp90.

Since we had some evidence that this SRR domain might display ATPase activity we decided it would be the best domain to utilize for our initial investigation. Therefore,

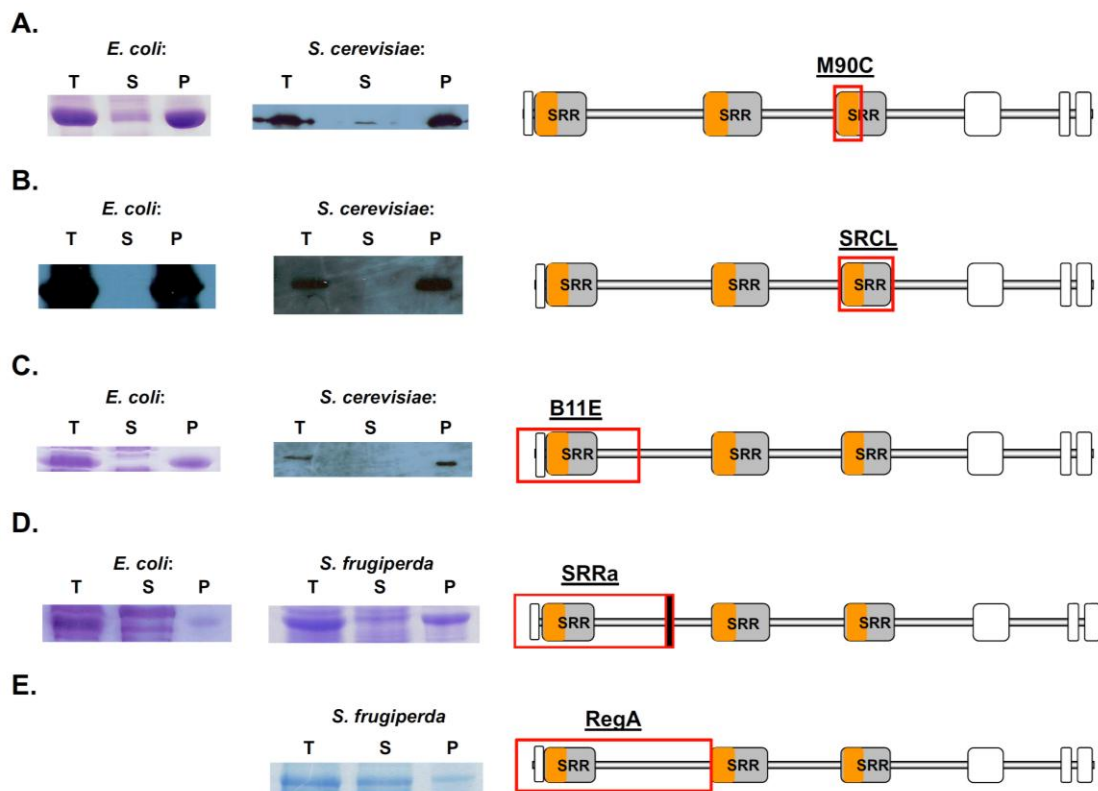


Figure 4.1: Analysis of the solubility of sacsin fragments.

The total cell extract (T) as well as the supernatant (S) and pellet (P) fractions after centrifugation for 30 minutes at 30,000 \times g are shown. The portion of sacsin included in each fragment and the name for that fragment are indicated with a red box. **(a)** M90C expressed in *E. coli* (left panel) and *S. cerevisiae* (right panel). **(b)** SRCL expressed in *E. coli* (left panel) and *S. cerevisiae* (right panel) **(c)** B11E expressed in *E. coli* (left panel) and *S. cerevisiae* (right panel). **(d)** SRRa expressed in *E. coli* (left panel) and *S. frugiperda* (right panel). **(e)** RegA expressed in *S. frugiperda* (right panel).

Since Hsp90 is a split ATPase and requires residues in the middle domain for catalysis, we reasoned that there may be structural determinants contained within the additional C-terminal sequence in the SRR which are required for proper folding (Ali et al., 2006; Meyer et al., 2003). Therefore we cloned a fragment that contains the entire third SRR in saccin (residues 2,524-2,924), which we termed SRCL. This fragment was 423 amino acid residues long and 47.9 kDa (Figure 4.1b). This construct was cloned into both pProEX for expression in *E. coli* and pESC-His for expression in *S. cerevisiae*. However, upon heterologous expression and solubility analysis, we found that there was no detectable soluble material by western blotting (Figure 4.1b). We were somewhat surprised at the extreme insolubility, since SRR domains appear to be independently evolving units found in multiple protein contexts (Figure 3.7); we had predicted that since this was an independent evolving unit, it may also be an independent folding unit.

During our analysis of the third SRR domain, several additional studies had been published which revealed the presence of human disease-causing mutations in all the SRR domains in human saccin (Table A.1). This genetic evidence indicated that each SRR was equally likely to retain catalytic activity. Thus, we decided to redirect our efforts towards the other SRR domains, since the third SRR had proven to be recalcitrant to heterologous expression. We reasoned that since the first SRR domain is the initial one to exit the ribosome, it may fold independently of the rest of the saccin molecule and thus be more amenable to heterologous expression. We generated a construct that commenced at the true N-terminus of saccin and extended well past the initial SRR domain (residues 1-698; Figure 4.1c). This would ensure that all of the folding determinants for this SRR domain would be available in the recombinant polypeptide. This fragment was termed B11E and was 725 amino acid residues long and 81.1 kDa. B11E was cloned into both

pProEX for expression in *E. coli* and pESC-His for expression in *S. cerevisiae*. However, upon heterologous expression and solubility analysis we found that there was no detectable soluble material by SDS-PAGE and western blotting, respectively (Figure 4.1c).

Since we had somewhat arbitrarily chosen the location of the stop codon for B11E, we reasoned that we may have truncated the protein at a location that is hazardous for proper folding. To determine a more rational site for the location of the stop codon, we consulted the ensemble genome browser (Kersey et al., 2009). We found that in murine saccin, there is an alternatively spliced transcript that contains all of the exons upstream of the giant exon and an additional exon downstream of the giant exon which is not present in the canonical saccin sequence. We anticipated that if this splice variant was expressed *in vivo* (as indicated by the ensemble database), then it may have a greater chance of folding during heterologous expression. This spliceoform was cloned and termed SRRa; it was 793 amino acid residues long and 88.6 kDa. The polypeptide encompasses the initial 731 residues of saccin as well as an additional 37 residues encoded by the alternatively spliced exon. We found a minor fraction of SRRa to be partially soluble (~15%) when expressed in *E. coli* (Figure 4.1d). Extensive efforts were made to purify this protein but there was no success; the soluble material was in too low abundance. Since a fraction of this protein appeared to fold in *E. coli*, we reasoned that altering the host system might be beneficial. Thus, we decided to pursue expression in a eukaryotic system.

Since we had not found *S. cerevisiae* to be advantageous compared to *E. coli* for expressing soluble fragments of saccin (Figure 4.1a-d), we decided test a different eukaryotic system. We employed an immortalized cell line, Sf9, from the fall armyworm

Spodoptera frugiperda, a species of moth. Sf9 cells have been extensively utilized to express and purify a large number of recombinant proteins, particularly recalcitrant proteins such as sacsin. For example, complexes of Hsp90 and its co-chaperones have been purified from Sf9 cells and utilized for structural studies (Vaughan et al., 2006). A modified baculovirus is the vector typically utilized to express recombinant protein in these cells. Baculovirus is a natural pathogen which infects *S. frugiperda* and utilizes the cellular machinery to express viral proteins at high abundance. The virus has been bioengineered for rapid and efficient recombinant protein production (bd biosciences). We generated a recombinant baculovirus containing the entire SRRa coding sequence and infected sf9 cells with it. However, solubility analysis of the recombinant protein indicated that no substantial improvement in solubility was achieved compared to *E. coli* (Figure 4.1d).

Next, we revisited our bioinformatics analysis of SRR domains to determine a different approach for generating soluble fragments. We had observed a strong tendency for SRR domains to occur with long regions of sequence on their C-terminal ends without identifiable domains (Figure 3.7). These were highly dissimilar, with only 4% identity in the intervening sequence regions in human sacsin (Figure 3.3). However, the amino acid content in these regions is quite large, and this long region is consistently present in SRR-containing proteins. The fact that these regions were retained during evolution across a broad range of species suggests that a selective pressure is acting on them and they may be required for proper folding and/or function of the SRR domain. We reasoned that the SRR domain plus the intervening sequence between it and the next SRR might comprise a functional unit.

Therefore, we cloned a portion of murine saccin that encompassed the N-terminus to the beginning of the second SRR domain (amino acids 1-1,456). This fragment was 1,481 amino acid residues long (including the poly-histidine tag and TEV cleavage site) and 167 kDa. Since this construct was so large, we decided to avoid *E. coli* as the expression host and utilize Sf9 cells for heterologous expression initially, because these cells are generally more amenable to production of large proteins. When we performed solubility analysis of the recombinant protein, we found that the majority was soluble (~85%; Figure 4.1e). The finding that this portion of saccin is highly soluble provides further evidence that it does indeed compose the functional unit of saccin. Previously, we had considered each individual SRR domain to be an autonomously evolving unit and likely an autonomous folding unit. However, the high degree of insolubility of the various constructs we attempted to express led us to reconsider that perhaps it is the SRR domain plus the intervening sequence that comprises a functional unit. Thus, we decided to divide saccin into functional “regions” (Figure 4.2). We named the first region, Region A (RegA, Figure 4.1e); this region extends from the N-terminus of saccin to the beginning of the second SRR domain; it contains the Ubl domain and the first SRR domain. Region B extends from the beginning of the second SRR domain to the beginning of the third SRR domain. Region C extends from the beginning of the third SRR domain to the beginning of the XPCB domain. Region D is different from the other

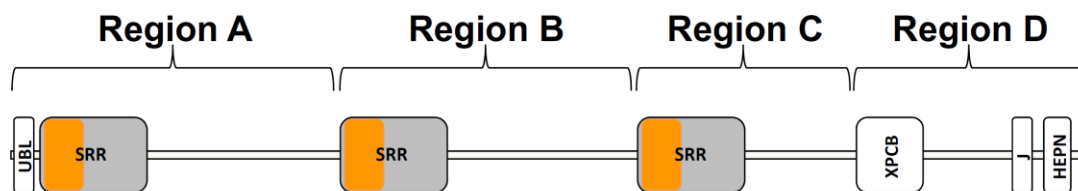


Figure 4.2: Functional regions of saccin.

Saccin can be conceptually divided into four regions as shown.

regions in that it does not contain an SRR domain. Rather, it contains the XPCB, J, and HEPN domains. This region may comprise a functional unit, as the XPCB domain may be a client interaction site and the J domain may interact with the chaperone Hsp70. Thus, Region D may act as a functional Hsp40 co-chaperone, which contains a substrate binding domain and a J domain, but in the opposite orientation compared to the canonical Hsp40.

Since we now possessed a highly soluble fragment of saccin, we wished to examine the effects of a human disease-causing mutation. We selected the D168Y mutation that was identified in the Hsp90-like region of the first SRR domain (Vermeer et al., 2008). This mutation is found in a patient from the Netherlands and leads to the same ARSACS phenotype as any other mutation. We identified the analogous residue in yeast Hsp90 and determined that it is a lysine (K86) (Ali et al., 2006). Thus, the presence of a charged side chain is conserved at this position, albeit the opposite charge. The side chain of this residue in the crystal structure appears to be solvent exposed and no

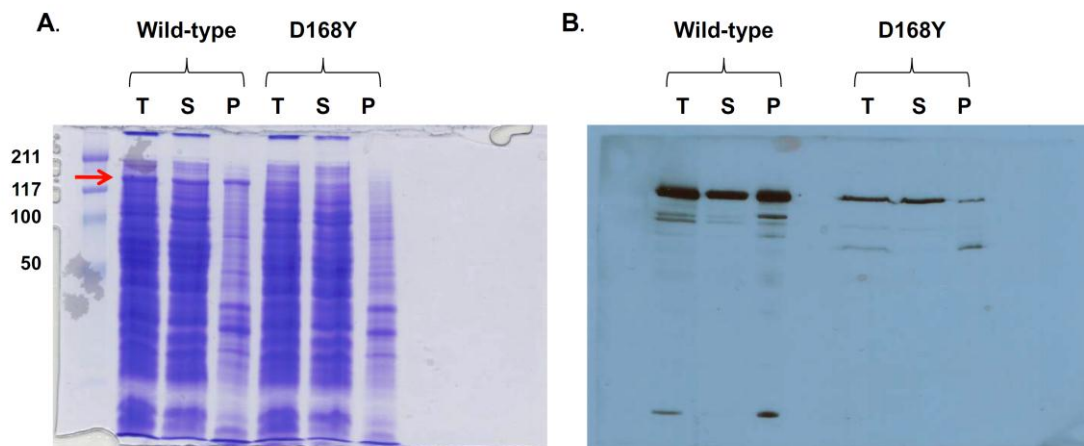


Figure 4.3: Solubility analysis of wild-type and mutant RegA.

A total cell extract (T) was prepared by sonication. The supernatant (S) and pellet (P) fractions were prepared by centrifugation at 20,000 x *g* for 20 minutes at 4 °C. **(a)** SDS-PAGE (10%) analysis of sf9 cells infected with bvRegA. The band that corresponds to RegA is marked with a red arrow. **(b)** Western blot of an identical gel as that in panel **a** probed with an antibody against the poly-histidine tag.

To test this hypothesis, we generated a version of RegA with this mutation and analyzed its solubility. After producing a recombinant baculovirus and infecting Sf9 cells, SDS-PAGE was performed and its solubility was analyzed. This demonstrated that the mutant protein is produced at much lower abundance (Figure 4.3a). This is likely due to a lower viral titer for the mutant construct. In order to definitively identify the recombinant protein and visualize its solubility relative to wild-type, we performed a western blot utilizing an antibody against the poly-histidine tag (Figure 4.3b). This analysis revealed that the mutant protein is as soluble as wild-type. In fact, it appears that it is more soluble than wild-type in Figure 4.3b; however, the film was slightly overexposed in the wild-type lanes in order to visualize the mutant protein bands. Thus, this mutation does not appear to lead to global misfolding and aggregation. A more detailed analysis of the effects of the mutation required purification of both wild-type and mutant proteins and examination of their biochemical activities.

B. PURIFICATION OF A SOLUBLE FRAGMENT OF SACSIN

RegA of Sacsin Can Be Purified by Immobilized Metal Affinity Chromatography

In order to determine the biochemical activities of sacsin, we set out to purify it from the Sf9 cells. The RegA construct included a poly-histidine tag at the N-terminus which allowed the use of immobilized metal affinity chromatography (IMAC) for purification. This purification process is based on the high affinity binding between stretches of histidine residues and a divalent metal ion which is chelated to an immobilized resin. After binding and washing, the recombinant protein can be eluted with imidazole, which competes with histidine for binding to the metal.

We performed a pilot purification with a 1 ml HisTrap nickel column (GE Healthcare). The samples were analyzed by SDS-PAGE (Figure 4.4a). A single band was

the major product in the elution fractions and this band was of the appropriate mobility for RegA. In order to confirm that this band was indeed RegA, a western blot was performed with an antibody raised against the poly-histidine tag (anti-his₆, Millipore). An identical gel was run and transferred to a nitrocellulose membrane and probed with the anti-his₆ antibody (Figure 4.4b). Immunoreactivity was prominent for the band we suspected was RegA, confirming its identity. We also confirmed its identity by mass spectroscopy: the band was excised with and submitted to the Mass Spectroscopy Core Facility (UTMB) for molecular identification, which revealed that it was indeed derived from murine saccin.

RegA Saturates IMAC Columns at Low Abundance

We achieved purification of RegA through IMAC, yet the yield was quiet low (Figure 4.4). The elution lanes were estimated to contain ~1 microgram (μg) of protein and 16 microliters (μl) of sample volume were loaded in each lane, corresponding to a

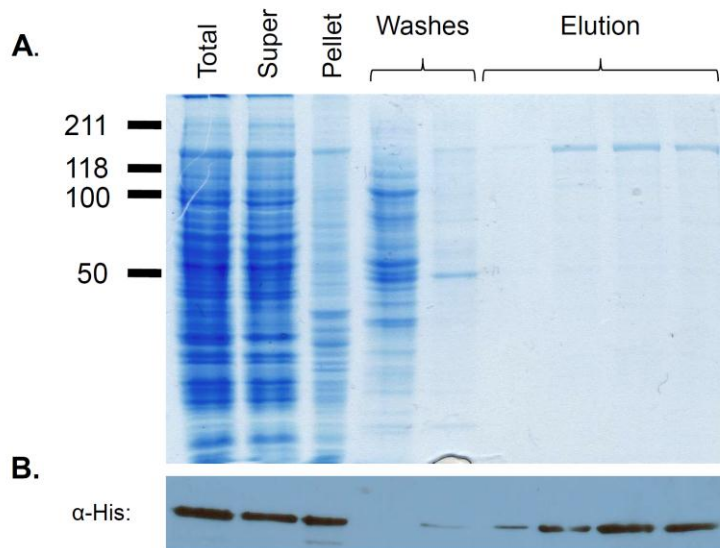


Figure 4.4: IMAC purification of RegA.

RegA was purified over a 1 ml HisTrap nickel column. **(a)** SDS-PAGE (10%) of fractions of the purification. **(b)** Western blot of an identical gel probed with an antibody against the poly-histidine tag.

concentration of 0.06 mg/ml in each fraction. The fractions were 750 μ l each for a total yield of 0.14 μ g recombinant protein. When compared with the amount of protein loaded onto the column, this represented a minor fraction. For example, the band that corresponded to RegA in the supernatant fraction (which was the input to the column) appeared to be \sim 2 μ g of protein and 8 μ l of sample were loaded for a concentration of 0.25 mg/ml in the input. A total of 20 ml of input were loaded for a total of 5 mg of protein. Thus, the recovery was less than 3% of the total recombinant protein produced in the Sf9 cells (recovered: 0.14 mg; input: 5 mg).

Subsequent analysis of another purification revealed that the majority of the material was present in the flow-through of the initial capture step and did not bind to the immobilized nickel ions on the column (Figure 4.5a). Certain poly-histidine tagged proteins possess higher affinity for metal ions other than nickel (Hemdan et al., 1989). Therefore, we attempted purification of RegA by chromatography on a column charged with a different ion. Experiments with copper and cobalt ions did not show any improvement in the affinity of RegA for the column (data not shown). Next, we considered that perhaps RegA may require more time to efficiently bind to the nickel ions. Thus, we looped the flow-through back onto the column three times; however, there was no substantial improvement in recovery (data not shown). Then, we considered that perhaps RegA would bind to the resin more efficiently with more thorough mixing. The input was directly added to the nickel resin (batch format) and mixed extensively. This manipulation did not result in any additional recovery (data not shown).

Finally, we wondered if RegA was saturating the resin. Typically 1 ml of HisTrap resin will efficiently capture \sim 40 mg of poly-histidine tagged protein, so we had not considered this a realistic possibility earlier. Even if the column was saturated at ten fold

lower levels the capacity would still be approximately equal to the protein input (input: ~5 mg). To test this hypothesis we performed a purification of RegA on both a 1 ml and a 5 ml nickel column.

A purification of RegA was performed on a 1 ml column and analyzed by SDS-PAGE (Figure 4.5a). A similar amount of RegA was obtained as in previous purifications (*e.g.* compare Figure 4.4 and Figure 4.5a), and the majority of the protein was present in

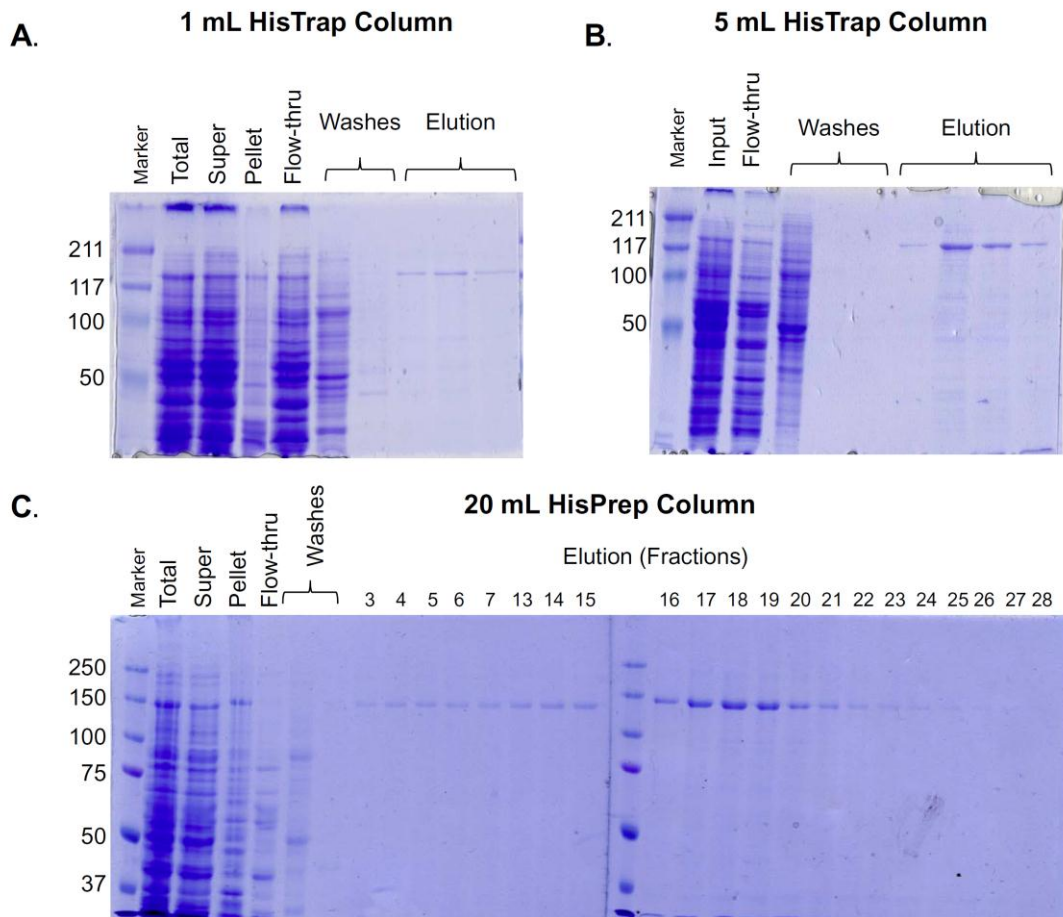


Figure 4.5: RegA saturates nickel resin at low concentrations.

(a) SDS-PAGE (10%) of purification of RegA from Sf9 cells on a 1 ml HisTrap nickel column. (b) SDS-PAGE (10%) of purification on a 5 ml HisTrap nickel column. The input for this purification is the flow-through from panel a. (c) SDS-PAGE (10%) from a purification on a 20 ml HisPrep nickel column. The fraction numbers correspond to those in Figure 4.6b.

the flow-through (compare super and flow-through in Figure 4.5a). Next, we removed the flow-through from this purification and utilized it as the input to a 5 ml HisTrap column following the same protocol as we had for the 1 ml column (Figure 4.5b). Importantly, the flow-through was reduced to almost 50% of the input, indicating that a substantial increase in the fraction-bound occurred. Additionally, the elution fractions possessed substantially more RegA from the 5 ml column compared to the 1 ml column. This indicates that RegA is saturating the resin at a very low concentration. We had recovered only about 0.14 mg of protein from the 1 ml column, corresponding to a saturation of 0.14 mg per ml of resin compared to the usual saturation of 40 mg per mL of resin with a typical poly-histidine tagged protein. Thus, RegA saturated the column at ~285 fold lower concentrations than a typical protein. We currently do not have an explanation for this unique behavior.

The improvement in yield from a 5 ml nickel column indicated that saturation of the column was a major obstacle in our purification scheme. Indeed, approximately half of the input to the 5 ml column remained in the flow-through. This was an improvement over the 1 ml column, yet represented a loss of approximately half of the recombinant protein. Thus, we attempted a purification utilizing a 20 ml HisPrep nickel column (GE Healthcare). This purification was performed exactly as with the 1 ml column, except that larger volumes were typically required for efficient washing of the column. When we utilized a 20 ml nickel column, we could not identify any RegA in the flow-through by SDS-PAGE (Figure 4.5c), indicating that the majority of the poly-histidine tagged protein was efficiently binding to the column. Consistently, we recovered substantially more purified RegA in the elution fractions from the 20 ml than either the 1 ml or the 5 ml columns.

Thus, the standard purification of RegA utilized a 20 ml HisPrep column for a single purification step. Two major drawbacks occurred with this protocol: large volumes

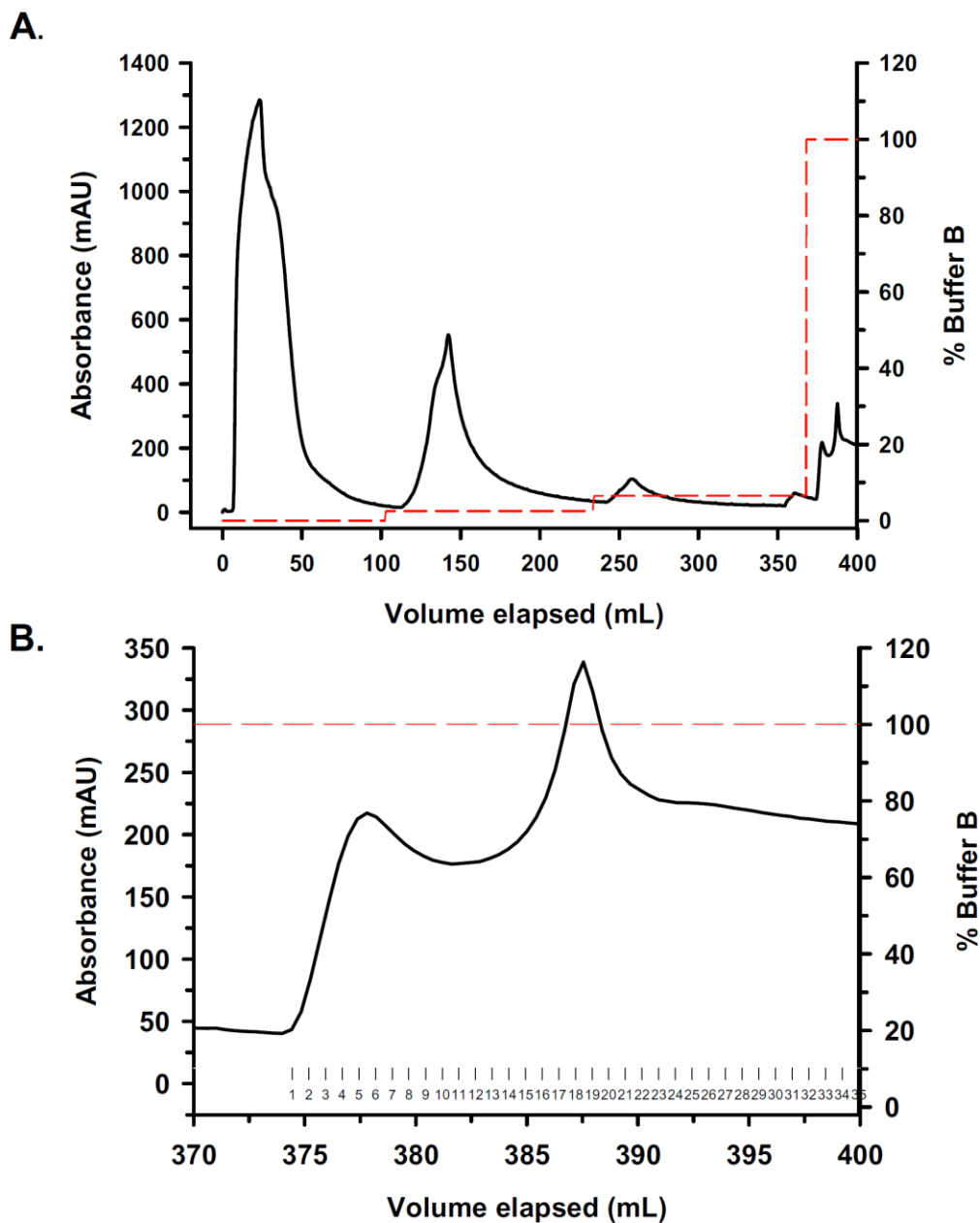


Figure 4.6: Chromatogram of RegA purification.

(a) Typical chromatogram of a purification of RegA over a 20 ml HisPrep nickel column. (b) Close-up of the elution portion of panel a. The fraction number corresponds to those in figure 4.5c.

of buffer were required and the elution profile was quite large. Since the column possessed such a large volume, there is a substantial amount of non-specific protein binding. These contaminants could be effectively removed with washing, but it required a large volume (*e.g.* more than 5 column volumes or 100 ml, see Figure 4.6a). The broad elution profile was a consequence of the large column volume as well. RegA eluted at low levels for approximately 30 fractions each 750 μ l for a total volume of ~20 ml (Figure 4.5c). The peak elution typically occurred in 4 to 6 fractions (Figures 4.5c and 4.6b). This allowed for a substantially higher yield of protein at useful concentrations relative to lower volume columns, however, there was still a substantial amount of RegA that was too dilute for typical biochemical experiments.

The final manipulation required before storing or utilizing RegA for experiments was to remove the imidazole and add a cryoprotectant. Fractions containing high concentrations of RegA were desalted utilizing a Sephadex G-25 column (GE Healthcare). Glycerol was added to the elution fraction to a final of 10% v/v and these were flash frozen in liquid nitrogen and stored at -80 °C. Concentrations were determined by estimation from an SDS-PAGE gel run against a standard curve, by absorbance at 280 nm and by the Bradford method (Bio-Rad). In every experiment (see following chapters), it was ensured that all control proteins contained identical buffer composition (including final glycerol content)

Attempts were made to concentrate these fractions. A standard ultra-filtration concentration apparatus (GE Healthcare) did not result in increased concentration of RegA. It appeared that RegA adhered to the filtration membrane. However, when Tween-20 was included at 0.5%, there was substantial recovery of concentrated RegA. Thus, this detergent may block the interaction sites on the membrane or stick to RegA and prevent it

from interacting with the membrane. However, this manipulation was not utilized for any of the experiments performed in this dissertation.

C. BIOCHEMICAL PROPERTIES OF SACSIN

RegA Migrates in the Void Volume on a Size Exclusion Column

Our initial investigation into the biophysical properties of RegA was size exclusion chromatography. In order to provide insight into the shape and oligomeric status of RegA, we utilized a Superdex 200 5/150 gel filtration column (GE Healthcare). This 3 ml column can be run with as little as 50 μ l of sample and provides high resolution for globular proteins between 10 and 600 kDa. The column was equilibrated with Buffer A and calibrated with standard proteins (GE Healthcare, Figure 4.7). The void volume is represented by dextran blue 2000 and was reached at 1.22 ml; the exclusion limit of this

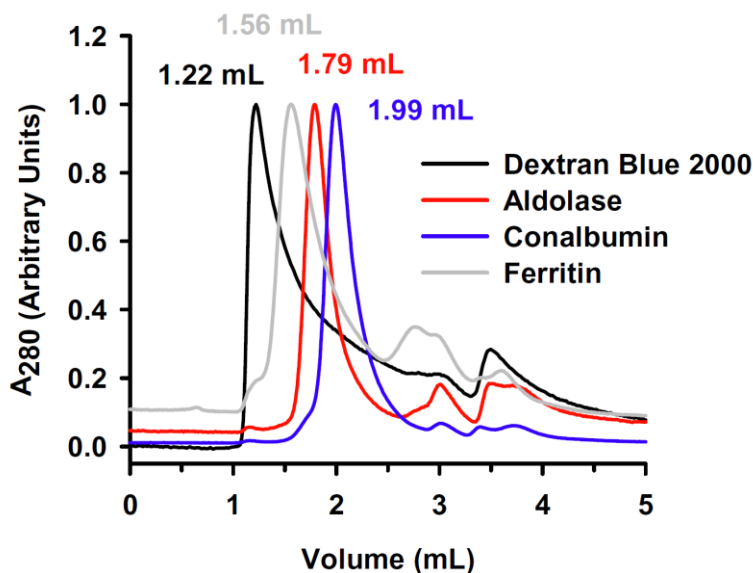


Figure 4.7: Gel filtration calibration.

The indicated molecules were each individually loaded on a Superdex 200 5/150 column. The absorption values were normalized to the maximum value in each series and plotted *versus* the volume elapsed. The volume of the peak is indicated above that peak in the corresponding color. Dextran blue 2000 represents the void volume (>1,300 kDa). The molecular weights of the other components are: Ferritin, 440 kDa; Aldolase, 158 kDa; Conalbumin, 75 kDa.

column is 1,300 kDa. The column volume at which the other standards elute is marked above their peaks in Figure 4.7.

Purified RegA of saccin was loaded onto the column in a 50 μ l sample volume. The sample was injected at 0.3 ml/min and a single peak was observed at 1.16 ml, consistent with the void volume (Figure 4.8a). A small salt peak was observed at 3.0 ml,

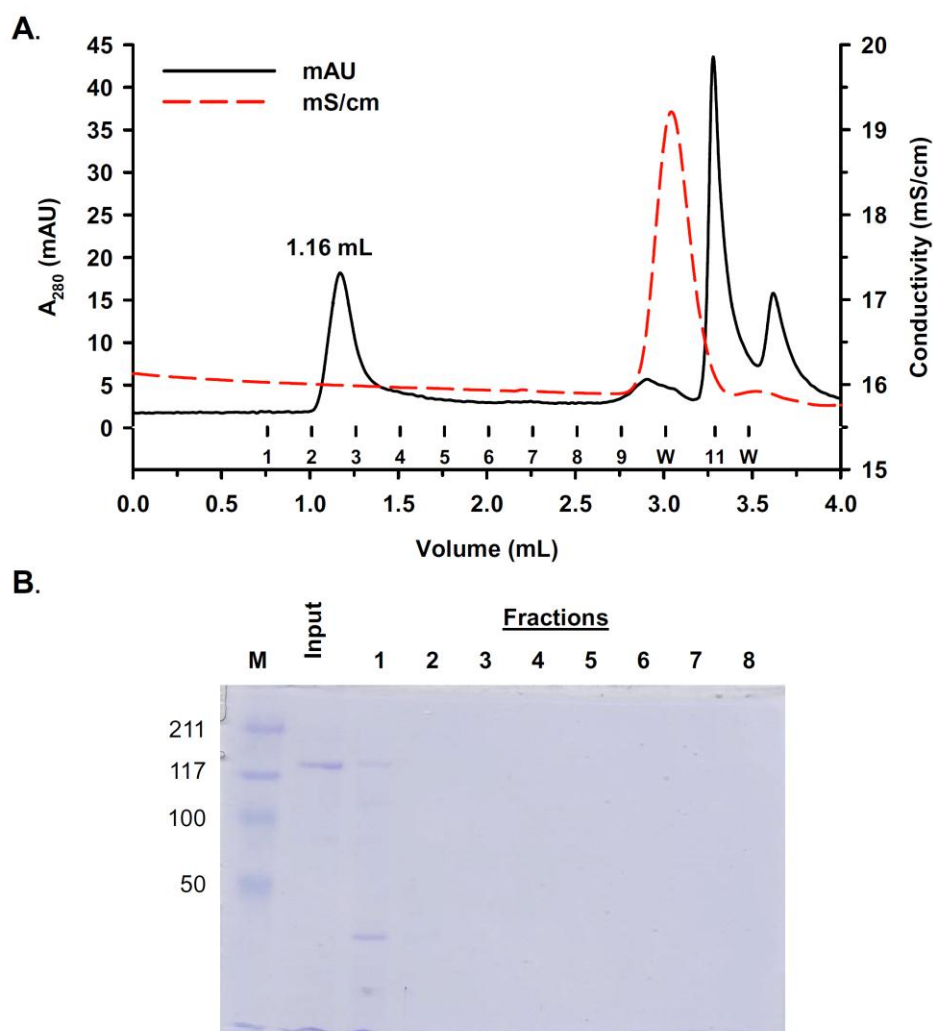


Figure 4.8: Size Exclusion Chromatography of RegA.

(a) Superdex 200 5/150 chromatography of RegA. Fractions were collected as indicated. “W” indicates fractions sent to waste. **(b)** SDS-PAGE (10%) of the input to the column as well as the fractions. The fractions (250 μ L) were precipitated in methanol/chloroform and resuspended in a volume suitable for SDS-PAGE.

corresponding to the total bed volume of the column. Fractions were collected throughout the purification (250 µl each). These were precipitated with methanol/chloroform and resuspended in a volume suitable for analysis by SDS-PAGE. The input to the column along with each fraction was separated by SDS-PAGE. This analysis revealed that all of the RegA was present in the void volume (Figure 4.8b). A low molecular weight contaminant is also present in the void volume. This may represent a proteolytic fragment of RegA as it was not observed in the input.

If RegA migrated as a globular protein, then it would need to be larger than 1,300 kDa in order to remain in the excluded volume. Since this protein is 167 kDa, it must be either oligomeric or highly elongated and flexible. The N-terminus of Hsp90 is a dimerization interface, thus it is possible that RegA may be in a dimeric state of 334 kDa. A molecule of this size would still be separated from the exclusion volume on this column if it were globular. Thus, RegA is either highly flexible/elongated or forming even larger oligomeric structures. We suggest RegA is a flexible molecule based on its similarity to Hsp90, which is a highly flexible and elongated dimer (Figure 1.5). Additionally, RegA contains a stretch of nearly 1,000 amino acid residues without identifiable domains that may not be form a globular structure (intervening sequence #1, Figure 3.3). If RegA is assuming a highly elongated and dimeric structure, then it would migrate at a much larger molecular weight than expected during size exclusion chromatography.

RegA of Sacsin Possesses ATPase Activity

Since the similar region in Hsp90 possesses ATPase activity, we wished to determine if RegA was ATPase active as well. In order to determine the potential nucleotide hydrolysis activity of RegA, we utilized the malachite green assay

(Bioassays). In the presence of inorganic phosphate, the malachite green reagent absorbs light at 595 nm. 1.0 μ M RegA was incubated with 1 mM ATP at 37 °C for 30 minutes and the presence of inorganic phosphate was determined. This experiment revealed that

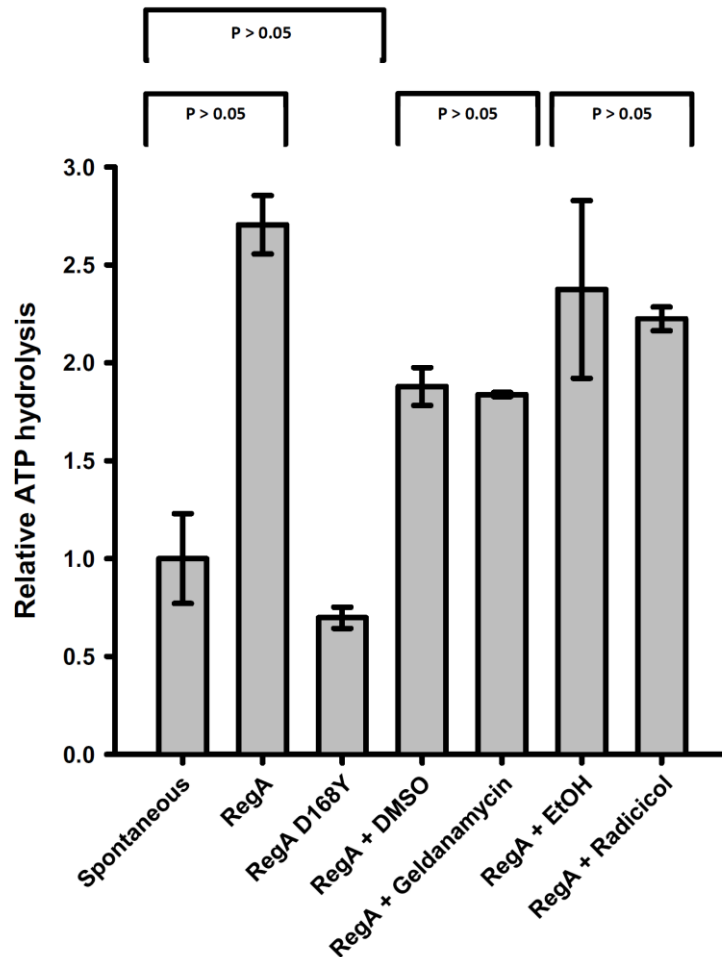


Figure 4.9: RegA of saccin is ATPase active.

ATPase activity was measured using the malachite green assay. Reactions were carried out with 1 μ M protein, 1 mM ATP, 5 mM $MgCl_2$, 30 mM Tris 7.4 and 150 mM NaCl at 37 °C for 30 minutes. Geldanamycin and radicicol were included at 60 μ M (or an equivalent amount (v/v) of DMSO or ethanol respectively). Values shown are background-subtracted and normalized to spontaneous ATPase hydrolysis.

Reactions were performed in triplicate and are shown as mean \pm SEM. A t-test was performed to determine the significance of the results. Reprinted from The Journal of Molecular Biology, 400 / 4, John F. Anderson, Efrain Siller, Jose M. Barral, The Saccin Repeating Region (SRR): A Novel Hsp90-Related Supra-Domain Associated with Neurodegeneration, 665-674, Copyright (2010), with permission from Elsevier.

A useful property of several classes of Hsp90 proteins is the susceptibility of their ATPase activity to inhibition by ansamycin antibiotics such as geldanamycin and radicicol (Grenert et al., 1997; Panaretou et al., 1998). If RegA were sensitive to these drugs, they would provide a valuable tool for studying its function. Therefore, we performed another ATPase assay in the presence of these drugs and their solvents. We found that RegA is not inhibitable by either of these drugs (Figure 4.9). While unusual, there is at least one example of an Hsp90 molecule that is similarly insensitive to geldanamycin. Careful experiments by David and colleagues revealed that the cytosolic *C. elegans* Hsp90 is not affected by this drug (David et al., 2003). These findings indicate that this saccin SRR domain has retained its capacity to hydrolyze ATP, however it possesses sufficient structural divergence that it is no longer sensitive to these drugs, which may be of importance in the design of studies of saccin function.

Since our initial hypothesis that the D168Y human disease-causing mutation was a folding mutant proved to be incorrect (Figure 4.3), we decided to assess whether the presence of this mutation had any measurable effect on the ability of this protein to hydrolyze ATP. Our previous experiments had revealed that when over-produced in Sf9 cells, the mutant protein remains soluble (Figure 4.3). Thus, we were able to purify this mutant in the same manner as wild-type. We assessed its capacity to function as an ATPase utilizing the malachite green assay. We found that the D168Y mutation completely abrogated the ability of this protein to hydrolyze ATP (Figure 4.9). This result served as a valuable control for the activity demonstrated by 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. Thus, ATP hydrolysis is a specific function of

the SRR domain that is critical for its function. Our results suggest that this mutation leads to loss of a specific biochemical function of one of the SRR domains in sacsin. Since the disease displays a clear recessive phenotype, the fact that the other two SRR domains are unable to confer function suggests that the activities of SRR domains are not simply redundant. Thus it is likely that they may function in a synergistic or cooperative fashion.

Wild-Type and Mutant RegA Are Spectroscopically Different

Since the mutant SRR is not globally misfolded, yet is not ATPase active, we decided to determine if we could detect more subtle conformational differences that may lead to a loss of nucleotide binding/hydrolysis. We performed fluorescence

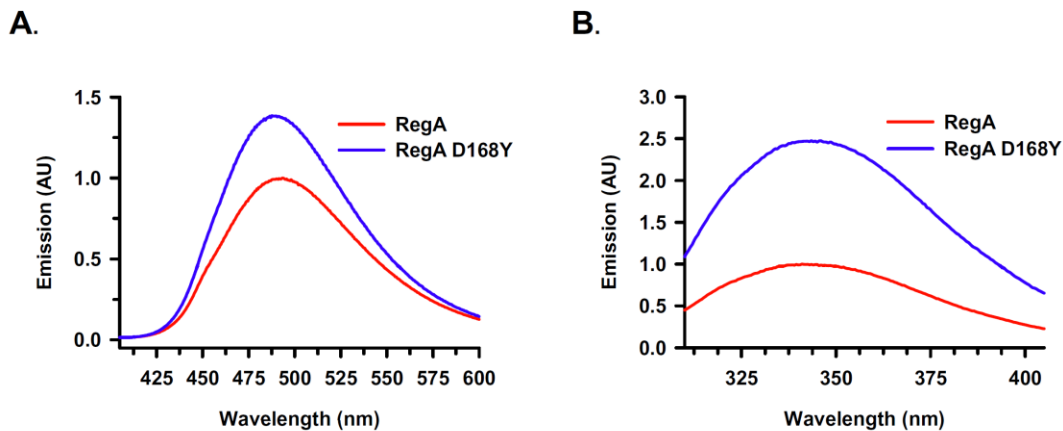


Figure 4.10: Spectroscopic effects of a human mutation in RegA.

(a) The emission spectrum of *bis*-ANS was determined in the presence of either wild type or mutant RegA. 100 nM protein was combined with 3 μ M *bis*-ANS in 30 mM Tris 7.4 and 150 mM NaCl. The samples were excited at 390 and emission spectra recorded from 400 to 600 nm. The background spectrum of *bis*-ANS was subtracted.

(b) Tryptophan emission spectra of wild type and mutant SRR were obtained with 1 μ M protein in 30 mM Tris 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. Reprinted from The Journal of Molecular Biology, 400 / 4, John F. Anderson, Efrain Siller, Jose M. Barral, The Sacsin Repeating Region (SRR): A Novel Hsp90-Related Supra-Domain Associated with Neurodegeneration, 665-674, Copyright (2010), with permission from Elsevier.

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*-ANS fluorescence (Hawe et al., 2008). Interestingly, there was a nearly 50% increase in *bis*-ANS emission with mutant sacsin compared to wild-type (Figure 4.10a). This may reflect the alteration of surface charge with the mutation (loss of a negative charge and replacement with a hydrophobic residue) or it may reflect a conformational shift in the protein. In order to differentiate between these two possibilities, we determined the tryptophan fluorescence spectra of the two proteins (Figure 4.10b). The excitation wavelength was set to 295 nm to avoid exciting tyrosine residues, allowing comparison between the two constructs (since the mutant has an additional tyrosine). This spectrum revealed a ~2.5-fold increase in fluorescence for the mutant, indicating that the local environment of at least some of the tryptophan residues is markedly different. Circular dichroism spectroscopy would be valuable in determining alterations in secondary structure in the mutant protein. However, technical challenges limit our ability to purify abundant recombinant protein and to concentrate the protein we have purified. Thus, the requirement of this technique for more highly concentrated protein limits its usefulness in our application.

We suggest that the mutation leads to an alternate conformation, as sensed by the intrinsic tryptophan fluorescence changes which likely exposes a hydrophobic patch, as indicated by the increased *bis*-ANS binding. Thus, this abnormal conformation is able to fold sufficiently to remain soluble, but is not competent to hydrolyze ATP. These findings suggest that ATP hydrolysis is necessary for sacsin function, as this mutation leads to essentially the same clinical phenotype as genomic deletions of the protein that do not allow the production of any sacsin protein (Breckpot et al., 2008).

Chapter 5: Sacsin Is Both a Chaperone and a Co-Chaperone

Sacsin possesses regions with similarity to both molecular chaperones and co-chaperones. Three SRR domains are present which contain similarity to the nucleotide binding domain of the chaperone Hsp90 (Chapter 3). We found that at least one of these domains is ATPase active (Chapter 4). Additionally, there is a segment of similarity to the J domain of the co-chaperone Hsp40 near the C-terminus (Chapter 2). Based on this information, we hypothesized that sacsins may be a novel molecular chaperone with built-in co-chaperone modules. We analyzed the potential of RegA of sacsins to act as a chaperone and found that indeed it can assist in the folding of a model protein. Likewise, we tested the J domain for co-chaperone activity and found it to be active as well.

A. SACSIN IS A CHAPERONE

Within each SRR, there is a region of similarity to the N-terminal domain of Hsp90. In addition to displaying ATPase activity, the N-terminus of Hsp90 has been demonstrated to possess chaperone activity independently of the rest of the molecule (Scheibel et al., 1998; Young et al., 1997) and we found that RegA of sacsins demonstrated ATPase activity (Chapter 4). The presence of an ATPase-active region with similarity to a portion of Hsp90 in this SRR led us to consider whether these domains might constitute chaperone modules.

RegA Acts as a Chaperone Towards Firefly Luciferase

In order to directly examine the chaperone activity of sacsins, we utilized purified RegA and the model client firefly luciferase (FLuc). FLuc has been extensively utilized to characterize protein folding properties of multiple chaperones, including Hsp90 and Hsp70/DnaK (Szabo et al., 1994; Yonehara et al., 1996). FLuc is the enzyme responsible

for the luminescence in the tail of the firefly. This enzyme catalyzes oxygen and ATP-dependent reactions that result in the adenylation of luciferin to adeny-luciferin followed by the oxidation of adeny-luciferin to oxyluciferin with the concomitant emission of a photon at 562 nm (Viviani, 2002).

FLuc is a valuable model client for chaperone studies due to its high propensity to aggregate when diluted from denaturant, its strong dependence on chaperones for efficient refolding (particularly the Hsp70/Hsp40 system) and the availability of a

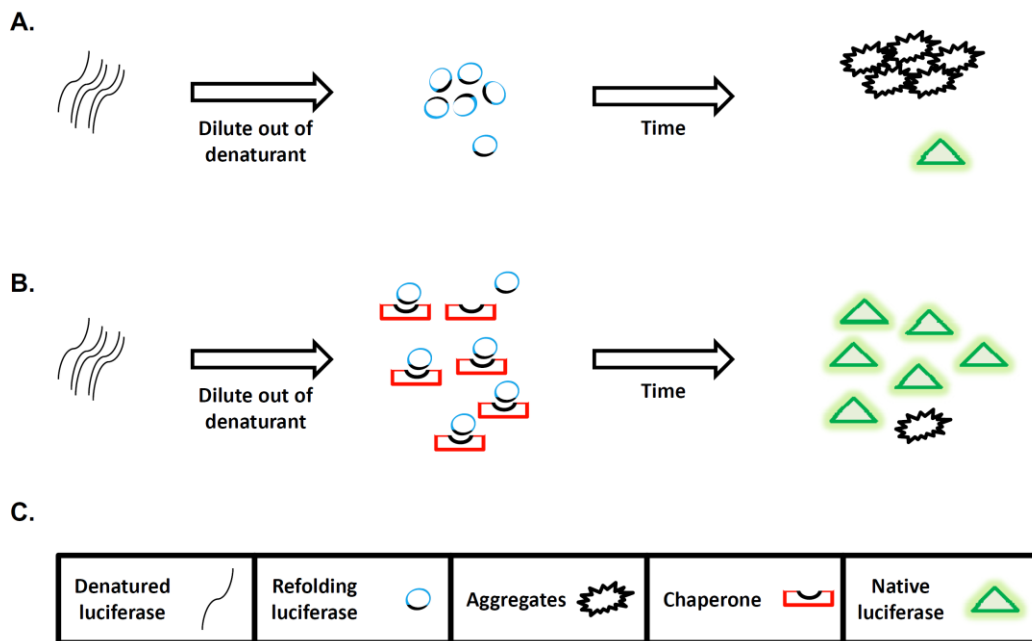


Figure 5.1: Scheme for chaperone-assisted refolding of FLuc.

(a) Scheme for unsupplemented refolding of FLuc. FLuc is unfolded in a strong denaturant. Refolding is initiated by a 100-fold dilution of the denaturant. As folding proceeds, FLuc exposes hydrophobic surfaces (represented as black regions on the blue circles) which are prone to aggregation. Over time, these self-associate forming misfolded aggregates with little protein folding to the native state. **(b)** Scheme for chaperone-assisted refolding of FLuc. Refolding is initiated as in panel **a**. When a chaperone is present it can bind through hydrophobic regions (represented as a black region in the chaperone) to the client. This prevents the self-association and aggregation of the refolding FLuc. Over time, the majority of molecules are able to refold with aggregation as only a minor side reaction. **(c)** The symbols used in the figure are defined.

sensitive and convenient assay of acquisition of the native state (bioluminescence) (Szabo et al., 1994). The experimental paradigm for a typical refolding assay is demonstrated schematically in Figure 5.1. FLuc is denatured in 6.0 M guanidinium chloride (GdmCl) followed by dilution out of denaturant into mixtures containing various purified components. In the presence of 6.0 M GdmCl, FLuc is largely unfolded, which is represented as curved lines; upon a 100-fold dilution, the GdmCl concentration is reduced to 60 mM, an environment compatible with refolding. Folding can be monitored as a function of time by determining the luminescence of the refolding reaction, since only natively folded FLuc molecules are competent to perform catalysis. Dilution of FLuc from denaturant into a buffer without chaperones results in the aggregation of the majority of the molecules (Figure 5.1a) (Herbst et al., 1998). The FLuc molecules in the process of folding are represented as blue circles; these species tend to expose hydrophobic patches, represented as a black portion of the blue circle. The hydrophobic patches may self-associate leading to aggregation (Figure 5.1a). However, If certain chaperones are included in the refolding reaction, then the aggregation will be diminished and the majority of FLuc molecules will be allowed to reach the native state (Figure 5.1b). The chaperone is represented as a red square with a region of hydrophobic surface area represented in black. This is represented as complementary to the hydrophobic patches on the refolding FLuc molecules since it is the site where the chaperone binds to the client. Thus, the chaperone can transiently and productively interact with the hydrophobic portions of the client, decreasing aggregation and enhancing yield of the native protein.

There are two important controls for FLuc refolding assays: the spontaneous FLuc refolding curve (negative control) and the bacterial Hsp70/Hsp40 (DnaK/DnaJ/GrpE)-

assisted refolding curve (positive control) (Szabo et al., 1994). Refolding was initiated by 100-fold dilution of FLuc from a buffer containing 6.0 M GdmCl. At various time points an aliquot of the refolding reaction was removed and the luminescence of the sample was determined in a luminometer (Sirius). This value was plotted as a percent of an equivalent concentration of native FLuc (which was never denatured) luminescence versus time. Less than 10% of FLuc molecules refolded in the absence of chaperones (Figure 5.2). However, when diluted into a buffer containing DnaK/DnaJ/GrpE (K/J/E) and ATP nearly all of the FLuc molecules refold (Figure 5.2). These values are consistent with the reported efficiency of FLuc folding under spontaneous conditions and chaperone-assisted conditions (Szabo et al., 1994).

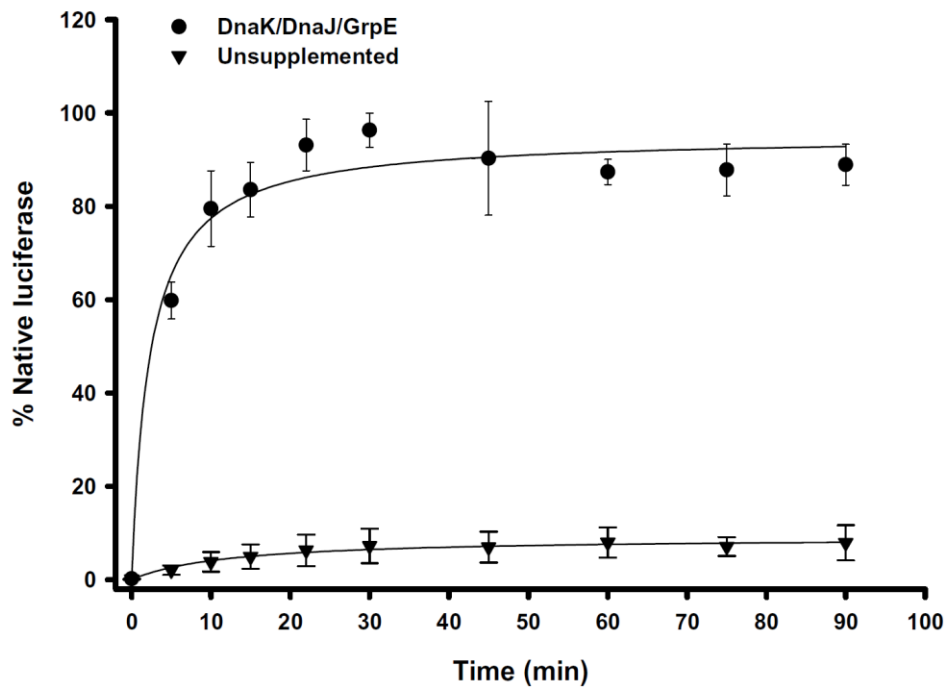


Figure 5.2 Chaperone-assisted refolding of FLuc.

FLuc was denatured in 6.0 M GdmCl and refolding was initiated by 100-fold dilution. In the unsupplemented reaction, <10% of the molecules refolded. In a reaction supplemented with DnaK/DnaJ/GrpE (10 μ M/2 μ M/6 μ M respectively) and ATP, the majority (>90%) of FLuc molecules refolded.

In order to assess the chaperone activity of RegA, we set up a similar reaction to those of Figure 5.2 with RegA in the refolding buffer as the putative chaperone. FLuc was denatured at a concentration of 5 μ M, followed by 100-fold dilution into buffer to a final concentration of 50 nM. RegA was included at 500 nM; these concentrations were chosen so RegA would be in a 10-fold molar excess to FLuc. In the absence of other

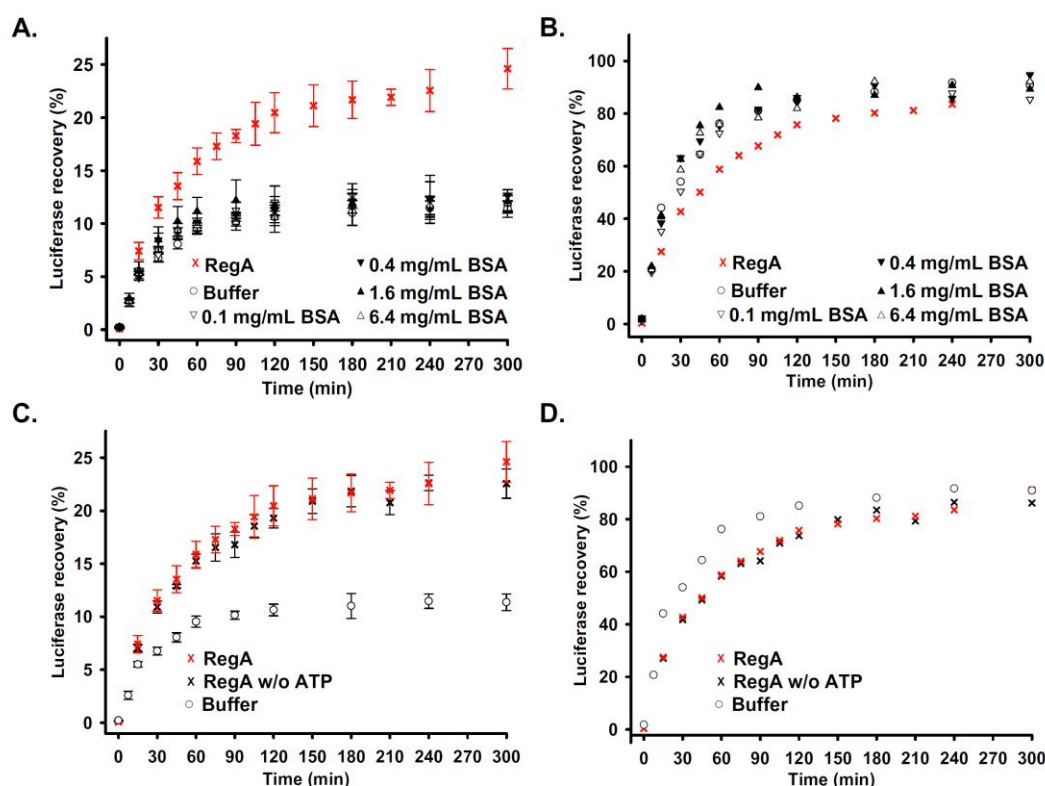


Figure 5.3: RegA of saccin is a chaperone.

(a) Plot depicting the amount of FLuc (50 nM final) refolded upon 1:100 dilution from GdmCl in the presence of 500 nM RegA or the indicated concentrations of BSA, as determined by the activity of an equivalent concentration of native luciferase. Error bars are the mean and standard deviation of three independent experiments. (b) Plot depicting the data in panel a normalized as follows: each series of data points were fit to a hyperbolic equation and the value of each data point was divided by the maximum value of the fit for that series. (c) Plot depicting FLuc refolding reactions performed as in panel a in the presence of RegA without ATP. RegA with ATP and buffer control are plotted for reference. (d) Plot depicting the data in panel c, normalized as indicated for panel b.

components, FLuc recovered ~12% of its activity (Figures 5.3a and 5.4a). However, when diluted into buffer supplemented with the 10-fold molar excess of RegA, the fraction of FLuc that acquired the native state increased approximately two-fold. Compared to the effect of K/J/E this is quite modest; however, this increase was significant, reproducible, and unlikely to be the result of non-specific binding, as it was not observed when reactions were supplemented with excess bovine serum albumin (BSA) (Figures 5.3a and 5.4a). BSA is a stringent control for chaperone assays due to its propensity to bind to proteins through its abundant hydrophobic surface patches (Freeman and Morimoto, 1996; Jakob et al., 1995) and was incapable of assisting in the refolding of FLuc even when present at a 64-fold excess by mass to RegA (~190-fold molar excess).

Since many chaperones undergo structural transitions which are regulated by ATP binding and hydrolysis we wished to determine whether the observed chaperone activity

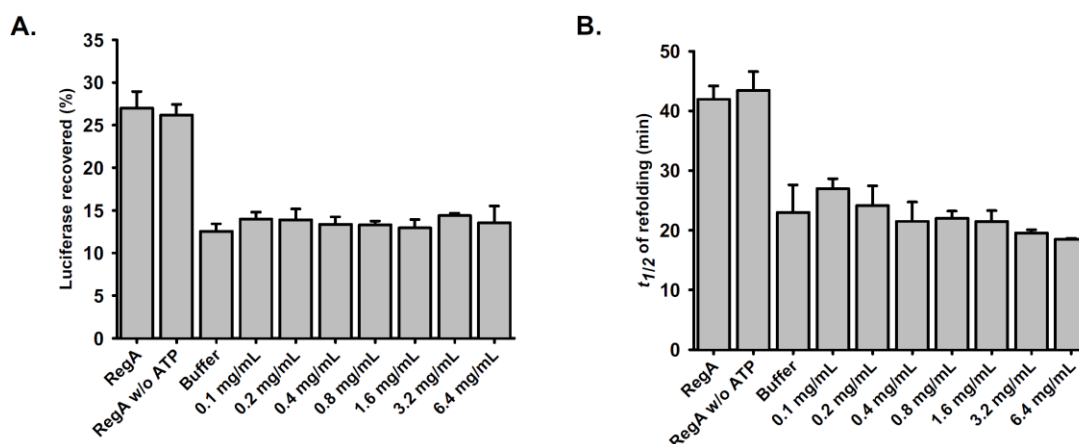


Figure 5.4: Quantification of the chaperone activity of RegA.

(a) Histogram summarizing the maximum percent of FLuc recovered in the various experiments shown in Figure 5.3, plotted as the mean and standard deviation as determined from a fit to a hyperbolic equation. (b) Histogram summarizing the $t_{1/2}$ values of FLuc refolding reactions in the various experiments shown in Figure 5.3, plotted as the mean and standard deviation as determined from the fit to a hyperbolic equation.

of RegA was similarly regulated as we had found that RegA possessed ATPase activity (Chapter 4). Therefore, we compared refolding reactions carried out in the presence or absence of ATP in the refolding buffer. We found that the refolding curves were essentially super-imposable, indicating that this activity of saccin is not regulated by ATP (Figures 5.3c and 5.4a).

In addition to enhancing the yield of refolded protein, molecular chaperones may also delay the velocity of a refolding reaction, due to the iterative binding and release of unfolded and/or partially folded intermediates (Agashe et al., 2004). Thus, we examined our results in terms of refolding rates in more detail. We plotted our data as the percent of the maximum luciferase refolded as a function of time, in order to visualize the relative refolding rates (Figures 5.3b and 5.3d). A qualitative examination of the curves revealed that the reactions containing RegA were substantially slower than those with excess BSA or buffer alone. In order to obtain a quantitative approximation of the refolding rates, we fit our data to a hyperbolic equation. Although this is clearly an oversimplification of the refolding process of FLuc (Herbst et al., 1998), it allowed us to compare across reactions the times required to achieve 50% of the refolded total, a term we are calling here $t_{1/2}$. The total yields of refolded luciferase and their $t_{1/2}$ are plotted in Figures 5.4a and 5.4b. These results demonstrate that there is a clear increase in yield (~12% *versus* ~25%) and a substantial delay in FLuc refolding ($t_{1/2}$ of 20 *versus* 40 min) in the presence of RegA, neither of which are affected by excess BSA. The observed delay in rate and increase in yield of refolded FLuc in the presence of RegA indicate that this region of saccin possesses chaperone activity.

RegA Prevents Luciferase from Aggregating During Refolding

Since K/J/E can assist ~90% of FLuc molecules to refold whereas RegA can only assist ~25%, we decided to investigate the fate of the fraction of FLuc that did not regain its activity (*i.e.*, the remaining ~75% in Figure 5.3a). In order to determine whether this fraction aggregated or was maintained in a soluble state, we initiated refolding experiments as in Figure 5.3 and analyzed the solubility of FLuc by centrifugation. Thirty minutes after FLuc was diluted out of denaturant, aliquots were removed and the material was separated into soluble (super) and insoluble (pellet) fractions. An aliquot of the total mixture was removed for reference (total). In order to ensure that all samples (in triplicate) were processed in an identical manner, all fractions were spotted onto a single nitrocellulose membrane, immunoblotted for FLuc simultaneously and quantified by densitometry. The majority of FLuc diluted into buffer became insoluble after 30 minutes of incubation (Figure 5.5), consistent with its strong chaperone dependence. FLuc diluted into buffer containing excess BSA remained partially soluble (~50 – 60%) (Figure 5.5), which was not unexpected, as BSA exhibits a high propensity to bind non-specifically *via* hydrophobic interactions to numerous ligands, including partially folded proteins (Varshney et al., 2010). Importantly, the majority (~80%) of FLuc molecules in the RegA-supplemented reaction remained in the soluble fraction (Figure 5.5), indicating that RegA is capable of interacting with unfolded and partially folded FLuc chains and preventing their aggregation. Presence or absence of ATP in the refolding mixtures did not result in any significant changes in the behaviors observed (Figure 5.5), indicating that the prevention of FLuc aggregation by RegA is not ATP dependent.

RegA Maintains Luciferase in a Folding-Competent State

Our findings above that RegA only moderately assists the refolding of FLuc, yet

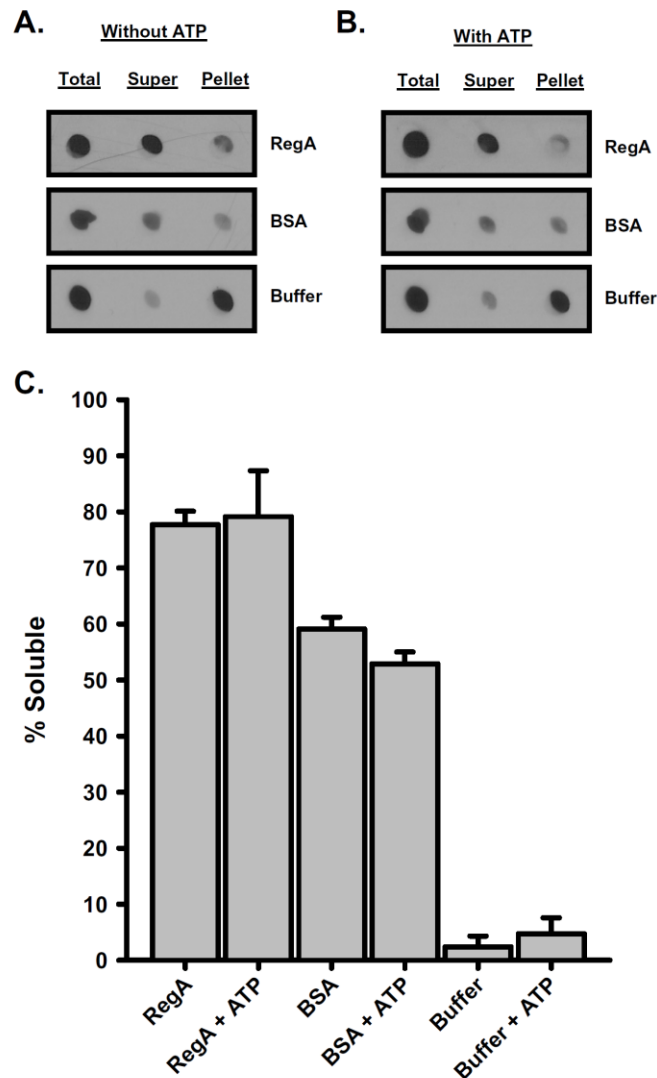


Figure 5.5: RegA of saccin maintains FLuc soluble.

(a) Dot-blot analysis of the amount of material present in the total, soluble (super) and insoluble (pellet) fractions of FLuc refolding reactions after 30 minutes of incubation and subsequent separation by centrifugation (20,000 x g for 30 minutes) and immunoblotting (see materials and methods). (b) Dot-blot analysis as in a, but refolding reactions were carried out in the presence of 3 mM ATP. (c) Histogram displaying the quantification of soluble material as determined by densitometry and plotted as the mean and standard deviation of three independent experiments.

is highly capable of maintaining it in a soluble state led us to ask whether the main function of this region of sarsin is to maintain partially folded clients in a folding competent state and subsequently deliver them to other regions within this large molecule (*e.g.*, more C-terminal SRRs) and/or to additional chaperone systems. This type of chaperone activity has been referred to as “holdase” activity, since the client is held in a folding competent state and then released to other chaperone systems for efficient

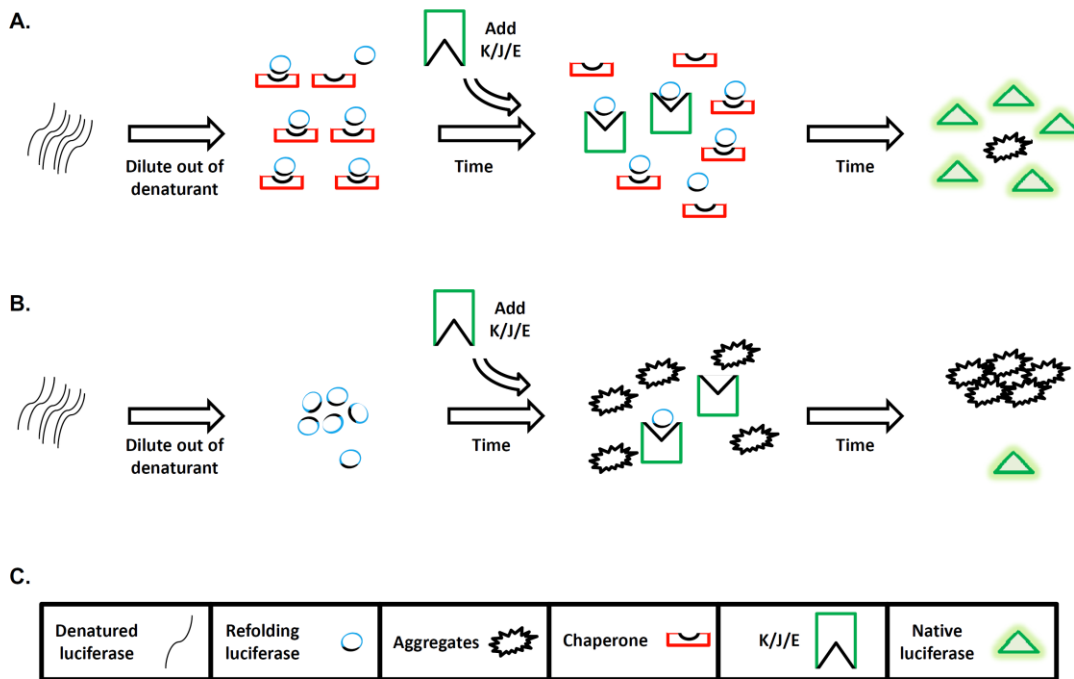


Figure 5.6: Scheme for holdase experiments.

(a) Scheme for chaperone-assisted holdase experiments. FLuc is denatured and refolded by dilution from denaturant into the presence of a chaperone with holdase activity. This chaperone will bind to the refolding FLuc and prevent aggregation of the client and hold it in a folding competent state. After a pre-determined amount of time, the DnaK/DnaJ/GrpE chaperone system is added to the reaction. This chaperone system will efficiently capture and allow the rapid refolding of FLuc. **(b)** Scheme for unsupplemented holdase experiments. FLuc is denatured and refolded as in panel **a**, but in the absence of a chaperone. After a pre-determined amount of time, the DnaK/DnaJ/GrpE chaperone system is added to the reaction. In this situation, this chaperone system is not able to efficiently enhance the refolding of FLuc, since most of the molecules have aggregated and are no longer folding competent. **(c)** The symbols used in the figure are defined.

refolding. Hsp90 has been demonstrated to possess this type of chaperone activity towards several model clients, including FLuc (Yonehara et al., 1996). If this were the case, RegA of sarsin would need to bind folding-competent clients and release them within biologically relevant time scales, which would then be assisted by downstream chaperones. On the other hand, if clients were being kept soluble by binding irreversibly to RegA, or if RegA were unable to discriminate between terminally misfolded and folding competent clients, one would expect that subsequent chaperone systems would have a diminished capacity to assist them during their folding.

In order to distinguish between these possibilities, we utilized the bacterial Hsp70 system, DnaK, DnaJ and GrpE (K/J/E) which is known to be capable of refolding FLuc to very high yields (*i.e.*, ~90%, Figure 5.2) (Szabo et al., 1994). The experimental design is demonstrated schematically in Figure 5.6. We reasoned that if RegA was productively binding to refolding FLuc molecules but not able to assist their complete refolding on its own, then the large fraction of FLuc which remained soluble would recover substantial activity upon the addition of K/J/E (Figure 5.6a). On the other hand, non-productive binding would not yield any significant increase in activity upon addition of K/J/E (Figure 5.6b), which would be expected for the large fraction of FLuc being maintained soluble by BSA (Figure 5.5).

FLuc refolding reactions were initiated as above, in the presence of a six-fold molar excess of RegA or equivalent amounts of BSA (by weight, corresponding to a ~3-fold molar excess). After 30 minutes of incubation, K/J/E (10 μ M, 2 μ M, 6 μ M, respectively) were added to the reactions and light emission was measured over time (Figure 5.7a). A substantial amount of reactivated luciferase was recovered in the RegA-supplemented reactions (~60%), whereas no additional active luciferase was recovered in

the reactions supplemented with excess BSA (Figures 5.7a and 5.7d). Thus, even though both RegA and BSA are capable of maintaining a large fraction of FLuc in a soluble state, only RegA is capable of releasing folding-competent intermediates in a biologically relevant time scale. Since we saw such a substantial effect of RegA in this assay, we

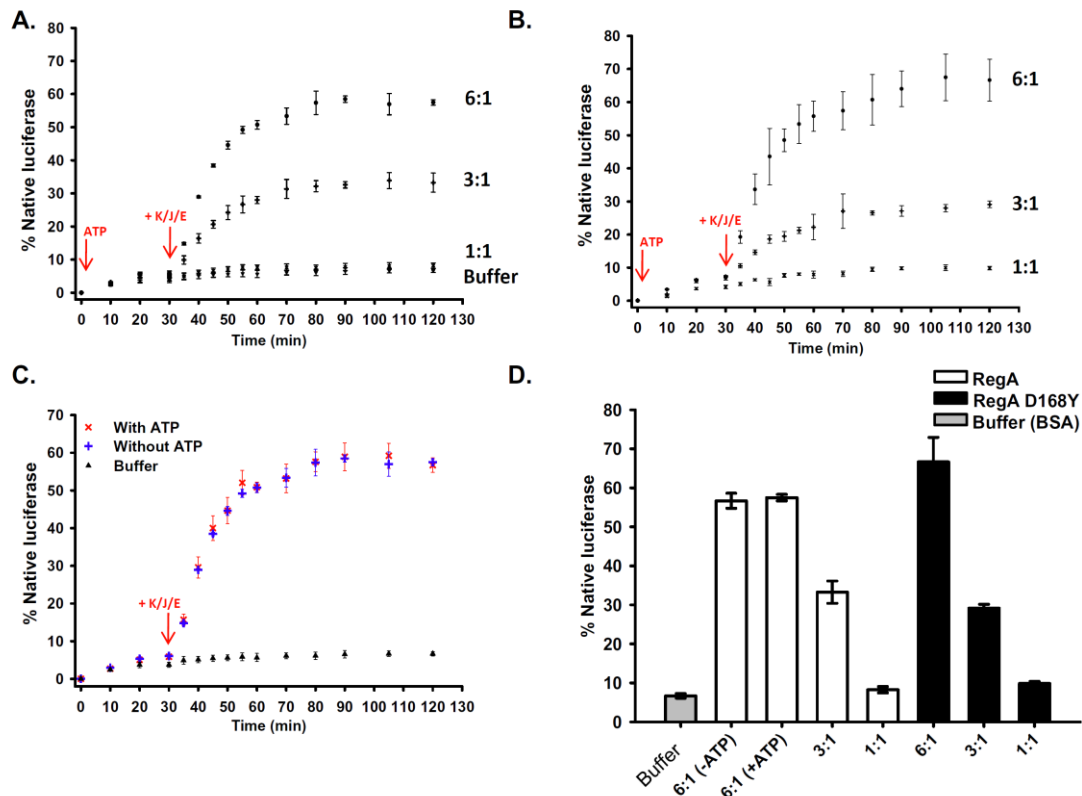


Figure 5.7: RegA maintains FLuc in a folding-competent state.

(a) Plot depicting the amount of FLuc refolded in an order-of-addition experiment as follows: FLuc (10 μ M) was denatured and allowed to refold by 1:100 dilution from GdmCl in the presence of 600, 300 or 100 nM RegA or an equivalent concentration of BSA (w/w). At 30 minutes, a cocktail of K/J/E was added to a final concentration of 10 μ M, 2 μ M and 6 μ M respectively. The data is plotted as mean and standard deviations of three independent experiments. (b) An identical experiment was performed as in panel a, except RegA D168Y was utilized instead of wild-type. (c) Plot of an identical experiment as in panel a, except ATP was initially excluded from the refolding reaction and added simultaneously with the K/J/E cocktail at thirty minutes to the RegA (600 nM)-supplemented reaction. The data points of RegA with ATP and BSA experiments from panel a are included for reference. (d) Histogram summarizing the maximum FLuc recovered (percent of a native control) for the indicated conditions (mean and standard deviation).

The function of the K/J/E system is dependent on ATP hydrolysis, and thus this nucleotide cannot be omitted after addition of these chaperones. However, we tested whether presence or absence of ATP had any effect on RegA function by including it or omitting it during the 30 minutes of incubation prior to addition of K/J/E. The refolding curves of these experiments are super-imposable (Figure 5.7c), demonstrating that RegA does not require ATP for the chaperone activity documented by this assay.

RegA Does Not Prevent Aggregation of Luciferase under Thermal Stress

In addition to its activity in protein refolding, Hsp90 can also prevent the inactivation of several model clients when exposed to thermal stress (Jakob et al., 1995). Since RegA demonstrated chaperone activity in a protein refolding paradigm, we reasoned that it may be active in this paradigm as well. To test this possibility we exposed native FLuc to elevated temperatures (42 °C) in the presence or absence of RegA and monitored activity over time. There was a rapid decline in the activity FLuc when exposed to high temperatures, and RegA had no effect on this decline (Figure 5.8a). Next,

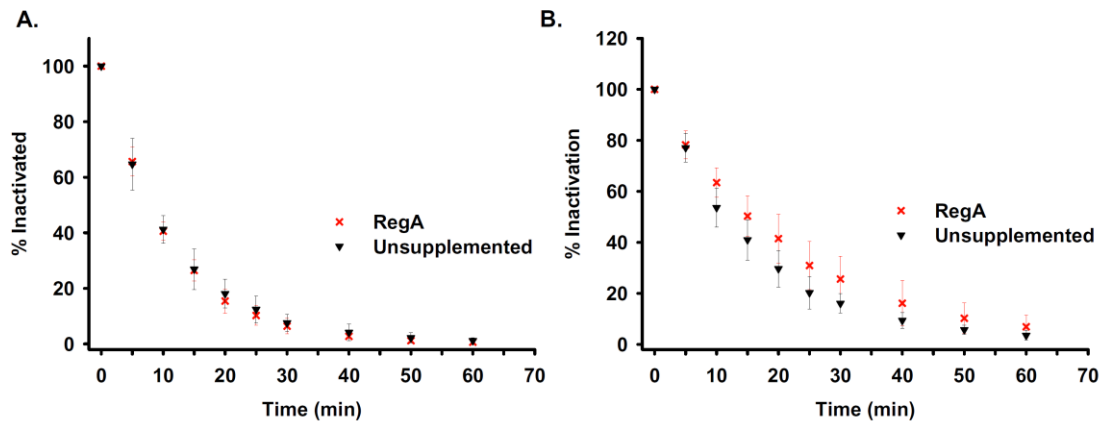


Figure 5.8: RegA does not prevent the thermal inactivation of FLuc.

(a) FLuc (100 nM) was incubated at 42 °C in the presence or absence of RegA (500 nM). Aliquots were removed at various time points and FLuc activity was determined. The activity was normalized to the same concentration of native luciferase at room temperature. (b) An identical experiment was performed as in panel a, except ATP was included at 5 mM.

we repeated the experiment in the presence of ATP to determine if RegA may become active in this paradigm in the presence of nucleotide. There was a slight trend towards delay of aggregation in the presence of ATP, but the difference was not significant (Figure 5.8b). This trend is likely due to a chemical chaperone effect of ATP; since FLuc does possess an ATP binding pocket, this nucleotide is likely capable of stabilizing the native state. Thus, it appears that even though RegA has a strong capacity to assist during the refolding of FLuc, it is not competent to prevent the aggregation of FLuc. An important caveat in the interpretation of this experiment is that RegA itself may unfold under these conditions. We have not systematically defined the folding and unfolding pathways of RegA, under normal or elevated temperatures.

RegA Has a High Affinity for Client Protein

Since the K/J/E system and RegA both interact with unfolded/partially folded FLuc, we reasoned that we could use this system to qualitatively estimate the relative affinity of RegA for FLuc. We predicted that, if DnaK and RegA were both included at equivalent concentrations in the same refolding reaction, the behavior of the refolding reaction would be governed by the chaperone with the greatest affinity for FLuc (fast and efficient refolding for K/J/E *versus* slower and less efficient refolding for RegA). Therefore, we decreased the concentration of the components of the K/J/E system to where DnaK is equal to RegA (while maintaining the ratio of DnaJ and GrpE as in all other refolding reactions). The combination of both RegA and K/J/E in a FLuc refolding reaction caused a considerable change in the rate and yield of refolded FLuc compared to K/J/E alone (Figure 5.9). Significantly, the resulting curve instead resembled the curve for RegA alone (Figure 5.3a), whereas the combination of BSA (at equivalent concentrations to RegA) and K/J/E did not have any effects on the behavior of the

refolding reaction. Taken together, these results suggest that the affinity of RegA for unfolded/partially folded FLuc is stronger than the non-specific interaction between BSA and FLuc and of similar or greater magnitude than the affinity between FLuc and the K/J/E system.

Interestingly, the fraction of refolded FLuc in the mixture containing RegA and K/J/E did not reach the same level as the reaction containing K/J/E alone. We interpret this result as effective competition between RegA and K/J/E at these low chaperone concentrations, due to a substantially higher affinity of RegA for FLuc. Consistent with these results, the binding of RegA to FLuc could be overcome by increasing the concentration of K/J/E in the refolding reaction: a 20-fold excess of K/J/E relative to RegA was able to efficiently and rapidly capture and assist the FLuc being maintained soluble by RegA (Figure 5.7a). These results indicate that in the cell, where the

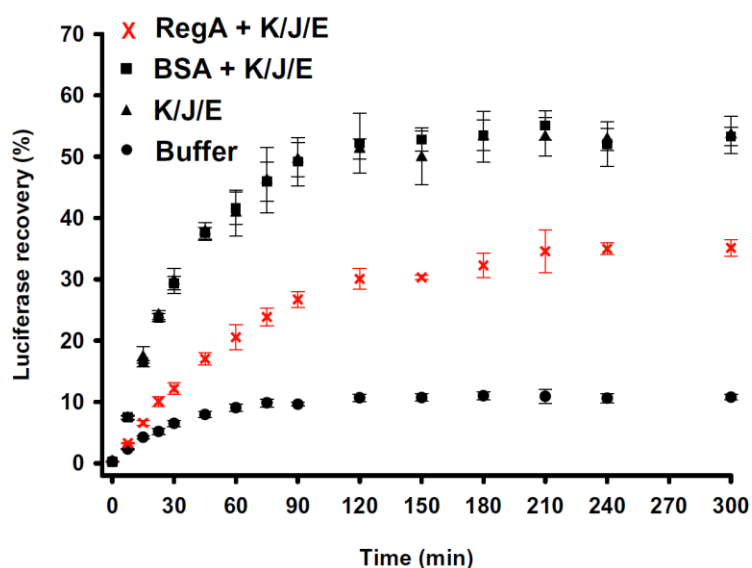


Figure 5.9: RegA has a high affinity for client protein.

Graph displaying the refolding data for denatured FLuc (5 μ M) after 1:100 dilution from GdmCl under the indicated conditions. The concentrations of the components were as follows: RegA, 500 nM; DnaK, 500 nM; DnaJ, 100 nM; GrpE, 300 nM; BSA, 0.1 mg/ml. The data are presented as mean and standard deviation of three independent experiments.

concentration of Hsp70 family members is in substantial excess to that of sarsin (Li et al., 2004; Parfitt et al., 2009) these systems may efficiently cooperate during the folding of client polypeptides.

B. SACSIN IS A CO-CHAPERONE

The presence of a co-chaperone module (the J domain) close to the C-terminus of sarsin indicates a likely role for sarsin in regulating the activity of other chaperones. Indeed, the J domain of sarsin was shown to be functional in a genetic complementation assay (Parfitt et al., 2009). Therefore, sarsin may possess an integrated co-chaperone activity among its multiple domains that allows for efficient and productive coordination among multiple chaperone systems for accomplishing protein folding.

RegA Does Not Depend on DnaJ to Cooperate with DnaK

The participation of DnaJ in the K/J/E-assisted folding cycle consists of two

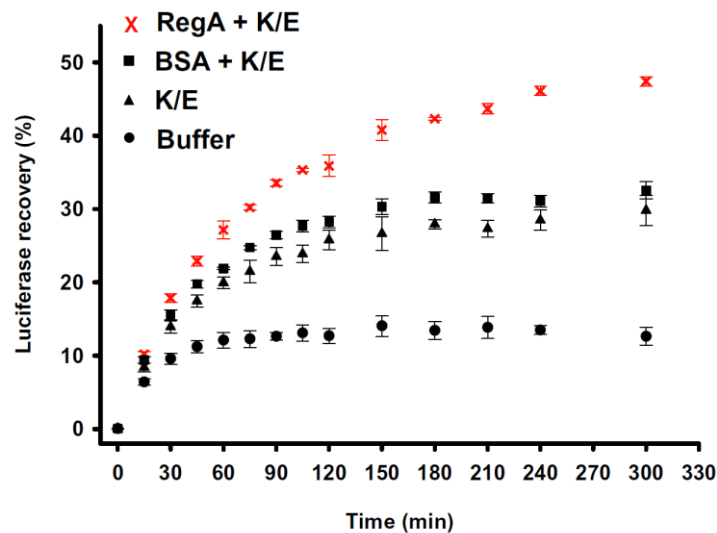


Figure 5.10: RegA does not depend on DnaJ to cooperate with DnaK.

Graph displaying the refolding data for denatured FLuc (5 μ M) after 1:100 dilution from GdmCl under the indicated conditions. The concentrations of the components were as follows: RegA, 500 nM; DnaK, 10 μ M; GrpE, 6 μ M; BSA, 0.1 mg/ml. The data are presented as mean and standard deviation of three independent experiments.

related activities: client delivery to DnaK (*via* its client binding domain) and stimulation of the ATPase activity of DnaK (*via* its J domain) (Szabo et al., 1994). Since saccin contains a functional J domain and at least one region that can interact with unfolded/partially folded FLuc (RegA), we wondered if saccin could perhaps operate as a functional DnaJ-type co-chaperone. We decided to test whether RegA could functionally operate as the client delivery module of a DnaJ co-chaperone by performing experiments in the presence and absence of bacterial DnaJ. If the cooperation of RegA with K/J/E during FLuc refolding was strictly dependent on the presence of DnaJ in the mixtures, then it would be less likely that saccin can function as a DnaJ co-chaperone. To test this, we began by examining the behavior of a FLuc refolding experiment in the presence of DnaK and GrpE but omitting DnaJ. As expected, DnaK and GrpE alone (without DnaJ) did not efficiently assist FLuc during its refolding (Figure 5.10). However, the presence of RegA in the DnaK/GrpE (K/E) mixture substantially enhanced the yield of refolded FLuc (~27% *versus* ~47%) (Figure 5.10). Consistent with the results presented throughout this chapter, substitution of RegA by BSA in the same experimental design did not result in any increases in activity beyond those observed for K/E or K/J/E (Figure 5.10). These results confirm that folding competent clients may be delivered from RegA to DnaK and suggest that full-length saccin, which contains a J domain, may be ideally situated to deliver clients to members of the Hsp70 family in the eukaryotic cytosol, where it has been localized.

The J Domain of Saccin Is Functional

RegA appears to possess a capacity to deliver client proteins to DnaK (Figure 5.10) and we found convergent evolutionary pathways leading to the occurrence of J domains in same polypeptide as SRRs (Figure 3.7). Thus, we wished to assess whether

the J domain from human sacsin conserved the well documented capacity of J domains to stimulate the intrinsic ATPase rates of the general molecular chaperone Hsp70 (Minami et al., 1996).

Our initial investigation into this possibility was computational. We searched the PDB structural database with the sacsin J domain and were surprised to find that its structure had been solved by NMR spectroscopy. The coordinates were deposited in the PDB but the structure was never published (1IUR, Kobayashi et al., unpublished). A comparison between human sacsin J domain structure and the bacterial DnaJ J domain structure reveals that they appear remarkably similar (Figures 5.11a and 5.11b). The

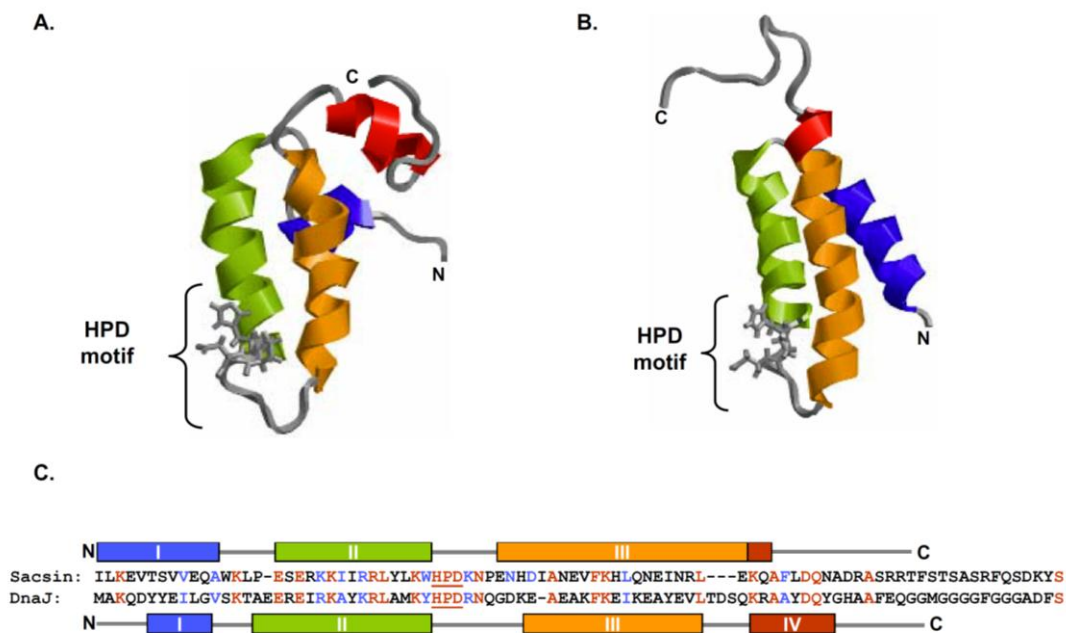


Figure 5.11: Similarity between the J domain of sacsin and other J domains.

(a) Ribbon representation of the solution structure of the J domain from *E. coli* DnaJ (1XBL). The HPD motif is denoted in ball-and-stick representation. (b) Ribbon representation of the solution structure of the J domain from human sacsin (1IUR). The HPD motif is denoted in ball-and-stick representation. (c) Pairwise sequence alignment of the J domains from *E. coli* DnaJ and human sacsin. The helices are marked and numbered above or below the alignment. The HPD motif is underlined. Identical residues are in red, and similar ones in blue.

bacterial J domain is composed of four helices with two main helices on the face of the protein orienting the catalytically critical HPD motif in its characteristic conformation. A similar topology of the helices is observed in the J domain of saccin. We performed a pair-wise sequence alignment between these two domains and mapped the secondary structure elements to the alignment. We found that the first three helices of the bacterial J domain are in the same location as in the J domain of saccin. Interestingly, the fourth helix in DnaJ appears to be fused into the third helix in saccin. Importantly, the HPD

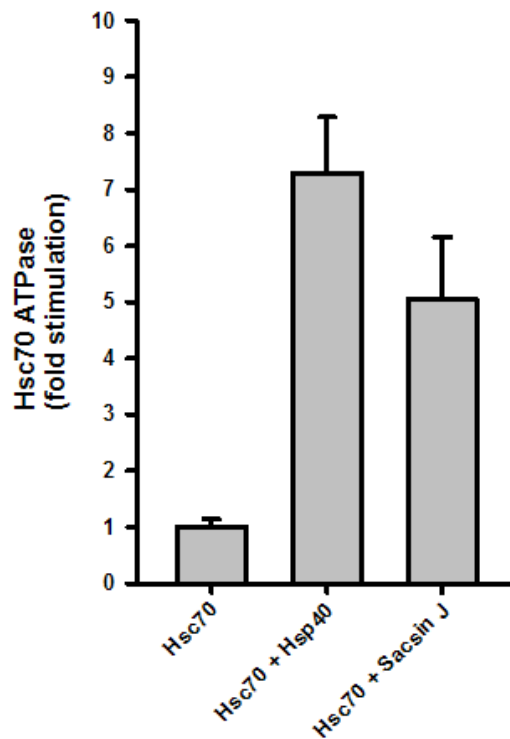


Figure 5.12: The J domain of saccin is functional.

ATPase activity of bovine Hsc70 alone or in the presence of bovine Hsp40 or murine saccin J domain. Reactions were performed with components at the following concentrations: Hsc70, 1 μ M; Hsp40, 2 μ M; saccin J domain, 4 μ M. The reaction buffer contained 10 mM MOPS-KOH 7.2, 50 mM KCl, 10 mM MgCl₂, and 1 mM ATP. Reactions were performed for 10 minutes at 30 °C, in triplicate, background-subtracted and normalized to Hsc70. Values shown are mean \pm SEM. Significance was determined using the t-test. We wish to thank Jason Chandler for technical assistance with this experiment.

motif is conserved in both sequence and structure, it is found in a similar orientation in the loop between helices 2 and 3.

In order to confirm that the J domain of saccin is active we wished to perform biochemical experiments. Thus, we cloned and purified a recombinant version of the J domain from mouse saccin in *E. coli*. This construct encompassed the entire J domain (amino acid residues 4308-4396) and was 115 residues long and 13.8 kDa with its poly-histidine tag. We performed ATPase assays of Hsc70 utilizing the malachite green assay. The baseline activity of Hsc70 was set to a value of 1. When Hsp40 was included in the reaction, this activity increased ~7-fold. When the same experiment was performed with the J domain of saccin we found that the activity increased ~5-fold (Figure 5.12). Thus, the J domain in saccin is nearly as effective as the full length human Hsp40 in increasing the ATPase activity of Hsp70; indeed, the difference between Hsp40 and the J domain of saccin was not significantly different. This experiment indicates that the saccin J domain is functional, in agreement with previous findings that it is capable of rescuing certain phenotypes of the genetic deletion of the Hsp40 homolog DnaJ in *E. coli* (Parfitt et al., 2009).

C. CONCLUSIONS

In this chapter we provide evidence that saccin functions as a molecular chaperone. RegA of saccin was moderately capable of directly assisting during the refolding of FLuc and was highly capable of maintaining it in a state that can be productively acted upon by bacterial Hsp70 family members. The capacity of saccin to bind and release folding competent client proteins within biologically relevant time scales, its ability to cooperate with additional chaperones, and the presence of a

functional J domain demonstrate that this unusually large protein contains both chaperone and co-chaperone modules within the same polypeptide chain.

Chapter 6: Potential Client Proteins of Sacsin

Chaperone and co-chaperone proteins recognize a diverse set of client proteins as substrates. Some recognize only a limited number of proteins that have attained a partially native structure, while others are generally promiscuous and bind to many unfolded and partially folded proteins (Frydman, 2001). Virtually nothing is known about the possible substrates that sacsins may recognize while acting as a chaperone/co-chaperone. Knowledge of these clients would shed light on the pathogenesis of ARSACS and provide additional therapeutic targets. Therefore, we sought to identify sacsins client proteins through pull-down experiments and analysis of sacsins knockout mouse brains.

A. SACSIN INTERACTS WITH MAP1A

The J Domain Construct of Sacsins Pulls Down MAP1A

Although the C-terminal domain of DnaJ has been reported to be capable of delivering client proteins, several recent reports have suggested that the N-terminal J domain itself can also function in this capacity (Bird et al., 2006). Thus, we began our characterization of sacsins substrates by searching for proteins which bind to the isolated J domain of sacsins, which we know is correctly folded and functional (Figure 5.12).

We designed an experiment with purified recombinant sacsins J domain as bait for a pull down from mouse brain and cerebellar extracts. Since this protein contains a poly-histidine tagged, we utilized a nickel-nitriloacetic acid (Ni-NTA) agarose resin (Novagen) to immobilize the J domain. For a non-specific binding control, a similarly sized and tagged protein, green fluorescent protein (GFP), was utilized. Sacsins-interacting proteins were identified by incubating mouse brain and cerebellar extracts with purified J domain or GFP immobilized on Ni-NTA agarose beads in lysis buffer.

After extensive washing, proteins were eluted with imidazole and analyzed by SDS-PAGE. One band appeared to be present in lanes containing sacsin J domain and not in lanes containing the control protein (Figure 6.1).

The band was excised and submitted for mass spectrometric analysis (UTMB Biomolecular Resource Facility) which revealed that this band corresponded to the microtubule associated protein 1A (MAP1A). High molecular weight microtubule associated proteins (MAPs) are required for the *in vivo* polymerization of tubulin into microtubules as well as for regulation of microtubule interactions with other components of the cytoskeleton (*e.g.* actin filaments). In particular, MAP1A is widely expressed in central nervous system neurons, co-localizing with microtubule arrays. MAP1A also interacts with other cytoskeletal components important in neuronal structure and function, including actin filaments (Maccioni and Cambiazo, 1995). Additionally, MAP1A is involved in activity-dependent dendritic remodeling (Szebenyi et al., 2005);

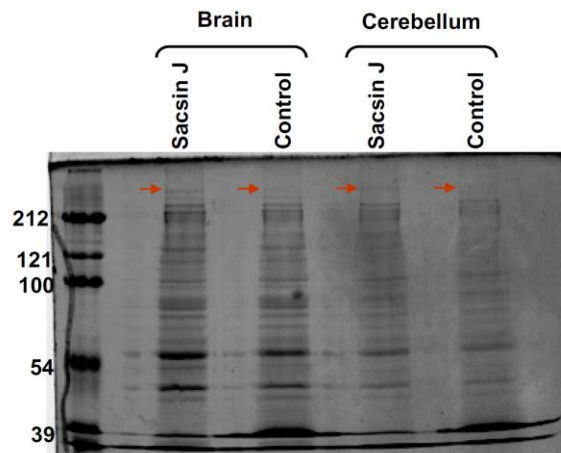


Figure 6.1: The J domain construct of sacsin pulls down MAP1A.

Sacsin J domain was utilized as bait in a pull-down from mouse brain or cerebellar extracts utilizing Ni-NTA agarose beads. The control protein utilized was a similarly tagged GFP. The elution was analyzed by SDS-PAGE. The marker is shown on the far left with the approximate molecular weights (in kDa). The red arrow denotes a band present in the sacsin J domain lanes but not the control lanes. This band was excised and identified by mass spectroscopy.

thus, a plausible hypothesis is that misfolding of MAP1A due to a deficiency of saccin may lead to some of the cellular degeneration observed in the cerebella of ARSACS patients.

MAP1A and Saccin Co-Immunoprecipitate

Since we had identified an interaction with MAP1A based on a pull-down with a fragment of saccin, we wished to determine if the interaction could be confirmed by a different method utilizing immobilized MAP1A. Therefore, we performed a co-immunoprecipitation experiment. An anti-MAP1A antibody (Santa Cruz Biotechnology) coupled to protein A beads (Sigma) was used to isolate endogenous MAP1A-interacting proteins from mouse cerebellar lysates. As a control, protein A beads derivatized with equivalent amounts of an unrelated antibody were utilized. The complexes were then eluted and immunoblotted with an antibody against mouse saccin (Santa Cruz Biotechnology) (Figure 6.2). A band present only in the input lane and the lane

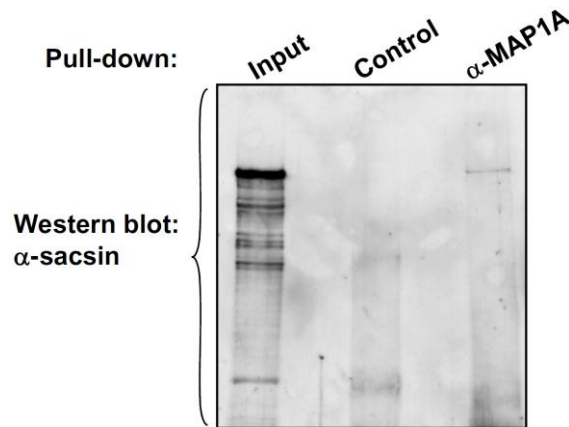


Figure 6.2: Co-immunoprecipitation of saccin and MAP1A.

A Map1a antibody was utilized as bait in a co-immunoprecipitation experiment. This antibody was immobilized on Protein A agarose beads and then incubated with a mouse cerebellar extract. An antibody against an unrelated protein of similar mass (inositol triphosphate receptor) was utilized as a control. After washing and eluting, the experiment was analyzed by western blot for saccin. A prominent band at the correct molecular weight was observed in the input and Map1a pull-down lanes, but not in the control.

corresponding to the MAP1A antibody confirmed that saccin is present in a complex with MAP1A in these extracts.

MAP1A and Saccin Co-Localize in the Cerebellum

Our results show that saccin can exist in a complex with MAP1A in brain and cerebellar extracts. However, in order for this interaction to be relevant *in vivo*, these two proteins must be present in the same cellular and sub-cellular locations simultaneously. To determine if this was indeed the case, we collaborated with Dr. Luis V. Colom (UT Brownsville) who performed immunofluorescence microscopy on rat cerebellar sections with the antibodies utilized in our previous experiments (Figure 6.3) (see materials and methods).

We observed particularly prominent reactions for both proteins in Purkinje cells. The diffuse distribution throughout the soma, dendrites and axons suggests that these proteins are not restricted to a particular organelle, but rather that they are present as soluble, cytoplasmic proteins. Importantly, the observed localization of both MAP1A and saccin were very similar to that previously reported in the literature (Parfitt et al., 2009;

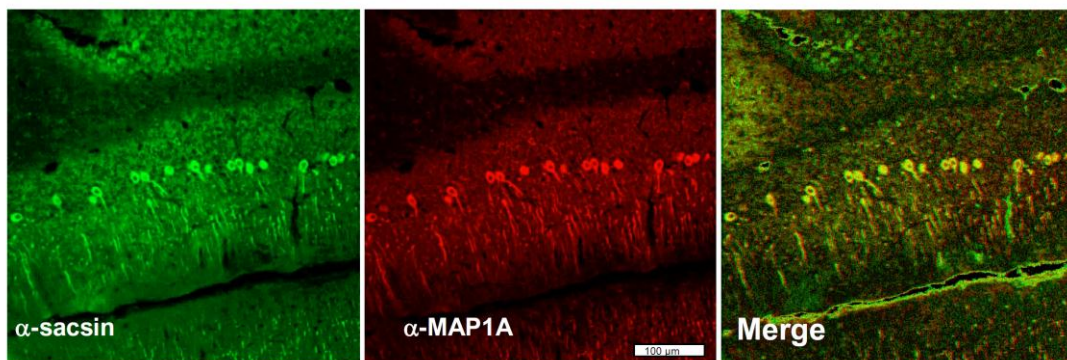


Figure 6.3: Map1a and saccin co-localize in the cerebellum.

A section of rat cerebellum was stained with both saccin and map1 antibodies followed by incubation with secondary antibodies conjugated to rhodamine or fluorescein to allow visualization of the expression pattern of both proteins in the same section. Staining was performed by M. T. Castaneda in the laboratory of L. V. Colom (UT Brownsville).

Szebenyi et al., 2005). Thus, it is likely that these proteins are available to interact with one another in the cell as they occupy approximately the same region of the neurons in the cerebellum.

RegA of Sacsin Does Not Interact with MAP1A

Since we had identified MAP1A as an interactor of saccin, we considered whether it may be a client. We had identified RegA of saccin as possessing chaperone activity (Chapter 5). Thus, we reasoned that MAP1A would likely interact with RegA if it were a client of this protein. To test this possibility, we utilized RegA for pull-down assays from a partially purified microtubule associated protein (Map) preparation from bovine brain

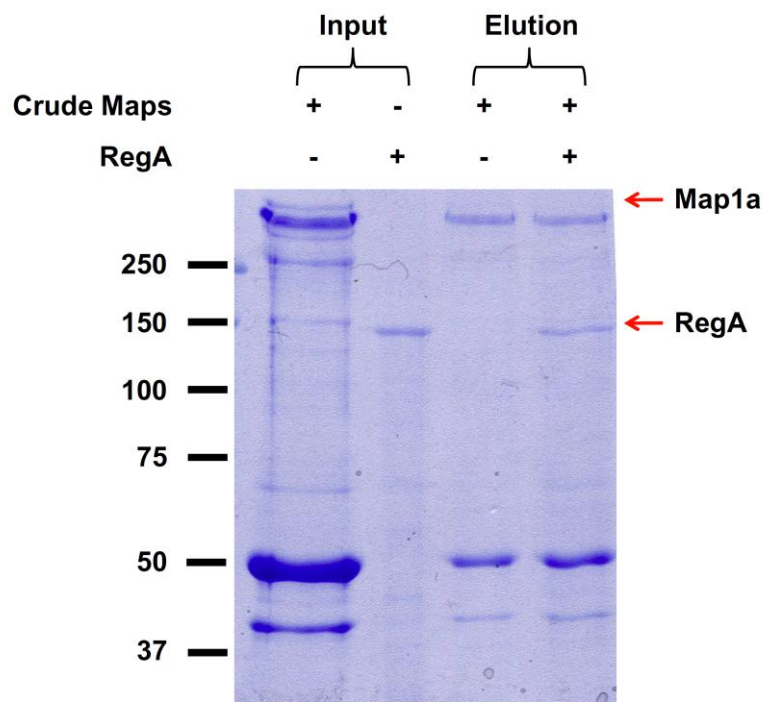


Figure 6.4: RegA of saccin does not interact with MAP1A, as assessed by a pull-down assay.

RegA of saccin was incubated with a crude preparation of Maps followed by immobilization on Ni-NTA agarose beads. The binding proteins were eluted with imidazole. The input and elution fractions were separated on SDS-PAGE and are shown. The location of Map1a and RegA are denoted with a red arrow. The approximate location of the molecular weight markers is shown on the far left.

B. ANALYSIS OF SACSIN KNOCKOUT MOUSE BRAIN

We considered it unlikely that MAP1A is a client of saccin since it does not interact with the SRR domain. Rather, saccin and MAP1A may interact for another purpose, such as sub-cellular localization. Additionally, saccin is not active in a prevention of thermally-induced aggregation chaperone assay, so it is difficult to develop an assay to test for potential clients (Figure 5.8). Therefore, we sought to develop a new approach to determining the clients of saccin. Since chaperones assist proteins to properly fold, it is expected that clients of a chaperone may aggregate when the chaperone is absent (Chapman et al., 2006); alternatively, it is possible that the clients will be degraded when they misfold. Consistent with our finding that saccin is a chaperone, autopsy reports have indicated the presence of inclusions in human patients (Richter, 1996). We suggest that these may represent clients of saccin. Therefore, we considered that the ideal experiment for identifying clients of saccin would be to purify aggregates from an ARSACS disease model; the most common model for human genetic diseases is a knockout mouse. We discussed this possibility with our collaborator in Montreal, Canada, Dr. Bernard Brais (Université de Montréal). He confirmed that his team had generated a saccin knockout mouse and this model did indeed possess cytoplasmic inclusions in the Purkinje cells of the cerebellum by light microscopy (personal communication). Additionally, the knockout mouse appeared to recapitulate the disease phenotype accurately. The cerebella from these mice would be an ideal substrate to identify clients of saccin. Our collaborators generously agreed to supply us with brain samples from these mice for biochemical studies.

Our approach was to homogenize the tissue under native conditions and perform a low speed centrifugation to separate the nuclei and cell debris from the cytoplasm and a

higher speed centrifugation to separate the insoluble cytoplasmic proteins from the soluble portions of the cytosol. The steps performed in this fractionation are outlined in Figure 6.5. Benzonase was included in the brain extract to ensure the complete removal of DNA. In order to maximally disrupt the lipid bilayers Triton X-100 was included at 1.0%. The final manipulation was to precipitate all samples in methanol/chloroform to remove organic impurities which interfere with precise separation on SDS-PAGE. We predicted that any aggregates would be present in the pellet fractions, most likely the P2 fraction.

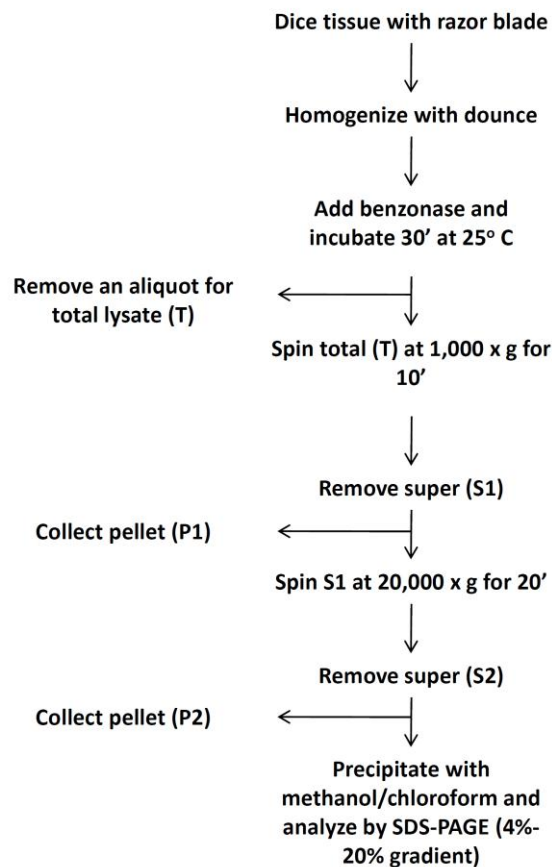


Figure 6.5: Protocol for fractionation of saccin knockout mouse brain.

We performed a fractionation on knockout mouse cerebellum and wild-type littermate controls according to this protocol; heterozygote samples were also available but these were not included in our initial analysis. The various fractions were analyzed by 4-20% gradient SDS-PAGE. When the gel was examined, we could not identify any

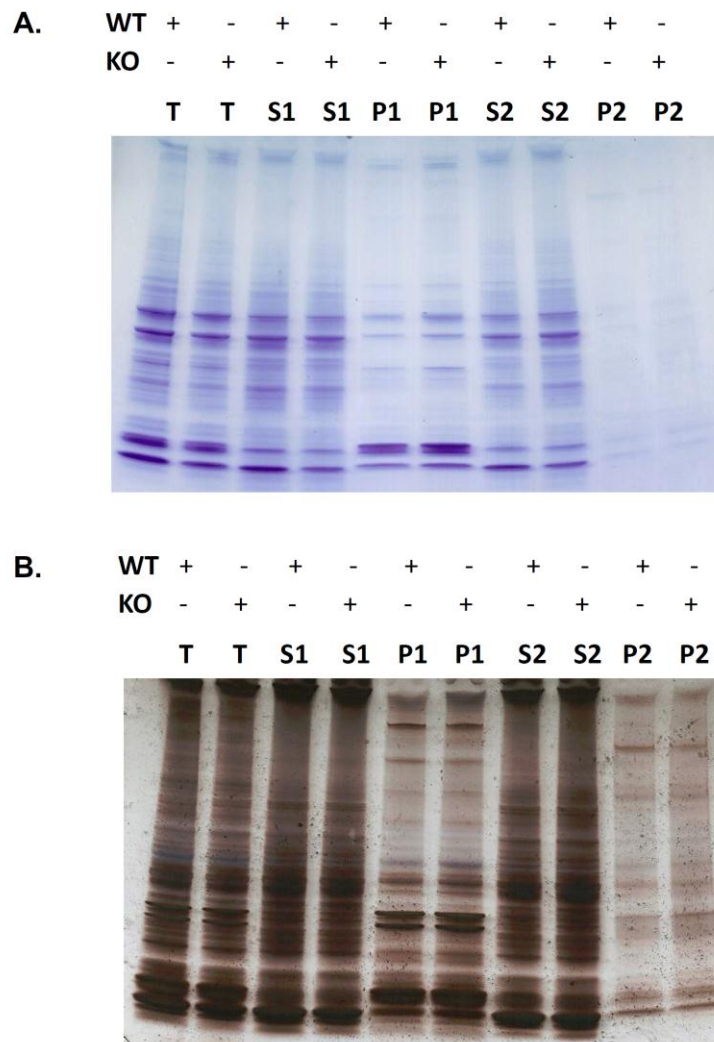


Figure 6.6: SDS-PAGE of saccin knockout mouse brain fractions.

Saccin knockout mouse brains and wild-type littermate brains were subjected to the fractionation protocol in Figure 6.5. **(a)** 4-20% gradient SDS-PAGE analysis of the fractions stained by Coomassie blue staining. **(b)** Silver stain of the same gel in panel **a**.

bands which were enriched in the knockout compared to the wild-type samples (Figure 6.6a). Since we could not visualize any differences on Coomassie blue-stained SDS-PAGE gel, we decided to perform silver staining, since this method is a more sensitive technique than Coomassie blue staining. The drawback to silver staining is that it is not compatible with mass spectroscopy which is our downstream application to determine the molecular identity of any identified bands. However, silver staining also did not demonstrate any differences (Figure 6.6b). Thus, it appears that the aggregates are not present or our methods were not adequate to detect them.

There are several conditions in our method that could be modified in future experiments to increase the possibility of determining the clients of sacsins. The inclusion of Triton X-100 at 1.0% may artificially solubilize the aggregates. Since Triton X-100 is a relatively mild, non-ionic detergent we had used this large concentration to increase the disruption of lipid bilayers, however, this may be at the expense of also disrupting aggregates. The fastest centrifugation step we utilized was a 20,000 x g spin for 20 minutes. While this is sufficient to isolate most large aggregates, it may not be sufficient to isolate smaller aggregates. Since we do not have any information on the nature of the aggregates in the sacsins knockout mouse, this centrifugation step may not be capable of sedimenting them. The inclusion of an ultracentrifugation step may reveal a difference between wild-type and knockout mouse samples. A final manipulation which may interfere with our ability to detect aggregates is the methanol/chloroform extraction. We noticed that some material precipitated by this method would not resuspend in the SDS sample buffer. While the amount of this material appeared to be similar between wild-type and knockout brains, it may contain the proteins of interest. Refining our protocol to

generate clean SDS-PAGE gels without this step may improve our ability to detect aggregates.

These technical drawbacks may be important in increasing the sensitivity to detect aggregates, but the major factor is likely the age of the mice. The inclusions visualized by light microscopy by our colleagues were noticed during their assessment of 200 day old mice. However, the mouse brains we were supplied with were from 120 day old mice. Since ARSACS is a progressive disease, it may take more time for the misfolded proteins to develop into aggregates that are detectable by SDS-PAGE. Future experiments should include mice of older ages.

C. CONCLUSIONS

We identified MAP1A as an interactor of salsin. This site of interaction appeared to be the J domain and not any domains present in RegA. Since the only site that we have assigned chaperone activity to is the SRR domain and MAP1A did not interact with it we consider it unlikely to be a client of salsin. Indeed, the J domain was found to be the site of interaction, and this region is a co-chaperone site which has not been demonstrated to possess chaperone activity. We expanded our search for clients by examining the insoluble material in a salsin knockout mouse brain. With our current methods we could not identify any aggregated proteins which were unique to salsin knockout mouse brain.

SECTION 3: DISCUSSION

The objective of this study was to determine the function of the saccin protein. The gene encoding saccin is mutated in the hereditary ataxia, ARSACS. We hypothesized that saccin is a novel molecular chaperone with built in co-chaperone modules. Our sequence analysis revealed the presence of a novel supra-domain in saccin (SRR) and our biochemical analysis confirmed that this domain possesses both ATPase activity and chaperone activity. We also found a region of co-chaperone activity in saccin (the J domain).

Chapter 7: The SRR Supra-Domain

We identified a novel supra-domain in saccin which we termed the saccin repeating region (SRR domain) due to its presence three times along the N-terminus of saccin. A supra-domain is defined as two or more domains which may exist in isolation but tend to move together as an evolutionary unit (Vogel et al., 2004). It is thought that the juxtaposition of two or more domains in a supra-domain provides for a common function that is not accomplished by either domain in isolation. The supra-domain in saccin contains a region of similarity to the nucleotide binding domain of Hsp90 and an additional region of conserved sequence which does not share similarity to Hsp90 (Figure 3.4). Indeed, we could only identify this region adjacent to a region of similarity to Hsp90 in other proteins. Thus, the SRR is a supra-domain in that its combination of domains is found in many different protein contexts, but is unique from other supra-domains in that one of its constituents does not appear to be able to exist in isolation. The region of similarity to a molecular chaperone in the SRR domain suggests that this module may be involved in protein folding-related activities.

The lack of occurrence of the C-terminal portion of the SRR in isolation suggests that its function may be specifically tied to the region of similarity to Hsp90 (*i.e.* it may not possess a function independent of this region). This may be related to the finding that Hsp90 is a split ATPase, requiring a basic residue in the middle domain to form the catalytic center (Ali et al., 2006; Meyer et al., 2003). Our sequence analysis led us to propose that a highly conserved basic residue in the C-terminal portion of the SRR domain may fulfill this function (Figure 3.5). Perhaps a major function of this C-terminal portion of the SRR is to complete the catalytic center in the region of similarity to Hsp90.

We determined that the first SRR in saccin is indeed ATPase active (Figure 4.9), consistent with our suggestion that the C-terminal portion may contain additional determinants for the catalytic center. Also, we generated a version of RegA that contains a human disease-causing mutation in the region of similarity to Hsp90 in the first SRR (RegA D168Y). This protein was soluble and could be purified by the same protocol as wild-type. We demonstrated that RegA D168Y is incompetent to hydrolyze ATP (Figure 4.9). Likewise, there are numerous other mutations throughout the SRR domains in saccin which may affect their ATPase activity as well (Figure 2.6, Table A.1). The D168Y protein appears not to be a grossly misfolded mutant *per se*, in that it is soluble under the conditions examined. However, we did find that the mutant protein adopts a different conformation than wild-type. This conformation contains more extensive surface hydrophobicity and is structured sufficiently different from wild-type to lead to an increase in the quantum yield of many of the tryptophan residues (Figure 4.10). In Hsp90, the conformation of the dimer is regulated by the nucleotide state of the N-terminal domain (Figure 1.6). Thus, the loss of ATP hydrolysis activity in RegA D168Y

may indicate that this SRR is locked into a conformation that is not competent to hydrolyze ATP.

In human sacsin, we found large tracts of sequence separating the SRR domains (Figure 3.3); our phylogenetic analysis revealed that this configuration is universal, as it is found in all SRR-containing proteins identified (Figure 3.7). This strong tendency for the sequence C-terminal to the SRR domain to be devoid of identifiable domains for up to 1000 residues or more is unusual. There is weak self-similarity in these regions, but they appear to be rapidly diverging. Based on the conservation of this configuration, we suggest that there may be a functional purpose for the SRR domains to be separated by these large regions of sequence which contain few constraints. We currently do not have evidence on what this function may be. However, this is a substantial amount of protein to only represent the vestigial remains of a repetition event.

This hypothesis is supported by our attempts to purify various fragments of sacsin containing the SRR domain by itself (Figure 4.1). Empirically, we and others have noted that smaller fragments of multidomain proteins tend to have a greater probability of folding properly on heterologous expression than larger fragments (Chang et al., 2005). However, we observed the opposite trend during our attempts to express the SRR domain. Expression of a truncated version of the SRR (M90C) or a full-length SRR (SRCL) resulted in nearly complete insolubility. As we generated longer constructs containing additional amounts of the intervening sequence, we were surprised to observe an increase in solubility. The entire RegA construct was ~85% soluble while the M90C construct was nearly 100% insoluble. This is interesting, since M90C is only 24 kDa, while RegA is 167 kDa. This provides evidence that the functional unit of the SRR domain likely contains these portions of intervening sequence. We suggest that it is

helpful to divide sacsin into a modular structure composed of four regions (Figure 4.2). Regions A-C contain an SRR at their N-terminus and a region of intervening sequence up to the next identifiable domain. Region D is composed of the XPCB, J, and HEPN domains, all the way to the C-terminus.

Why are there repeating regions along the sacsin sequence? 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 tetratricopeptide repeat (TPR) and ankyrin repeats (Andrade et al., 2001). 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 in different protein contexts tend to be larger (>100 amino acid residues) and are generally functional by themselves. Examples include the thioredoxin domain (Gruber et al., 2006) and the immunoglobulin fold (Barclay, 2003). The SRR domain is a repeating element that falls into the second class. There are several examples of proteins which contain a single SRR domain, but there is a strong tendency for SRR domains to be present in repeating arrangements ranging from two to four in a single polypeptide (Figure 3.7).

Indeed, we determined that the generation of a three repeat sacsin homolog in *S. purpuratus* occurred by an independent duplication event of the middle SRR domain (group B) (Figure 3.8). Likewise, we found that many plant species contain a three repeat protein without the components found in Region D of sacsin. This arrangement likely occurred independently from that found in animals, since no common ancestor contains three repeats. There appears to be a functional advantage for a polypeptide to contain multiple SRR domains; yet, it is difficult to speculate on the possibilities without additional data on SRR function. An attractive hypothesis is that the SRR domains may

comprise an “assembly-line” of protein folding, wherein each SRR interacts with polypeptides at different stages in the folding process. This hypothesis follows from the finding that folding proteins often flux through different chaperone systems (Young et al., 2004). Perhaps each SRR in sacsin is tuned to different folding species ranging from fully unfolded (similar to the function of DnaK) to nearly folded (similar to the function of Hsp90). This hypothesis would also be consistent with the presence of a Ubl domain in sacsin (Parfitt et al., 2009); terminally misfolded proteins may be “triaged” by the various SRR domains and this information could be relayed *via* the Ubl domain to the proteasome. Thus, we suggest that there may be cooperativity or synergy among the various SRR domains at some level. A detailed analysis of the ability of each SRR to interact with various refolding intermediates would be required to confirm this hypothesis. Similarly, each SRR domain may have different substrate specificities and assist in the folding of distinct proteins involved in the pathways disrupted by the disease. Identification of the client proteins of sacsin would be extremely advantageous to understand the relationships among the SRRs and the proteins that necessitate their assistance.

Our phylogenetic analysis also indicates that the SRR domain is found in polypeptides which contain a diverse arrangement of other domains. The relationship between the chaperone activity of the SRR domain and the activities of the other associated domains is not clear. In a fascinating case of convergent evolution, we found a three repeat plant protein that *also* contains a J domain in a configuration reminiscent of human sacsin. However, this protein does not possess the XPCB or HEPN domains and there is no common ancestor with a similar arrangement; thus, this likely occurred through convergent evolution. There appears to be a biochemical advantage for a

polypeptide to contain both a J domain and an SRR domain. There is extensive cooperation among chaperone networks that allows for flux of client proteins through chaperones with slightly different protein folding activities. Indeed functional interactions between J domain containing proteins and Hsp90 has been demonstrated (Dittmar et al., 1998; Kimura et al., 1995). Thus the presence of these two modules in sacsin provides a physical link between the various chaperone networks.

Chapter 8: The Protein Folding Activities of Sacsin

We determined that RegA of sacsins possesses chaperone activity (Chapter 5). Since RegA contains one of the three SRRs along the sacsins molecule and all of these retain considerable similarity to the nucleotide binding domain of Hsp90, these domains likely represent sites of chaperone activity. In Hsp90, the nucleotide binding domain regulates a complex cycle of conformational adjustments that are thought to be related to the chaperone activity of the full-length dimer (Hessling et al., 2009). Additionally, the conformational fluctuations occur around a fixed point that is formed from the constitutive dimerization center contained in the C-terminal domain of Hsp90 (Pearl and Prodromou, 2006). However, Hsp90 is more than an ATPase; it utilizes this activity to govern its protein folding activities. Thus, the C-terminal portion of the SRR domain likely has an additional function(s) beyond its potential role in catalysis. In this regard, it may be insightful to compare its putative functions to the middle domain of Hsp90. In addition to its role in ATP hydrolysis, the middle domain of Hsp90 is a site for client binding (in addition to the N-terminal domain) and also possesses chaperone activity (Chapter 5). Perhaps the C-terminal portion of the SRR accomplishes a similar function. Indeed we did find that the first SRR in sacsins possesses chaperone activity. Thus, the similarity of the SRR to Hsp90 extends beyond simple sequence similarity and ATPase activity and we demonstrated that it functions similarly in terms of chaperone activity.

We examined the multiple sequence alignment of the C-terminal portion of the SRR to attempt to gain insight into its potential role in protein folding. We noted that the majority of conserved residues are hydrophobic and are contained in predicted β -strands (Figure 3.5). This finding raises at least two possibilities which are consistent with the

chaperone function of the SRR domain. These structural elements may lie in the hydrophobic core of the fold and provide a scaffold for the more divergent loop regions that are solvent exposed. If this region is a client binding site, then the divergence in these loops may represent a means for fine-tuning client specificity among the various SRR domain containing protein. Alternatively, these conserved hydrophobic residues may themselves be protein binding sites. Since chaperones interact with clients through hydrophobic patches, it is reasonable to suggest that the conserved hydrophobic residues form a client binding pocket. However, it is speculative to assign client binding to any specific region in the SRR, indeed this action could be localized to the long intervening sequence contained in RegA. Nevertheless, the function of the two components of the SRR supra-domain appear to be intimately intertwined as demonstrated by the lack of occurrence of the C-terminal portion in isolation and the extreme insolubility of the N-terminal portion when expressed alone (Figure 4.1).

Importantly, the molecular chaperone activity described here for RegA is subtly different from that of Hsp90. Hsp90 is an abundant molecular chaperone that, in the eukaryotic cytosol, is involved in the maturation of a growing list of clients, which notably include protein kinases and steroid hormone receptors (Pearl and Prodromou, 2006). Additionally, along with UNC45, it has been shown to be involved in the folding of myosin heads (Barral et al., 2002). The chaperone activity of Hsp90 has been documented in multiple experimental paradigms. For example, it has been shown to prevent the aggregation of various model proteins (including β -galactosidase and FLuc) and maintain them in a state competent for subsequent action by other molecular chaperones (Freeman and Morimoto, 1996; Yonehara et al., 1996). RegA of sarsin is similar to Hsp90 in this regard. However, in contrast to the Hsp70/Hsp40 system, Hsp90

has not been demonstrated to be directly capable of increasing the refolding yield of several model clients (Freeman and Morimoto, 1996). Thus, in this regard, RegA of saccin was different from Hsp90, in that it was capable of increasing the refolding efficiency of FLuc. Thus, saccin possesses a “protective” chaperone activity similar to that of Hsp90 *in vitro*. In addition it displays a distinct, albeit modest, “folding” chaperone action not present in Hsp90.

The cellular consequences of this moderate direct chaperone activity are not clear currently. The degree of this “folding” activity is much lower than that seen for DnaK (~25% *versus* ~90%). However, as described in the previous chapter, the SRR domain present in RegA is only one of three SRRs present along the full length saccin molecule. Its intermediate capacity to directly assist in the refolding of FLuc may reflect the fact that full activity may necessitate cooperation or synergy among the various SRR domains of saccin. That is, successive copies of SRRs may display a considerably stronger chaperone activity as reflected by this type of assays. This is consistent with our finding that the vast majority of SRRs present in eukaryotes occur in multiples, and convergent evolution led to this characteristic arrangement of consecutive SRRs along the same polypeptide chain (Figure 3.8). Additionally, the presence of a functional J domain in the full length saccin molecule suggests that SRRs are perhaps not designed to function independently, but rather in concert with other chaperone systems in the cell, most likely with members of the Hsp70 family. Consistently, we found that RegA can efficiently deliver unfolded FLuc to DnaK and can cooperate with DnaK independently of DnaJ (Figures 5.7 and 5.10).

We showed that RegA of saccin possesses ATPase activity and that a human disease-causing mutation completely abrogates this activity (Figure 4.9). This suggests

that this protein segment is structured and that the novel chaperone-like properties described in chapter 5 are not the result of non-specific binding through hydrophobic regions that could be present if this segment were unstructured. Our extensive controls with BSA support our conclusions that the observed properties of RegA cannot be attributed to irreversible, non-specific binding. However, we found that none of the newly identified properties of saccin are ATP-dependent. Although this was at first unexpected, the extensively documented properties of Hsp90 reveal that this general chaperone also does not depend on ATP binding or hydrolysis for its *in vitro* chaperone functions (Jakob et al., 1995; Johnson et al., 2000; Wiech et al., 1992). However, Hsp90 does require nucleotide for its activity *in vivo* (Obermann et al., 1998; Ostrovsky et al., 2009; Panaretou et al., 1998). This apparent discrepancy may be related to the complex regulation of chaperone networks in the cell, which depend on multiple associations between general chaperones and co-chaperones (Grenert et al., 1997; Johnson et al., 2007). Thus, for Hsp90, co-chaperone interactions, structural transitions and nucleotide binding and hydrolysis are part of a complex cycle that promotes efficient folding *in vivo* (Hessling et al., 2009; Mickler et al., 2009). However, these activities are not all required for certain *in vitro* experimental paradigms of Hsp90 chaperone function.

Similar processes may operate on saccin, where abrogation of ATPase activity is critical for its *in vivo* function, as a human disease-causing mutation results in abrogation of this function (Figure 4.9). Thus, saccin likely collaborates with other modules involved in protein quality control in the cell, either within its own polypeptide, or with downstream members of the folding and proteolytic machinery of the cell. Future work may reveal functional partners of saccin, likely including members of the eukaryotic

Hsp70 family, as well as potential client polypeptides that necessitate sasin activity to acquire and/or maintain their structure and promote neuronal survival.

A critical aspect of the chaperone function of sasin that remains to be determined is the identity of its client proteins. We identified MAP1A as an interactor of sasin (Chapter 6); however, this protein did not interact with the chaperone active RegA fragment and thus we considered it unlikely to be a client (at least for the first SRR in sasin). Alternatively, the MAP1A interaction may be a localization strategy. This protein is a member of the neuronal cytoskeleton and is found at discrete sites throughout the complex architecture of the Purkinje cells of the cerebellum (Brenman et al., 1998). Perhaps MAP1A localizes sasin to regions of susceptibility to protein misfolding. We attempted to identify client proteins from the sasin knockout mouse model. Our initial attempts were unsuccessful, but further experiments will likely reveal the identity of these proteins.

Chapter 9: Conclusions and Perspectives

CONCLUSIONS

Sacsin is a large multi-domain protein which contains regions of similarity to both molecular chaperones and co-chaperones. Our analysis of sacsins sequence revealed the presence of unique supra-domains we termed sacsins repeating regions (SRR), present three times in human sacsins and animal sacsins-like homologs. We have assigned both ATPase activity and molecular chaperone activity to the SRR. We also demonstrated that the co-chaperone like region of sacsins, the J domain, possesses *bona fide* co-chaperone activity. Thus, sacsins represents a new class of chaperone with both chaperone and co-chaperone activities within a single polypeptide chain. Finally, we demonstrated that sacsins interacts with a component of the neuronal cytoskeleton, MAP1A.

PERSPECTIVES

Several outstanding issues related to sacsins function require further investigation. The explanation for the repeating arrangement of the SRR domain is unclear. Experiments directed at determining the precise function of each SRR domain may shed light on this arrangement; for example, these may possess differing substrate specificities or may engage in cooperative or synergistic protein folding activities. Generating various constructs of each SRR individually or in combination and performing chaperone and ATPase activity experiments with these in cis or in trans configurations could shed light on these issues. Additionally, the role of ATP hydrolysis in sacsins function requires further investigation. Based on our sequence and biochemical analyses with wild-type and D168Y RegA, we suggest that SRRs necessitate nucleotide hydrolysis for their function, provided by the common Hsp90 ATPase domain, which, when coupled to the

unique adjacent sequence, gives rise to a novel activity related to protein quality control. However, we found that the chaperone activity of sarsin *in vitro* does not require ATP and the mutant is functional in these experiments as well. Studies focusing on the interactors of sarsin may help in elucidating the role of ATP hydrolysis for its function, as Hsp90 utilizes its ATPase activity to regulate structural transitions that are related to the formation of protein heterocomplexes (Grenert et al., 1999). Once protein cofactors are identified, the interactions with these should be examined for ATP dependence. Additionally, cellular studies on the protein folding activities of sarsin may shed light on the role of nucleotide hydrolysis, as Hsp90 does not require ATP to assist in protein folding *in vitro*, but does require it for function *in vivo* (Obermann et al., 1998). Generating ATPase deficient and wild-type constructs to transfect into a sarsin deficient cell line would allow the determination of the role of the ATPase activity of sarsin *in vivo*. Finally, the pathogenesis of ARSACS will likely become clearer when client proteins that necessitate folding assistance from sarsin are identified. The knockout mouse represents an invaluable resource for determining the identity of these clients and if explored in more detail will likely yield positive results.

SECTION 4: MATERIALS AND METHODS

Chapter 10: Methods

A. BIOINFORMATICS ANALYSIS

Sacsin homologs were initially identified through a web-server based BLAST 2.2.14 (Altschul et al., 1990) used to search the non-redundant (NR) sequence annotation database. Pair-wise sequence alignments were constructed utilizing the BLAST two sequences method. Internal repeats were identified by using the web-based algorithm Dotlet (Junier and Pagni, 2000). All SRR domains in the NR database were identified utilizing a PSI-BLAST search with the C-terminal portion of the third SRR in murine sacsin and the default parameters. This portion of the SRR was utilized to avoid a bias we observed towards full length Hsp90 homologs if the entire SRR was utilized as the query. All multiple sequence alignments were constructed utilizing the E-INS-i strategy in MAFFT (Katoh and Toh, 2008). Secondary structure prediction was performed by submitting the alignment of all SRR domains to the Jpred server (Cole et al., 2008). Phylogenetic tree construction was conducted *via* the neighbor-joining method in MEGA4 (Tamura et al., 2007) and is presented as the bootstrap consensus tree from 1000 replicates. Protein structures were visualized and manipulated utilizing PyMOL v0.99rc6 (Delano scientific).

B. BUFFER COMPOSITIONS

Lysis buffer	30 mM Tris-HCl [pH 7.4 @ 4 °C]
	150 mM NaCl
	1X protease inhibitors without EDTA [Amresco]

Buffer A	30 mM Tris-HCl [pH 7.4 @ 4 °C] 150 mM NaCl
Buffer B	30 mM Tris-HCl [pH 7.4 @ 4 °C] 150 mM NaCl 1 M imidazole pH 7.4
PEM	100 mM PIPES pH 6.6 1 mM EGTA 1 mM MgSO ₄
Separating gel	7.5, 10, 12, or 15% Acrylamide 375 mM Tris, pH 8.8 (RT) 0.1% SDS 0.5 M Sucrose
Stacking gel	4.0% Acrylamide 125 mM Tris pH 6.8 (RT) 0.1% SDS
5X SDS sample buffer	310 mM Tris pH 6.8 (RT) 50% Glycerol 5% SDS 5% β-mercaptoethanol 0.1% Bromophenol blue
10X SDS running buffer	250 mM Tris 1.9 M Glycine 1.0% SDS
Coomassie blue stain	0.025% Coomassie brilliant blue R250 40% Methanol 7% Acetic acid
Coomassie destain	40% Methanol 7% Acetic acid
Stripping buffer	100 mM β-mercaptoethanol 2.0% SDS 62.5 mM Tris pH 6.7 (RT)
LB media	10 g Tryptone 5 g Yeast extract 10 g NaCl

	1 ml 1 M NaOH 15 g Agar (if preparing plates) 1 L final volume (with H ₂ O)
YPAD media	0.0075% l-adenine hemisulfate salt 1.0% Yeast extract 2.0% Peptone 2.0% Dextrose 2.0% Agar (if preparing plates)
SD media	6.7 g Yeast nitrogen base (w/o amino acids) 20 g Dextrose 1.3 g Amino acid dropout powder (w/o histidine) 20 g Agar (if preparing plates) 1 L final volume (with H ₂ O)
SG media	6.7 g Yeast nitrogen base (w/o amino acids) 20 g Galactose 1.3 g Amino acid dropout powder (w/o histidine) 20 g Agar (if preparing plates) 1 L final volume
Histidine dropout powder	40 mg/ml Adenine sulfate 20 mg/ml L-arginine (HCl) 100 mg/ml L-aspartic acid 100 mg/ml L-glutamic acid 60 mg/ml L-leucine 30 mg/ml L-lysine 20 mg/ml L-methionine 50 mg/ml L-phenylalanine 375 mg/ml L-serine 200 mg/ml L-threonine 40 mg/ml L-tryptophan 30 mg/ml L-tyrosine 150 mg/ml L-valine 20 mg/ml L-uracil
1X TBST	50 mM Tris pH 8.0 (RT) 150 mM NaCl 0.05% Tween-20
1X PBS	8 g NaCl 0.2 g KCl 1.44 g Na ₂ HPO ₄

	0.24 g KH_2PO_4 1 L final volume
Blocking solution	1X TBST 5.0% powdered milk
50X TAE	242 g Tris base 57.1 ml Glacial acetic acid 100 ml 0.5 M EDTA 1 L final volume
10X luciferase buffer	250 mM HEPES-KOH pH 7.4 500 mM KOAc 50 mM MgOAc
Luciferase stop buffer	25 mM Tris-Phosphate pH 7.8 2 mM DTT 2 mM EDTA 10% glycerol 1% triton X-100 1 mg/ml BSA

C. CLONING

All sacsin constructs were cloned from murine chromosomal DNA (C5/BL6) or cDNA plasmids (RIKEN). Phusion was utilized for DNA amplification according to manufacturer recommendations (Finnzymes). Primers were obtained from Sigma. All restriction digestions were performed according to the manufacturer recommendations (New England Biolabs; Fermentas). Five-prime phosphate moieties were removed from the vectors utilizing shrimp alkaline phosphatase (Fermentas). Ligations were performed overnight at 4 °C with 20-100 ng of dephosphorylated vector and a 4-7 fold molar excess of insert and T4 DNA ligase (New England Biolabs). Transformations were accomplished utilizing chemically competent DH5 α *E. coli* cells according to manufacturer recommendations (Invitrogen). The D168Y mutant in RegA was generated by ordering a synthetic mutant construct (DNA 2.0) which overlapped unique restriction

sites in the vector. pProEX (Invitrogen) was utilized for expression of constructs in *E. coli*. pESC-HIS (Agilent Technologies) was utilized for expression of constructs in *S. cerevisiae*. pVL1392 (bd biosciences) was utilized for generating recombinant baculovirus for expression of constructs in *S. frugiperda*.

Plasmid DNA was prepared utilizing the QIAprep spin miniprep kit (Qiagen). PCR products were processed utilizing the QIAquick PCR purification kit (Qiagen). DNA was purified was agarose gels utilizing the QIAquick gel extraction kit (Qiagen). Agarose gels were separated in a Horizon 58 apparatus (Whatman) and visualized under ultraviolet light. DNA concentrations were determined by absorbance at 260 nm utilizing the nanodrop absorbance spectrophotometer (Thermo scientific).

A recombinant baculovirus containing RegA (bvRegA) was generated by co-transfecting a pVL1392 plasmid containing RegA with baculogold DNA (bd biosciences) into Sf9 insect cells according to the manufacturer protocol. The bvRegA virus contained in the supernatant of the co-transfection was amplified by adding 1 ml to 2×10^7 Sf9 cells in 30 ml media in a 15 cm plate and incubating 7 days. The amplification was performed a total of three times. A large volume/high titer stock was generated by infecting 500 ml of Sf9 cells (5×10^5 cells/ml) with 12 mls of the third amplification for 7 days in a spinner flask. After 7 days, the cells were removed by centrifugation and the high-titer stock contained in the supernatant was aliquoted into 50 ml falcon tubes and stored in the dark at 4 °C.

D. PURIFICATION OF PROTEINS

RegA

A stock of Sf9 cells were maintained in logarithmic phase in a spinner flask at 25 °C. This stock was utilized to seed twenty fifteen-centimeter plates with 2×10^7 cells per

plate in 30 ml Insect Xpress (Lonza) with 10% fetal bovine serum (FBS) (Gibco) and 1% penicillin/streptomycin (Gibco). The cells were allowed to attach to the plate for at least thirty minutes. After adhering tightly, the cells were infected with a high-titer stock of bvRegA (1 ml per plate) by mixing in the virus drop-wise and gently rotating the plate for thorough mixing. Protein production was accomplished by incubating for 72 hours at 25 °C. The cells were then harvested by gently sloughing them off with a serological pipette in the media they were incubated in. The cells were collected by centrifugation at 300 x g for 15 minutes at 4 °C in 50 ml falcon tubes. The media was discarded and the cells were used immediately or stored at -80 °C.

The cell pellets from 20 plates were resuspended in 1 ml lysis buffer (30 mM Tris-HCl [pH 7.4 @ 4 °C], 150 mM NaCl, 1X protease inhibitors without EDTA [Amresco]) per plate (20 ml total per purification). Lysis was accomplished by sonnication. The cell suspension was placed in a 50 ml Falcon test-tube in a bucket of wet ice and a tip sonnicator (Fisher) was immersed in the suspension and turned to power setting of 10-15 for ~10 seconds. Ten pulses of sonnication were sufficient to fully disrupt the cells. The cell debris and insoluble material were removed by centrifugation at 30,000 x g for 30 minutes at 4 °C. The supernatant was the input for IMAC.

A 20 ml HisTrap FF column (GE Healthcare) was equilibrated Buffer A (30 mM Tris-HCl [pH 7.4 @ 4 °C], 150 mM NaCl), washed with Buffer B (Buffer A and 1 M imidazole pH 7.4), then re-equilibrated in Buffer A using an AKTA chromatography system (GE Healthcare). The supernatant was loaded onto the column utilizing a sample pump at a flow rate of 1.0 ml/minute. The flow-through was collected utilizing a fraction collector. The column was washed extensively with Buffer A at the maximum flow rate compatible with the pressure limit for the column until baseline absorbance values were

achieved. Once baseline was reached, the buffer was switched to 2.5% Buffer B for a 25 mM imidazole wash. The absorbance peaks were collected and the column was extensively washed until baseline was achieved. Once baseline was reached, the buffer was switched to 6.5% Buffer B for a 65 mM imidazole wash. The absorbance peaks were collected and the column was extensively washed until baseline was achieved. Once baseline was achieved, the flow rate was reduced to 1.0 mL/minute and the buffer was switched to 100% Buffer B for a 1.0 M imidazole elution and 750 μ L fractions were collected.

Fractions containing high concentrations of RegA were pooled in pairs for a total volume of 1.5 ml. These were then desalted over a 5 ml HiTrap Sephadex G-25 desalting column (GE Healthcare) equilibrated in Buffer A. The column was loaded with a 2.0 mL sample loop at a flow rate of 5.0 ml/minute. A 2.0 ml fraction was collected beginning 1.5 ml from the start of sample injection. This fraction contained the RegA peak and excluded the imidazole. Glycerol was added to the elution fraction to a final of 10% v/v (*i.e.* 222 μ l 100% glycerol per 2 ml sample). The samples were aliquoted and flash frozen in liquid nitrogen and stored at -80 °C.

DnaK

E. coli DnaK was purified from BL21 *E. coli* cells overexpressing the wild type protein, as described (Szabo et al., 1994). Harvested cells were resuspended in 50 mM MOPS pH 7.6, 20 mM KCl, 1 mM EDTA, 1 mM DTT, 1X protease inhibitors w/o EDTA (Amresco) and subsequently disrupted in an EmulsiFlex C3 (Avestin) and clarified by centrifugation at 30,000 x g for 30 minutes at 4 °C. The supernatant was loaded on a DE52 column and eluted with a linear gradient of KCl in the lysis buffer. Fractions containing purified DnaK were collected, pooled, concentrated by

ultrafiltration. Glycerol was added to 10% (v/v) and aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

DnaJ

E. coli DnaJ was purified from BL21 *E. coli* cells overexpressing the wild type protein as described (Szabo et al., 1994). Harvested cells were resuspended in 20 mM Tris pH 6.8, 1 mM DTT, 0.6% Brij-58, 1X protease inhibitors w/o EDTA (Amresco) and disrupted in an EmulsiFlex C3 (Avestin) and clarified by centrifugation at 30,000 x g for 30 minutes at 4 °C. The supernatant was loaded onto a HiTrap SP column (GE Healthcare) pre-equilibrated with 20 mM Tris pH 6.8, 1 mM DTT, 0.6% Brij-58. After extensive washing with equilibration buffer, elution was accomplished with a linear gradient to 20 mM Tris pH 6.8, 1 mM DTT, 1 M KCl. Fractions containing pure DnaJ were pooled and glycerol was added to 10% (v/v). Aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

GrpE

A His₆-tagged version of *E. coli* GrpE was purified as described from BL21 *E. coli* cells (Szabo et al., 1994). Cells were harvested and lysed in 1X PBS, 0.1% Triton X-100, 1X protease inhibitors w/o EDTA (Amresco) and disrupted in an EmulsiFlex C3 (Avestin) and clarified by centrifugation at 30,000 x g for 30 minutes at 4 °C. The supernatant was loaded on a HisTrap column (GE Healthcare) and washed with 1X PBS containing 50 mM imidazole and subsequently eluted with 1X PBS containing 0.5 M imidazole. Fractions containing pure GrpE were collected and desalted into 1X PBS. Glycerol was added to 10% (v/v) and aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

Sacsin J Domain

Murine J domain was cloned as a His₆-tagged construct followed by a TEV protease cleavage site and residues 4309-4395 of the mouse sacsin homolog into *E. coli* expression vector pProEX. Protein expression was induced with 1 mM IPTG at 37 °C for 5 hours. Recombinant protein was purified utilizing a HisTrap Nickel column and Mono S cation exchange column (GE Healthcare).

MAP1A

Bovine brains were obtained from a local butcher shop. Two-hundred grams of gray matter was homogenized in 250 ml PEM with 1X protease inhibitors and 1 mM PMSF using a potter homogenizer. The extract was centrifuged at 30,000 x *g* for 15 minutes (4 °C) and the pellet discarded. The supernatant was centrifuged at 180,000 x *g* for 90 minutes (4 °C) and the pellet discarded. Microtubules were polymerized by adding taxol to 20 μM and GTP to 1 mM and incubating at 37 °C for 15 minutes. The microtubules and associated proteins were centrifuged at 100,000 x *g* for 30 minutes at 30 °C through a cushion of 5% sucrose in PEM with 20 μM taxol and 1 mM GTP. The supernatant was discarded. The pellet was washed in 1/8 volume of PEM with 20 μM taxol and 1 mM GTP by incubating with agitation for 15 minutes at 37 °C. The microtubules were sedimented by centrifugation at 30,000 x *g* for 30 minutes at 30 °C and the supernatant was discarded. The microtubule associated proteins were dissociated by resuspending the washed microtubule pellet in 1 volume (1/8 original volume) of PEM with 1 mM GTP, 20 μM taxol, 350 mM NaCl, and 10 μM poly-L-aspartic acid and incubating at 37 °C for 15 minutes. The microtubules were sedimented by centrifugation at 30,000 x *g* for 30 minutes at 30 °C. The pellet was discarded and a crude MAP preparation was obtained from the supernatant.

E. CELL CULTURE

E. coli cells were grown on Luria-Bertani media (LB) with the appropriate antibiotics at 37 °C; in liquid cultures, shaking was included at 300 RPM. The antibiotics utilized were, ampicillin (100 µg/ml final), kanamycin (50 µg/ml final), streptomycin (100 µg/ml final). All cloning was performed in DH5α cells (F- ϕ 80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (*r_k*-, *m_k*+) *phoA supE44 λ- thi-1 gyrA96 relA1*), while protein expression was accomplished in BL21 cells (F- *ompT hsdSB* (rB-mB-) *gal dcm rne131*(DE3)).

S. cerevisiae cells were grown in YPAD agar or broth at 30 °C, if in liquid cultures, shaking was included at 300 RPM. Protein expression was performed in YPH499 cells (*ura3-52 lys 2-801^{amber} ade2-101^{ochre} trp1-Δ63 his3-Δ200 keu2-Δ1*). Induction was accomplished by growing the cells in SD media to the density desired, gently spinning the cells and washing with water, then resuspending in SG media.

S. frugiperda Sf9 insect cells were obtained from bd biosciences and grown in Insect Xpress media with 10% FBS and 1% penicillin/streptomycin. Cells were maintained in logarithmic phase in 75 cm² flasks at 25 °C with passaging every 3-4 days (*e.g.* when cells reached confluency, they were passaged to 1/3 confluency). When grown in suspension, cell cultures were initiated at 5 x 10⁵ cells/ml and allowed to grow to 2 x 10⁶ cells/ml before diluting back.

F. SDS-PAGE

SDS-PAGE was performed utilizing the mini-PROTEAN system (Bio-Rad). The Separating gel was cast by combining 3.5 ml separating solution with 35 µl 10% APS and 1.7 µl TEMED. This solution was added to the gel and overlaid with 70% ethanol until fully polymerized. The stacking gel was cast by combining 1.5 ml stacking solution

with 15 μ l 10% APS and 1.5 μ l TEMED. Samples for SDS-PAGE were prepared by adding sample buffer to a final of 1X and incubating at 95 °C for 10 minutes, then spinning at 20,000 x *g* for 10 minutes.

G. WESTERN BLOTS

SDS-PAGE was performed as usual. The gel was transferred to a nitrocellulose membrane utilizing the iBlot dry blotting device (Invitrogen) according to the manufacturer recommendations. Blocking was performed at 25 °C for 1.0 hours with shaking (or overnight at 4 °C with shaking). The membrane was incubated with primary antibody for 1.0 hours at 25 °C with shaking, followed by three washes with TBST, 15 minutes each. The membrane was incubated with secondary antibody for 1.0 hours at 25 °C with shaking, followed by three washes with TBST, 15 minutes each. The reaction was developed utilizing the extended chemiluminescence reagent (Amersham) according to the manufacturer recommendations. Film was exposed and developed in a darkroom.

H. FLUORESCENCE SPECTROSCOPY

The emission spectra of *bis*-ANS was determined in the presence of either wild type or mutant SRR. 100 nM protein was combined with 3 μ M *bis*-ANS in 30 mM Tris 7.4 and 150 mM NaCl. The samples were excited at 390 and emission spectra recorded from 400 to 600 nm. The background spectrum of *bis*-ANS was subtracted. Tryptophan emission spectra of wild type and mutant SRR were obtained with 1 μ M protein in 30 mM Tris 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.

I. IMMUNOPRECIPITATIONS

Sacsin interacting proteins were identified by incubating mouse brain and cerebellar extracts (10 and 15 g/l total protein, respectively) with purified DnaJ domain (80 μ M) or GFP (80 μ M) immobilized on Ni-NTA agarose beads in lysis buffer (20 mM HEPES-KOH pH 7.4, 150 mM NaCl, 0.1% Triton X-100, 10% glycerol, 1X EDTA-free protease inhibitors [Amresco]). After extensive washing with lysis buffer, proteins were eluted with 500 mM imidazole and analyzed by SDS-PAGE.

Co-immunoprecipitations were performed with lysates as above, utilizing 50 μ l of anti-MAP1A antibody (0.2 g/l; Santa Cruz Biotechnology) immobilized on protein-A beads, washing extensively, and eluted by boiling in SDS-PAGE sample buffer. Samples were run on SDS-PAGE and electroblotted onto nitrocellulose membranes (Bio-Rad). Western blotting was performed with a primary anti-Sacsin antibody (0.25 g/l; bd biosciences) followed by a secondary HRP-linked anti-mouse IgG antibody.

J. LUCIFERASE REFOLDINGS

Luciferase from *Photinus pyralis* was obtained from Sigma. The lyophilized powder was resuspended in 0.5 M Tris-succinate pH 7.7. Five or 10 μ M luciferase was denatured in 6.0 M GdmCl and 5 mM DTT for 30 minutes at room temperature and then kept on ice. The standard refolding reaction was carried out in 1X luciferase buffer with 3 mM ATP, 5 mM DTT, and 0.1 mg/mL BSA. Measurements were made by diluting an aliquot of the refolding reaction 1:100 in stop buffer, then 1:10 in assay buffer (Promega) followed by measuring luminescence in a luminometer (Sirius) with a delay time of 1.0 seconds and measurement time of 2.0 seconds.

K. DOT BLOT ANALYSIS

Luciferase refolding reactions were performed as above. At 30 minutes, an aliquot was removed from each reaction (and kept as a reference for the total reaction) and the remainder was centrifuged for 30 minutes at 20,000 x g at 4 °C. The supernatant was removed and 1 volume of 2X SDS sample buffer was added to 1 volume of total and super samples. The pellet was resuspended in an equivalent volume of 1X SDS sample buffer. The reactions were spotted onto a nitrocellulose membrane utilizing the Bio-Dot microfiltration apparatus (Bio-Rad). After extensive washing, the membrane was removed and blocked with 5% milk in TBS with Tween-20. The membrane was then treated as a traditional western-blot and probed with anti-luciferase (1:2,500; Millipore) and HRP-conjugated goat-anti-mouse (1:5,000; Pierce), followed by incubation with ECL (GE Healthcare) and exposure to film.

L. IMMUNOFLUORESCENCE

Male Sprague-Dawley rats (250-350 g) were perfused with phosphate-buffered saline (PBS) and subsequently fixative (4% paraformaldehyde, 10% picric acid, 0.1% glutaraldehyde) with an overnight cryoprotectant including 30% sucrose. Cerebellar slices (30 µm) were blocked with blocking solution (5% donkey serum, 0.05% Triton X-100 in PBS) for 1 h at room temperature, washed with PBS, incubated with primary antibody anti-MAP1A (same as above, 1:200) or anti-Sacs1 antibody (same as above, 1:200) in blocking solution overnight at room temperature, washed with PBS, incubated with secondary antibody donkey anti-goat IgG-rhodamine (1:100) or donkey anti-mouse IgG-fluorescein (1:100) in blocking solution 2 h at room temperature, washed in PBS, mounted in mounting medium with anti-bleaching agent Fluorsave (Calbiochem). Images

were obtained by using an Olympus 300 confocal system equipped with a 603 objective lens, argon, HeNe (R) and HeNe (G) lasers, and Fluoview software (Olympus).

M. ATPASE ASSAYS

ATPase assays were performed in a 96-well format utilizing a plate reader (BMG Labtech). Reactions were performed as described in the main text in a microfuge tube utilizing a thermomixer (eppendorf). At the completion of the reaction, 80 μ l was removed and combined with 20 μ l of the malachite green reagent (Bioassays) and mixed thoroughly by repeated pipetting. The incubated at room temperature for 10 minutes. The absorbance values at 600 nm were determined. These values were background subtracted and/or normalized to a standard curve.

N. METHANOL/CHLOROFORM PRECIPITATION

Methanol/chloroform precipitations were performed by combining 1 volume of methanol and $\frac{1}{4}$ volume of chloroform with 1 volume of protein (e.g. 500 μ l and 125 μ l to 500 μ l respectively). After mixing well in a vortex, these were centrifuged at max speed for 5 minutes. The aqueous phase (upper phase) was discarded, taking care to leave the interface intact (1~100 μ l of liquid). Two-thirds volume of methanol (e.g. 350 μ l) was added. After mixing well in a vortex, these were centrifuged at max speed for 5 minutes. The supernatant was discarded and the pellet was incubated at 37 °C for ~10 minutes until dry (with cap open). The pellet was resuspend in a suitable volume of 1X SDS sample buffer and analyzed by SDS-PAGE.

SECTION 5: APPENDICES

Table A.1: Mutations identified in human saccin

Mutation	Amino Acid Change	Domain	Mutation Type	Origin	Reference
44365delT	C72 fsX76		Frameshift (del)	Belgium	(Baets et al., 2010)
57029delA	N161 fsX175	SRRa (90-like)	Frameshift (del)	Japan	(Kamada et al., 2008)
G57049T	D168Y	SRRa (90-like)	Missense	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
57147_57152 delAACAGG	I200 fsX206	SRRa (90-like)	Frameshift (del)	Italy	(Terracciano et al., 2009)
C57149A	T201K	SRRa (90-like)	Missense	Belgium	(Baets et al., 2010)
C59688T	R272C	SRRa (90-like)	Missense	Canada	(Guernsey et al., 2010)
C59796T	L308F	SRRa (C-term)	Missense	Japan	(Takado et al., 2007)
C59835T	R321X	SRRa (C-term)	Nonsense	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
60058-60067 delGTAACAGTGT	C395 fsX407	SRRa (C-term)	Frameshift (del)	Japan	(Ouyang et al., 2006)
G60349A	W492X	SRRa (C-term)	Nonsense	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
C60481T	P536L		Missense	France	(Anheim et al., 2009)
T60541C	L556P		Missense	Morocco	(Baets et al., 2010)
60934delA	D687 fsX713		Frameshift (del)	Japan	(Ouyang et al., 2006)
C61698T	R728X		Nonsense	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
T74015C	L802P		Missense	Japan	(Kamada et al., 2008)
T74581C	C991R		Missense	Belgium	(Baets et al.,

Mutation	Amino Acid Change	Domain	Mutation Type	Origin	Reference
					2010)
T74771C	F1054S		Missense	Japan	(Shimazaki et al., 2005)
74938insA	I1110 fsX1111		Frameshift (ins)	Tunisia	(El Euch-Fayache et al., 2003)
75031insAC	I1141 fsX1150		Frameshift (ins)	Belgium	(Baets et al., 2010)
75195delT	I1195 fsX1206		Frameshift (del)	Tunisia	(El Euch-Fayache et al., 2003)
T75542A	M1311K		Missense	Belgium	(Ouyang et al., 2007)
C75643T	Q1345X		Nonsense	Japan	(Okawa et al., 2006)
75643insC	Q1345 fsX1349		Frameshift (ins)	Italy	(Criscuolo et al., 2004)
C75718T	Q1370X		Nonsense	Italy	(Grieco et al., 2004)
T75802C	C1398R		Missense	Turkey	(Richter et al., 2004)
G76334C	R1575P	SRRb (90-like)	Missense	Serbia	(Baets et al., 2010)
A76370G	H1587R	SRRb (90-like)	Missense	Belgium	(Baets et al., 2010)
G76567T	E1653X	SRRb (C-term)	Nonsense	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
C76735T	Q1709X	SRRb (C-term)	Nonsense	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
A76753T	K1715X	SRRb (C-term)	Nonsense	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
76811-76812 delAG	E1734 fsX1736	SRRb (C-term)	Frameshift (del)	Japan	(Yamamoto et al., 2005)
C77239T	R1877X	SRRb (C-term)	Nonsense	Italy	(Anesi et al., 2011)
T77446C	W1946R		Missense	Tunisia	(El Euch-Fayache et al., 2003)

Mutation	Amino Acid Change	Domain	Mutation Type	Origin	Reference
77598-77599 delCT	D1996 fsX1999		Frameshift (del)	Japan	(Shimazaki et al., 2005)
77611-77614 delAGAA	R2002 fsX2013		Frameshift (del)	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
77703-77705 delTTC	2032delS		Deletion	Serbia	(Baets et al., 2010)
77782delA	S2058 fsX2076		Frameshift (del)	Japan	(Yamamoto et al., 2005)
C77965T	R2119X		Nonsense	Japan	(Hara et al., 2007)
78444insA	K2279 fsX2290		Frameshift (ins)	Italy	(Grieco et al., 2004)
T78731C	L2374S		Missense	Italy	(Terracciano et al., 2009)
78860-78864 delCAGAA	T2417 fsX2429		Frameshift (del)	Italy	(Grieco et al., 2004)
C78886T	R2426X		Nonsense	Belgium	(Baets et al., 2010)
78984delT	L2458 fsX2474		Frameshift (del)	Belgium	(Baets et al., 2010)
C79114T	R2502X		Nonsense	Canada	(Engert et al., 2000)
C79283T	A2558V	SRRc (90-like)	Missense	France	(Anheim et al., 2009)
C79717T	R2703C	SRRc (90-like)	Missense	Spain	(Criscuolo et al., 2005)
C80003A	P2798Q	SRRc (C-term)	Missense	Morocco	(Baets et al., 2010)
80011-80013 delCAA	2801delQ	SRRc (C-term)	Deletion	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
80403delA	K2931 fsX2952		Frameshift (del)	Japan	(Hara et al., 2005)
80454delT	P2948 fsX2952		Frameshift (del)	Canada	(Engert et al., 2000)
T81352C	W3248R		Missense	Japan	(Ogawa et al., 2004)
81521-81522 delTC	L3304 fsX3317		Frameshift (del)	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
81908delC	T3433 fsX2458		Frameshift (del)	Turkey	(Richter et al.,

Mutation	Amino Acid Change	Domain	Mutation Type	Origin	Reference
					2004)
T82052P	L3481P		Missense	UK	(Vermeer et al., 2009; Vermeer et al., 2008)
C82516T	R3636X		Nonsense	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
G82517A	R3636Q		Missense	The Netherlands	(Baets et al., 2010)
T82544C	L3645P		Missense	Belgium	(Baets et al., 2010)
C82564A	P3652T		Missense	Belgium	(Baets et al., 2010)
T82568C	F3653S		Missense	Belgium	(Baets et al., 2010)
82844-82845 delTT	L3745 fsX3746		Frameshift (del)	Belgium	(Baets et al., 2010)
82875-82876 delAT	I3755 fsX3762		Frameshift (del)	Hungary	(Baets et al., 2010)
C82984T	R3792X		Nonsense	Tunisia	(Bouhlal et al., 2008)
C83317T	R3903X		Nonsense	Canada	(Guernsey et al., 2010)
83439-83442 delAGTT	L3943 fsX3949		Frameshift (del)	Turkey	(Richter et al., 2004)
C83770T	Q4054X		Nonsense	The Netherlands	(Vermeer et al., 2008)
G83830C	A4074P		Missense	Tunisia	(El Euch-Fayache et al., 2003)
84461-84464 delAGAG	R4283 fsX4306		Frameshift (del)	Tunia	(Bouhlal et al., 2008)
C84583T	R4325X	DnaJ	Nonsense	Japan	(Yamamoto et al., 2006)
G84602A	R4331Q	DnaJ	Missense	The Netherlands	(Vermeer et al., 2009; Vermeer et al., 2008)
G84637A	E4343K	DnaJ	Missense	Belgium	(Baets et al., 2010)
C84742T	R4378X	DnaJ	Nonsense	Italy	(Anesi et al., 2011)

Mutation	Amino Acid Change	Domain	Mutation Type	Origin	Reference
A85133C	K4508T	HEPN	Missense	Belgium	(Baets et al., 2010)
A85255G	N4549D	HEPN	Missense	Turkey	(Richter et al., 2004)
SACS deletion	584 kb (2 genes)		Deletion	Canada	(McMillan et al., 2009)
SACS deletion	1.54 Mb (6 genes)		Deletion	Belgium, Italy	(Breckpot et al., 2008; Terracciano et al., 2009)
<p>The DNA numbering is according to AL157766 complementary. The protein numbering is according to NP_055178.</p>					

Kingdom	Organism	Name in Figure 3	Accession ID	Other ID's	Daltons	Amino Acids	SRR	SRR1	SRR2	SRR3	SRR4	XPCB	DnaJ	HEPN	UBL	SSL2	RING	BTB	HET	zf-C2HC plant	RpsA	Peptidase C14	RNFB	NADB_Rossmann	Helicase c	DEXDc	Transglut core	STKc_CDK9 like	AAT I	AAA 4
Animalia	<i>Ailuropida melanoleuca</i>	Amelanoleuca_I	EFB22146	GI:281346562	521938	4582	3	95-509	1457-1923	2525-2924		1	1	1	1															
	<i>Bos taurus</i>	Btaurus_I	XP_586753	GI:76631591	517642	4575	3	87-500	1450-1916	2518-2917		1	1	1	1															
	<i>Branchiostoma floridae</i>	Bfloridae_I	XP_002599568	GI:260809550	556595	4931	3	17-403	1334-1762	2309-2733		1																		
		Bfloridae_II	XP_002589277	GI:260788479	492834	4368	3	12-395	1328-1752	2303-2720		1	1	1																
		Bfloridae_III	XP_002606415	GI:260822048	510858	4521	3	126-523	1433-1866	2481-2899		1	1	1																
	<i>Canis familiaris</i>	Cafmiliaris_I	XP_534533	GI:73993520	518141	4549	3	87-501	1450-1916	2518-2917		1	1	1	1															
	<i>Danio rerio</i>	Drerio_I	XP_001921714	GI:189528701	510880	4488	3	2-416	1366-1834	2435-2834		1	1	1																
		Drerio_III	XP_001920014	GI:189525657	266830	2361	1	711-1020																						
		Drerio_III	XP_682866	GI:189523736	419815	3669	3	14-411	1058-1315	1885-2273		1	1	1																
	<i>Equus caballus</i>	Ecaballus_I	XP_001490776	GI:194221780	524395	4612	3	125-540	1487-1953	2555-2954		1	1	1	1															
	<i>Gallus gallus</i>	Ggallus_I	XP_417138	GI:118085040	514712	4524	3	35-449	1398-1867	2468-2867		1	1	1																
	<i>Homo sapiens</i>	Hsapiens_I	NP_055178	GI:163659918	521126	4579	3	90-505	1455-1921	2523-2922		1	1	1	1															
	<i>Hydra magnipapillata</i>	Hmagnipapillata_I	XP_002155553	GI:221129249	183945	1581	1	83-404																						
		Hmagnipapillata_II	XP_002157472	GI:221128489	329550	2867	2	429-843	1397-1790			1	1	1																
		Hmagnipapillata_III	XP_002166082	GI:221102659	430939	3739	3	1-382	1029-1413	1967-2361		1	1	1																
	<i>Macaca mulatta</i>	Mmulatta_I	XP_001089964	GI:109120189	514417	4579	3	90-505	1455-1921	2523-2922		1	1	1	1															
	<i>Mus musculus</i>	Mmusculus_I	Q9JLC8	GI:122066080	520684	4582	3	90-504	1457-1923	2525-2924		1	1	1	1															
	<i>Nematostella vectensis</i>	Nvectensis_I	XP_001636175	GI:156392678	506227	4477	3	39-431	1457-1894	2473-2875		1	1	1																
	<i>Ornithorhynchus anatinus</i>	Oanatinus_I	XP_001519222	GI:149598998	539479	4777	3	288-702	1652-2119	2720-3119		1	1	1	1															
	<i>Pan troglodytes</i>	Ptroglyodytes_I	XP_001174272	GI:114649156	549939	4851	3	362-777	1727-2193	2795-3194		1	1	1																
	<i>Rattus norvegicus</i>	Rnorvegicus_I	XP_001063390	GI:109502540	536248	4738	3	246-660	1613-2079	2681-3080		1	1	1	1															
	<i>Schistosoma mansoni</i>	Smansoni_I	XP_002573657	GI:256074692	348488	3070	1	1015-1390																						
	<i>Strongylocentrotus purpuratus</i>	Spurpuratus_I	XP_787004	GI:72057201	75826	668	1	47-442																						
		Spurpuratus_II	XP_001200196	GI:115665280	439356	3846	3	1-339	884-1289	1874-2264		1	1	1																
	<i>Taeniopygia guttata</i>	Tguttata_I	XP_002186744	GI:224077406	552620	4976	4	12-412	1287-1684	2478-2878	3345-3728				2															
		Tguttata_II	XP_002186552	GI:224077314	464977	4198	2	929-1326	2493-2876					1																
	<i>Tetraodon nigroviridis</i>	Tnigroviridis_I	CAG02848	GI:47215794	527227	4658	3	185-601	1533-2001	2602-3001		1	1	1	1															
	<i>Trichoplax</i>	Tadhaerens_I	XP_002107908	GI:195996079	168616	1467	1	38-																						

Kingdom	Organism	Name in Figure 3	Accession ID	Other ID's	Daltons	Amino Acids	SRR	SRR1	SRR2	SRR3	SRR4	XPCB	DnaJ	HEPN	UBL	SSL2	RING	BTB	HET	zf-C2HC plant	RpsA	Peptidase C14	RNFB	NADB_Rossmann	Helicase c	DEXDc	Transglut core	STKc_CDK9 like	AAT I	AAA 4
	<i>adhaerens</i>							363																						
		Tadhaerens_III	XP_002108695	GI:195997653	165322	1421	1	33-389																						
		Tadhaerens_III	XP_002108697	GI:195997657	65549	558	1	32-358																						
		Tadhaerens_IV	XP_002108698	GI:195997659	165998	1419	1	32-358																						
Archaea	<i>Haloquadratum walsbyi</i>		YP_658972	GI:110669161	104461	916	1	8-384																						
	<i>Methanococcus marisnigri</i>		YP_001047385	GI:126179420	188523	1625	1	27-344																						
Bacteria	<i>Algoriphagus</i>		ZP_01718023	GI:126645479	217974	1880	1	217-610																						
	<i>Alkaliphilus metalliredigens</i>		YP_001319446	GI:150389397	192970	1675	1	9-361														1								
	<i>Anabaena variabilis</i>		YP_320509	GI:75812892	95557	828	1	5-449																						
	<i>Bacteroides capillosus</i>		ZP_02038342	GI:154500304	123587	1069	1	12-334																						
	<i>Bacteroides intestinalis</i>		ZP_03015341	GI:189466556	182515	1554	1	12-344																						
	<i>Bacteroides ovatus</i>		ZP_02068428	GI:160887425	214312	1862	1	464-866													1									
	<i>Beggiatoa</i>		ZP_01999887	GI:153870496	119092	1020	1	20-349																						
	<i>Brevibacterium linens</i>		ZP_05915904	GI:260907582	177814	1609	1	29-368																	1	1				
	<i>Clostridium asparagiforme</i>		ZP_03760867	GI:225405678	124105	1103	1	17-392																						
	<i>Cyanotheca</i>		ZP_03139802	GI:196241037	295147	2593	1	24-422																						
	<i>Dehalococcoides</i>		YP_307289	GI:73748050	118532	1023	1	11-326																						
	<i>Desulfotobacterium hafniense</i>		YP_002457135	GI:219666700	181033	1584	1	7-349																		1				
	<i>Desulfotomaculum acetoxidans</i>		YP_003191024	GI:258514802	192477	1672	1	9-361														1								1
	<i>Dokdonia donghaensis</i>		ZP_01049227	GI:86130627	174968	1500	1	10-340																						
	<i>gamma proteobacterium NOR5-3</i>		ZP_05128578	GI:254516519	120667	1083	1	15-336																						
	<i>Gemmata obscuriglobus</i>		ZP_02730558	GI:168698281	204996	1863	1	34-463																						
	<i>Geodermatophilus obscurus</i>		ZP_03891247	GI:227408016	77237	731	1	35-380																						
	<i>Haliangium ochraceum</i>		YP_003270158	GI:262198949	271609	2543	1	188-532																			1			
	<i>Klebsiella pneumoniae</i>		YP_002237605	GI:206577566	166113	1477	1	61-362																						
	<i>Listeria monocytogenes</i>		ZP_05251627	GI:254878917	174617	1507	1	24-355																						
	<i>Lyngbya aestuarii</i>		ZP_01623418	GI:119491399	192570	1680	1	12-428																						

Kingdom	Organism	Name in Figure 3	Accession ID	Other ID's	Daltons	Amino Acids	SRR	SRR1	SRR2	SRR3	SRR4	XPCB	DnaJ	HEPN	UBL	SSL2	RING	BTB	HET	zf-C2HC plant	RpsA	Peptidase C14	RNFB	NADB_Rossmann	Helicase c	DEXDc	Transglut core	STKc_CDK9 like	AAT I	AAA 4
	<i>marine gamma proteobacterium HTCC2148</i>		ZP_05095019	GI:254481776	117154	1032	1	18-340																						
	<i>Marinobacter</i>		ZP_01738589	GI:126667620	169257	1516	1	26-346																						
	<i>Methylococcus capsulatus</i>		YP_113238	GI:53805067	64519	583	1	7-339																						
	<i>Nitrobacter hamburgensis</i>		YP_578895	GI:92119166	143753	1306	1	30-352								1														
	<i>Nocardiopsis dassonvillei</i>		ZP_04331404	GI:229204943	114186	1084	1	26-377																						
	<i>Parvularcula bermudensis</i>		ZP_01017066	GI:84702491	161072	1418	1	21-321																						
	<i>Pectobacterium carotovorum</i>		ZP_03829157	GI:227115501	193201	1675	1	28-371														1								1
	<i>Prevotella veroralis</i>		ZP_05858383	GI:260592925	266562	2288	1	209-563																						
	<i>Pseudomonas aeruginosa</i>		YP_002439960	GI:218891094	177805	1603	1	11-343																						
	<i>Psychroflexus torquis</i>		ZP_01254470	GI:91217511	146032	1254	1	20-361																						
	<i>Rhodopseudomonas palustris</i>		NP_947564	GI:39935288	158131	1412	1	30-327																						
	<i>Roseobacter</i>		ZP_01903629	GI:149915101	273937	2459	1	1-398																						
	<i>Shewanella</i>		YP_735335	GI:113971542	192910	1687	1	19-362														1								1
	<i>Shewanella oneidensis</i>		NP_717736	GI:24373693	63431	554	1	26-347																						
	<i>Solibacter usitatus</i>		YP_822957	GI:116620801	180310	1586	1	17-328																	1					
	<i>Sphingomonas wittichii</i>		YP_001260706	GI:148553124	156896	1414	1	31-331																						
			YP_001260213	GI:148550774	211367	1934	1	125-482																						
	<i>Streptomyces clavuligerus</i>		ZP_05005534	GI:254390316	196907	1801	1	27-436																						
	<i>Streptomyces lividans</i>		ZP_05527832	GI:256789401	165149	1526	1	53-362																						
	<i>Syntrophus aciditrophicus</i>		YP_461435	GI:85859233	140074	1210	1	84-383																						
	<i>Thermobifida fusca</i>		YP_288298	GI:72160641	110600	1029	1	28-383																						
	<i>Verminephrobacter eiseniae</i>		YP_997375	GI:121609568	46244	412	1	1-321																						
	<i>Verrucomicrobiae bacterium</i>		ZP_05057070	GI:254443594	168276	1483	1	12-328																						
	<i>Xanthobacter autotrophicus</i>		YP_001417878	GI:154246920	157045	1406	1	31-328																						
Fungi	<i>Ajellomyces capsulatus</i>		EEH05180	GI:225556893	178790	1578	1	84-422																						
	<i>Aspergillus fumigatus</i>		XP_751403	GI:159125686	183303	1608	1	46-381																						
			XP_751426	GI:70993156	172397	1516	1	30-349																						
			EDP48569	GI:159123450	180986	1590	1	38-																						

Kingdom	Organism	Name in Figure 3	Accession ID	Other ID's	Daltons	Amino Acids	SRR	SRR1	SRR2	SRR3	SRR4	XPCB	DnaJ	HEPN	UBL	SSL2	RING	BTB	HET	zf-C2HC plant	RpsA	Peptidase C14	RNFB	NADB_Rossmann	Helicase c	DEXDc	Transglut core	STKc_CDK9 like	AAT I	AAA 4
								379																						
	<i>Aspergillus nidulans</i>		XP_680549	GI:67900586	210332	1882	1	22-333																						
	<i>Aspergillus niger</i>		XP_001396587	GI:145250147	194838	1743	1	9-453																						
			XP_001395613	GI:145246728	216456	1917	1	13-383											1											
			XP_001398736	GI:145254753	192843	1689	1	31-438																						
	<i>Botryotinia fuckeliana</i>		XP_001550315	GI:154299794	292039	2580	2	10-383	1150-1525									1												
			XP_001546118	GI:154291063	162966	1440	1	39-377																						
	<i>Chaetomium globosum</i>		XP_001226495	GI:116201367	170834	1525	1	45-382																						
	<i>Coccidioides posadasii</i>		EER28290	GI:240110131	236187	2074	1	1-278											1											
	<i>Coprinopsis cinerea</i>		XP_001840246	GI:169867330	297512	2639	2	3-385	1277-1657									1												
			XP_001840262	GI:169867362	293380	2626	2	1-378	1273-1651									1												
	<i>Cryptococcus neoformans</i>		XP_778245	GI:134106469	188537	1697	1	1-469																						
	<i>Gibberella zeae</i>		XP_388273	GI:46127439	253110	2219	1	104-445											1											
			XP_390655	GI:46137929	214121	1889	1	41-374																						
			XP_382153	GI:46110190	184079	1612	1	32-376																						
	<i>Microsporum canis</i>		EEQ32626	GI:238842964	192576	1719	1	9-457																						
	<i>Monosiga brevicollis</i>		XP_001744867	GI:167521057	381273	3466	1	1320-1650																						
	<i>Nectria haematococca</i>		EEU33955	GI:256720532	168453	1482	1	46-345																						
			EEU46732	GI:256733385	197611	1747	1	62-379																						
			EEU37858	GI:256724488	255181	2274	1	70-380											1											
			EEU43645	GI:256730293	252517	2237	1	49-340											1											
			EEU48268	GI:256734922	197136	1738	1	66-378																						
			EEU45274	GI:256731925	233658	2071	1	47-340											1											
	<i>Neosartorya fischeri</i>		XP_001261825	GI:119483844	186826	1634	1	46-381																						
			XP_001266639	GI:119499764	183228	1609	1	11-327																						
	<i>Paracoccidioides brasiliensis</i>		EEH16790	GI:225678506	196286	1763	1	14-460																						
	<i>Penicillium chrysogenum</i>		XP_002566594	GI:255951655	156039	1376	1	25-298																						
	<i>Penicillium mameffeii</i>		XP_002147342	GI:212534372	187574	1645	1	24-350																						
	<i>Podospora anserina</i>		XP_001907983	GI:171686084	48227	420	1	1-294																						
			CAP65496	GI:170940269	184507	1640	1	60-																						

Kingdom	Organism	Name in Figure 3	Accession ID	Other ID's	Daltons	Amino Acids	SRR	SRR1	SRR2	SRR3	SRR4	XPCB	DnaJ	HEPN	UBL	SSL2	RING	BTB	HET	zf-C2HC plant	RpsA	Peptidase C14	RNFB	NADB_Rossmann	Helicase c	DEXDc	Transglut core	STKc_CDK9 like	AAT I	AAA 4
	<i>Sclerotinia sclerotiorum</i>		XP_001590508	GI:156049083	291232	2571	2	381-10-328	1158-1534									1												
	<i>Uncinocarpus reesii</i>		XP_002543539	GI:258569471	188226	1694	1	15-458																						
			XP_002584819	GI:258568150	182125	1609	1	23-335																						
	<i>Ustilago maydis</i>		XP_761281	GI:71022101	207546	1873	1	20-507																						
	<i>Verticillium albo-atrum</i>		EEY15539	GI:261353111	165024	1457	1	33-355																						
			EEY15701	GI:261353273	168694	1505	1	30-342																						
			EEY19489	GI:261357061	177824	1583	1	48-376																						
Plantae	<i>Arabidopsis thaliana</i>		NP_197702	GI:15237223	527251	4706	3	10-419	1382-1823	2773-3183							1													
			NP_190446	GI:15229022	215803	1899	1	215-562												1										
			CAB36839	GI:4455304	240109	2137	1	594-912																						
			NP_193111	GI:186511742	306348	2729	1	1183-1504																						
	<i>Micromonas pusilla</i>		EEH54087	GI:226456787	571841	5511	3	16-440	1477-1920	3098-3643							1							1						
	<i>Oryza officinalis</i>		CBG76273	GI:260446991	559030	5010	3	12-422	1290-1738	2648-3054			1										1							
	<i>Oryza sativa (Indica Cultivar)</i>		CAJ86102	GI:90399053	530206	4737	3	12-422	1385-1833	2782-3190							1													
			EEC69267	GI:218186840	205332	1815	1	27-379																						
			EEC76364	GI:218193937	212259	1909	1	1187-1500																						
			EAY81073	GI:125534525	192481	1695	1	19-370																						
	<i>Oryza sativa (Japonica Cultivar)</i>		EAZ18499	GI:125577277	190122	1674	1	19-370																						
			NP_001067964	GI:115485641	190167	1674	1	21-370																						
			EEE53198	GI:222617066	205392	1815	1	27-379																						
			ABA93879	GI:77551082	191065	1707	1	9-380																						
			EEE52167	GI:222616035	189435	1690	1	1-363																						
			ABF99416	GI:108711621	296961	2655	1	1100-1439																						
			EEE60127	GI:222625995	315955	2821	1	1180-1519																						
			ABF99415	GI:108711620	302363	2702	1	1147-1486																						
			AAO72367	GI:29150358	289406	2592	1	1189-1511																						
			EEE61898	GI:222629766	519050	4635	3	12-422	1385-1833	2782-3146							1													
			CAE03243	GI:38344927	522090	4666	3	12-404	1367-1815	2764-3172							1													

Kingdom	Organism	Name in Figure 3	Accession ID	Other ID's	Daltons	Amino Acids	SRR	SRR1	SRR2	SRR3	SRR4	XPCB	DnaJ	HEPN	UBL	SSL2	RING	BTB	HET	zf-C2HC plant	RpsA	Peptidase C14	RNFB	NADB_Rossmann	Helicase c	DEXDc	Transglut core	STKc_CDK9 like	AAT I	AAA 4
			NP_001067963	GI:115485639	197112	1757	1	9-380																						
			ABA98229	GI:77555433	207073	1831	1	27-395																						
	<i>Physcomitrella patens</i>		XP_001765706	GI:168026372	513101	4642	3	13-409	1373-1809	2751-3169							1													
			XP_001777863	GI:168050836	217751	1962	1	320-647																						
	<i>Phytophthora infestans</i>		EEY57574	GI:262099522	404720	3635	1	2131-2448																						
			EEY57113	GI:262099061	536624	4807	3	12-423	1522-1891	2848-3281							1												1	
	<i>Populus trichocarpa</i>		XP_002322285	GI:224136262	192262	1692	1	19-351																						
			XP_002318757	GI:224122118	194662	1713	1	19-371																						
			XP_002322287	GI:224136270	166520	1467	1	19-371																						
			XP_002307173	GI:224083922	323015	2870	2	19-434	1397-1844																					
	<i>Ricinus communis</i>		XP_002518058	GI:255554036	319322	2833	1	1263-1579																						
			XP_002511120	GI:255540111	258638	2299	1	638-961																					1	
			XP_002527141	GI:255572407	526971	4704	3	18-426	1388-1834	2797-3208							1													
	<i>Sorghum bicolor</i>		XP_002463708	GI:242032627	292337	2610	1	1077-1386																						
			XP_002443173	GI:242085496	221108	1948	1	27-405																						
			XP_002450788	GI:242071023	198920	1755	1	34-396																						
			XP_002449538	GI:242068523	192843	1699	1	24-374																						
			XP_002449539	GI:242068525	184031	1620	1	24-416																						
			XP_002447315	GI:242074758	527321	4709	3	15-367	1330-1778	2764-3171							1													
	<i>Vitis vinifera</i>		CAN78188	GI:147775835	169787	1488	1	22-346																						
			CBI17221	GI:270306163	268432	2388	1	1115-1415																						
			XP_002277154	GI:225423757	546615	4868	3	21-430	1392-1837	2795-3197							1													
			CBI27138	GI:270239905	443313	3960	3	21-430	910-1246	2147-2560							1													
			CBI15722	GI:270226988	98841	875	1	22-370																						
			XP_002284314	GI:225456431	194556	1712	1	22-370																						
			XP_002280614	GI:225456428	196873	1725	1	22-370																						
			CAN62566	GI:147787343	195403	1725	1	58-405																						
			XP_002283535	GI:225458934	204121	1801	1	40-387																						

Kingdom	Organism	Name in Figure 3	Accession ID	Other ID's	Daltons	Amino Acids	SRR	SRR1	SRR2	SRR3	SRR4	XPCB	DnaJ	HEPN	UBL	SSL2	RING	BTB	HET	zf-C2HC_plant	RpsA	Peptidase C14	RNFB	NADB_Rossmann	Helicase_c	DEXDc	Transglut_core	STKc_CDK9_like	AAT_I	AAA_4
Protozoa	<i>Babesia bovis</i>		XP_001608749	GI:156082529	289105	2581	1	715-1085																						
	<i>Chlamydomonas reinhardtii</i>		XP_001690195	GI:159463930	292459	2852	2	12-309	569-917																					
	<i>Cryptosporidium muris</i>		XP_002139897	GI:209876910	281327	2423	1	17-560																						
			XP_002139334	GI:209875783	264579	2286	1	711-1063																						
	<i>Cryptosporidium parvum</i>		XP_626083	GI:66357810	236910	2022	1	667-1004																						
			XP_627068	GI:66359780	246065	2120	1	20-463																						
	<i>Dictyostelium Discoideum</i>		XP_647015	GI:66827321	307353	2661	2	6-419	1290-1718																					
			XP_647016	GI:66827323	309734	2673	2	6-419	1309-1735																					
			XP_647014	GI:66827321	307589	2660	2	6-419	1292-1719																					
	<i>Phaeodactylum tricornutum</i>		XP_002176458	GI:219114579	150993	1348	1	20-389																						
	<i>Polysphondylium pallidum</i>		EFA85625	GI:281211463	296496	2582	2	12-414	1206-1556					1																
			EFA85624	GI:281211462	314437	2733	2	11-436	1293-1688					1																
	<i>Toxoplasma gondii</i>		XP_002371897	GI:237845199	649352	6036	1	1978-2589																						

Table A.2: Characteristics of SRR containing proteins.

The names and descriptive characteristics of all proteins identified as containing SRR domains are provided. The exact identities of the domains included in Figure 2.8 are indicated for cross-reference. The locations of identified SRR domains are provided as amino acid number counting from the N-terminus. The CDD nomenclature for the other domains is given in Table A.3.

Abbreviation	Name	Family
Transglut_core	Transglutaminase-like superfamily	pfam01841
AAT_I	Aspartate aminotransferase superfamily	cd01494
STKc_CDK9_like	catalytic domain of Cyclin-Dependent protein Kinase 9-like Serine/Threonine Kinases	cd07840
NADB_Rossmann	Rossmann-fold NAD(P)(+)-binding proteins	cl09931
XPCB	Xeroderma Pigmentosum Group C Binding Protein	N/A
rnfB	electron transport complex, RnfABCDGE type, B subunit	TIGRO1944
SSL2	DNA or RNA helicases of superfamily II	COG1061
DnaJ	DnaJ domain	pfam00226
RING	RING-finger domain	cd00162
Helicase_C	Helicase conserved C-terminal domain	pfam00271
HEPN	Higher Eukaryotes and Prokaryotes Nucleotide-binding domain	pfam05168
AAA_4	Divergent AAA domain	pfam04326
DEXDc	DEAD-like helicases superfamily	smart00487
BTB	BTB/POZ domain	pfam00651
Peptidase_C14	Caspase domain	pfam00656
zf-C2HC_plant	unknown function	pfam07567
HET	Heterokaryon incompatibility protein	pfam06985
RpsA	Ribosomal protein S1	COG0539
UBL	Ubiquitin like	cd01769

Table A.3: Nomenclature for other domains in SRR containing proteins. The abbreviations, full names, and family identifications for the various domains listed in Figure 2.7 and Table A.2 are given.

Bibliography

- Abagyan, R. A., and Batalov, S. (1997). Do aligned sequences share the same fold? *J Mol Biol* 273, 355-368.
- Agashe, V. R., Guha, S., Chang, H. C., Genevieux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F. U., and Barral, J. M. (2004). Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* 117, 199-209.
- Agorogiannis, E. I., Agorogiannis, G. I., Papadimitriou, A., and Hadjigeorgiou, G. M. (2004). Protein misfolding in neurodegenerative diseases. *Neuropathol Appl Neurobiol* 30, 215-224.
- Ali, M. M., Roe, S. M., Vaughan, C. K., Meyer, P., Panaretou, B., Piper, P. W., Prodromou, C., and Pearl, L. H. (2006). Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* 440, 1013-1017.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.
- Ambudkar, S. V., Kimchi-Sarfaty, C., Sauna, Z. E., and Gottesman, M. M. (2003). P-glycoprotein: from genomics to mechanism. *Oncogene* 22, 7468-7485.
- Andrade, M. A., Perez-Iratxeta, C., and Ponting, C. P. (2001). Protein repeats: structures, functions, and evolution. *J Struct Biol* 134, 117-131.
- Anesi, L., de Gemmis, P., Pandolfo, M., and Hladnik, U. (2011). Two Novel Homozygous SACS Mutations in Unrelated Patients Including the First Reported Case of Paternal UPD as an Etiologic Cause of ARSACS. *J Mol Neurosci* 43, 346-349.
- Anfinsen, C. B., and Scheraga, H. A. (1975). Experimental and theoretical aspects of protein folding. *Adv Protein Chem* 29, 205-300.
- Angelini, M., Cannata, S., Mercaldo, V., Gibello, L., Santoro, C., Dianzani, I., and Loreni, F. (2007). Missense mutations associated with Diamond-Blackfan anemia affect the assembly of ribosomal protein S19 into the ribosome. *Hum Mol Genet* 16, 1720-1727.
- Anheim, M., Fleury, M., Monga, B., Laugel, V., Chaigne, D., Rodier, G., Ginglinger, E., Boulay, C., Courtois, S., Drouot, N., Fritsch, M., Delaunoy, J. P., *et al.* (2009). Epidemiological, clinical, paraclinical and molecular study of a cohort of 102 patients affected with autosomal recessive progressive cerebellar ataxia from Alsace, Eastern France: implications for clinical management. *Neurogenetics* 11, 1-12.
- Antonellis, A., Ellsworth, R. E., Sambuughin, N., Puls, I., Abel, A., Lee-Lin, S. Q., Jordanova, A., Kremensky, I., Christodoulou, K., Middleton, L. T., Sivakumar, K., Ionasescu, V., *et al.* (2003). Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. *Am J Hum Genet* 72, 1293-1299.
- Antonellis, A., Lee-Lin, S. Q., Wasterlain, A., Leo, P., Quezado, M., Goldfarb, L. G., Myung, K., Burgess, S., Fischbeck, K. H., and Green, E. D. (2006). Functional analyses of glycyl-tRNA synthetase mutations suggest a key role for tRNA-charging enzymes in peripheral axons. *J Neurosci* 26, 10397-10406.

Arsene, F., Tomoyasu, T., and Bukau, B. (2000). The heat shock response of *Escherichia coli*. *Int J Food Microbiol* *55*, 3-9.

Ashburner, M., and Bonner, J. J. (1979). The induction of gene activity in *Drosophila* by heat shock. *Cell* *17*, 241-254.

Avila, J., Soares, H., Fanarraga, M. L., and Zabala, J. C. (2008). Isolation of microtubules and microtubule proteins. *Curr Protoc Cell Biol Chapter 3*, Unit 3 29.

Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. *Science* *234*, 179-186.

Bachmair, A., and Varshavsky, A. (1989). The degradation signal in a short-lived protein. *Cell* *56*, 1019-1032.

Badhai, J., Frojmark, A. S., Razzaghian, H. R., Davey, E., Schuster, J., and Dahl, N. (2009). Posttranscriptional down-regulation of small ribosomal subunit proteins correlates with reduction of 18S rRNA in RPS19 deficiency. *FEBS Lett* *583*, 2049-2053.

Baets, J., Deconinck, T., Smets, K., Goossens, D., Van den Bergh, P., Dahan, K., Schmedding, E., Santens, P., Rasic, V. M., Van Damme, P., Robberecht, W., De Meirleir, L., *et al.* (2010). Mutations in SACS cause atypical and late-onset forms of ARSACS. *Neurology* *75*, 1181-1188.

Ballinger, C. A., Connell, P., Wu, Y., Hu, Z., Thompson, L. J., Yin, L. Y., and Patterson, C. (1999). Identification of CHIP, a novel tetratricopeptide repeat-containing protein that interacts with heat shock proteins and negatively regulates chaperone functions. *Mol Cell Biol* *19*, 4535-4545.

Barclay, A. N. (2003). Membrane proteins with immunoglobulin-like domains--a master superfamily of interaction molecules. *Semin Immunol* *15*, 215-223.

Bardwell, J. C., and Craig, E. A. (1984). Major heat shock gene of *Drosophila* and the *Escherichia coli* heat-inducible *dnaK* gene are homologous. *Proc Natl Acad Sci U S A* *81*, 848-852.

Barral, J. M., Broadley, S. A., Schaffar, G., and Hartl, F. U. (2004). Roles of molecular chaperones in protein misfolding diseases. *Semin Cell Dev Biol* *15*, 17-29.

Barral, J. M., Hutagalung, A. H., Brinker, A., Hartl, F. U., and Epstein, H. F. (2002). Role of the myosin assembly protein UNC-45 as a molecular chaperone for myosin. *Science* *295*, 669-671.

Bartlett, A. I., and Radford, S. E. (2009). An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat Struct Mol Biol* *16*, 582-588.

Bekris, L. M., Mata, I. F., and Zabetian, C. P. (2010). The genetics of Parkinson disease. *J Geriatr Psychiatry Neurol* *23*, 228-242.

Bertelsen, E. B., Chang, L., Gestwicki, J. E., and Zuiderweg, E. R. (2009). Solution conformation of wild-type *E. coli* Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc Natl Acad Sci U S A* *106*, 8471-8476.

Bhushan, S., Gartmann, M., Halic, M., Armache, J. P., Jarasch, A., Mielke, T., Berninghausen, O., Wilson, D. N., and Beckmann, R. (2010). α -Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol* *17*, 313-317.

Bird, J. G., Sharma, S., Roshwalb, S. C., Hoskins, J. R., and Wickner, S. (2006). Functional analysis of CbpA, a DnaJ homolog and nucleoid-associated DNA-binding protein. *J Biol Chem* 281, 34349-34356.

Bloom, B. S., de Pourville, N., and Straus, W. L. (2003). Cost of illness of Alzheimer's disease: how useful are current estimates? *Gerontologist* 43, 158-164.

Boocock, G. R., Morrison, J. A., Popovic, M., Richards, N., Ellis, L., Durie, P. R., and Rommens, J. M. (2003). Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat Genet* 33, 97-101.

Borkovich, K. A., Farrelly, F. W., Finkelstein, D. B., Taulien, J., and Lindquist, S. (1989). hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol* 9, 3919-3930.

Bouchard, J. P., Barbeau, A., Bouchard, R., and Bouchard, R. W. (1978). Autosomal recessive spastic ataxia of Charlevoix-Saguenay. *Can J Neurol Sci* 5, 61-69.

Bouchard, J. P., Barbeau, A., Bouchard, R., and Bouchard, R. W. (1979a). Electromyography and nerve conduction studies in Friedreich's ataxia and autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS). *Can J Neurol Sci* 6, 185-189.

Bouchard, J. P., Richter, A., Mathieu, J., Brunet, D., Hudson, T. J., Morgan, K., and Melancon, S. B. (1998). Autosomal recessive spastic ataxia of Charlevoix-Saguenay. *Neuromuscul Disord* 8, 474-479.

Bouchard, M., and Langlois, G. (1999). Orthopedic management in autosomal recessive spastic ataxia of Charlevoix-Saguenay. *Can J Surg* 42, 440-444.

Bouchard, R. W., Bouchard, J. P., Bouchard, R., and Barbeau, A. (1979b). Electroencephalographic findings in Friedreich's ataxia and autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS). *Can J Neurol Sci* 6, 191-194.

Bouhlal, Y., Zouari, M., Kefi, M., Ben Hamida, C., Hentati, F., and Amouri, R. (2008). Autosomal recessive ataxia caused by three distinct gene defects in a single consanguineous family. *J Neurogenet* 22, 139-148.

Bouhouche, A., Benomar, A., Bouslam, N., Chkili, T., and Yahyaoui, M. (2006). Mutation in the epsilon subunit of the cytosolic chaperonin-containing t-complex peptide-1 (Cct5) gene causes autosomal recessive mutilating sensory neuropathy with spastic paraplegia. *J Med Genet* 43, 441-443.

Brandt, F., Carlson, L. A., Hartl, F. U., Baumeister, W., and Grunewald, K. (2010). The three-dimensional organization of polyribosomes in intact human cells. *Mol Cell* 39, 560-569.

Brandt, F., Etchells, S. A., Ortiz, J. O., Elcock, A. H., Hartl, F. U., and Baumeister, W. (2009). The native 3D organization of bacterial polysomes. *Cell* 136, 261-271.

Breckpot, J., Takiyama, Y., Thienpont, B., Van Vooren, S., Vermeesch, J. R., Ortibus, E., and Devriendt, K. (2008). A novel genomic disorder: a deletion of the SACS gene leading to spastic ataxia of Charlevoix-Saguenay. *Eur J Hum Genet* 16, 1050-1054.

Brehmer, D., Rudiger, S., Gassler, C. S., Klostermeier, D., Packschies, L., Reinstein, J., Mayer, M. P., and Bukau, B. (2001). Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange. *Nat Struct Biol* 8, 427-432.

Brenman, J. E., Topinka, J. R., Cooper, E. C., McGee, A. W., Rosen, J., Milroy, T., Ralston, H. J., and Bredt, D. S. (1998). Localization of postsynaptic density-93 to dendritic microtubules and interaction with microtubule-associated protein 1A. *J Neurosci* *18*, 8805-8813.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* *21*, 167-195.

Bryngelson, J. D., and Wolynes, P. G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* *84*, 7524-7528.

Buchberger, A. (2002). From UBA to UBX: new words in the ubiquitin vocabulary. *Trends Cell Biol* *12*, 216-221.

Buchner, J., Grallert, H., and Jakob, U. (1998). Analysis of chaperone function using citrate synthase as nonnative substrate protein. *Methods Enzymol* *290*, 323-338.

Bukau, B., Weissman, J., and Horwich, A. (2006). Molecular chaperones and protein quality control. *Cell* *125*, 443-451.

Burdette, A. J., Churchill, P. F., Caldwell, G. A., and Caldwell, K. A. (2010). The early-onset torsion dystonia-associated protein, torsinA, displays molecular chaperone activity in vitro. *Cell Stress Chaperones* *2010*, 19.

Camilloni, C., Sutto, L., Provati, D., Tiana, G., and Broglio, R. A. (2008). Early events in protein folding: Is there something more than hydrophobic burst? *Protein Sci* *17*, 1424-1433.

Canterini, S., Mangia, F., and Fiorenza, M. T. (2005). Thg-1 pit gene expression in granule cells of the developing mouse brain and in their synaptic targets, mature Purkinje, and mitral cells. *Dev Dyn* *234*, 689-697.

Cardoso, F. (2009). Huntington disease and other choreas. *Neurol Clin* *27*, 719-736, vi.

Casari, G., De Fusco, M., Ciarmatori, S., Zeviani, M., Mora, M., Fernandez, P., De Michele, G., Filla, A., Coccozza, S., Marconi, R., Durr, A., Fontaine, B., *et al.* (1998). Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* *93*, 973-983.

Caughey, B., and Lansbury, P. T. (2003). Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* *26*, 267-298.

Chadli, A., Graham, J. D., Abel, M. G., Jackson, T. A., Gordon, D. F., Wood, W. M., Felts, S. J., Horwitz, K. B., and Toft, D. (2006). GCUNC-45 is a novel regulator for the progesterone receptor/hsp90 chaperoning pathway. *Mol Cell Biol* *26*, 1722-1730.

Chan, E. M., Young, E. J., Ianzano, L., Munteanu, I., Zhao, X., Christopoulos, C. C., Avanzini, G., Elia, M., Ackerley, C. A., Jovic, N. J., Bohlega, S., Andermann, E., *et al.* (2003). Mutations in NHLRC1 cause progressive myoclonus epilepsy. *Nat Genet* *35*, 125-127.

Chang, H. C., Kaiser, C. M., Hartl, F. U., and Barral, J. M. (2005). De novo folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria. *J Mol Biol* *353*, 397-409.

Chapman, E., Farr, G. W., Usaite, R., Furtak, K., Fenton, W. A., Chaudhuri, T. K., Hondorp, E. R., Matthews, R. G., Wolf, S. G., Yates, J. R., Pypaert, M., and Horwich, A.

L. (2006). Global aggregation of newly translated proteins in an *Escherichia coli* strain deficient of the chaperonin GroEL. *Proc Natl Acad Sci U S A* *103*, 15800-15805.

Cheetham, M. E., and Caplan, A. J. (1998). Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress Chaperones* *3*, 28-36.

Chen, G., Duran, G. E., Steger, K. A., Lacayo, N. J., Jaffrezou, J. P., Dumontet, C., and Sikic, B. I. (1997). Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein, altered phenotype, and resistance to cyclosporins. *J Biol Chem* *272*, 5974-5982.

Cheung, J. C., and Deber, C. M. (2008). Misfolding of the cystic fibrosis transmembrane conductance regulator and disease. *Biochemistry* *47*, 1465-1473.

Chiang, A. P., Beck, J. S., Yen, H. J., Tayeh, M. K., Scheetz, T. E., Swiderski, R. E., Nishimura, D. Y., Braun, T. A., Kim, K. Y., Huang, J., Elbedour, K., Carmi, R., *et al.* (2006). Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). *Proc Natl Acad Sci U S A* *103*, 6287-6292.

Ciechanover, A. (2005). Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol* *6*, 79-87.

Clegg, H. V., Itahana, K., and Zhang, Y. (2008). Unlocking the Mdm2-p53 loop: ubiquitin is the key. *Cell Cycle* *7*, 287-292.

Cmejla, R., Cmejlova, J., Handrkova, H., Petrak, J., and Pospisilova, D. (2007). Ribosomal protein S17 gene (RPS17) is mutated in Diamond-Blackfan anemia. *Hum Mutat* *28*, 1178-1182.

Cole, C., Barber, J. D., and Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* *36*, W197-201.

Connell, P., Ballinger, C. A., Jiang, J., Wu, Y., Thompson, L. J., Hohfeld, J., and Patterson, C. (2001). The co-chaperone CHIP regulates protein triage decisions mediated by heat-shock proteins. *Nat Cell Biol* *3*, 93-96.

Craig, E. A., McCarthy, B. J., and Wadsworth, S. C. (1979). Sequence organization of two recombinant plasmids containing genes for the major heat shock-induced protein of *D. melanogaster*. *Cell* *16*, 575-588.

Crews, L., and Masliah, E. (2010). Molecular mechanisms of neurodegeneration in Alzheimer's disease. *Hum Mol Genet* *19*, R12-20.

Criscuolo, C., Banfi, S., Orio, M., Gasparini, P., Monticelli, A., Scarano, V., Santorelli, F. M., Perretti, A., Santoro, L., De Michele, G., and Filla, A. (2004). A novel mutation in SACS gene in a family from southern Italy. *Neurology* *62*, 100-102.

Criscuolo, C., Sacca, F., De Michele, G., Mancini, P., Combarros, O., Infante, J., Garcia, A., Banfi, S., Filla, A., and Berciano, J. (2005). Novel mutation of SACS gene in a Spanish family with autosomal recessive spastic ataxia. *Mov Disord* *20*, 1358-1361.

Cuervo, A. M., and Dice, J. F. (1998). Lysosomes, a meeting point of proteins, chaperones, and proteases. *J Mol Med* *76*, 6-12.

Danilova, N., Sakamoto, K. M., and Lin, S. (2008). Ribosomal protein S19 deficiency in zebrafish leads to developmental abnormalities and defective erythropoiesis through activation of p53 protein family. *Blood* *112*, 5228-5237.

David, C. L., Smith, H. E., Raynes, D. A., Pulcini, E. J., and Whitesell, L. (2003). Expression of a unique drug-resistant Hsp90 ortholog by the nematode *Caenorhabditis elegans*. *Cell Stress Chaperones* 8, 93-104.

De Braekeleer, M., Giasson, F., Mathieu, J., Roy, M., Bouchard, J. P., and Morgan, K. (1993). Genetic epidemiology of autosomal recessive spastic ataxia of Charlevoix-Saguenay in northeastern Quebec. *Genet Epidemiol* 10, 17-25.

De Felice, F. G., Velasco, P. T., Lambert, M. P., Viola, K., Fernandez, S. J., Ferreira, S. T., and Klein, W. L. (2007). Abeta oligomers induce neuronal oxidative stress through an N-methyl-D-aspartate receptor-dependent mechanism that is blocked by the Alzheimer drug memantine. *J Biol Chem* 282, 11590-11601.

Di Bella, D., Lazzaro, F., Brusco, A., Plumari, M., Battaglia, G., Pastore, A., Finardi, A., Cagnoli, C., Tempia, F., Frontali, M., Veneziano, L., Sacco, T., *et al.* (2010). Mutations in the mitochondrial protease gene AFG3L2 cause dominant hereditary ataxia SCA28. *Nat Genet* 42, 313-321.

Diaz, G. A., Gelb, B. D., Ali, F., Sakati, N., Sanjad, S., Meyer, B. F., and Kambouris, M. (1999). Sanjad-Sakati and autosomal recessive Kenny-Caffey syndromes are allelic: evidence for an ancestral founder mutation and locus refinement. *Am J Med Genet* 85, 48-52.

Dionne, J., Wright, G., Barber, H., Bouchard, R., and Bouchard, J. P. (1979). Oculomotor and vestibular findings in autosomal recessive spastic ataxia of Charlevoix-Saguenay. *Can J Neurol Sci* 6, 177-184.

Dittmar, K. D., Banach, M., Galigniana, M. D., and Pratt, W. B. (1998). The role of DnaJ-like proteins in glucocorticoid receptor.hsp90 heterocomplex assembly by the reconstituted hsp90.p60.hsp70 foldosome complex. *J Biol Chem* 273, 7358-7366.

Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem Sci* 24, 329-332.

Dobson, C. M. (2004). Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* 15, 3-16.

Doherty, L., Sheen, M. R., Vlachos, A., Choesmel, V., O'Donohue, M. F., Clinton, C., Schneider, H. E., Sieff, C. A., Newburger, P. E., Ball, S. E., Niewiadomska, E., Matysiak, M., *et al.* (2010). Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. *Am J Hum Genet* 86, 222-228.

Dollins, D. E., Warren, J. J., Immormino, R. M., and Gewirth, D. T. (2007). Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. *Mol Cell* 28, 41-56.

Draptchinskaia, N., Gustavsson, P., Andersson, B., Pettersson, M., Willig, T. N., Dianzani, I., Ball, S., Tchernia, G., Klar, J., Matsson, H., Tentler, D., Mohandas, N., *et al.* (1999). The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat Genet* 21, 169-175.

Dulyaninova, N. G., Malashkevich, V. N., Almo, S. C., and Bresnick, A. R. (2005). Regulation of myosin-IIA assembly and Mts1 binding by heavy chain phosphorylation. *Biochemistry* 44, 6867-6876.

Dumouлина, M., and Bader, R. (2006). A short historical survey of developments in amyloid research. *Protein Pept Lett* 13, 213-217.

Dupre, N., Bouchard, J. P., Brais, B., and Rouleau, G. A. (2006). Hereditary ataxia, spastic paraparesis and neuropathy in the French-Canadian population. *Can J Neurol Sci* 33, 149-157.

Dutta, R., and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci* 25, 24-28.

Easton, D. P., Kaneko, Y., and Subjeck, J. R. (2000). The hsp110 and Grp1 70 stress proteins: newly recognized relatives of the Hsp70s. *Cell Stress Chaperones* 5, 276-290.

Ebert, B. L., Pretz, J., Bosco, J., Chang, C. Y., Tamayo, P., Galili, N., Raza, A., Root, D. E., Attar, E., Ellis, S. R., and Golub, T. R. (2008). Identification of RPS14 as a 5q-syndrome gene by RNA interference screen. *Nature* 451, 335-339.

Edvardson, S., Shaag, A., Kolesnikova, O., Gomori, J. M., Tarassov, I., Einbinder, T., Saada, A., and Elpeleg, O. (2007). Deleterious mutation in the mitochondrial arginyl-transfer RNA synthetase gene is associated with pontocerebellar hypoplasia. *Am J Hum Genet* 81, 857-862.

El Euch-Fayache, G., Lalani, I., Amouri, R., Turki, I., Ouahchi, K., Hung, W. Y., Belal, S., Siddique, T., and Hentati, F. (2003). Phenotypic features and genetic findings in sarsin-related autosomal recessive ataxia in Tunisia. *Arch Neurol* 60, 982-988.

Ellis, R. J., and Minton, A. P. (2003). Cell biology: join the crowd. *Nature* 425, 27-28.

Engert, J. C., Berube, P., Mercier, J., Dore, C., Lepage, P., Ge, B., Bouchard, J. P., Mathieu, J., Melancon, S. B., Schalling, M., Lander, E. S., Morgan, K., *et al.* (2000). ARSACS, a spastic ataxia common in northeastern Quebec, is caused by mutations in a new gene encoding an 11.5-kb ORF. *Nat Genet* 24, 120-125.

Engert, J. C., Dore, C., Mercier, J., Ge, B., Betard, C., Rioux, J. D., Owen, C., Berube, P., Devon, K., Birren, B., Melancon, S. B., Morgan, K., *et al.* (1999). Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS): high-resolution physical and transcript map of the candidate region in chromosome region 13q11. *Genomics* 62, 156-164.

Erlandsen, H., Canaves, J. M., Elsliger, M. A., von Delft, F., Brinen, L. S., Dai, X., Deacon, A. M., Floyd, R., Godzik, A., Grittini, C., Grzechnik, S. K., Jaroszewski, L., *et al.* (2004). Crystal structure of an HEPN domain protein (TM0613) from *Thermotoga maritima* at 1.75 Å resolution. *Proteins* 54, 806-809.

Evans, K. J., Gomes, E. R., Reisenweber, S. M., Gundersen, G. G., and Luring, B. P. (2005). Linking axonal degeneration to microtubule remodeling by Spastin-mediated microtubule severing. *J Cell Biol* 168, 599-606.

Evgrafov, O. V., Mersiyanova, I., Irobi, J., Van Den Bosch, L., Dierick, I., Leung, C. L., Schagina, O., Verpoorten, N., Van Impe, K., Fedotov, V., Dadali, E., Auer-Grumbach, M., *et al.* (2004). Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy. *Nat Genet* 36, 602-606.

Farrar, J. E., Nater, M., Caywood, E., McDevitt, M. A., Kowalski, J., Takemoto, C. M., Talbot, C. C., Jr., Meltzer, P., Esposito, D., Beggs, A. H., Schneider, H. E., Grabowska, A., *et al.* (2008). Abnormalities of the large ribosomal subunit protein, Rpl35a, in Diamond-Blackfan anemia. *Blood* 112, 1582-1592.

- Ferreiro, D. U., Hegler, J. A., Komives, E. A., and Wolynes, P. G. (2007). Localizing frustration in native proteins and protein assemblies. *Proc Natl Acad Sci U S A* *104*, 19819-19824.
- Fersht, A. R. (2000). Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc Natl Acad Sci U S A* *97*, 1525-1529.
- Filip, A. M., Klug, J., Cayli, S., Frohlich, S., Henke, T., Lacher, P., Eickhoff, R., Bulau, P., Linder, M., Carlsson-Skwirut, C., Leng, L., Bucala, R., *et al.* (2009). Ribosomal protein S19 interacts with macrophage migration inhibitory factor and attenuates its pro-inflammatory function. *J Biol Chem* *284*, 7977-7985.
- Flaherty, K. M., McKay, D. B., Kabsch, W., and Holmes, K. C. (1991). Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein. *Proc Natl Acad Sci U S A* *88*, 5041-5045.
- Freeman, B. C., and Morimoto, R. I. (1996). The human cytosolic molecular chaperones hsp90, hsp70 (hsc70) and hsp71 have distinct roles in recognition of a non-native protein and protein refolding. *Embo J* *15*, 2969-2979.
- Frid, P., Anisimov, S. V., and Popovic, N. (2007). Congo red and protein aggregation in neurodegenerative diseases. *Brain Res Rev* *53*, 135-160.
- Frosk, P., Weiler, T., Nysten, E., Sudha, T., Greenberg, C. R., Morgan, K., Fujiwara, T. M., and Wrogemann, K. (2002). Limb-girdle muscular dystrophy type 2H associated with mutation in TRIM32, a putative E3-ubiquitin-ligase gene. *Am J Hum Genet* *70*, 663-672.
- Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem* *70*, 603-647.
- Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassilyev, D. G., and Yokoyama, S. (2000). Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase. *Cell* *103*, 793-803.
- Gazda, H. T., Grabowska, A., Merida-Long, L. B., Latawiec, E., Schneider, H. E., Lipton, J. M., Vlachos, A., Atsidaftos, E., Ball, S. E., Orfali, K. A., Niewiadomska, E., Da Costa, L., *et al.* (2006). Ribosomal protein S24 gene is mutated in Diamond-Blackfan anemia. *Am J Hum Genet* *79*, 1110-1118.
- Gemignani, F., and Marbini, A. (2001). Charcot-Marie-Tooth disease (CMT): distinctive phenotypic and genotypic features in CMT type 2. *J Neurol Sci* *184*, 1-9.
- Gentry, M. S., Worby, C. A., and Dixon, J. E. (2005). Insights into Lafora disease: malin is an E3 ubiquitin ligase that ubiquitinates and promotes the degradation of laforin. *Proc Natl Acad Sci U S A* *102*, 8501-8506.
- Gerwig, M., Kruger, S., Kreuz, F. R., Kreis, S., Gizewski, E. R., and Timmann, D. (2010). Characteristic MRI and fundoscopic findings help diagnose ARSACS outside Quebec. *Neurology* *75*, 2133.
- Gleiter, S., and Bardwell, J. C. (2008). Disulfide bond isomerization in prokaryotes. *Biochim Biophys Acta* *1783*, 530-534.
- Gomez, C. M. (2004). ARSACS goes global. *Neurology* *62*, 10-11.

Graf, C., Stankiewicz, M., Kramer, G., and Mayer, M. P. (2009). Spatially and kinetically resolved changes in the conformational dynamics of the Hsp90 chaperone machine. *Embo J* 28, 602-613.

Gragerov, A., Zeng, L., Zhao, X., Burkholder, W., and Gottesman, M. E. (1994). Specificity of DnaK-peptide binding. *J Mol Biol* 235, 848-854.

Grammatikakis, N., Lin, J. H., Grammatikakis, A., Tsiachlis, P. N., and Cochran, B. H. (1999). p50(cdc37) acting in concert with Hsp90 is required for Raf-1 function. *Mol Cell Biol* 19, 1661-1672.

Greer, P. L., Hanayama, R., Bloodgood, B. L., Mardinly, A. R., Lipton, D. M., Flavell, S. W., Kim, T. K., Griffith, E. C., Waldon, Z., Maehr, R., Ploegh, H. L., Chowdhury, S., *et al.* (2010). The Angelman Syndrome protein Ube3A regulates synapse development by ubiquitinating arc. *Cell* 140, 704-716.

Grenert, J. P., Johnson, B. D., and Toft, D. O. (1999). The importance of ATP binding and hydrolysis by hsp90 in formation and function of protein heterocomplexes. *J Biol Chem* 274, 17525-17533.

Grenert, J. P., Sullivan, W. P., Fadden, P., Haystead, T. A., Clark, J., Mimnaugh, E., Krutzsch, H., Ochel, H. J., Schulte, T. W., Sausville, E., Neckers, L. M., and Toft, D. O. (1997). The amino-terminal domain of heat shock protein 90 (hsp90) that binds geldanamycin is an ATP/ADP switch domain that regulates hsp90 conformation. *J Biol Chem* 272, 23843-23850.

Grieco, G. S., Malandrini, A., Comanducci, G., Leuzzi, V., Valoppi, M., Tessa, A., Palmeri, S., Benedetti, L., Pierallini, A., Gambelli, S., Federico, A., Pierelli, F., *et al.* (2004). Novel SACS mutations in autosomal recessive spastic ataxia of Charlevoix-Saguenay type. *Neurology* 62, 103-106.

Gruber, C. W., Cemazar, M., Heras, B., Martin, J. L., and Craik, D. J. (2006). Protein disulfide isomerase: the structure of oxidative folding. *Trends Biochem Sci* 31, 455-464.

Grynberg, M., Erlandsen, H., and Godzik, A. (2003). HEPN: a common domain in bacterial drug resistance and human neurodegenerative proteins. *Trends Biochem Sci* 28, 224-226.

Guernsey, D. L., Dube, M. P., Jiang, H., Asselin, G., Blowers, S., Evans, S., Ferguson, M., Macgillivray, C., Matsuoka, M., Nightingale, M., Rideout, A., Delatycki, M., *et al.* (2010). Novel mutations in the sartin gene in ataxia patients from Maritime Canada. *J Neurol Sci* 288, 79-87.

H'Mida-Ben Brahim, D., M'Zahem, A., Assoum, M., Bouhlal, Y., Fattori, F., Anheim, M., Ali-Pacha, L., Ferrat, F., Chaouch, M., Lagier-Tourenne, C., Drouot, N., Thibaut, C., *et al.* (2011). Molecular diagnosis of known recessive ataxias by homozygosity mapping with SNP arrays. *J Neurol* 258, 56-67.

Haber, E., and Anfinsen, C. B. (1961). Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. *J Biol Chem* 236, 422-424.

Hansen, J. J., Durr, A., Courneau-Rebeix, I., Georgopoulos, C., Ang, D., Nielsen, M. N., Davoine, C. S., Brice, A., Fontaine, B., Gregersen, N., and Bross, P. (2002). Hereditary spastic paraplegia SPG13 is associated with a mutation in the gene encoding the mitochondrial chaperonin Hsp60. *Am J Hum Genet* 70, 1328-1332.

- Hara, K., Onodera, O., Endo, M., Kondo, H., Shiota, H., Miki, K., Tanimoto, N., Kimura, T., and Nishizawa, M. (2005). Sacsin-related autosomal recessive ataxia without prominent retinal myelinated fibers in Japan. *Mov Disord* 20, 380-382.
- Hara, K., Shimbo, J., Nozaki, H., Kikugawa, K., Onodera, O., and Nishizawa, M. (2007). Sacsin-related ataxia with neither retinal hypermyelination nor spasticity. *Mov Disord* 22, 1362-1363.
- Hardy, J., and Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 297, 353-356.
- Hartl, F. U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295, 1852-1858.
- Hartl, F. U., and Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* 16, 574-581.
- Hartmann-Petersen, R., and Gordon, C. (2004). Integral UBL domain proteins: a family of proteasome interacting proteins. *Semin Cell Dev Biol* 15, 247-259.
- Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., Ogata, H., and Ohta, T. (2001). The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. *J Biol Chem* 276, 14537-14540.
- Hawe, A., Sutter, M., and Jiskoot, W. (2008). Extrinsic fluorescent dyes as tools for protein characterization. *Pharm Res* 25, 1487-1499.
- Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Samson, D., Artiguenave, F., Davoine, C. S., Cruaud, C., Durr, A., Wincker, P., Brottier, P., Cattolico, L., *et al.* (1999). Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia. *Nat Genet* 23, 296-303.
- Hebert, L. E., Scherr, P. A., Bienias, J. L., Bennett, D. A., and Evans, D. A. (2003). Alzheimer disease in the US population: prevalence estimates using the 2000 census. *Arch Neurol* 60, 1119-1122.
- Heiss, N. S., Knight, S. W., Vulliamy, T. J., Klauck, S. M., Wiemann, S., Mason, P. J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet* 19, 32-38.
- Hemdan, E. S., Zhao, Y. J., Sulkowski, E., and Porath, J. (1989). Surface topography of histidine residues: a facile probe by immobilized metal ion affinity chromatography. *Proc Natl Acad Sci U S A* 86, 1811-1815.
- Henn, I. H., Bouman, L., Schlehe, J. S., Schlierf, A., Schramm, J. E., Wegener, E., Nakaso, K., Culmsee, C., Berninger, B., Krappmann, D., Tatzelt, J., and Winklhofer, K. F. (2007). Parkin mediates neuroprotection through activation of IkappaB kinase/nuclear factor-kappaB signaling. *J Neurosci* 27, 1868-1878.
- Herbst, R., Gast, K., and Seckler, R. (1998). Folding of firefly (*Photinus pyralis*) luciferase: aggregation and reactivation of unfolding intermediates. *Biochemistry* 37, 6586-6597.
- Hessling, M., Richter, K., and Buchner, J. (2009). Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nat Struct Mol Biol* 16, 287-293.
- Hoffmeyer, S., Burk, O., von Richter, O., Arnold, H. P., Brockmoller, J., John, A., Cascorbi, I., Gerloff, T., Roots, I., Eichelbaum, M., and Brinkmann, U. (2000).

Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proc Natl Acad Sci U S A* 97, 3473-3478.

Hohfeld, J., and Jentsch, S. (1997). GrpE-like regulation of the hsc70 chaperone by the anti-apoptotic protein BAG-1. *Embo J* 16, 6209-6216.

Holzbeierlein, J. M., Windsperger, A., and Vielhauer, G. (2010). Hsp90: a drug target? *Curr Oncol Rep* 12, 95-101.

Hong, L., Piao, Y., Han, Y., Wang, J., Zhang, X., Du, Y., Cao, S., Qiao, T., Chen, Z., and Fan, D. (2005). Zinc ribbon domain-containing 1 (ZNRD1) mediates multidrug resistance of leukemia cells through regulation of P-glycoprotein and Bcl-2. *Mol Cancer Ther* 4, 1936-1942.

Huber, C., Dias-Santagata, D., Glaser, A., O'Sullivan, J., Brauner, R., Wu, K., Xu, X., Pearce, K., Wang, R., Uzielli, M. L., Dagoneau, N., Chemaitilly, W., *et al.* (2005). Identification of mutations in CUL7 in 3-M syndrome. *Nat Genet* 37, 1119-1124.

Hulmes, D. J. (2002). Building collagen molecules, fibrils, and suprafibrillar structures. *J Struct Biol* 137, 2-10.

Hurst, J. A., and Baraitser, M. (1989). Johanson-Blizzard syndrome. *J Med Genet* 26, 45-48.

Ionasescu, V., Searby, C., Sheffield, V. C., Roklina, T., Nishimura, D., and Ionasescu, R. (1996). Autosomal dominant Charcot-Marie-Tooth axonal neuropathy mapped on chromosome 7p (CMT2D). *Hum Mol Genet* 5, 1373-1375.

Irobi, J., Van Impe, K., Seeman, P., Jordanova, A., Dierick, I., Verpoorten, N., Michalik, A., De Vriendt, E., Jacobs, A., Van Gerwen, V., Vennekens, K., Mazanec, R., *et al.* (2004). Hot-spot residue in small heat-shock protein 22 causes distal motor neuropathy. *Nat Genet* 36, 597-601.

Iwai, K., Yamanaka, K., Kamura, T., Minato, N., Conaway, R. C., Conaway, J. W., Klausner, R. D., and Pause, A. (1999). Identification of the von Hippel-lindau tumor-suppressor protein as part of an active E3 ubiquitin ligase complex. *Proc Natl Acad Sci U S A* 96, 12436-12441.

Jakob, U., Lilie, H., Meyer, I., and Buchner, J. (1995). Transient interaction of Hsp90 with early unfolding intermediates of citrate synthase. Implications for heat shock in vivo. *J Biol Chem* 270, 7288-7294.

Jakob, U., Scheibel, T., Bose, S., Reinstein, J., and Buchner, J. (1996). Assessment of the ATP binding properties of Hsp90. *J Biol Chem* 271, 10035-10041.

Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry* 79, 368-376.

Janov, A. J., Leong, T., Nathan, D. G., and Guinan, E. C. (1996). Diamond-Blackfan anemia. Natural history and sequelae of treatment. *Medicine (Baltimore)* 75, 77-78.

Jiang, J., Prasad, K., Lafer, E. M., and Sousa, R. (2005). Structural basis of interdomain communication in the Hsc70 chaperone. *Mol Cell* 20, 513-524.

Jin, J., Davis, J., Zhu, D., Kashima, D. T., Leroueil, M., Pan, C., Montine, K. S., and Zhang, J. (2007). Identification of novel proteins affected by rotenone in mitochondria of dopaminergic cells. *BMC Neurosci* 8, 67.

- Johanson, A., and Blizzard, R. (1971). A syndrome of congenital aplasia of the alae nasi, deafness, hypothyroidism, dwarfism, absent permanent teeth, and malabsorption. *J Pediatr* 79, 982-987.
- Johnson, B. D., Chadli, A., Felts, S. J., Bouhouche, I., Catelli, M. G., and Toft, D. O. (2000). Hsp90 chaperone activity requires the full-length protein and interaction among its multiple domains. *J Biol Chem* 275, 32499-32507.
- Johnson, J. L., Halas, A., and Flom, G. (2007). Nucleotide-dependent interaction of *Saccharomyces cerevisiae* Hsp90 with the cochaperone proteins Sti1, Cpr6, and Sba1. *Mol Cell Biol* 27, 768-776.
- Johnson, J. L., and Toft, D. O. (1994). A novel chaperone complex for steroid receptors involving heat shock proteins, immunophilins, and p23. *J Biol Chem* 269, 24989-24993.
- Jordanova, A., Irobi, J., Thomas, F. P., Van Dijck, P., Meerschaert, K., Dewil, M., Dierick, I., Jacobs, A., De Vriendt, E., Guerguelcheva, V., Rao, C. V., Tournev, I., *et al.* (2006). Disrupted function and axonal distribution of mutant tyrosyl-tRNA synthetase in dominant intermediate Charcot-Marie-Tooth neuropathy. *Nat Genet* 38, 197-202.
- Jung, T., Catalgol, B., and Grune, T. (2009). The proteasomal system. *Mol Aspects Med* 30, 191-296.
- Junier, T., and Pagni, M. (2000). Dotlet: diagonal plots in a web browser. *Bioinformatics* 16, 178-179.
- Kabani, M., McLellan, C., Raynes, D. A., Guerriero, V., and Brodsky, J. L. (2002). HspBP1, a homologue of the yeast Fes1 and Sls1 proteins, is an Hsc70 nucleotide exchange factor. *FEBS Lett* 531, 339-342.
- Kaiser, C. M., Chang, H. C., Agashe, V. R., Lakshmipathy, S. K., Etchells, S. A., Hayer-Hartl, M., Hartl, F. U., and Barral, J. M. (2006). Real-time observation of trigger factor function on translating ribosomes. *Nature* 444, 455-460.
- Kamada, S., Okawa, S., Imota, T., Sugawara, M., and Toyoshima, I. (2008). Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS): novel compound heterozygous mutations in the SACS gene. *J Neurol* 255, 803-806.
- Kamionka, M., and Feigon, J. (2004). Structure of the XPC binding domain of hHR23A reveals hydrophobic patches for protein interaction. *Protein Sci* 13, 2370-2377.
- Kapon, R., Nevo, R., and Reich, Z. (2008). Protein energy landscape roughness. *Biochem Soc Trans* 36, 1404-1408.
- Karplus, M., and Weaver, D. L. (1976). Protein-folding dynamics. *Nature* 260, 404-406.
- Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9, 286-298.
- Katsanis, N., Beales, P. L., Woods, M. O., Lewis, R. A., Green, J. S., Parfrey, P. S., Ansley, S. J., Davidson, W. S., and Lupski, J. R. (2000). Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. *Nat Genet* 26, 67-70.
- Kawahara, M., and Kuroda, Y. (2000). Molecular mechanism of neurodegeneration induced by Alzheimer's beta-amyloid protein: channel formation and disruption of calcium homeostasis. *Brain Res Bull* 53, 389-397.
- Kawakami, S., and Hagiwara, H. (2008). A taxonomic revision of two dictyostelid species, *Polysphondylium pallidum* and *P. album*. *Mycologia* 100, 111-121.

Kelley, W. L. (1998). The J-domain family and the recruitment of chaperone power. *Trends Biochem Sci* 23, 222-227.

Kersey, P. J., Lawson, D., Birney, E., Derwent, P. S., Haimel, M., Herrero, J., Keenan, S., Kerhornou, A., Koscielny, G., Kahari, A., Kinsella, R. J., Kulesha, E., *et al.* (2009). Ensembl Genomes: extending Ensembl across the taxonomic space. *Nucleic Acids Res* 38, D563-569.

Kiefhaber, T., Rudolph, R., Kohler, H. H., and Buchner, J. (1991). Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. *Biotechnology (N Y)* 9, 825-829.

Kim, B., Ryu, K. S., Kim, H. J., Cho, S. J., and Choi, B. S. (2005). Solution structure and backbone dynamics of the XPC-binding domain of the human DNA repair protein hHR23B. *Febs J* 272, 2467-2476.

Kim, J., Lowe, T., and Hoppe, T. (2008). Protein quality control gets muscle into shape. *Trends Cell Biol* 18, 264-272.

Kim, R. B., Fromm, M. F., Wandel, C., Leake, B., Wood, A. J., Roden, D. M., and Wilkinson, G. R. (1998). The drug transporter P-glycoprotein limits oral absorption and brain entry of HIV-1 protease inhibitors. *J Clin Invest* 101, 289-294.

Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., and Gottesman, M. M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 315, 525-528.

Kimura, Y., Yahara, I., and Lindquist, S. (1995). Role of the protein chaperone YDJ1 in establishing Hsp90-mediated signal transduction pathways. *Science* 268, 1362-1365.

Kioka, N., Tsubota, J., Kakehi, Y., Komano, T., Gottesman, M. M., Pastan, I., and Ueda, K. (1989). P-glycoprotein gene (MDR1) cDNA from human adrenal: normal P-glycoprotein carries Gly185 with an altered pattern of multidrug resistance. *Biochem Biophys Res Commun* 162, 224-231.

Kishino, T., Lalande, M., and Wagstaff, J. (1997). UBE3A/E6-AP mutations cause Angelman syndrome. *Nat Genet* 15, 70-73.

Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., and Shimizu, N. (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 392, 605-608.

Ko, H. S., Kim, S. W., Sriram, S. R., Dawson, V. L., and Dawson, T. M. (2006). Identification of far upstream element-binding protein-1 as an authentic Parkin substrate. *J Biol Chem* 281, 16193-16196.

Ko, H. S., von Coelln, R., Sriram, S. R., Kim, S. W., Chung, K. K., Pletnikova, O., Troncoso, J., Johnson, B., Saffary, R., Goh, E. L., Song, H., Park, B. J., *et al.* (2005). Accumulation of the authentic parkin substrate aminoacyl-tRNA synthetase cofactor, p38/JTV-1, leads to catecholaminergic cell death. *J Neurosci* 25, 7968-7978.

Komar, A. A. (2009). A pause for thought along the co-translational folding pathway. *Trends Biochem Sci* 34, 16-24.

Konstantinova, I. M., Tsimokha, A. S., and Mittenberg, A. G. (2008). Role of proteasomes in cellular regulation. *Int Rev Cell Mol Biol* 267, 59-124.

Kristjansdottir, S., Lindorff-Larsen, K., Fieber, W., Dobson, C. M., Vendruscolo, M., and Poulsen, F. M. (2005). Formation of native and non-native interactions in ensembles of

denatured ACBP molecules from paramagnetic relaxation enhancement studies. *J Mol Biol* 347, 1053-1062.

Kubota, H. (2009). Quality control against misfolded proteins in the cytosol: a network for cell survival. *J Biochem* 146, 609-616.

Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324, 255-258.

Kwon, Y. T., Reiss, Y., Fried, V. A., Hershko, A., Yoon, J. K., Gonda, D. K., Sangan, P., Copeland, N. G., Jenkins, N. A., and Varshavsky, A. (1998). The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc Natl Acad Sci U S A* 95, 7898-7903.

Laberge, C. (1969). Hereditary tyrosinemia in a French Canadian isolate. *Am J Hum Genet* 21, 36-45.

Landsverk, M. L., Li, S., Hutagalung, A. H., Najafov, A., Hoppe, T., Barral, J. M., and Epstein, H. F. (2007). The UNC-45 chaperone mediates sarcomere assembly through myosin degradation in *Caenorhabditis elegans*. *J Cell Biol* 177, 205-210.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K., and Hartl, F. U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* 356, 683-689.

Larbrisseau, A., Vanasse, M., Brochu, P., and Jasmin, G. (1984). The Andermann syndrome: agenesis of the corpus callosum associated with mental retardation and progressive sensorimotor neuronopathy. *Can J Neurol Sci* 11, 257-261.

Latour, P., Thauvin-Robinet, C., Baudelet-Mery, C., Soichot, P., Cusin, V., Faivre, L., Locatelli, M. C., Mayencon, M., Sarcey, A., Broussolle, E., Camu, W., David, A., *et al.* (2009). A major determinant for binding and aminoacylation of tRNA(Ala) in cytoplasmic Alanyl-tRNA synthetase is mutated in dominant axonal Charcot-Marie-Tooth disease. *Am J Hum Genet* 86, 77-82.

LaVoie, M. J., Ostaszewski, B. L., Weihofen, A., Schlossmacher, M. G., and Selkoe, D. J. (2005). Dopamine covalently modifies and functionally inactivates parkin. *Nat Med* 11, 1214-1221.

Leger-Silvestre, I., Caffrey, J. M., Dawaliby, R., Alvarez-Arias, D. A., Gas, N., Bertolone, S. J., Gleizes, P. E., and Ellis, S. R. (2005). Specific Role for Yeast Homologs of the Diamond Blackfan Anemia-associated Rps19 Protein in Ribosome Synthesis. *J Biol Chem* 280, 38177-38185.

Leroy, E., Boyer, R., Auburger, G., Leube, B., Ulm, G., Mezey, E., Harta, G., Brownstein, M. J., Jonnalagada, S., Chernova, T., Dehejia, A., Lavedan, C., *et al.* (1998). The ubiquitin pathway in Parkinson's disease. *Nature* 395, 451-452.

Levav-Cohen, Y., Haupt, S., and Haupt, Y. (2005). Mdm2 in growth signaling and cancer. *Growth Factors* 23, 183-192.

Li, J., Qian, X., and Sha, B. (2003). The crystal structure of the yeast Hsp40 Ydj1 complexed with its peptide substrate. *Structure* 11, 1475-1483.

Li, K. W., Hornshaw, M. P., Van Der Schors, R. C., Watson, R., Tate, S., Casetta, B., Jimenez, C. R., Gouwenberg, Y., Gundelfinger, E. D., Smalla, K. H., and Smit, A. B. (2004). Proteomics analysis of rat brain postsynaptic density. Implications of the diverse

protein functional groups for the integration of synaptic physiology. *J Biol Chem* 279, 987-1002.

Lim, J., Hao, T., Shaw, C., Patel, A. J., Szabo, G., Rual, J. F., Fisk, C. J., Li, N., Smolyar, A., Hill, D. E., Barabasi, A. L., Vidal, M., *et al.* (2006). A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 125, 801-814.

Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C. M., Poulsen, F. M., and Vendruscolo, M. (2004). Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. *J Am Chem Soc* 126, 3291-3299.

Litt, M., Kramer, P., LaMorticella, D. M., Murphey, W., Lovrien, E. W., and Weleber, R. G. (1998). Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Hum Mol Genet* 7, 471-474.

Locke, M., Tinsley, C. L., Benson, M. A., and Blake, D. J. (2009). TRIM32 is an E3 ubiquitin ligase for dysbindin. *Hum Mol Genet* 18, 2344-2358.

Lohr, N. J., Molleston, J. P., Strauss, K. A., Torres-Martinez, W., Sherman, E. A., Squires, R. H., Rider, N. L., Chikwava, K. R., Cummings, O. W., Morton, D. H., and Puffenberger, E. G. (2010). Human ITCH E3 ubiquitin ligase deficiency causes syndromic multisystem autoimmune disease. *Am J Hum Genet* 86, 447-453.

Love, S. (2005). Neuropathological investigation of dementia: a guide for neurologists. *J Neurol Neurosurg Psychiatry* 76 Suppl 5, v8-14.

Macario, A. J., and Conway de Macario, E. (2007). Chaperonopathies and chaperonotherapy. *FEBS Lett* 581, 3681-3688.

Maccioni, R. B., and Cambiasso, V. (1995). Role of microtubule-associated proteins in the control of microtubule assembly. *Physiol Rev* 75, 835-864.

Magen, D., Georgopoulos, C., Bross, P., Ang, D., Segev, Y., Goldsher, D., Nemirovski, A., Shahar, E., Ravid, S., Luder, A., Heno, B., Gershoni-Baruch, R., *et al.* (2008). Mitochondrial hsp60 chaperonopathy causes an autosomal-recessive neurodegenerative disorder linked to brain hypomyelination and leukodystrophy. *Am J Hum Genet* 83, 30-42.

Mapa, K., Sikor, M., Kudryavtsev, V., Waegemann, K., Kalinin, S., Seidel, C. A., Neupert, W., Lamb, D. C., and Mokranjac, D. (2010). The conformational dynamics of the mitochondrial Hsp70 chaperone. *Mol Cell* 38, 89-100.

Martin, M. H., Bouchard, J. P., Sylvain, M., St-Onge, O., and Truchon, S. (2007). Autosomal recessive spastic ataxia of Charlevoix-Saguenay: a report of MR imaging in 5 patients. *AJNR Am J Neuroradiol* 28, 1606-1608.

Masutani, C., Sugawara, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., van der Spek, P. J., Bootsma, D., and *et al.* (1994). Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *Embo J* 13, 1831-1843.

Mayer, M. P. (2010). Gymnastics of molecular chaperones. *Mol Cell* 39, 321-331.

Mayer, M. P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci* 62, 670-684.

Mayer, M. P., Schroder, H., Rudiger, S., Paal, K., Laufen, T., and Bukau, B. (2000). Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat Struct Biol* 7, 586-593.

McCallum, C. D., Do, H., Johnson, A. E., and Frydman, J. (2000). The interaction of the chaperonin tailless complex polypeptide 1 (TCP1) ring complex (TRiC) with ribosome-bound nascent chains examined using photo-cross-linking. *J Cell Biol* 149, 591-602.

McGowan, K. A., Li, J. Z., Park, C. Y., Beaudry, V., Tabor, H. K., Sabnis, A. J., Zhang, W., Fuchs, H., de Angelis, M. H., Myers, R. M., Attardi, L. D., and Barsh, G. S. (2008). Ribosomal mutations cause p53-mediated dark skin and pleiotropic effects. *Nat Genet* 40, 963-970.

McKenzie, S. L., Henikoff, S., and Meselson, M. (1975). Localization of RNA from heat-induced polysomes at puff sites in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 72, 1117-1121.

McMillan, H. J., Carter, M. T., Jacob, P. J., Laffan, E. E., O'Connor, M. D., and Boycott, K. M. (2009). Homozygous contiguous gene deletion of 13q12 causing LGMD2C and ARSACS in the same patient. *Muscle Nerve* 39, 396-399.

Meetei, A. R., de Winter, J. P., Medhurst, A. L., Wallisch, M., Waisfisz, Q., van de Vrugt, H. J., Oostra, A. B., Yan, Z., Ling, C., Bishop, C. E., Hoatlin, M. E., Joenje, H., *et al.* (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* 35, 165-170.

Mercier, J., Prevost, C., Engert, J. C., Bouchard, J. P., Mathieu, J., and Richter, A. (2001). Rapid detection of the sasin mutations causing autosomal recessive spastic ataxia of Charlevoix-Saguenay. *Genet Test* 5, 255-259.

Meyer, P., Prodromou, C., Hu, B., Vaughan, C., Roe, S. M., Panaretou, B., Piper, P. W., and Pearl, L. H. (2003). Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol Cell* 11, 647-658.

Mickler, M., Hessling, M., Ratzke, C., Buchner, J., and Hugel, T. (2009). The large conformational changes of Hsp90 are only weakly coupled to ATP hydrolysis. *Nat Struct Mol Biol* 16, 281-286.

Minami, Y., Hohfeld, J., Ohtsuka, K., and Hartl, F. U. (1996). Regulation of the heat-shock protein 70 reaction cycle by the mammalian DnaJ homolog, Hsp40. *J Biol Chem* 271, 19617-19624.

Mogk, A., Haslberger, T., Tessarz, P., and Bukau, B. (2008). Common and specific mechanisms of AAA+ proteins involved in protein quality control. *Biochem Soc Trans* 36, 120-125.

Mrissa, N., Belal, S., Hamida, C. B., Amouri, R., Turki, I., Mrissa, R., Hamida, M. B., and Hentati, F. (2000). Linkage to chromosome 13q11-12 of an autosomal recessive cerebellar ataxia in a Tunisian family. *Neurology* 54, 1408-1414.

Muchowski, P. J. (2002). Protein misfolding, amyloid formation, and neurodegeneration: a critical role for molecular chaperones? *Neuron* 35, 9-12.

Muchowski, P. J., and Wacker, J. L. (2005). Modulation of neurodegeneration by molecular chaperones. *Nat Rev Neurosci* 6, 11-22.

Nagamine, K., Peterson, P., Scott, H. S., Kudoh, J., Minoshima, S., Heino, M., Krohn, K. J., Lalioti, M. D., Mullis, P. E., Antonarakis, S. E., Kawasaki, K., Asakawa, S., *et al.* (1997). Positional cloning of the APECED gene. *Nat Genet* *17*, 393-398.

Nangle, L. A., Zhang, W., Xie, W., Yang, X. L., and Schimmel, P. (2007). Charcot-Marie-Tooth disease-associated mutant tRNA synthetases linked to altered dimer interface and neurite distribution defect. *Proc Natl Acad Sci U S A* *104*, 11239-11244.

Narla, A., and Ebert, B. L. (2010). Ribosomopathies: human disorders of ribosome dysfunction. *Blood* *115*, 3196-3205.

Nascimento, R. M., Otto, P. A., de Brouwer, A. P., and Vianna-Morgante, A. M. (2006). UBE2A, which encodes a ubiquitin-conjugating enzyme, is mutated in a novel X-linked mental retardation syndrome. *Am J Hum Genet* *79*, 549-555.

Netzer, W. J., and Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* *388*, 343-349.

Ng, I. O., Liu, C. L., Fan, S. T., and Ng, M. (2000). Expression of P-glycoprotein in hepatocellular carcinoma. A determinant of chemotherapy response. *Am J Clin Pathol* *113*, 355-363.

Ng, J. M., Vermeulen, W., van der Horst, G. T., Bergink, S., Sugasawa, K., Vrieling, H., and Hoeijmakers, J. H. (2003). A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein. *Genes Dev* *17*, 1630-1645.

Nillegoda, N. B., Theodoraki, M. A., Mandal, A. K., Mayo, K. J., Ren, H. Y., Sultana, R., Wu, K., Johnson, J., Cyr, D. M., and Caplan, A. J. (2010). Ubr1 and ubr2 function in a quality control pathway for degradation of unfolded cytosolic proteins. *Mol Biol Cell* *21*, 2102-2116.

Nolting, B., and Agard, D. A. (2008). How general is the nucleation-condensation mechanism? *Proteins* *73*, 754-764.

Nureki, O., Vassilyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T. L., Schimmel, P., and Yokoyama, S. (1998). Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* *280*, 578-582.

Obermann, W. M., Sonderrmann, H., Russo, A. A., Pavletich, N. P., and Hartl, F. U. (1998). In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. *J Cell Biol* *143*, 901-910.

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

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

Olzscha, H., Schermann, S. M., Woerner, A. C., Pinkert, S., Hecht, M. H., Tartaglia, G. G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F. U., and Vabulas, R. M. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* *144*, 67-78.

Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* *48*, 545-600.

Ostrovsky, O., Makarewich, C. A., Snapp, E. L., and Argon, Y. (2009). An essential role for ATP binding and hydrolysis in the chaperone activity of GRP94 in cells. *Proc Natl Acad Sci U S A* *106*, 11600-11605.

Otey, C. A., and Carpen, O. (2004). Alpha-actinin revisited: a fresh look at an old player. *Cell Motil Cytoskeleton* *58*, 104-111.

Ouyang, Y., Segers, K., Bouquiaux, O., Wang, F. C., Janin, N., Andris, C., Shimazaki, H., Sakoe, K., Nakano, I., and Takiyama, Y. (2007). Novel SACS mutation in a Belgian family with saccin-related ataxia. *J Neurol Sci* *21*, 21.

Ouyang, Y., Takiyama, Y., Sakoe, K., Shimazaki, H., Ogawa, T., Nagano, S., Yamamoto, Y., and Nakano, I. (2006). Saccin-related ataxia (ARSACS): expanding the genotype upstream from the gigantic exon. *Neurology* *66*, 1103-1104.

Ozelius, L. J., Hewett, J. W., Page, C. E., Bressman, S. B., Kramer, P. L., Shalish, C., de Leon, D., Brin, M. F., Raymond, D., Corey, D. P., Fahn, S., Risch, N. J., *et al.* (1997). The early-onset torsion dystonia gene (DYT1) encodes an ATP-binding protein. *Nat Genet* *17*, 40-48.

Panaretou, B., Prodromou, C., Roe, S. M., O'Brien, R., Ladbury, J. E., Piper, P. W., and Pearl, L. H. (1998). ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. *Embo J* *17*, 4829-4836.

Panaretou, B., Siligardi, G., Meyer, P., Maloney, A., Sullivan, J. K., Singh, S., Millson, S. H., Clarke, P. A., Naaby-Hansen, S., Stein, R., Cramer, R., Mollapour, M., *et al.* (2002). Activation of the ATPase activity of hsp90 by the stress-regulated cochaperone aha1. *Mol Cell* *10*, 1307-1318.

Parfitt, D. A., Michael, G. J., Vermeulen, E. G., Prodromou, N. V., Webb, T. R., Gallo, J. M., Cheetham, M. E., Nicoll, W. S., Blatch, G. L., and Chapple, J. P. (2009). The ataxia protein saccin is a functional co-chaperone that protects against polyglutamine-expanded ataxin-1. *Hum Mol Genet* *18*, 1556-1565.

Parvari, R., HersHKovitz, E., Grossman, N., Gorodischer, R., Loeys, B., Zecic, A., Mortier, G., Gregory, S., Sharony, R., Kambouris, M., Sakati, N., Meyer, B. F., *et al.* (2002). Mutation of TBCE causes hypoparathyroidism-retardation-dysmorphism and autosomal recessive Kenny-Caffey syndrome. *Nat Genet* *32*, 448-452.

Pauli-Magnus, C., and Kroetz, D. L. (2004). Functional implications of genetic polymorphisms in the multidrug resistance gene MDR1 (ABCB1). *Pharm Res* *21*, 904-913.

Pearl, L. H., and Prodromou, C. (2001). Structure, function, and mechanism of the Hsp90 molecular chaperone. *Adv Protein Chem* *59*, 157-186.

Pearl, L. H., and Prodromou, C. (2006). Structure and mechanism of the hsp90 molecular chaperone machinery. *Annu Rev Biochem* *75*, 271-294.

Peattie, D. A., Harding, M. W., Fleming, M. A., DeCenzo, M. T., Lippke, J. A., Livingston, D. J., and Benasutti, M. (1992). Expression and characterization of human FKBP52, an immunophilin that associates with the 90-kDa heat shock protein and is a component of steroid receptor complexes. *Proc Natl Acad Sci U S A* *89*, 10974-10978.

Pedersen, L. C., Benning, M. M., and Holden, H. M. (1995). Structural investigation of the antibiotic and ATP-binding sites in kanamycin nucleotidyltransferase. *Biochemistry* *34*, 13305-13311.

Pelham, H. R. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell* 46, 959-961.

Pellecchia, M., Montgomery, D. L., Stevens, S. Y., Vander Kooi, C. W., Feng, H. P., Gierasch, L. M., and Zuiderweg, E. R. (2000). Structural insights into substrate binding by the molecular chaperone DnaK. *Nat Struct Biol* 7, 298-303.

Phillips, J. J., Yao, Z. P., Zhang, W., McLaughlin, S., Laue, E. D., Robinson, C. V., and Jackson, S. E. (2007). Conformational dynamics of the molecular chaperone Hsp90 in complexes with a co-chaperone and anticancer drugs. *J Mol Biol* 372, 1189-1203.

Picard, D. (2006). Chaperoning steroid hormone action. *Trends Endocrinol Metab* 17, 229-235.

Pickart, C. M., and Eddins, M. J. (2004). Ubiquitin: structures, functions, mechanisms. *Biochim Biophys Acta* 1695, 55-72.

Plaxco, K. W., Simons, K. T., and Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277, 985-994.

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

Purvis, I. J., Bettany, A. J., Santiago, T. C., Coggins, J. R., Duncan, K., Eason, R., and Brown, A. J. (1987). The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J Mol Biol* 193, 413-417.

Quaderi, N. A., Schweiger, S., Gaudenz, K., Franco, B., Rugarli, E. I., Berger, W., Feldman, G. J., Volta, M., Andolfi, G., Gilgenkrantz, S., Marion, R. W., Hennekam, R. C., *et al.* (1997). Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nat Genet* 17, 285-291.

Ramakrishnan, V. (2010). Unraveling the structure of the ribosome (Nobel Lecture). *Angew Chem Int Ed Engl* 49, 4355-4380.

Ramser, J., Ahearn, M. E., Lenski, C., Yariz, K. O., Hellebrand, H., von Rhein, M., Clark, R. D., Schmutzler, R. K., Lichtner, P., Hoffman, E. P., Meindl, A., and Baumbach-Reardon, L. (2008). Rare missense and synonymous variants in UBE1 are associated with X-linked infantile spinal muscular atrophy. *Am J Hum Genet* 82, 188-193.

Rezaei, N., Sabbaghian, M., Liu, Z., and Zenker, M. (2011). Eponym : Johanson-Blizzard syndrome. *Eur J Pediatr* 2010, 17.

Richter, A., Morgan, K., Bouchard, J. P., Poirier, J., Mercier, J., Gosselin, F., and Melancon, S. B. (1993). Clinical and molecular genetic studies on autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS). *Adv Neurol* 61, 97-103.

Richter, A., Rioux, J. D., Bouchard, J. P., Mercier, J., Mathieu, J., Ge, B., Poirier, J., Julien, D., Gyapay, G., Weissenbach, J., Hudson, T. J., Melancon, S. B., *et al.* (1999). Location score and haplotype analyses of the locus for autosomal recessive spastic ataxia of Charlevoix-Saguenay, in chromosome region 13q11. *Am J Hum Genet* 64, 768-775.

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

Richter, K., Haslbeck, M., and Buchner, J. (2010). The heat shock response: life on the verge of death. *Mol Cell* 40, 253-266.

Richter, K., Muschler, P., Hainzl, O., and Buchner, J. (2001). Coordinated ATP hydrolysis by the Hsp90 dimer. *J Biol Chem* 276, 33689-33696.

Richter, K., Muschler, P., Hainzl, O., Reinstein, J., and Buchner, J. (2003). Sti1 is a non-competitive inhibitor of the Hsp90 ATPase. Binding prevents the N-terminal dimerization reaction during the atpase cycle. *J Biol Chem* 278, 10328-10333.

Richter, M. K., Bouchard JP., Mathieu J., Lamrache J., Rioux J., Hudson T., Melancon SB. (1996). ARSACS: Possibly a lysosomal storage disease? *Am J Hum Genet Suppl* 59, A379.

Rise, M. L., Hall, J. R., Rise, M., Hori, T. S., Browne, M. J., Gamperl, A. K., Hubert, S., Kimball, J., Bowman, S., and Johnson, S. C. (2010). Impact of asymptomatic nodavirus carrier state and intraperitoneal viral mimic injection on brain transcript expression in Atlantic cod (*Gadus morhua*). *Physiol Genomics* 42, 266-280.

Robledo, S., Idol, R. A., Crimmins, D. L., Ladenson, J. H., Mason, P. J., and Bessler, M. (2008). The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *Rna* 14, 1918-1929.

Ross, C. A., and Tabrizi, S. J. (2011). Huntington's disease: from molecular pathogenesis to clinical treatment. *Lancet Neurol* 10, 83-98.

Rudiger, S., Buchberger, A., and Bukau, B. (1997a). Interaction of Hsp70 chaperones with substrates. *Nat Struct Biol* 4, 342-349.

Rudiger, S., Germeroth, L., Schneider-Mergener, J., and Bukau, B. (1997b). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *Embo J* 16, 1501-1507.

Sadqi, M., Lapidus, L. J., and Munoz, V. (2003). How fast is protein hydrophobic collapse? *Proc Natl Acad Sci U S A* 100, 12117-12122.

Saito, H., and Uchida, H. (1977). Initiation of the DNA replication of bacteriophage lambda in *Escherichia coli* K12. *J Mol Biol* 113, 1-25.

Sakono, M., and Zako, T. (2010). Amyloid oligomers: formation and toxicity of Abeta oligomers. *Febs J* 277, 1348-1358.

Sanjad, S. A., Sakati, N. A., Abu-Osba, Y. K., Kaddoura, R., and Milner, R. D. (1991). A new syndrome of congenital hypoparathyroidism, severe growth failure, and dysmorphic features. *Arch Dis Child* 66, 193-196.

Saunders, R., and Deane, C. M. (2010). Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res* 2010, 8.

Scheibel, T., Neuhofen, S., Weikl, T., Mayr, C., Reinstein, J., Vogel, P. D., and Buchner, J. (1997). ATP-binding properties of human Hsp90. *J Biol Chem* 272, 18608-18613.

Scheibel, T., Siegmund, H. I., Jaenicke, R., Ganz, P., Lilie, H., and Buchner, J. (1999). The charged region of Hsp90 modulates the function of the N-terminal domain. *Proc Natl Acad Sci U S A* 96, 1297-1302.

Scheibel, T., Weikl, T., and Buchner, J. (1998). Two chaperone sites in Hsp90 differing in substrate specificity and ATP dependence. *Proc Natl Acad Sci U S A* 95, 1495-1499.

Scheper, G. C., van der Klok, T., van Andel, R. J., van Berkel, C. G., Sissler, M., Smet, J., Muravina, T. I., Serkov, S. V., Uziel, G., Bugiani, M., Schiffmann, R., Krageloh-Mann, I., *et al.* (2007). Mitochondrial aspartyl-tRNA synthetase deficiency causes

leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation. *Nat Genet* 39, 534-539.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* 101, 199-210.

Schmid, F. X. (2001). Prolyl isomerases. *Adv Protein Chem* 59, 243-282.

Schoser, B. G., Frosk, P., Engel, A. G., Klutzny, U., Lochmuller, H., and Wrogemann, K. (2005). Commonality of TRIM32 mutation in causing sarcotubular myopathy and LGMD2H. *Ann Neurol* 57, 591-595.

Seelig, A., Blatter, X. L., and Wohnsland, F. (2000). Substrate recognition by P-glycoprotein and the multidrug resistance-associated protein MRP1: a comparison. *Int J Clin Pharmacol Ther* 38, 111-121.

Selkoe, D. J. (2001). Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev* 81, 741-766.

Selkoe, D. J. (2008). Soluble oligomers of the amyloid beta-protein impair synaptic plasticity and behavior. *Behav Brain Res* 192, 106-113.

Sharp, P. M., and Li, W. H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15, 1281-1295.

Shaw, B. F., and Valentine, J. S. (2007). How do ALS-associated mutations in superoxide dismutase 1 promote aggregation of the protein? *Trends Biochem Sci* 32, 78-85.

Shiau, A. K., Harris, S. F., Southworth, D. R., and Agard, D. A. (2006). Structural Analysis of *E. coli* hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* 127, 329-340.

Shimazaki, H., Sakoe, K., Nijima, K., Nakano, I., and Takiyama, Y. (2007). An unusual case of a spasticity-lacking phenotype with a novel SACS mutation. *J Neurol Sci* 255, 87-89.

Shimazaki, H., Takiyama, Y., Sakoe, K., Ando, Y., and Nakano, I. (2005). A phenotype without spasticity in sartin-related ataxia. *Neurology* 64, 2129-2131.

Shimura, H., Hattori, N., Kubo, S., Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K., and Suzuki, T. (2000). Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nat Genet* 25, 302-305.

Siller, E., Dezwaan, D. C., Anderson, J. F., Freeman, B. C., and Barral, J. M. (2010). Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency. *J Mol Biol* 2010, 4.

Skre, H. (1974). Genetic and clinical aspects of Charcot-Marie-Tooth's disease. *Clin Genet* 6, 98-118.

Slavotinek, A. M., Stone, E. M., Mykityn, K., Heckenlively, J. R., Green, J. S., Heon, E., Musarella, M. A., Parfrey, P. S., Sheffield, V. C., and Biesecker, L. G. (2000). Mutations in MKKS cause Bardet-Biedl syndrome. *Nat Genet* 26, 15-16.

Sorensen, M. A., and Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in *Escherichia coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J Mol Biol* 222, 265-280.

Srivastava, M., Begovic, E., Chapman, J., Putnam, N. H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M. L., Signorovitch, A. Y., Moreno, M. A., *et al.* (2008). The *Trichoplax* genome and the nature of placozoans. *Nature* 454, 955-960.

Stebbins, C. E., Russo, A. A., Schneider, C., Rosen, N., Hartl, F. U., and Pavletich, N. P. (1997). Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. *Cell* 89, 239-250.

Steel, G. J., Fullerton, D. M., Tyson, J. R., and Stirling, C. J. (2004). Coordinated activation of Hsp70 chaperones. *Science* 303, 98-101.

Steitz, T. A. (2010). From the structure and function of the ribosome to new antibiotics (Nobel Lecture). *Angew Chem Int Ed Engl* 49, 4381-4398.

Stirling, P. C., Bakhoun, S. F., Feigl, A. B., and Leroux, M. R. (2006). Convergent evolution of clamp-like binding sites in diverse chaperones. *Nat Struct Mol Biol* 13, 865-870.

Stone, D. L., Slavotinek, A., Bouffard, G. G., Banerjee-Basu, S., Baxevas, A. D., Barr, M., and Biesecker, L. G. (2000). Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome. *Nat Genet* 25, 79-82.

Street, T. O., Bolen, D. W., and Rose, G. D. (2006). A molecular mechanism for osmolyte-induced protein stability. *Proc Natl Acad Sci U S A* 103, 13997-14002.

Sullivan, W., Stensgard, B., Caucutt, G., Bartha, B., McMahon, N., Alnemri, E. S., Litwack, G., and Toft, D. (1997). Nucleotides and two functional states of hsp90. *J Biol Chem* 272, 8007-8012.

Szabo, A., Langer, T., Schroder, H., Flanagan, J., Bukau, B., and Hartl, F. U. (1994). The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system DnaK, DnaJ, and GrpE. *Proc Natl Acad Sci U S A* 91, 10345-10349.

Szebenyi, G., Bollati, F., Bisbal, M., Sheridan, S., Faas, L., Wray, R., Haferkamp, S., Nguyen, S., Caceres, A., and Brady, S. T. (2005). Activity-driven dendritic remodeling requires microtubule-associated protein 1A. *Curr Biol* 15, 1820-1826.

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

Takiyama, Y. (2007). Sacsinopathies: saccin-related ataxia. *Cerebellum* 28, 1-7.

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

Tanaka, K., Suzuki, T., and Chiba, T. (1998). The ligation systems for ubiquitin and ubiquitin-like proteins. *Mol Cells* 8, 503-512.

Tang, B. S., Zhao, G. H., Luo, W., Xia, K., Cai, F., Pan, Q., Zhang, R. X., Zhang, F. F., Liu, X. M., Chen, B., Zhang, C., Shen, L., *et al.* (2005). Small heat-shock protein 22 mutated in autosomal dominant Charcot-Marie-Tooth disease type 2L. *Hum Genet* 116, 222-224.

Tasaki, T., and Kwon, Y. T. (2007). The mammalian N-end rule pathway: new insights into its components and physiological roles. *Trends Biochem Sci* 32, 520-528.

Terracciano, A., Casali, C., Grieco, G. S., Orteschi, D., Di Giandomenico, S., Seminara, L., Di Fabio, R., Carrozzo, R., Simonati, A., Stevanin, G., Zollino, M., and Santorelli, F. M. (2009). An inherited large-scale rearrangement in SACS associated with spastic ataxia and hearing loss. *Neurogenetics* 10, 151-155.

Trockenbacher, A., Suckow, V., Foerster, J., Winter, J., Krauss, S., Ropers, H. H., Schneider, R., and Schweiger, S. (2001). MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. *Nat Genet* 29, 287-294.

Tsai, C. J., Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M., and Nussinov, R. (2008). Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J Mol Biol* 383, 281-291.

Uchida, D., Hatakeyama, S., Matsushima, A., Han, H., Ishido, S., Hotta, H., Kudoh, J., Shimizu, N., Doucas, V., Nakayama, K. I., Kuroda, N., and Matsumoto, M. (2004). AIRE functions as an E3 ubiquitin ligase. *J Exp Med* 199, 167-172.

Vainberg, I. E., Lewis, S. A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H. L., and Cowan, N. J. (1998). Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* 93, 863-873.

Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol* 180, 549-576.

Varshney, A., Sen, P., Ahmad, E., Rehan, M., Subbarao, N., and Khan, R. H. (2010). Ligand binding strategies of human serum albumin: how can the cargo be utilized? *Chirality* 22, 77-87.

Vaughan, C. K., Gohlke, U., Sobott, F., Good, V. M., Ali, M. M., Prodromou, C., Robinson, C. V., Saibil, H. R., and Pearl, L. H. (2006). Structure of an Hsp90-Cdc37-Cdk4 complex. *Mol Cell* 23, 697-707.

Vermeer, S., Meijer, R. P., Hofste, T. G., Bodmer, D., Bosgoed, E. A., Cremers, F. P., Kremer, B. H., Knoers, N. V., and 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, R. P., Pijl, B. J., Timmermans, J., Cruysberg, J. R., Bos, M. M., Schelhaas, H. J., van de Warrenburg, B. P., Knoers, N. V., Scheffer, H., and Kremer, B. (2008). ARSACS in the Dutch population: a frequent cause of early-onset cerebellar ataxia. *Neurogenetics* 9, 207-214.

Vicart, P., Caron, A., Guicheney, P., Li, Z., Prevost, M. C., Faure, A., Chateau, D., Chapon, F., Tome, F., Dupret, J. M., Paulin, D., and Fardeau, M. (1998). A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy. *Nat Genet* 20, 92-95.

Viola, K. L., Velasco, P. T., and Klein, W. L. (2008). Why Alzheimer's is a disease of memory: the attack on synapses by A beta oligomers (ADDLs). *J Nutr Health Aging* 12, 51S-57S.

Viviani, V. R. (2002). The origin, diversity, and structure function relationships of insect luciferases. *Cell Mol Life Sci* 59, 1833-1850.

Voellmy, R., and Boellmann, F. (2007). Chaperone regulation of the heat shock protein response. *Adv Exp Med Biol* 594, 89-99.

Vogel, C., Berzuini, C., Bashton, M., Gough, J., and Teichmann, S. A. (2004). Supra-domains: evolutionary units larger than single protein domains. *J Mol Biol* 336, 809-823.

Voisine, C., Pedersen, J. S., and Morimoto, R. I. (2010). Chaperone networks: tipping the balance in protein folding diseases. *Neurobiol Dis* 40, 12-20.

Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J., Jr. (2005). Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* 44, 7342-7372.

Walter, S., and Buchner, J. (2002). Molecular chaperones--cellular machines for protein folding. *Angew Chem Int Ed Engl* 41, 1098-1113.

Wandinger, S. K., Richter, K., and Buchner, J. (2008). The Hsp90 chaperone machinery. *J Biol Chem* 283, 18473-18477.

Wang, C., Ko, H. S., Thomas, B., Tsang, F., Chew, K. C., Tay, S. P., Ho, M. W., Lim, T. M., Soong, T. W., Pletnikova, O., Troncoso, J., Dawson, V. L., *et al.* (2005). Stress-induced alterations in parkin solubility promote parkin aggregation and compromise parkin's protective function. *Hum Mol Genet* 14, 3885-3897.

Wang, H., Kurochkin, A. V., Pang, Y., Hu, W., Flynn, G. C., and Zuiderweg, E. R. (1998). NMR solution structure of the 21 kDa chaperone protein DnaK substrate binding domain: a preview of chaperone-protein interaction. *Biochemistry* 37, 7929-7940.

Warner, J. R., Rich, A., and Hall, C. E. (1962). Electron Microscope Studies of Ribosomal Clusters Synthesizing Hemoglobin. *Science* 138, 1399-1403.

Welch, W. J. (2004). Role of quality control pathways in human diseases involving protein misfolding. *Semin Cell Dev Biol* 15, 31-38.

Wiech, H., Buchner, J., Zimmermann, R., and Jakob, U. (1992). Hsp90 chaperones protein folding in vitro. *Nature* 358, 169-170.

Wilkinson, B., and Gilbert, H. F. (2004). Protein disulfide isomerase. *Biochim Biophys Acta* 1699, 35-44.

Willig, T. N., Gazda, H., and Sieff, C. A. (2000). Diamond-Blackfan anemia. *Curr Opin Hematol* 7, 85-94.

Winklhofer, K. F., Tatzelt, J., and Haass, C. (2008). The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. *Embo J* 27, 336-349.

Wong, E. S., Tan, J. M., Wang, C., Zhang, Z., Tay, S. P., Zaiden, N., Ko, H. S., Dawson, V. L., Dawson, T. M., and Lim, K. L. (2007). Relative sensitivity of parkin and other cysteine-containing enzymes to stress-induced solubility alterations. *J Biol Chem* 282, 12310-12318.

Xia, Z., Webster, A., Du, F., Piatkov, K., Ghislain, M., and Varshavsky, A. (2008). Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. *J Biol Chem* 283, 24011-24028.

Xie, W., Nangle, L. A., Zhang, W., Schimmel, P., and Yang, X. L. (2007). Long-range structural effects of a Charcot-Marie-Tooth disease-causing mutation in human glycyl-tRNA synthetase. *Proc Natl Acad Sci U S A* 104, 9976-9981.

Xu, J., and Xia, J. (2006). Structure and function of PICK1. *Neurosignals* 15, 190-201.

Yamamoto, N., Matsubara, E., Maeda, S., Minagawa, H., Takashima, A., Maruyama, W., Michikawa, M., and Yanagisawa, K. (2007). A ganglioside-induced toxic soluble Abeta assembly. Its enhanced formation from Abeta bearing the Arctic mutation. *J Biol Chem* 282, 2646-2655.

Yamamoto, Y., Hiraoka, K., Araki, M., Nagano, S., Shimazaki, H., Takiyama, Y., and Sakoda, S. (2005). Novel compound heterozygous mutations in sarsin-related ataxia. *J Neurol Sci* 239, 101-104.

Yamamoto, Y., Nakamori, M., Konaka, K., Nagano, S., Shimazaki, H., Takiyama, Y., and Sakoda, S. (2006). Sarsin-related ataxia caused by the novel nonsense mutation Arg4325X. *J Neurol* 253, 1372-1373.

Yen, H. C., Xu, Q., Chou, D. M., Zhao, Z., and Elledge, S. J. (2008). Global protein stability profiling in mammalian cells. *Science* 322, 918-923.

Yonath, A. (2010). Polar bears, antibiotics, and the evolving ribosome (Nobel Lecture). *Angew Chem Int Ed Engl* 49, 4341-4354.

Yonehara, M., Minami, Y., Kawata, Y., Nagai, J., and Yahara, I. (1996). Heat-induced chaperone activity of HSP90. *J Biol Chem* 271, 2641-2645.

Young, J. C., Agashe, V. R., Siegers, K., and Hartl, F. U. (2004). Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol* 5, 781-791.

Young, J. C., Schneider, C., and Hartl, F. U. (1997). In vitro evidence that hsp90 contains two independent chaperone sites. *FEBS Lett* 418, 139-143.

Yura, T., Tobe, T., Ito, K., and Osawa, T. (1984). Heat shock regulatory gene (htpR) of *Escherichia coli* is required for growth at high temperature but is dispensable at low temperature. *Proc Natl Acad Sci U S A* 81, 6803-6807.

Zenker, M., Mayerle, J., Lerch, M. M., Tagariello, A., Zerres, K., Durie, P. R., Beier, M., Hulskamp, G., Guzman, C., Rehder, H., Beemer, F. A., Hamel, B., *et al.* (2005). Deficiency of UBR1, a ubiquitin ligase of the N-end rule pathway, causes pancreatic dysfunction, malformations and mental retardation (Johanson-Blizzard syndrome). *Nat Genet* 37, 1345-1350.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* 16, 274-280.

Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E., and Hendrickson, W. A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* 272, 1606-1614.

Vita

John F. Anderson was born on September 26, 1983 in Wichita, Kansas to Steven E. Anderson and Vicki F. Anderson. After graduating from Berean Academy in Elbing, Kansas, he attended Kansas State University where he graduated Cum Laude with a Bachelor of Science degree in biology. At Kansas State University he performed research in the laboratory of Dr. Helmut Hirt on the prevalence and spread of antibiotic resistance determinant in *Enterococcus* species. This work was funded by an award from the Kansas Idea Network of Biomedical Research Excellence where he was designated as a STAR trainee. Mr. Anderson's undergraduate research was published in Applied and Environmental Microbiology. At Kansas State University, Mr. Anderson was named the most promising student in the division of biology (2005). After graduation, Mr. Anderson enrolled in the combined M.D./Ph.D. program at the University of Texas Medical Branch (UTMB). He has received several honors in both academics and research during his time at UTMB. These include the Student Performing at Maximum in Gross Anatomy and Radiology award (2006), the "Who's Anatomically Correct" Grand Champion award (2006) and the Truman Graves Blocker Junior Research Scholar award (2007) for his performance in the first two years of medical school. During his graduate studies he received the Second Place Overall Poster Award at the National Student Research Forum (2007), the Seymour Fisher Award for Academic Excellence (2007), the 14th Annual Sealy Center for Structural Biology Poster Award (2008), the UTMB GSBS Associates Scholarship Award (2009 and 2010) and the William D. and Jean C. Willis Research Scholarship Award (2009). Mr. Anderson's graduate research has been published in the Journal of Molecular Biology and Protein and Peptide Letters. During graduate school

Mr. Anderson was involved with education and committee responsibilities. He served as a mentor for the Bromberg scholars program (2009-2010), was a member of the recruitment committee for the M.D./Ph.D. program (2008, 2009, 2011), the seminar organization committee for the M.D./Ph.D. program (2009-2010), a co-director for the National Student Research Forum (2010), and the vice-president of the M.D./Ph.D. student organization (2009-2010).

Permanent address: 2814 Seastrand Ln., Dickinson, TX, 77539

This dissertation was typed by John F. Anderson.