

University of Alberta

Purification and Characterization of SMARCAL1 Orthologs
in an *E. coli* Expression System

by

Michael Burkat

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biochemistry

© Michael Burkat
Edmonton, Alberta
Fall 2013

Permission is hereby granted to the University of Alberta Library to reproduce
single copies of this thesis and to lend or sell such copies for private, scholarly or
scientific research purposes only.

The author reserves all other publication and other rights in association with the
copyright in the thesis, and except as herein before provided, neither the thesis nor
any substantial portion thereof may be printed or otherwise reproduced in any
material form whatever without the author's prior written permission.

Abstract

SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily A-like 1 (SMARCAL1) is a SWI2/SNF2 chromatin remodeler. It reanneals stalled replication forks to maintain genome integrity. It contains HARP domains that are necessary for the annealing activity. This thesis pursues structural studies of SMARCAL1 to understand its molecular mechanism. We establish an efficient *E. coli* expression system for the production of *R. norvegicus* and *C. elegans* SMARCAL1 orthologs. High yield and purity of six *R. norvegicus* proteins was obtained. One full length and five truncation mutants were designed to conserve the helicase domain and at least one HARP. These purified proteins were characterized and compared to an insect cell expressed human SMARCAL1 protein. DNA binding was assessed and the purified proteins displayed the typical preferential binding to fork DNA substrates. However, we determined that the DNA-stimulated ATPase activity reached similar catalytic rates for both fork and double stranded DNA.

Acknowledgements

I would like to thank Dr. Mark Glover for being a great mentor and providing the required push I needed to complete this degree. I want to thank all my lab members that made my time in the lab a pleasant experience. A thanks for Dr. Nicolas Coquelle, which provided much of the input in optimization of the ATPase assay. A thanks goes to Dr. Ross Edwards, who was literally behind me during my time in the lab and has provided countless science and life advice. A special thanks goes to Curtis Hodges, who helped me get started on the SMARCAL1 project and is the first resource I consulted when attempting any experiment. A special thanks to Melissa Morrison who gave me infinite encouragement.

Contents

1	Introduction	1
1.1	Helicases	2
1.1.1	SWI2/SNF2 family of helicases	4
1.2	Annealing Helicases	6
1.3	Stalled Replication Forks	9
1.4	ZRANB3	10
1.5	Schimke Immunoosseous Displasia	12
1.5.1	SIOD Associated SMARCAL1 Mutations	13
1.6	SMARCAL1	15
1.6.1	History	15
1.6.2	Biological Function of SMARCAL1	15
1.6.3	SMARCAL1 and RPA interactions	17
1.6.4	SMARCAL1 Phosphorylation	18
1.6.5	SMARCAL1 Fork Remodeling Model	18
1.7	Thesis Objective	19
2	Materials and Methods	21
2.1	Materials	21
2.2	Methods	25
2.2.1	Cloning	25
2.2.2	Expression and Purification	26
2.2.3	DNA Annealing	28
2.2.4	Electrophoretic mobility shift assay	28
2.2.5	ATPase assay	29
2.2.6	Crystallization trials	29
3	SMARCAL1 Expression and Purification	30
3.1	Introduction	31

3.2	Results	32
3.2.1	SMARCAL1 orthologs used in our studies	32
3.2.2	SMARCAL1 Heterologous Expression and Purification . .	35
3.3	Discussion	39
4	Characterization of Purified SMARCAL1 proteins	41
4.1	Introduction	42
4.2	Results	43
4.2.1	SMARCAL1 DNA binding activity by EMSA	43
4.2.2	SMARCAL1 DNA-dependent ATPase activity	45
4.3	Discussion	50
5	SMARCAL1 Crystallography Trials	52
5.1	Introduction	53
5.2	Results	54
5.2.1	rSMARCAL1 _{Δ210}	54
5.2.2	rSMARCAL1 ₂₈₇₋₈₅₀	54
5.3	Discussion	57
6	Conclusions	58
6.1	Conclusions	59
Bibliography		61
References		62

List of Figures

1.1	Roles of conserved SF2 helicase motifs	3
1.2	Helicase binding and translocation models	5
1.3	SWI2/SNF2 family structural and sequence characteristics	7
1.4	Annealing Helicases	8
1.5	Stalled replication forks	11
1.6	SIOD Mutations	14
1.7	SMARCAL1 fork regression model	20
3.1	SMARCAL1 ortholog sequence alignment	33
3.2	SMARCAL1 orthologs	34
3.3	Retention volume of rSMARCAL1 protiens	36
3.4	rSMARCAL1 size exclusion chromatography elution fraction	37
3.5	cSMARCAL1 expression and purification	38
4.1	SMARCAL1 DNA binding ability by EMSA	44
4.2	ATP/NADH coupled assay reaction scheme	46
4.3	rSMARCAL1 _{WT} fork DNA progression curves	47
4.4	SAMRCAL1 fork DNA and dsDNA ATPase activity	48

List of Tables

4.1	Michaelis-Menten kinetic parameters	49
5.1	Summary of all crystallization trials	56

Abbreviations

AH2	Annealing Helicase 2
AMP-PNP	Adenylylimidodiphosphate
ATP	Adenosine Triphosphate
BER	Base Excision Repair
BLM	Bloom Helicase
BME	2-Mercaptoethanol
bp	Base Pairs
BSA	Bovine Serum Albumin
CLS	Canadian Light Source
D-loop	Displacement Loop
DLC	Dynamic Light Scattering
DSBs	Double Strand Breaks
dsDNA	Double Stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
EMSA	Electrophoretic Mobility Shift Assay
GST	Glutathione S-Transferase
h	Hour

HARP	HepA Related Protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid
HPL	HARP-like
HU	Hydroxyurea
L	Liter
LB	Luria-Bertani
MCM	Minichromosome Maintenance
MDa	Mega Dalton
mg	Milligram
min	Minute
mL	Millilitre
mM	Millimolar
MMR	Mismatch Repair
MW	Molecular Weight
NADH	-Nicotinamide adenine dinucleotide
NER	Nucleotide Excision Repair
Ni-NTA	Nickel Nitrilotriacetic Acid
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase chain reaction
PEP	Phospho(enol)pyruvate
PK/LDH	Pyruvate Kinase/Lactic Dehydrogenase
PTMs	Post-Transnational Modifications
RAD54L	Radiation Sensitive isolate 54-Like
RapA	RNA Polymerase-associated Protein

RBM	RPA Binding Motif
RecG	ATP-dependent DNA Helicase RecG
RECQL	ATP-dependent DNA Helicase Q4
Replication Protein A RPA	
RFU	Relative Fluorescence Units
s	Seconds
SAXS	Small Angle X-ray Scattering
SCEs	Sister Chromatid Exchanges
SDS-Page	Sodium Docearyl Sulphate Polyacrylamide Gel Electrophoresis
SF	Superfamily
SIOD	Schimke immunoosseous dysplasia
slDNA	Stem-loop DNA
SMARCA1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 1
SMARCAL1	SWI/SNF-related Matrix-associated Actin-dependent Regulator of Chromatin Subfamily A-like Protein 1
ssDNA	Single Stranded DNA
SWI/SNF	Switch/Sucrose Non-fermentable
µL	Microliter
V	Volt
WRN	Werner Helicase
ZRANB3	Zinc Finger, RAN-binding Domain Containing 3

Chapter 1

Introduction

1.1 Helicases

The structure of DNA and its role as the carrier of a cells genetic information made this molecule central to the description of life¹. Studies of DNA polymerases show single stranded DNA (ssDNA) can be used as a template for DNA synthesis. However, double stranded DNA (dsDNA) templates lack replicative activity². This is due to the nature of DNA, which establishes a double helix, where hydrogen bonds keep the two strands together¹. The discovery of the first *Escherichia coli* helicase in 1976³ lead to the realization that for replication of dsDNA, the bonds holding the two strands together need to be broken. This is achieved at replication forks through helicases, which can use the energy from adenosine triphosphate (ATP) hydrolysis to catalyze the separation of dsDNA in a processive manner⁴. To confirm dsDNA unwinding is a conserved mechanism of replication, scientists searched for helicases across all species. The first mammalian helicase was purified in 1985⁵.

With the enormity of helicases now identified, they are classified in a hierarchy of superfamilies composed of families, which are divided into groups. There are now six superfamilies of helicases. Superfamily (SF) 1 and 2 helicases contain a conserved RecA like helicase core domain. SF3 to SF6 helicases form hexameric ring structures⁶. SF1 and SF2 helicases differ by sequence characteristics, although they also differ in functional, structural, and mechanistic features such as unwinding polarity.

SF2 helicases show a conservation across 12-14 motifs (Figure 1.1 A-B)⁶. The helicase core forms two lobes, which coordinate ATP and nucleic acid binding (Figure 1.1 C-D). Motif Q is important for ATP recognition and catalysis⁷. Motifs I, II and IV are highly conserved, and are involved in nucleotide binding and hydrolysis in a pocket between the two helicase domains⁶. Motifs III and Va are thought to coordinate the process of hydrolysis and DNA unwinding⁶. Motifs Ia, Ib, Ic, IV, IVa, V, Vb, provide a DNA binding surface across both helicase lobes, and they form

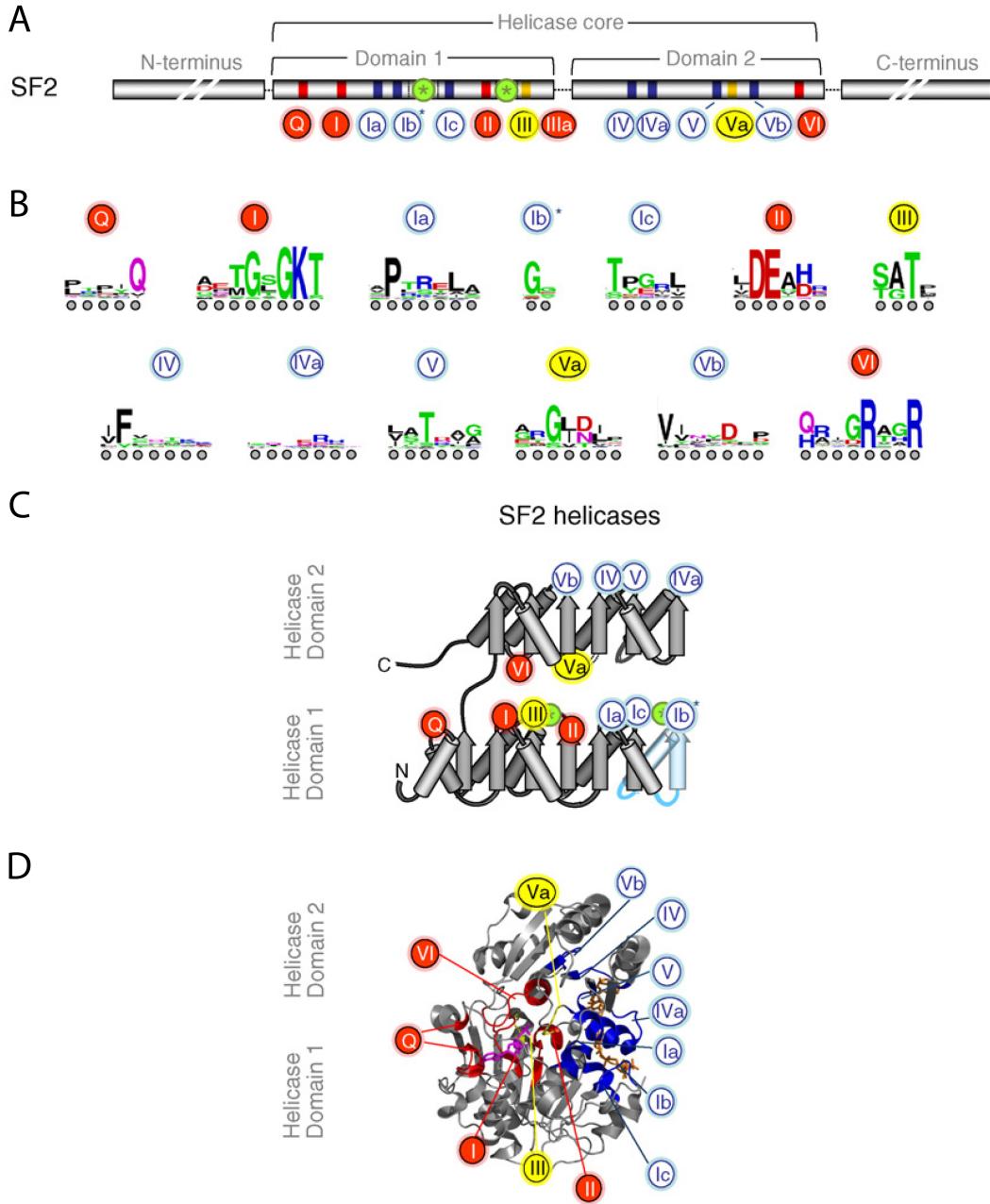


Figure 1.1: SF2 conserved motifs and their roles in ATP and nucleic acid binding. (A) SF2 helicases display 12-14 conserved motifs dispersed through two domains. They are colored dependent on their function: red, ATP binding; blue, DNA binding; yellow, coordination of ATP and DNA binding. (B) Sequence logos representing amino acid conservation among motifs. (C) Domain fold and positions of conserved motifs. Arrows represent β -strands, cylinders represent α -helices. (D) Positions of conserved motifs in the three-dimensional structure of Vasa helicase. An ATP analog is colored in magenta, and nucleic acid is colored in wheat⁶. Adapted from Fairman-Williams, M. E., Guenther, U.-P., and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* 20, 313–324.

direct contacts to the DNA phosphodiester backbone⁶. SF2 helicases are divided into 10 families dependent on sequence conservation across these motifs⁶.

1.1.1 SWI2/SNF2 family of helicases

In eukaryotic cells, DNA is compacted into a tight chromatin structure. However, this poses a problem for DNA replication, transcription and repair, where DNA must be accessible⁸. Eukaryotes have thus evolved proteins that catalyze chromatin remodeling reactions to disrupt and remodel protein nucleic acid interactions, which mediates accessibility. One family of such proteins takes its name from *Saccharomyces cerevisiae* genetic screens, where switch/sucrose non-fermentable (SWI/SNF) strains have mutations disrupting the SWI/SNF complex⁹. The SWI/SNF complex can regulate gene expression through the process of chromatin remodeling¹⁰. The yeast SWI/SNF complex is composed of up to 12 proteins, where the total molecular mass is above 1.5 MDa¹¹. At the core of the SWI/SNF complex there is a molecular motor that drives translocation along DNA strands to effect remodeling activity⁹. Sequence similarity to the SWI2/SNF2 motor protein is the basis of this family of helicases.

The SWI2/SNF2 proteins are not typical helicases, due to the lack of DNA unwinding activity, instead they act as chromatin remodeling ATPases, which can disrupt protein-nucleic acid interactions¹². SWI2/SNF2 helicases can bind to the minor groove of dsDNA and translocate along it (Figure 1.2 A)¹³. The mechanism of translocation is thought to have similarities to strand separation by helicases (Figure 1.2 B). This translocation mechanism is often referred to as the inchworm model¹⁴. This entails the sequential binding and slipping of helicase lobes, where movement is generated by an intermediate ATP hydrolysis step¹³.

The SWI2/SNF2 family has further sequence conservation additional to the typical RecA helicase motifs (Figure 1.3 C). A distinguishing sequence characteristic

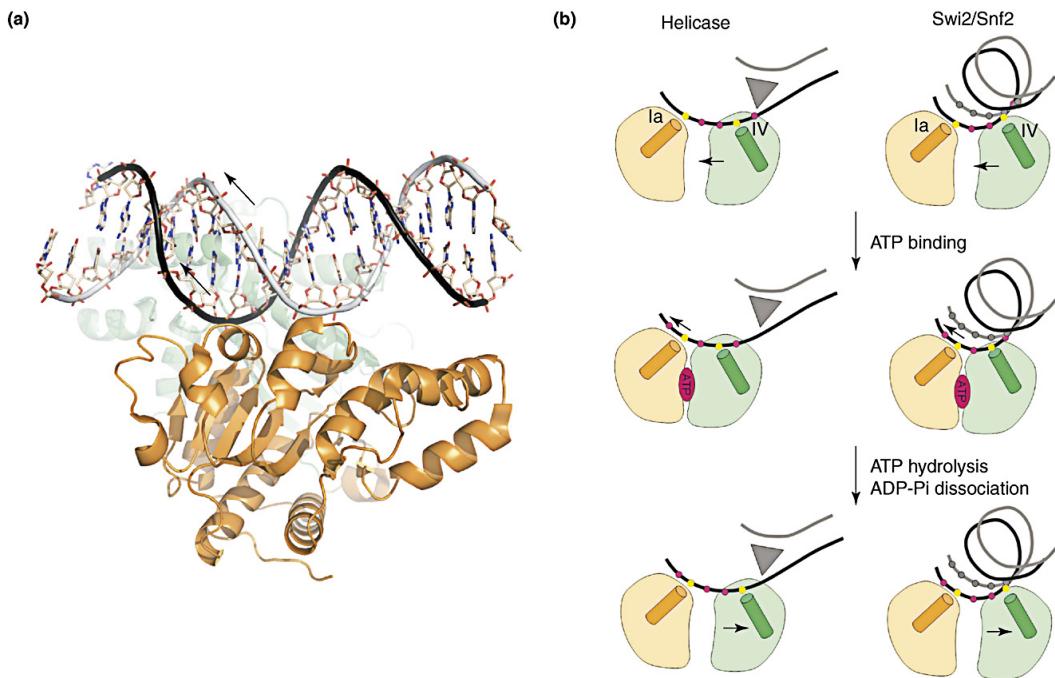


Figure 1.2: Helicase binding and translocation models. (A) Three-dimensional structure of *Sulfolobus solfataricus* SSO1653 bound to DNA. The helicase is shown bound to the minor groove of dsDNA. The N and C terminal RecA domains are colored orange and green respectively. (B) The inchworm model of movement along DNA is shown for an unwinding helicase and a SWI2/SNF2 helicase. The N and C terminal RecA domains are colored orange and green respectively. Conserved DNA binding motifs Ia and IV are shown. ATP binding may cause a slip in the first domain and shifting it closer to the second domain. ATP hydrolysis could then in turn push the second domain further¹³. Adapted from Hopfner, K.-P., and Michaelis, J. (2007) Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics. *Curr. Opin. Struct. Biol.* 17, 87–95.

of the SWI2/SNF2 family is a long insert of 160 amino acids or more between motifs III and IV of the RecA helicase lobes^{6,15}. This insert encodes two protrusions, a linker domain and provides an insertion region. C terminal to the second RecA lobe, SWI2/SNF2 helicases encode a brace (Figure 1.3 A, C). The presence of the insertion between the two helicase lobes does not seem to change the structure, but may instead affect the function (Figure 1.3 B)⁶.

The SWI2/SNF2 family is further divided into six groups, one of which is a distant group of three proteins: SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1 (SMARCAL1), zinc finger RAN-binding domain containing 3 (ZRANB3), and RNA polymerase-associated protein (RapA)¹⁵. Helicase motif conservation of the distant group differs by the absence of a conserved C block, which is between motifs III and IV¹⁵.

1.2 Annealing Helicases

The first evidence of strand annealing activity came from the *E. coli* RecA helicase in 1979¹⁶. Since then many other helicases exhibiting strand annealing activity have been identified¹⁷. Helicases with strand annealing activity are thought to be involved in DNA replication, repair, transcription and telomere metabolism¹⁷. The strand annealing domains have been mapped for many of these human proteins (Figure 1.4), but sequence alignment does not seem to show any conservation in these domains¹⁷.

The SWI2/SNF2 family contains four proteins that have been identified with annealing activity. These proteins show conserved HepA related protein (HARP) or HARP-like (HPL) domains¹⁸. These four proteins are: SMARCAL1, ZRANB3, radiation sensitive isolate 54-like (RAD54L) and SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 1 (SMARCA1). The

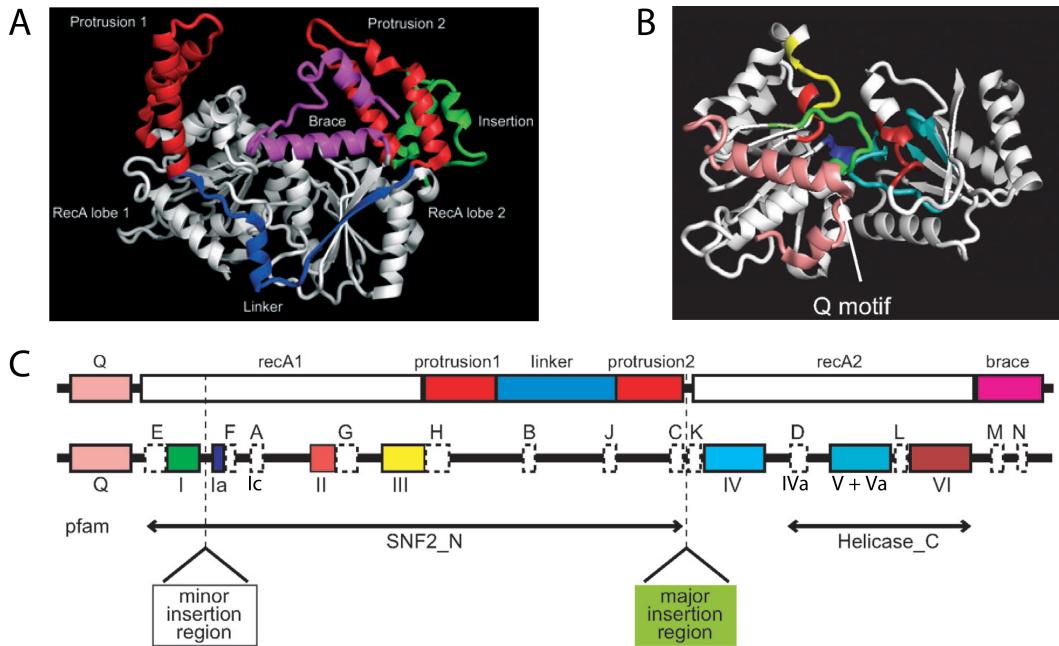


Figure 1.3: SWI2/SNF2 family structural and sequence characteristics. (A) Three-dimensional structure of zebrafish Rad54A. In gray are shown both RecA lobes, with additional SWI2/SNF2 structural features. Protrusions are colored red, the linker in blue, the brace in purple, and an insertion in green. (B) Conserved RecA motifs are colored: I in green, Ia in blue, II in red, III in yellow, IV in cyan, V and Va in teal and VI in brown. (C) Schematic representation of SWI2/SNF2 domains and conserved motifs, colored as in A and B¹⁵. Adapted from Flaus, A. (2006) Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res.* 34, 2887–2905.

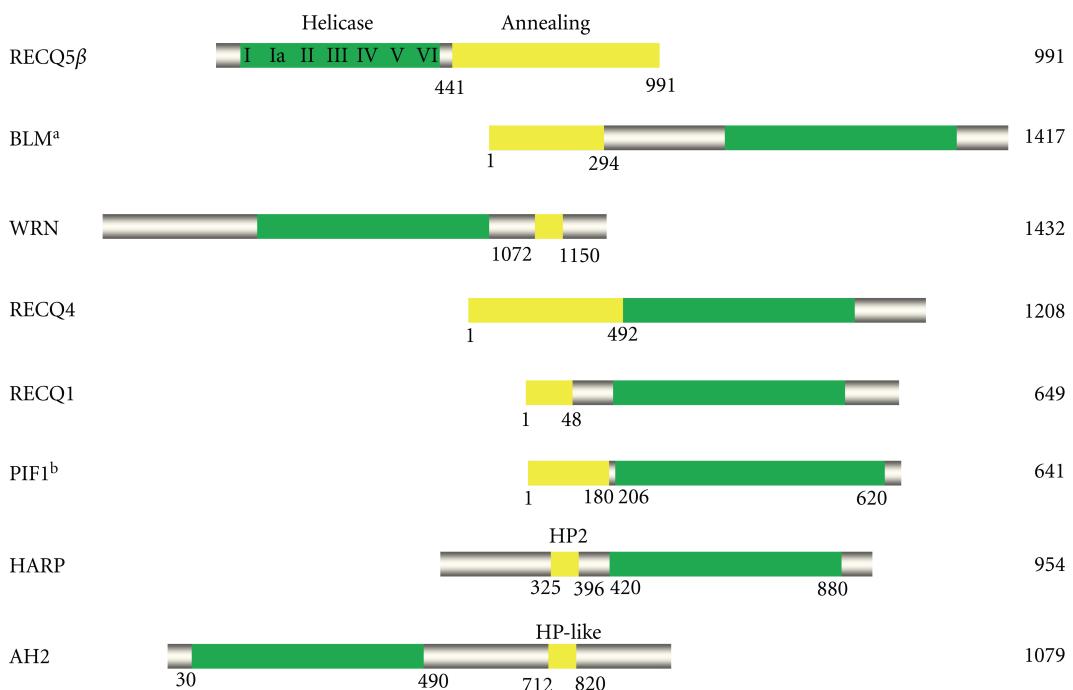


Figure 1.4: Human helicases with known strand annealing activity. The conserved helicase domain is defined in green. In yellow the domains which have been determined necessary for strand annealing activity are defined¹⁷. Adapted from Wu, Y. (2012) Unwinding and Rewinding: Double Faces of Helicase? *Journal of Nucleic Acids* 2012, 1–14.

dependence of annealing activity to these domains has been shown for SMARCAL1¹⁹ and ZRANB3¹⁸.

Within the RecQ family of SF2 helicases, the Bloom (BLM), Werner (WRN), ATP-dependent DNA helicase Q4 (RECQ4) and bacterial ATP-dependent DNA helicase RecG (RecG) helicases all exhibit strand annealing activity¹⁷. Autosomal recessive mutations in BLM, WRN and RECQ4 proteins give rise to Bloom's, Werner's, and Rothmund-Thomson syndromes respectively²⁰. These syndromes exhibit similar symptoms; they all show growth retardation, and all affect genetic stability of cells, leading to cancer predisposition²¹. However, beyond this the syndromes differ, indicating these helicases act non-redundantly²². Both BLM and WRN are recruited to stalled forks²⁰, additionally *in vitro* studies show they can regress DNA into a four-way junction commonly referred to as a chicken foot structures²³.

Many parallels can be made between RecQ family helicases and SMARCAL1. It has recently been proposed that SMARCAL1 may be a functional ortholog of RecG²⁴. RecG is a helicase that has been found to act on stalled replication forks to regress them into chicken foot structures and further exhibits fork reversal activity to restore the fork^{24–26}. The RecG model has been the basis for the newly proposed mechanism for SMARCAL1, described in section 1.6.5.

1.3 Stalled Replication Forks

DNA replication occurs during the S phase of the cell cycle. It is a very tightly controlled event to prevent over and under replication of the genome. When replication forks encounter DNA damage, low nucleotide pools, or DNA-protein complexes, forks can be stalled while obstructions are resolved. In the cases of damage or low nucleotides the minichromosome maintenance complex (MCM) helicase can

become uncoupled from the polymerase and generate long stretches of ssDNA. It is unclear if and how the MCM helicase is arrested in eukaryotic cells. Although, in *E. coli* unwinding generates ssDNA of several hundred base pairs (bp)²⁷. There are different models of fork resolution (Figure 1.5), although they are speculative since *in vivo* mammalian chicken foot structures have not been observed yet. However, a single molecule T4 replication system has shown fork regression and fork restart of these structures²⁸.

The template-switch mechanisms of repairing stalled forks allows for correct pairing of sister strands, whereas a double strand break mediated mechanism involving homologous recombination would have a higher rate of sister chromatid exchanges (SCEs), and thus lead to genetic instability³⁰. There is evidence of a fork restart mechanism, which has been seen through observation of DNA fibers. Additionally, cleavage and collapse of stalled forks is a time dependent event²⁹. Fork restart is controversial, due to the fact that once the MCM complex is removed, it may not be reloaded during S phase³¹. This leads to the idea that the ssDNA annealing and fork regression and reversal mechanisms may not be competent at restarting stalled forks, but instead stabilize them to prevent erroneous repair³¹. Replication can be completed in this way through dormant or late origin firing³².

1.4 ZRANB3

Both SMARCAL1 and ZRANB3 are distinct from other helicases due to their lack of unwinding activity. However, they exhibit strand annealing activity, thus they have been termed annealing helicases³³. Moreover ZRANB3 has been proposed to be renamed as annealing helicase 2 (AH2)³³. ZRANB3 can be recruited to polyubiquitinated proliferating cell nuclear antigen (PCNA), which is thought to act at stalled replication forks in promoting template switching and fork restart

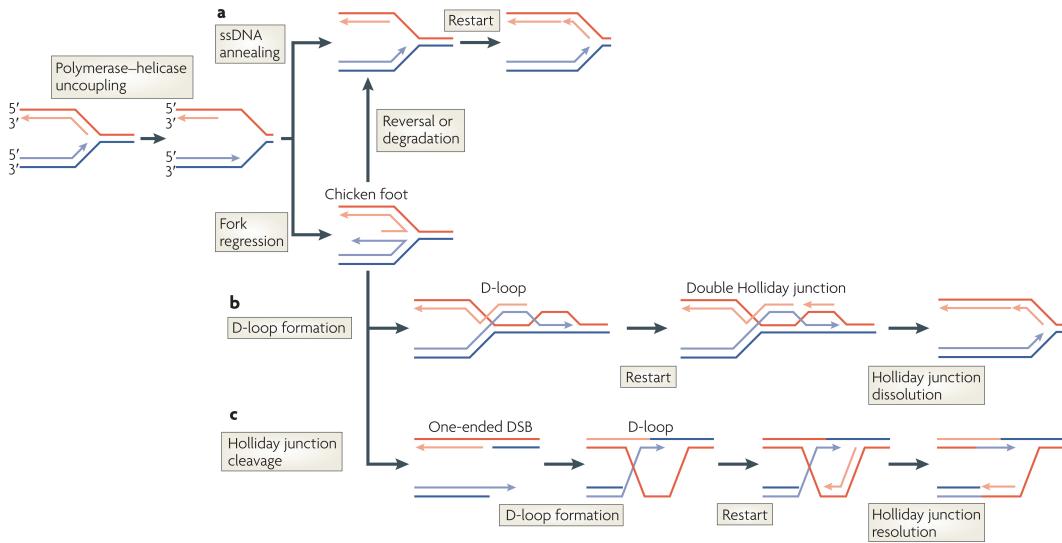


Figure 1.5: Stalled replication fork resolution paths. Following polymerase and helicase uncoupling, there are multiple ways to resolving a stalled replication fork. (a) ssDNA annealing can occur to regenerate a normal replication fork, further fork regression activity may produce a chicken foot structure, which can later be reversed or restored to a normal replication fork and lead to fork restart. (b) A chicken foot structure can initiate strand invasion forming a displacement loop (D-loop), which leads to fork restart and further Holliday junction resolution. (c) Unresolved chicken foot structures can lead to cleavage and DSB mediated repair, which can lead to fork restart, although at the cost of possibly producing SCEs²⁹. Adapted from Petermann, E., and Helleday, T. (2010) Pathways of mammalian replication fork restart. *Nat. Rev. Mol. Cell Biol.* 11, 683–687.

pathways³⁴. In agreement with this hypothesis, ZRANB3 depletion showed there was an increase in SCEs³⁴. ZRANB3 has not been associated with any disease and its role does not seem to be redundant to SMARCAL1¹⁷. Instead, they may act upon different stalled fork substrates²⁴.

1.5 Schimke Immunoosseous Displasia

A new disease was distinguished in 1971 based on a single case, and described as a disease of nephrotic syndrome and defective cellular immunity³⁵. It was later named Schimke immunoosseous displasia (SIOD) after the author of the original report, and it was found to be an autosomal recessive condition³⁶. SIOD symptoms, such as growth retardation can be detected even *in utero*, leading to death, within years of age³⁷. A milder form of the disease displays symptom onset in late childhood, and survival into early adulthood³⁸. Disease management includes kidney and bone marrow transplants, which rescues patients from renal disease and immunodeficiency respectively, although patients succumb to the overall progression of the disease, due to symptoms affecting the central nervous system^{37,39}.

The predisposition for cancer in SIOD patients has been analyzed in a cohort of 71 patients, four of which developed cancer⁴⁰. One osteosarcoma, and three B-cell non-Hodgekin lymphomas, although two were Epstein-Barr virus positive, thus attributable to immunodeficiency and opportunistic infection. Since then an additional case of a nasal high-grade, undifferentiated carcinoma has been noted in a SIOD patient⁴¹. Both reports note that there may not be a clear predisposition to cancer, due to the short life expectancy of patients, where cancer may not have had sufficient time to develop^{40,41}.

1.5.1 SIOD Associated SMARCAL1 Mutations

Through genetic screening of patients and parents, the homozygous loss of function of the SMARCAL1 gene was found to be responsible for SIOD³⁸. A summary of over 40 known SMARCAL1 mutations in SIOD patients was compiled in 2007⁴², and since then new mutations have been reported^{41,43} (Figure 1.6). The SMARCAL1 gene may be disrupted through missense, nonsense, insertion or deletion, and splice site mutations (Figure 1.6).

Missense mutations may lead to milder forms of disease^{42,44}, because they seem to diminish SMARCAL1 enzymatic activity, whereas other mutations generally suppress normal mRNA levels⁴⁴. Yet, the analysis of SMARCAL1 mutations, and correlation to disease shows that there is no clear genotype to phenotype association. One reason is that environmental factors may have an effect on disease penetrance⁴⁵. Experiments on SMARCAL1 knockout flies and mice showed there was normal growth and development, however when the animals were put under heat stress, the SMARCAL1 deficient groups died earlier and in greater numbers⁴⁵.

In SIOD patients, many point mutations are within the helicase domains (Figure 1.6). ATPase activity studies have been performed *in vitro* for some of these mutants, and interestingly it was found that some mutations decreased, and others increased activity, but for the higher activity mutants, the mRNA and protein levels within cells were lower, thus the higher activity may not be a physiologically relevant.

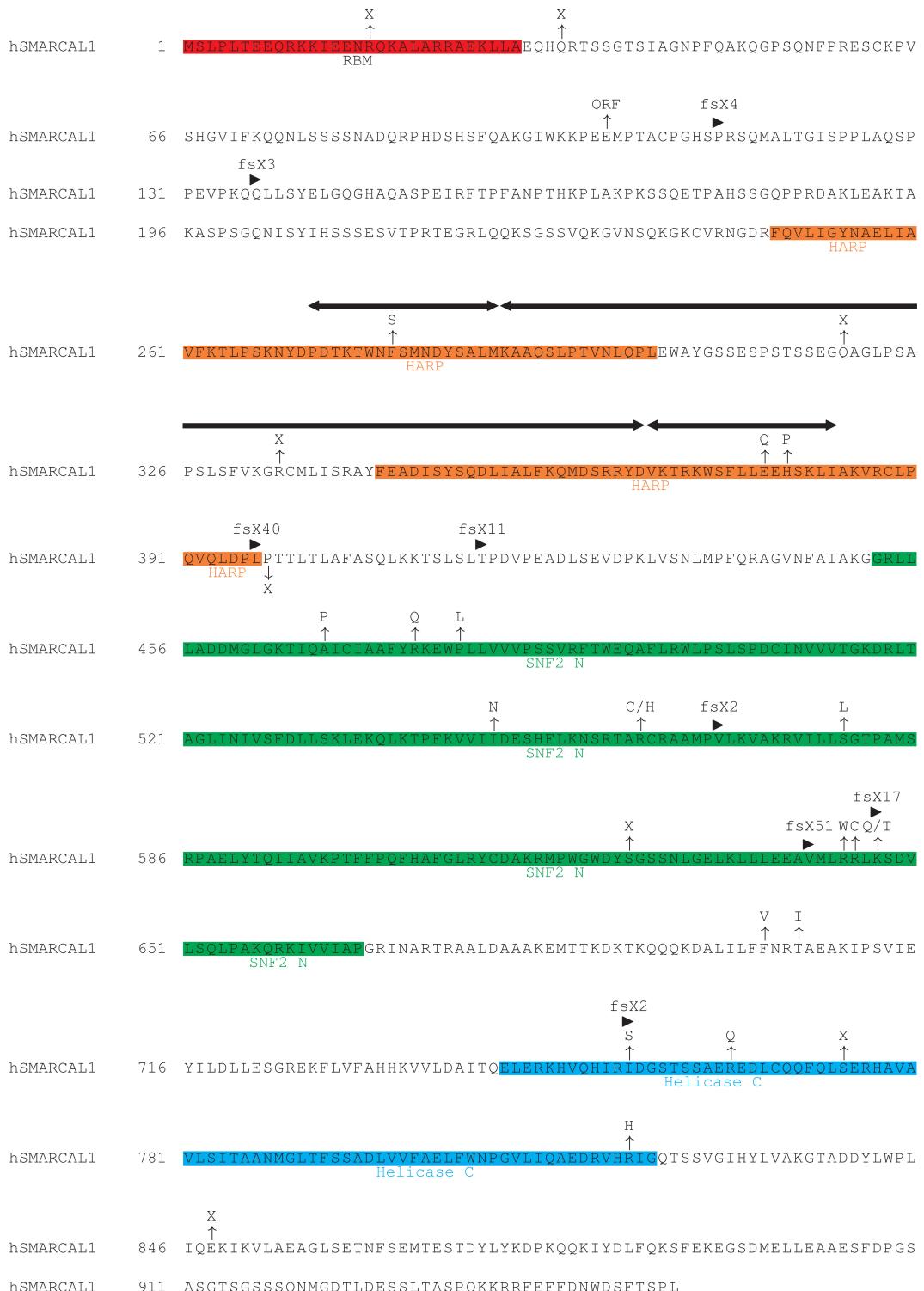


Figure 1.6: SIOD mutations described in literature⁴¹⁻⁴³ mapped to SMARCAL1. (↑) point mutation or early termination codon. (►) start of a frame shift mutation. (◀-►) Span of internal deletion mutants.

1.6 SMARCAL1

1.6.1 History

In 1986 a paper describing a calf thymus extract purification method of what they named DNA-dependent ATPase A was published. These authors noted that it had no deoxyribonuclease, DNA or RNA polymerizing, topoisomerase, DNA ligase, DNA unwinding, terminal deoxynucleotidyl transferase, polynucleotide kinase, autophosphorylating, DNA reanaturing, or strand exchange activities. However, they found it had ATPase activity when in presence of DNA fragments containing single stranded and double stranded regions such as recessed DNA ends, DNA hairpins and primer-template junctions. They developed antibodies⁴⁶ and with them, improved yields and purity by performing immunoaffinity purification⁴⁷. The sequence of DNA-dependent ATPase A was determined in 2000 by Edman degradation peptide sequencing, and they found it was part of the SWI2/SNF2 family of proteins, with orthologs in humans, mouse and *Caenorhabditis elegans*⁴⁸. This protein was also named HARP/SMARCAL1, based on the initial SWI/SNF human genome map locus designation^{49,50}. The HARP naming is rooted in HepA (helicaselike protein)⁵¹, which is now renamed RapA (RNA polymerase-associated protein)⁵² it is one of two prokaryotic protein in the SWI2/SNF2 family, and it is part of the same distant group as SMARCAL1 (Section 1.1)¹⁵.

1.6.2 Biological Function of SMARCAL1

During DNA replication helicase and polymerase uncoupling can occur when DNA damage is encountered⁵³. This generates long stretches of ssDNA coated by replication protein A (Replication Protein A), which in turn activate the ATR dependent checkpoint activation to prevent further replication origin firing, prevent entry into mitosis, and to activate repair proteins⁵⁴.

Following fork stalling, evidence shows that SMARCAL1 is involved in fork remodeling pathways that lead to fork restart. SMARCAL1 fork remodeling activity includes ssDNA annealing, fork regression and restoration (Figure 1.5)⁵⁵. When SMARCAL1 is depleted or SMARCAL1 mutations occur, there is an increase in the presence of ssDNA and DNA cleavage, which generates DSBs⁵⁶. DSBs can be repaired, although it may induce SCEs, where these lead to an increase of genomic instability⁵⁷.

SMARCAL1 orthologs have been identified in 41 organisms among: protists, plants, invertebrates, vertebrates, mammals and one fungus¹⁵. In mammals the HARP domain comes as a dual repeat, yet in other species SMARCAL1 proteins contain only one HARP domain¹⁹. Deletion of both HARP domains has shown they are necessary for *in vitro* annealing helicase activity and *in vivo* rescue of cells from γ -H2AX foci formation due to activation of the double strand break repair pathway^{19,58}. Alternate deletion mutants of the HARP domains within the human SMARCAL1 protein have shown that the first HARP domain is dispensable, but the second HARP domain is necessary for conservation of annealing helicase activity⁵⁹. It is yet unclear if the HARP domains are necessary for DNA binding. There are two cases where HARP repeats are deleted; one shows DNA binding and ATPase activity¹⁹, whereas the other cannot replicate these results⁵⁹.

Within SMARCAL1 the helicase domain is a molecular motor which uses energy from ATP to exert fork remodeling activity. Studies of the bovine SMARCAL1 ortholog, based on fluorescence quenching experiments to observe conformation changes, have proposed a mechanism of ATP hydrolysis. ATP binding first and DNA second creates an unproductive complex, whereas binding DNA first and ATP second forms a productive complex⁶⁰. They suggest that there may be a conformation change required for the unproductive complex to be converted to the productive one⁶⁰. These authors also determined there was binding of stem-loop

DNA (sIDNA), dsDNA, and ssDNA, which led them to suggest that the DNA binding pocket is composed of two subsites, one can bind the ssDNA and the other the dsDNA⁶⁰. Only when both sites are filled, such as when sIDNA or fork binding occurs, the conformation required for ATP hydrolysis is attained⁶⁰.

Further studies of the bovine SMARCAL1 ortholog have shown there was binding of DNA and ATP to a truncation mutant containing the HARP repeats and the SNF_N domain, but not the Helicase_C domain⁶¹. They have shown magnesium ions are not required for ATP and DNA binding, but have to be present for hydrolysis⁶¹. They also determined that adenine and ADP can bind to SMARCAL1, but the γ -phosphate of ATP is required to elicit an active conformation⁶¹.

1.6.3 SMARCAL1 and RPA interactions

RPA is a trimeric complex of 70 kDa, 32 kDa, and 14 kDa subunits, which are referred to as RPA70, RPA32, and RPA14, or often RPA1, RPA2, and RPA3 respectively. RPA can bind to ssDNA in a multistep mechanism of at least two modes, a weak 8-10 nucleotide binding, and a strong 30 nucleotide binding mode⁶². RPA is thus involved in DNA replication, DNA repair and telomere maintenance pathways⁶². RPA is central for DNA repair and its presence is required for nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR), and it plays a role in double strand breaks (DSBs)⁶². RPA has two known protein interaction sites, the A domain of RPA70 and the C domain of RPA32⁶³. Proteins can bind to either one alone, or be able to interact with both domains⁶³.

SMARCAL1 has been identified as a RPA binding protein^{56,64-66}. Experiments show that it can bind to RPA32 C domain, through a conserved motif located in the first 30 amino acids of SMARCAL1^{56,64}. Sequence alignment shows this is a RPA binding motif (RBM) also found in TIPIN, XPA, UNG2 and RAD52 other RPA32 interacting proteins^{56,64}. One account has shown that SMARCAL1 can also bind to

the RPA70 A domain⁶⁶. The interaction of SMARCAL1 and RPA is essential to the localization of SMARCAL1 to stalled replication forks, and *in vitro* experiments show that truncation of the RBM prevents the colocalization of SMARCAL1 to RPA foci^{56,64–66}.

1.6.4 SMARCAL1 Phosphorylation

SMARCAL1 is phosphorylated after hydroxyurea (HU) treatments^{56,66,67}. ATR, ATP and DNA-PK were able to phosphorylate SMARCAL1 *in vitro*⁵⁶, and only when all three are inactivated *in vivo* the phosphorylation is lost^{56,66}. Similarly caffeine treatment of cells, which inhibits ATM, ATR and many other kinases, was seen to prevent SMARCAL1 phosphorylation⁶⁷.

There are many potential consensus sites within SMARCAL1 that can be phosphorylated, and through mass spectroscopy three have been determined: S173, S652, and S919^{55,56}. Phosphorylation may have a regulatory role on SMARCAL1 activity at stalled replication forks, it was found that the S652 is a conserved residue within the linker region of the helicase domains, and that S652 phosphorylation lowered ATPase activity⁵⁵. Overexpression and phosphomimetic S652D SMARCAL1 show disregulation of fork processing, supporting the role of regulation.

1.6.5 SMARCAL1 Fork Remodeling Model

Early on it was shown that SMARCAL1 prefers structures presenting both ss-DNA and dsDNA regions, such as: DNA with regressed ends, DNA hairpins, and replication forks^{48,68,69}. Additionally its annealing helicase activity, together with observations of interaction with RPA and reduction of ssDNA in cells placed this protein at replication forks to stabilize and anneal these structures^{31,56,64–66,69}. Recent research has shown SMARCAL1 can also display fork regression⁵⁹, and determined there is stimulation of RPA bound leading gaps, whereas inhibition of RPA

bound lagging gaps²⁴. Single molecule studies have shown that SMARCAL1 can also catalyze branch migration and fork restoration activity on Holliday junctions²⁴.

This evidence can be put together into a model, where SMARCAL1 is not recruited in the presence of normal replication and lagging gaps from okazaki fragments. However, when a DNA lesion on a leading stand is encountered that can decouple the helicase and polymerase, it generates a leading strand gap of RPA bound ssDNA, which acts to recruit SMARCAL1. SMARCAL1 can catalyze ss-DNA annealing, fork regression and branch migration, to produce a chicken foot structure (Figure 1.7). Next the lesion is repaired, and SMARCAL1 can catalyze fork restoration, which leads to replication restart.

1.7 Thesis Objective

To understand the molecular mechanism of a protein, it is important to assess the protein structure, using techniques such as X-ray crystallography. A difficult stage of structural analysis is protein crystallization; it is a sensitive technique which requires high purity and quantity of the protein under study. This thesis investigates the heterologous expression of *R. norvegicus* and *C. elegans* SMARCAL1 orthologs in a bacterial expression system. Chapter 3 establishes an efficient purification protocol for the mammalian *R. norvegicus* SMARCAL1. Further, the approach of truncation was utilized in the search of a stable, high yield and high purity protein. Chapter 4 details the characterization of the purified proteins in order to confirm function, including the comparison to a human SMARCAL1 purified from insect cells. Structural analysis was pursued through crystallization trials, which are described in Chapter 5. Conclusions and final remarks are made in chapter 6.

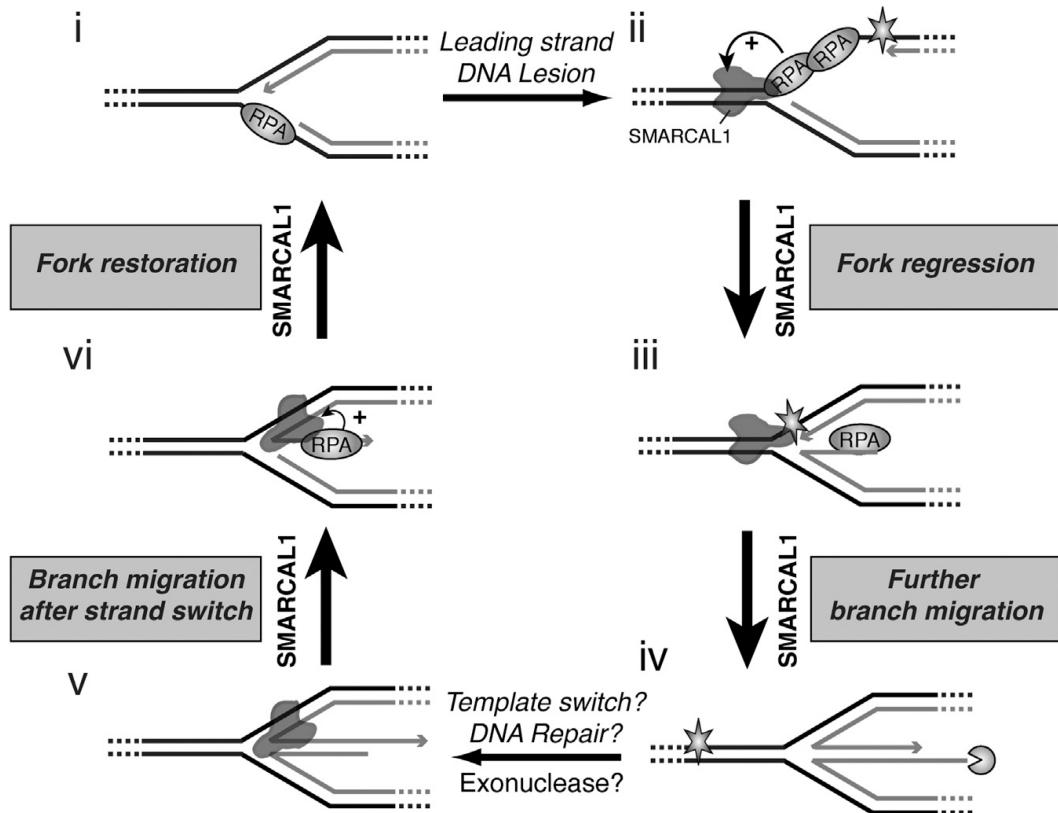


Figure 1.7: SMARCAL1 fork regression model. SMARCAL1 can selectively bind to leading strand gaps, and catalyze fork regression and branch migration past a DNA lesion, which can be later repaired. Further, SMARCAL1 can catalyze fork restorations to regenerate a normal fork that can restart replication.²⁴. Adapted from Btous, R., Couch, F. B., Mason, A. C., Eichman, B. F., Manosas, M., and Cortez, D. (2013) Substrate-selective repair and restart of replication forks by DNA translocases. *Cell Rep* 3, 1958–1969.

Chapter 2

Materials and Methods

2.1 Materials

Kits

QIAprep Spin Miniprep Kit (Qiagen)
QIAquick PCR Purification Kit (Qiagen)
QIAquick Nucleotide Removal Kit (Qiagen)
MiniPrep Plasmid Purification Kit (Truin Science)
Pierce Bicinchoninic Acid Protein Assay Kit (Thermo Scientific)

Enzymes

Restriction enzymes BamHI and NotI (Invitrogen)
Platinum Pfx DNA Polymerase (Invitrogen)
Phusion High Fidelity DNA Polymerase (Thermo Scientific)

Medium

Luria-Bertani (LB) broth Miller (Fisher Scientific)

Difco LB Agar (Fisher Scientific)

Antibiotics

Ampicillin Sodium Salt (Invitrogen)

Kanamycin Sulphate (Invitrogen)

ATP/NADH assay components

Pyruvate Kinase/Lactic Dehydrogenase (PK/LDH) enzymes from rabbit muscle

Phospho(enol)pyruvic acid monopotassium salt (PEP)

β -Nicotinamide adenine dinucleotide (NADH) reduced dipotassium salt

Adenosine 5'-triphosphate (ATP) disodium salt hydrate

All components were purchased from Sigma-Aldrich

Crystallography Suites

Classics I & II, JCSG+, MPD, PEG I & II, ProComplex, Nucleix (Qiagen)

Protein-Nucleic Acid Complex Crystal Screen (KeraFast)

Wizard I & II (Emerald Biosystems)

Solutions

Lysis Buffer	20 mM HEPES pH 7.3, 400 mM NaCl, 10% glycerol, 1 mM EDTA, 14 mM BME
Extraction Buffer	Lysis Buffer, 0.33X Halt Protease Inhibitor Cocktail (Thermo Scientific), .015g/L lysozyme (Sigma))
Buffer A	20 mM HEPES pH 7.3, 50 mM NaCl, 14 mM BME
Buffer B	20 mM HEPES pH 7.3, 1 M NaCl, 14 mM BME
Storage Buffer	20 mM HEPES pH 7.3, 300 mM NaCl, 10% glycerol, 14 mM BME
Crystallography Buffer	20 mM HEPES pH 7.3, 300 mM NaCl, 14 mM BME
GST Elution Buffer	20 mM HEPES pH 7.3, 400 mM NaCl, 20 mM L-Glutathione reduced, 1 mM EDTA, 14 mM BME
Binding Buffer (EMSA) 5X	100 mM HEPES, 500 mM NaCl, 25 mM MgCl ₂ , 15% glycerol, 1.25 mg/mL BSA, 0.25 mM EDTA, 25 mM BME
Binding Buffer (ATPase) 5X	100 mM HEPES, 500 mM NaCl, 25 mM MgCl ₂ , 1.25 mg/mL BSA, 0.25 mM EDTA, 25 mM BME

Cloning Primers

rSMARCAL1 _{WT} - FWD	GATCTGGAAGTTCTGTTCCAGGGGCC TGATGTCCTGCCACTTACGGAGG
rSMARCAL1 _{WT} - RV	TCAGTCACGATGCGGCCGCTCGAGTCGA CCTCAAAAGGGAGAGGTAAAGCTGTC
rSMARCAL1 _{Δ210} - FWD ¹	CTAGTCGGATCCGGGGATCGCTCCGGG TAAAGATCGGCTAC
rSMARCAL1 _{Δ210} - RV	CGATGCGGCCGCTCAAAAGGGAGAGGT AAAGCTGTCCAATT
rSMARCAL1 ₂₁₀₋₈₅₀ - RV ²	CGATGCGGCCGCTTACCCATCGTCTCG AAGGACTGCTGGAACAG
rSMARCAL1 ₂₁₀₋₈₃₁ - RV	CGATGCGGCCGCTTACTTGTAGAGGTAG TCAGTAGCTTCTGTCAT
rSMARCAL1 ₂₁₀₋₈₁₇ - RV	CGATGCGGCCGCTTAGGTCTCAGAAAGC CCAGCTTCCCCCAGAAC
rSMARCAL1 ₂₈₇₋₈₅₀ - FWD	CTAGTCGGATCCTTGTACAGGGAGAT GCATGCTCATTTCC
cSMARCAL1 _{WT} - FWD	CTAGCCCCGGGTATGGTGCTCACTGAT GAGCAACGACAGGCG
cSMARCAL1 _{WT} - RV ³	CGATGCGGCCGCTCAATTGGATACGT TTAGGAGCTGGAGA
cSMARCAL1 _{Δ109} - FWD	CTAGCCCCGGGTGGTATAGAATTAAG ATTGAATTCTATCCA

¹Same forward primers used for rSMARCAL1₂₁₀₋₈₅₀, rSMARCAL1₂₁₀₋₈₃₁, and rSMARCAL1₂₁₀₋₈₁₇.

²Same reverse primer used for rSMARCAL1₂₈₇₋₈₅₀.

³Same reverse primer used for cSMARCAL1_{Δ109}.

Assay Primers

21ds9ss_ top	CCAGTGAATTGTTGCTCGGTACCTGCTAAC
30ds_ DS	GTTAGCAGGTACCGAGCAACAATTCACTGG
21ds9ss_ fork	GACATTGATACCGAGCAACAATTCACTGG
ssDNA	CTCTCTCTCTCTCTCTCTCTCTCTCTCT
16ds9ss_ top	GAATTGTTGCTCGGTACCTGCTAAC
16ds9ss_ bottom	GACATTGATACCGAGCAACAATTTC
21ds5' _ bottom	TACCGAGCAACAATTCACTGG
16ds4ss_ top	GAATTGTTGCTCGGTACCTG
16ds4ss_ bottom	TTGATACCGAGCAACAATTTC
21ds4ss_ top	CCAGTGAATTGTTGCTCGGTACCTG
21ds4ss_ bottom	TTGATACCGAGCAACAATTCACTGG

Annealed DNA Substrates

dsDNA	21ds9ss_ top + 30ds_ DS
fork DNA	21ds9ss_ top + 21ds9ss_ fork
ssDNA	ssDNA
16ds4ss	16ds4ss_ top + 16ds4ss_ bottom
16ds9ss	16ds9ss_ top + 16ds9ss_ bottom
21ds4ss	21ds4ss_ top + 21ds4ss_ bottom
21ds9ss5'	21ds9ss_ top + 21ds5' _ bottom

2.2 Methods

2.2.1 Cloning

All six *R. norvegicus* and both *C. elegans* proteins were cloned into a pGEX-6P-1 (GE Healthcare) plasmid to form GST-fusion proteins. The first rat protein

cloned was rSMARCAL1 Δ 210, where we saw proteolytic degradation. The stable lower molecular weight fragments were analyzed by mass spectrometry, and defined stable regions that could be further cloned. The next rat truncation mutants generated were based on knowledge from stable degradation products, sequence alignments, and domain boundaries. All proteins except rSMARCAL1_{WT} were cloned by use of restriction enzymes between the BamHI and NotI sites of the plasmid. rSMARCAL1_{WT} GST-fusion protein was generated by overlap extension cloning, a restriction free method⁷⁰. Plasmids products were transformed into DH5 α *Escherichia coli* cells (Invitrogen). Plasmids were then isolated and transformed into BL21-Gold *E. coli* cells for protein expression. Cloning was verified for all rat constructs by sequencing.

2.2.2 Expression and Purification

Expression

BL21-Gold clones are ampicillin and kanamycin resistant, thus all culture media contained 100 mg/L amp and 50 mg/L kanamycin. Starter cultures were inoculated with clones and grown overnight at 37°C with shaking. 2 or 8 L of secondary cultures were prepared by adding starter culture to LB medium, and cells were grown at 25°C with shaking. Cells were grown to an A600 of 0.6 and induced with 0.05 mM IPTG at 18°C and grown overnight.

rSMARCAL1 Assay Purification

Cells were harvested by centrifugation for 12 min at 7,500 x g, and resuspended in extraction buffer. Cells were lysed by an Emulsiflex-C3 homogenizer (Avestin), then centrifuged for 45 min at 23,000 x g. Supernatants containing the soluble GST fusion proteins were mixed with glutathione agarose (Pierce) resin and incu-

bated 1 h 30 min on a nutator at 4 °C, and then loaded on to a 50 mL purification column and washed with lysis buffer. An HRV3C protease GST fusion protein was added and left overnight to cleave the GST-SMARCAL1 fusion proteins.

The free SMARCAL1 proteins were eluted with lysis buffer. The eluted fractions and a post-elution sample of beads were analyzed by SDS-PAGE to validate cleavage was completed. Fractions containing protein were concentrated and further purified on a Superdex 200 (GE Healthcare) column in storage buffer (20 mM HEPES pH 7.3, 300 mM NaCl, 10% glycerol, 14 mM BME). Purity was analyzed by SDS-PAGE, and fractions were concentrated to 0.5 - 1 mg/mL. Protein quantification was done using a Pierce BCA Assay Kit (Thermo Scientific)

rSMARCAL1 Crystallography Purification

Expression and purification protocol up to GST-elution is the same, however, the buffers were glycerol free and prior to size exclusion chromatography we performed additional cation exchange chromatography. GST-elution fractions were concentrated and diluted to 50 mM NaCl. Protein was loaded on to a SP Sepharose column (GE Healthcare) and washed with Buffer A. The bound protein was eluted using a NaCl gradient of buffer B, and the peak fractions were analyzed by SDS-PAGE. The fractions containing SMARCAL1 proteins were concentrated and further purified on a Superdex 200. Peak fractions were analyzed by SDS-PAGE and selected fractions concentrated and final products quantified and again analyzed by SDS-PAGE to determine final purity.

cSMARCAL1 Purification

Expression protocol and purification protocol up to GST-elution remained the same, however, the purification was different. Buffers used were the rSMARCAL1 crystallography glycerol free buffers. GST-bound cSMARCAL1 proteins were

eluted with GST elution buffer that contains reduced glutathione. GST-fusion tag was cleaved in the eluted fraction. cSMARCAL1_{WT} were concentrated and purified by size exclusion chromatography. Whereas cSMARCAL1_{Δ109} fractions were concentrated, diluted to 50 mM NaCl and purified by cation exchange chromatography and further size exclusion chromatography. Peaks were analyzed by SDS-PAGE.

hSMARCAL1

Human hSMARCAL1_{WT} protein purified from insect cells was a gift from Dr. Cortez.

2.2.3 DNA Annealing

DNA annealing was performed by a thermocycler. The program used for annealing heated the two DNA primers to 95°C for 5 min and cooled it to 25°C over a length of 25 min to generate the final annealed DNA substrates.

2.2.4 Electrophoretic mobility shift assay

Using radioactive γ -P32 ATP (PerkinElmer) we labeled the 21ds9ss_ top and ss-DNA primers. DNA annealing was then performed to form radiolabelled fork DNA, and dsDNA respectively. Purified SMARCAL1 proteins were incubated for 1 h at 4°C in a reaction volume of 10 μ L where the final concentrations were 1 nM radiolabelled oligo probe, 20 mM HEPES, 100 mM NaCl, 5 mM MgCl₂, 5% glycerol, 0.25 mg/mL BSA, 0.05 mM EDTA, 7 mM BME and with varying SMARCAL1 concentrations (0-250 nM). Samples were then run on non-denaturing 7% polyacrylamide gels at 50 V for 3 h at 4°C. Gels were exposed to a phosphor screen, which was scanned using a Typhoon PhosphorImager (GE Healthcare).

2.2.5 ATPase assay

The ATPase assay was performed under steady-state kinetic conditions by using an ATP/NADH coupled assay⁷¹, where ATP is regenerated through pyruvate kinase by the conversion of phosphoenol pyruvate into pyruvate, which in turn is converted by lactose dehydrogenase into lactate through NADH oxidation into NAD+. NADH fluorescence was measured by excitation at 355 nm and emission at 460 nm over a time of 10 to 15 min, where readings were taken each 13 s to generate progression curves. The components were combined in a black 384 well plate (PerkinElmer) at the assay start point in a final volume of 40 µL per well. The final concentrations of the assay was 1 mM PEP, 100 units/mL PK/LDH, .2 mM NADH, 20 mM HEPES, 100 mM NaCl, 5 mM MgCl², 5% glycerol, 0.25 mg/mL BSA, 0.05 mM EDTA, 7 mM BME, 5-10 nM SMARCAL1, 2 mM ATP, 0-10,000 nM oligo. Each assay was performed in triplicate and was comprised of a serial dilution (2.15) of DNA substrates in 11 concentrations and a no substrate control. We used a NADH standard curve prepared to convert the relative fluorescence units (RFU) given by the FLUOstar OPTIMA (BMG Labtech) plate reader into NADH concentrations. Initial rates were found by linear regression and plotted against substrate concentration using the Michaelis-Menten fit. All analysis was performed in GraphPad Prism.

2.2.6 Crystallization trials

Purified proteins for crystallization trials were at final concentrations between 6 and 2.5 mg/mL. The protein was set as sitting drop in INTELLI-PLATE 96-3 (Hampton Research) trays. Each reservoir contained 100 µL of crystallization conditions from screens. Each drop was made by depositing 1 µL of protein and mixing with 1 µL of crystallization conditions from the reservoir. Drops were typically observed after 1 day, 3 days, 1 week, 2 weeks 1 month, 2 months of growth.

Chapter 3

SMARCAL1 Expression and Purification

Contributions: Curtis Hodges has performed the expression and purification trials of human SMARCAL1. Curtis Hodges has completed the cloning of four *Rattus norvegicus* truncation mutants: rSMARCAL1 Δ 210, rSMARCAL1₂₁₀₋₈₅₀, rSMARCAL1₂₁₀₋₈₃₁, and rSMARCAL1₂₁₀₋₈₁₇. He has also established an expression and purification protocol for rSMARCAL1 Δ 210.

3.1 Introduction

There is an interest in the SMARCAL1 protein due to its involvement in pathways of stalled replication fork restart, and its mutation is a cause of SIOD. There are three domains: RBM, HARP, and SWI2/SNF2 helicase. The functions of these domains are described in the introductory chapter. It has been noted that for *in vitro* DNA binding, DNA stimulated ATPase and annealing activity the RBM domain is not necessary. Similar studies show the first HARP domain is also dispensable. SMARCAL1 orthologs in certain invertebrates contain only a single HARP.

Expression and purification in high yields and purity is a requirement for structural studies. A simple and cost effective method is heterologous protein expression in a bacterial host. However, biochemical studies of human SMARCAL1 are performed from insect or mammalian expression systems, which are more costly and time consuming. A human SMARCAL1 protein was previously tested in our lab, but failed to express in *E. coli*. However, to produce a SMARCAL1 protein we turned to orthologs from *R. norvegicus* and *C. elegans*. In order to optimize purity, yield and stability of the protein, we attempted purification of truncation mutants.

3.2 Results

3.2.1 SMARCAL1 orthologs used in our studies

To define regions of interest and conservation between human, *R. norvegicus* and *C. elegans* SMARCAL1 we performed a sequence alignment (Figure 3.1). Sequence identity of human SMARCAL1 to its rat and roundworm orthologs is 78% and 34% respectively. The *C. elegans* SMARCAL1 protein only contains one HARP domain. Sequence conservation is greatest within the two helicase domains, and most notably across all the SWI2/SNF2 conserved motifs. The RBM and HARP domains show good sequence conservation. Regions of poor conservation are situated between the RBM and HARP, between the two HARP domains, and finally at the N terminus.

Alignment information was used to generate six rat proteins and two *C. elegans* proteins (Figure 3.2). All the proteins were produced as glutathione S-transferase (GST) fusion proteins, where the GST tag was later cleaved during the purification process. The six *R. norvegicus* proteins we cloned are: full length rSMARCAL1_{WT} protein; rSMARCAL1_{Δ210}, in which the N-terminal RBM and predicted unstructured region was deleted; rSMARCAL1₂₁₀₋₈₅₀ in which a poorly conserved C-terminal region was also removed; rSMARCAL1₂₁₀₋₈₃₁ and rSMARCAL1₂₁₀₋₈₁₇ in which the C-terminal region is further truncated; and rSMARCAL1₂₈₇₋₈₅₀, in which the N-terminal HARP domain and a portion of the C-terminal tail was also deleted. Two *C. elegans* proteins are cloned, a full length cSMARCAL1_{WT} and cSMARCAL1_{Δ109} in which the RBM and the N-terminal unstructured region is deleted.

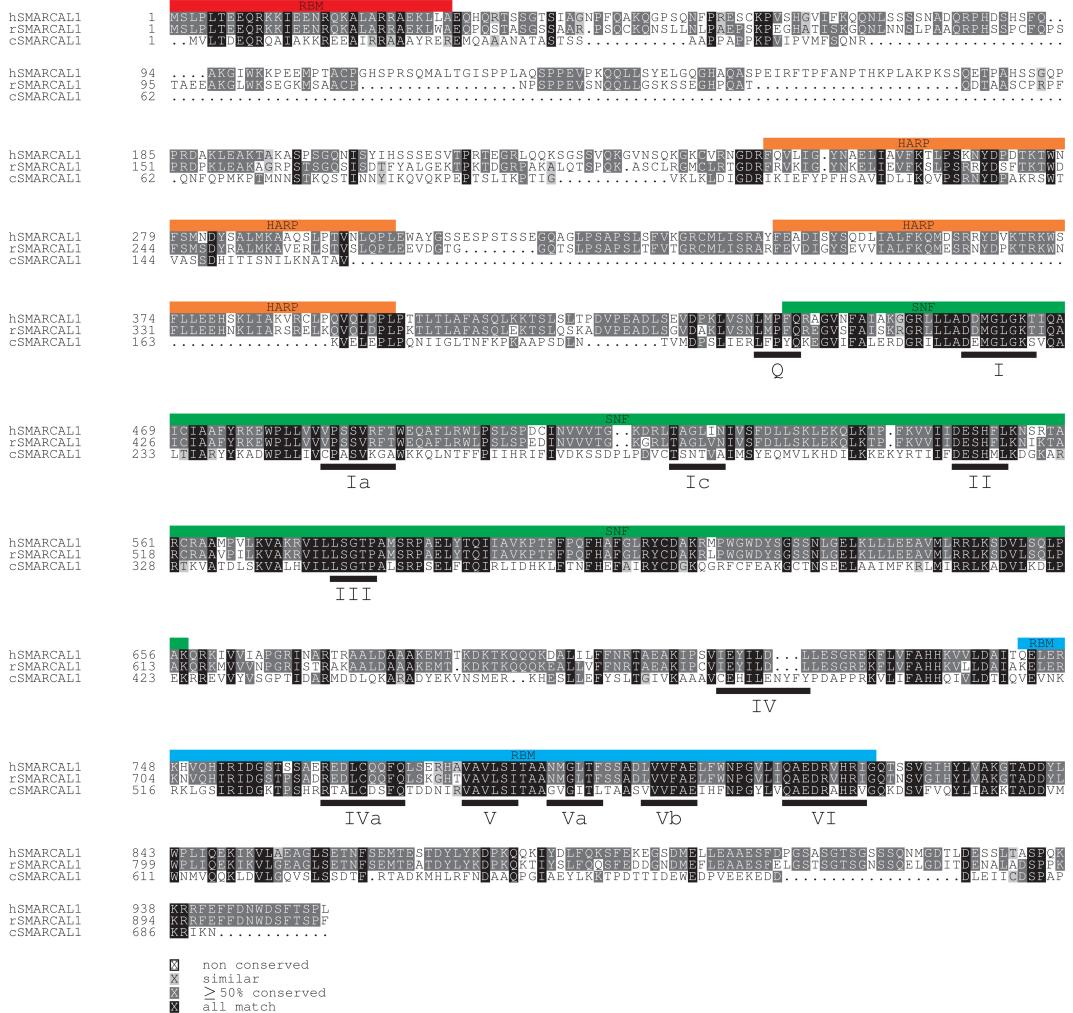


Figure 3.1: Sequence alignment of human, rat and roundworm SMARCAL1 proteins. Above the sequence the RBM, HARP, SNF2_N, and Helicase_C domains are annotated. Below the sequence SWI2/SNF2 conserved helicase motifs are defined. Amino acid conservation is shaded in gray as per the legend.

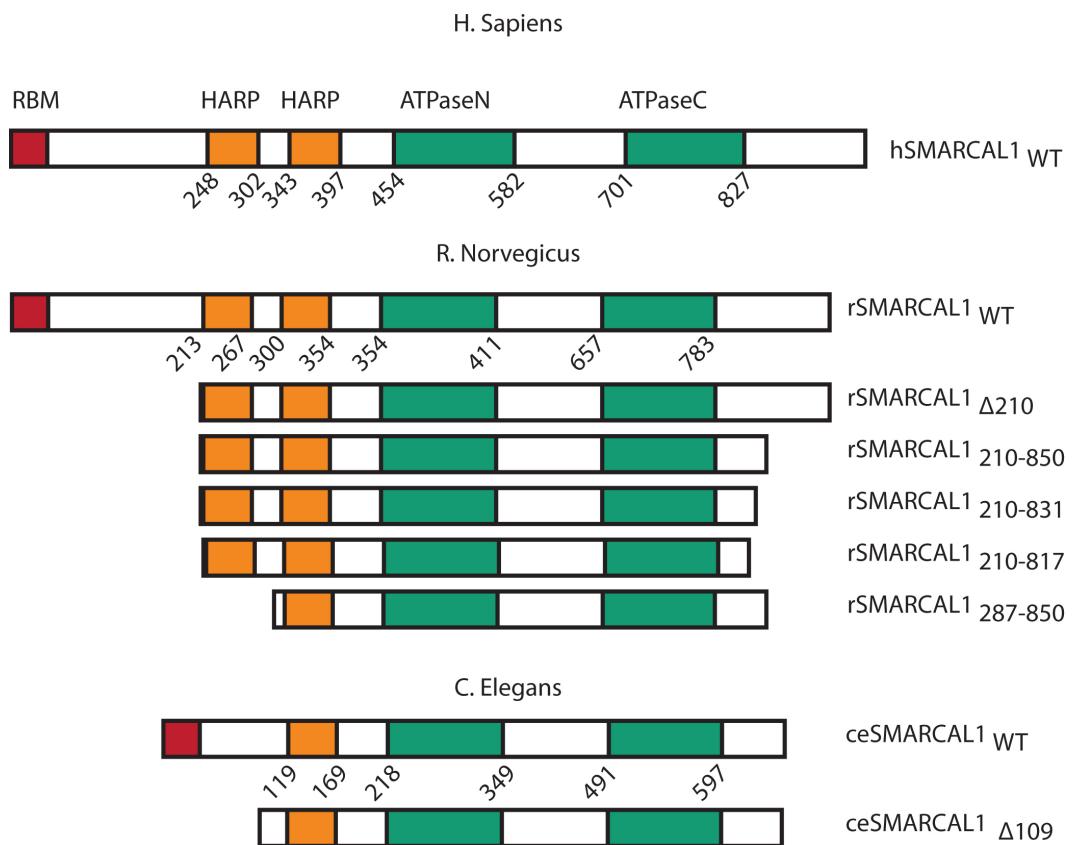


Figure 3.2: Diagram of proteins expressed with conserved domains illustrated.

3.2.2 SMARCAL1 Heterologous Expression and Purification

All of the SMARCAL1 proteins cloned were used in a bacterial expression system. We purified proteins in two methods, one for the use of proteins for characterization assays (Chapter 4), and another for the use in crystallography trials (Chapter 5).

One of the final steps of purification is size exclusion chromatography and separation from contaminants such as chaperones can be seen (Figure 3.3). Chaperones elute at a retention volume between 50 mL and 70 mL. Protein purity was assessed by SDS-PAGE over the elution peaks (Figure 3.4 A). Selected fractions were concentrated and the final purity of the proteins that were used for assays was determined (Figure 3.4 B). Proteins were produced in yields of: 0.73 mg/L for rSMARCAL1_{WT}; 0.21 mg/L for rSMARCAL1_{Δ210}; 0.72 mg/L for rSMARCAL1₂₁₀₋₈₅₀; 0.70 mg/L for rSMARCAL1₂₁₀₋₈₃₁; 0.73 mg/L for rSMARCAL1₂₁₀₋₈₁₇; and 0.94 mg/L for rSMARCAL1₂₁₀₋₈₅₀. There is an almost five fold increase in yield from rSMARCAL1_{Δ210} to rSMARCAL1₂₈₇₋₈₅₀.

Three proteins were purified for crystallization trials: rSMARCAL1_{Δ210}, rSMARCAL1₂₁₀₋₈₁₇ and rSMARCAL1₂₈₇₋₈₅₀. The distinctions being that proteins destined for crystallography were prepared in glycerol free buffers and had an additional ion exchange chromatography step. These proteins were produced in similar yields, although, they had reduced stability when stored at 4°C. Precipitation formed overnight for rSMARCAL1₂₁₀₋₈₁₇ and in a matter of weeks for rSMARCAL1_{Δ210} and rSMARCAL1₂₁₀₋₈₁₇.

The two *C. elegans* proteins were expressed and GST affinity purified (Figure 3.5). Eluted fractions showed poor cSMARCAL1 yields, but high amounts of chaperone copurification. Following GST cleavage and size exclusion chromatography no protein could be recovered.

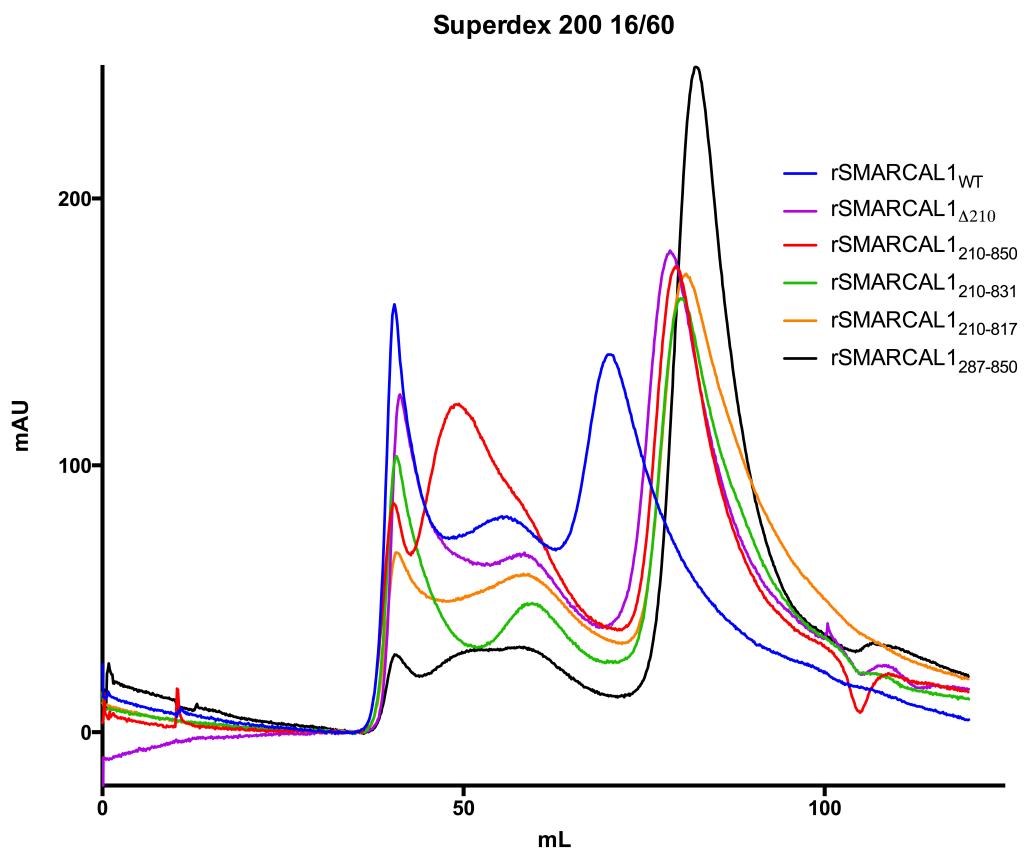


Figure 3.3: Retention volume of rSMARCAL1 proteins on an Superdex 200 16/60 chromatography column.

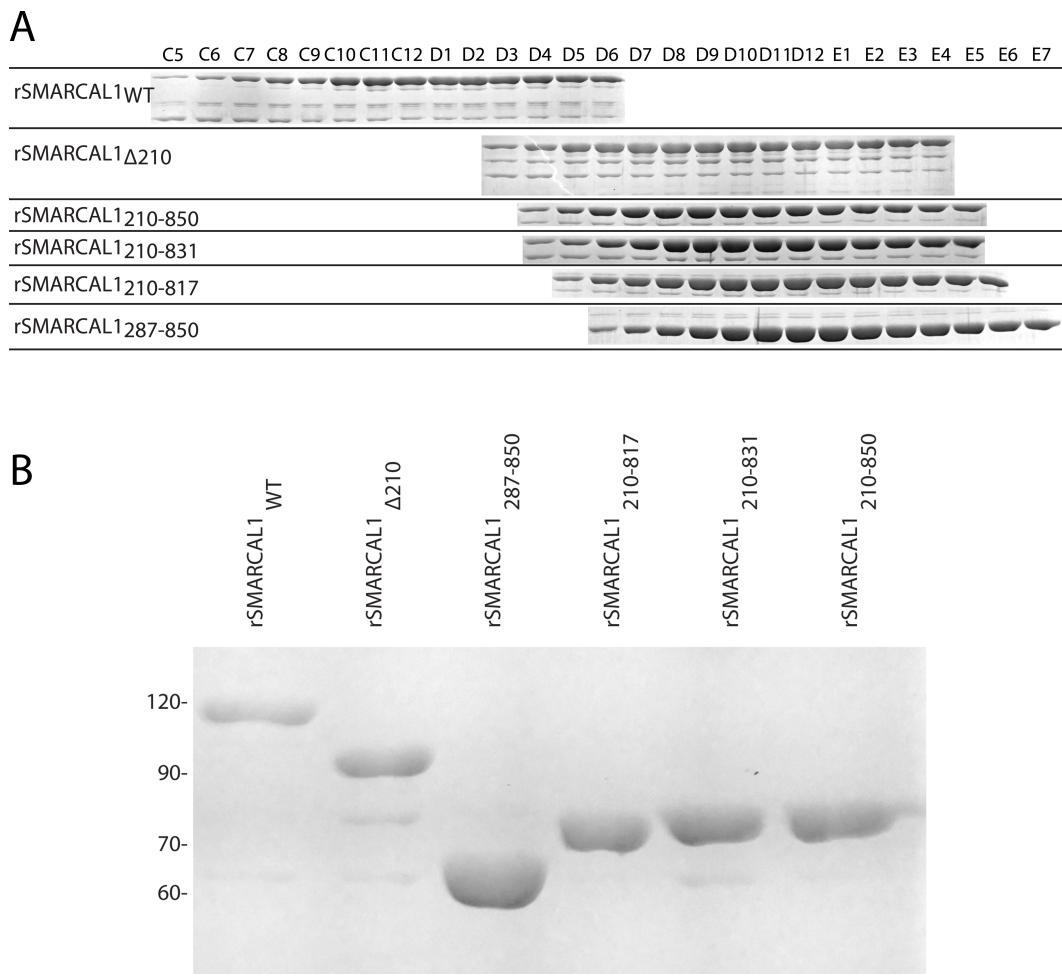


Figure 3.4: (A) Peak fractions collected were run on an SDS-PAGE for all six proteins for comparison of purity. (B) Peak fractions collected and spin concentrated, and the proteins used for assays were run on an SDS-PAGE to verify final purity.

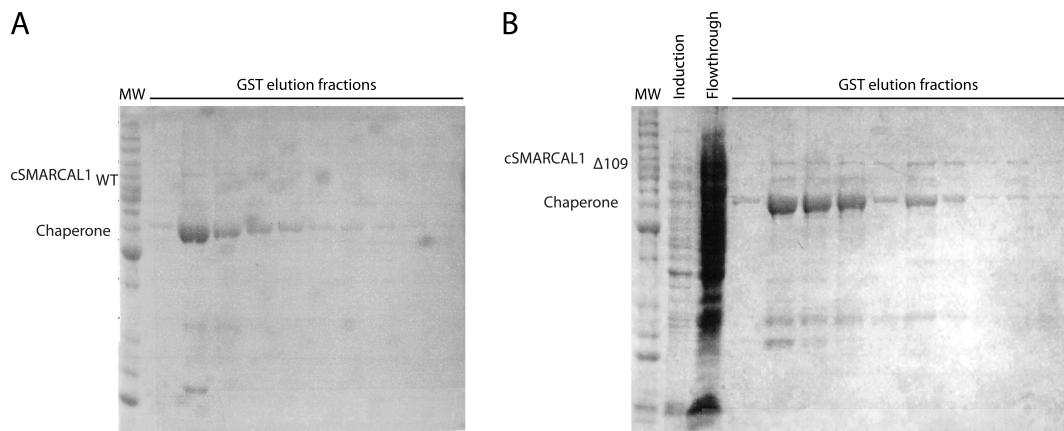


Figure 3.5: *C. elegans* expression and purification. (A) cSMARCAL1_{WT} fractions collected from a GST affinity elution. (B) cSMARCAL1_{Δ109} fractions collected from a GST affinity elution.

3.3 Discussion

In the context of structural biology *E. coli* is the most adopted expression system, it composes 89% of all PDB structures deposited⁷². Many advantages exist, though this can be mainly attributed to the ease of use, short time necessity, and a low cost. The major disadvantages being *E. coli* can't add post-translational modifications (PTMs) and it lacks an extensive chaperone machinery to facilitate protein folding⁷³. SMARCAL1 may not require PTMs as it has been found that phosphorylation at S652 can be inhibitory⁵⁵. However, the extensive chaperone folding machinery may be required, since expression and purification of a human SMARCAL1 protein was unsuccessful in our lab.

We were successful in purifying six *R. norvegicus* SMARCAL1 in an *E. coli* expression system, including a full length protein and five truncation mutants. Our protein truncations are expected to be functional as biochemical studies determined the RBM and first HARP domain of the human SMARCAL1 protein are dispensable for *in vitro* ssDNA annealing activity⁵⁹. Furthermore, these authors have determined the minimal enzymatic unit to be 325-870 containing the second HARP domain and ATPase domain⁵⁹. Our rSMARCAL1_{WT} and rSMARCAL1_{Δ210} purified proteins show a similar SDS-PAGE banding pattern suggesting the proteins degrade to a consistent lower molecular weight product. We think this degradation product is similar to our rSMARCAL1₂₈₇₋₈₅₀ protein and the human SMARCAL1 minimal enzymatic fragment. The copurification of a chaperone is treated as an indicator of inherently unfolded regions⁷⁴. Furthermore, rSMARCAL1₂₈₇₋₈₅₀ is the protein for which we obtain highest yield, purity, and lowest chaperone copurification.

We investigated the possibility that *C. elegans* SMARCAL1 could be a highly stable protein, on the basis that it only contains one HARP. It also has an overall shorter sequence, thus it could be a naturally occurring minimal enzymatic frag-

ment. However, we were unsuccessful in obtaining any amounts of purified protein. Additionally, the large amounts of chaperone collected and low cSMARCAL1 protein yields indicate unfolded proteins were expressed.

Chapter 4

Characterization of Purified SMARCAL1 proteins

Contributions: Dr. David Cortez provided insect cell purified human SMAR-CAL1 protein.

4.1 Introduction

Purified proteins must be characterized, to ensure they are functional and that conclusions drawn from structural studies are accurate. Many SMARCAL1 DNA binding and ATPase activity studies are found in literature, and an extensive list of substrates has been tested. DNA binding studies by EMSA have shown SMARCAL1 selectively binding to a fork DNA structure, whereas dsDNA and ssDNA does not show binding^{59,69}. However, studies of a fragment of the bovine SMARCAL1 ortholog by means of fluorescence quenching techniques show it can bind to fork DNA, dsDNA and ssDNA⁶⁰. The binding mechanism is still ambiguous, however, a hypothesis suggests there are two binding sites where one site can accommodate ssDNA and the other the dsDNA⁶⁰. Furthermore, it is proposed that both sites must be occupied to effect hydrolysis⁶⁰. This is supported by the fact that SMARCAL1 ATPase activity is seen for structures containing both ssDNA and dsDNA regions, although SMARCAL1 ATPase activity has not been observed for ssDNA alone⁶⁰. However, very low levels of ATPase activity in the presence of dsDNA are seen in one study of human SMARCAL1⁶⁹.

In this chapter we describe DNA binding and ATPase activity for our purified SMARCAL1 proteins, and determine effects of the truncations mutants produced. As a comparison point in our biochemical studies we have included a human SMARCAL1 protein produced in an insect cell expression system.

4.2 Results

4.2.1 SMARCAL1 DNA binding activity by EMSA

We characterized our SMARCAL1 proteins for DNA binding using an electrophoretic mobility shift assays (EMSA). All six rat SMARCAL1 protein produced were assayed and additionally we included a human SMARCAL1 insect cell purified protein. The proteins were tested binding to three different DNA substrates: fork DNA, dsDNA, ssDNA (Figure 4.1 A-C). SMARCAL1 binding produces a shift of the radioactively labeled probe to a higher weight species, when fork DNA substrate is employed (Figure 4.1 A). The human hSMARCAL1_{WT} protein displays similar affinities as all our rat proteins with the exception of rSMARCAL1_{WT}. A weak shift can be seen for dsDNA (Figure 4.1 B) and almost none can be detected from ssDNA (Figure 4.1 C). The higher protein concentrations seem to produce some aggregation, as can be seen from DNA being retained in the wells.

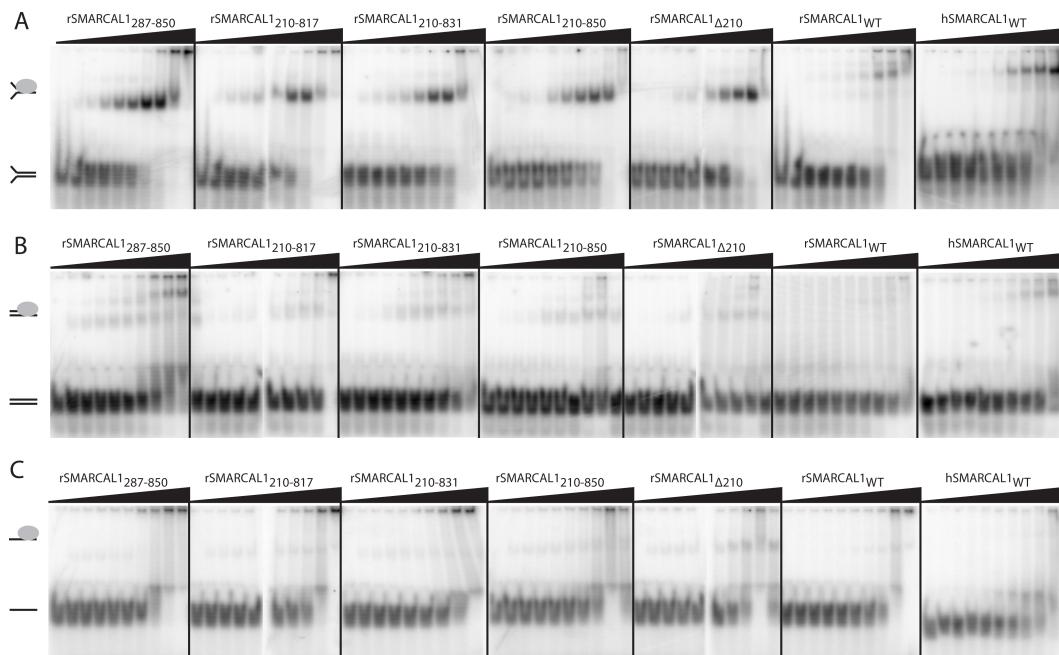


Figure 4.1: Comparison of rSMARCAL1 proteins and hSMARCAL1WT for DNA binding to DNA substrates by electrophoretic mobility shift assay. DNA concentration was kept constant at 1 nM. Protein concentration was varied from 0 to 256 nM for rat SMARCAL1, and 0 to 64 nM for human SMARCAL1. Three substrates were tested: A) fork DNA (21 bp dsDNA arm and two 9 nt ssDNA arms). B) double stranded DNA (30 bp) . C) single stranded DNA (30 bp).

4.2.2 SMARCAL1 DNA-dependent ATPase activity

The ATPase activity was determined by an ATP/NADH coupled assay. This assay allows for steady state kinetics measurements, since the ATP is constantly regenerated through reactions leading to a final conversion of NADH to NAD⁺ (Figure 4.2). Progression curves were collected in triplicate and the relative fluorescence units (RFU) were converted to NADH by the use of a standard curve (Figure 4.3). Linear regression curves were plotted to obtain the initial velocities for each sample. We assayed all six rat SMARCAL1 proteins and the human SMARCAL1 insect cell purified protein. Proteins were tested at a constant concentration, and varying amounts of DNA substrates. Three DNA substrates were assayed for each protein: fork DNA, dsDNA and ssDNA.

The initial reaction rates were plotted against DNA concentrations (Figure 4.4). As expected we see ATPase activity from fork DNA, and a lack of activity from ssDNA. However, our results show dsDNA stimulated ATPase activity, which hasn't been observed in prior research. A comparison of the extracted Michaelis-Menten parameters for fork DNA and dsDNA shows a similar k_{cat} for each protein is reached (Table 4.1). However a comparison of the affinities for fork DNA and shows a 5-23 fold increase in K_M for dsDNA. A comparison of the human hSMARCAL1_{WT} to our rat rSMARCAL1_{WT} full length proteins shows similar k_{cat} , although there may be a loss of discrimination of fork DNA to dsDNA. The fold change is diminished from 23 to 14 for hSMARCAL1_{WT} and rSMARCAL1_{WT} respectively. The protein that displays both the highest k_{cat} and K_M is rSMARCAL1₂₈₇₋₈₅₀. This may be due to the loss of the first HARP which may induce some disregulation. We were unable to extract k_{cat} and K_M parameters for ssDNA due to lack of ATPase activity, even though there may be weak binding observable by EMSA.

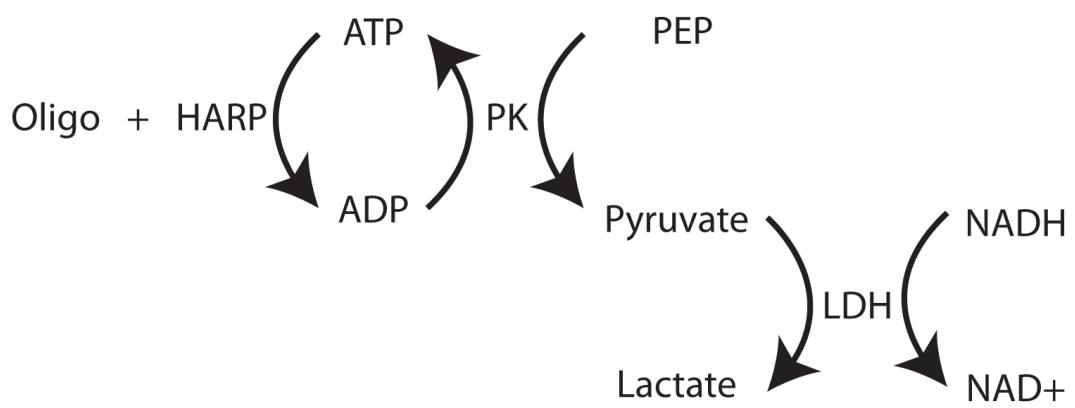


Figure 4.2: ATP/NADH coupled assay reaction scheme. We used the oligo to start the reaction. ATP is regenerated through pyruvate kinase by the conversion of phosphoenol pyruvate into pyruvate. Pyruvate is converted by lactose dehydrogenase into lactate through NADH oxidation into NAD⁺. NADH is a fluorescent molecule, however, NAD⁺ is not. Therefore this assay monitors a decrease in fluorescence over time.

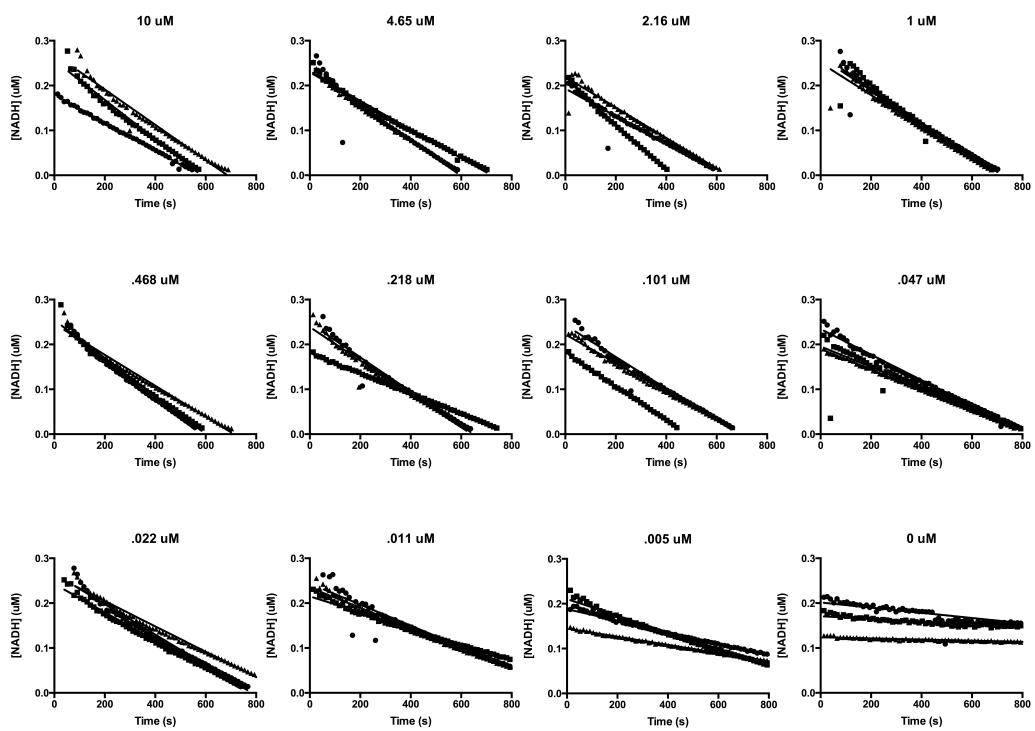


Figure 4.3: Progression curves collected for rSMARCAL1_{WT} ATPase assay where RFU was transformed to NADH concentration. The linear regression is plotted to the data to obtain the initial rates of reactions.

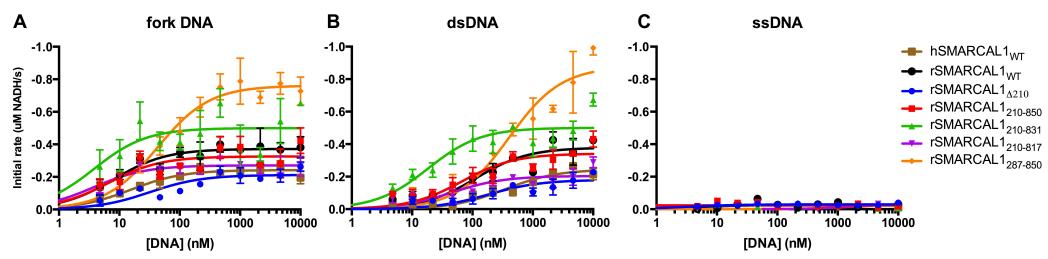


Figure 4.4: SAMRCAL1 fork DNA and dsDNA ATPase activity. The initial rates obtained from progression curved of all seven SMARCAL1 proteins were plotted against substrate concentration. The data was fitted to the Michaelis-Menten equation and graphed in a logarithmic scale for DNA concentration

Table 4.1: Michaelis-Menten kinetic parameters

Protein	k_{cat} (s ⁻¹)			K_M (nM)			k_{cat}/K_M (nM ⁻¹ x s ⁻¹)	
	fork DNA	dsDNA	fork DNA	dsDNA	fold change	fork DNA	dsDNA	
hSMARCAL1 WT	48 ± 2	49 ± 3	15 ± 3	347 ± 86	23	3.1 ± 0.7	0.1 ± 0.04	
rSMARCAL1 WT	37 ± 1	38 ± 2	8 ± 2	113 ± 27	14	4.6 ± 1.1	0.3 ± 0.08	
rSMARCAL1 Δ210	21 ± 1	18 ± 1	32 ± 10	159 ± 56	5	0.7 ± 0.2	0.1 ± 0.04	
rSMARCAL1Δ210-850	33 ± 2	34 ± 2	7 ± 2	67 ± 17	10	4.9 ± 1.8	0.5 ± 0.13	
rSMARCAL1Δ210-831	50 ± 3	50 ± 2	4 ± 2	19 ± 4	5	13.0 ± 6.4	2.7 ± 0.66	
rSMARCAL1Δ210-817	27 ± 1	20 ± 1	4 ± 1	45 ± 15	12	7.3 ± 1.9	0.5 ± 0.16	
rSMARCAL1Δ287-850	76 ± 2	88 ± 4	43 ± 6	402 ± 79	9	1.8 ± 0.3	0.2 ± 0.04	

4.3 Discussion

SMARCAL1 has a known DNA binding preference for fork DNA over ssDNA and dsDNA by EMSA, which translates to a similar preference in ATPase activity^{48,59,69}. All of our purified proteins were tested for DNA binding and ATPase activity in respect to three substrates: fork DNA, dsDNA, and ssDNA.

It is unfortunate our rat rSMARCAL1_{WT} protein has the lowest fork DNA binding affinity of the seven tested proteins. Although, this may be due to the aggregation that can be seen in the well, which correlates with the increasing protein concentrations. rSMARCAL1_{WT} behaves better in the ATP/NADH coupled assay, due to the very low protein concentrations needed.

The ATPase assay to confirm the functionality of our purified proteins and compare them to human SMARCAL1 expressed from insect cells, turned out to draw interesting results. We defined a new interaction of SMARCAL1 with dsDNA. However, it is not surprising, because the SWI2/SNF2 family of proteins acts as ds-DNA translocases¹². The SWI2/SNF2 translocases are known to bind to the minor groove of DNA in order to track along it¹³. SMARCAL1 contains the conserved motifs representative of this family as previously described (Section 1.1.1). The dsDNA dependent SMARCAL1 ATPase activity was previously seen, although to a very small extent. We have employed higher concentrations of dsDNA than previously tested^{59,69}.

The use of a human SMARCAL1 insect cell expressed protein has been useful in determining that the ATPase activity seen in presence of dsDNA is not due to the use of a bacterial expression system nor a feature of the rat ortholog. Thus ruling these out as a confounding factors. Furthermore, a comparison of our rat rSMARCAL1_{WT} to the human hSMARCAL1_{WT} validates that the proteins have similar activities and a structural analysis of the rat SMARCAL1 proteins is warranted. We determined that for wild type SMARCAL1 proteins the affinity of fork

DNA is 14 to 23 fold greater than that for dsDNA. The difference in fold change of human to rat proteins may be due to differences in sequence conservation related to the species. Although, it may also be due to slight problems in folding emerging from the bacterial expression system.

The SMARCAL1 binding model where dsDNA and ssDNA bind to two different sites⁶⁰ is consistent with our findings. We see small amounts of dsDNA and ssDNA binding on our EMSA assays. The finding that dsDNA elicits ATPase activity suggests SMARCAL1 may have a similar DNA translocase mechanism as found in the SWI2/SNF2 family¹³. The annealing activity is specific to the action of the HARP domains^{19,59}. Chimeric proteins, where the SMARCAL1 HARP domains were fused to SWI2/SNF2 helicase domains display efficient annealing activity¹⁹. The HARP domains may indeed be the second proposed site of binding which direct the specificity to ssDNA. Studies of the HARP domains alone show that they can bind fork DNA to a limited extent⁵⁹. Supporting results show fork stimulated ATPase activity in the absence of both HARP domains, thus suggesting the conserved helicase is sufficient to elicit ATPase activity¹⁹.

Chapter 5

SMARCAL1 Crystallography Trials

Contributions: Curtis Hodges has performed crystallography trials of the rSMARCAL1 Δ 210 protein prior to the start of my research, although, his trials are not reported here. G2 Wizard optimization grid was prepared by Curtis Hodges.

5.1 Introduction

The mechanism of SMARCAL1 fork remodelling and annealing helicase activities are still poorly understood. However, recent single molecule studies have determined a preference for leading strand gaps and DNA binding in an asymmetric manner. Furthermore, SMARCAL1 structural studies have been performed by small angle x-ray scattering (SAXS). A fragment of human SMARCAL1 containing the second HARP and helicase domains has supplied some structural insight. Modeling SMARCAL1 to SAXS curves displays a conserved SSORAD54 helicase domain linked to a HARP domain in an elongated form⁵⁹. The HARP domains were successfully modeled by replacing them with the PUR- α DNA binding domain⁵⁹. This domain is a valid homology model, because it has known ssDNA and dsDNA binding⁵⁹.

We sought to expand the structural knowledge of SMARCAL1 by X-ray crystallography. However, protein crystallization is often not straight forward. In this chapter I describe multiple crystallization trials and the rationale behind the different attempts.

5.2 Results

5.2.1 rSMARCAL1 Δ_{210}

Protein was expressed and purified for crystallography trials. Due to the low yield of rSMARCAL1 Δ_{210} protein only one tray was set using the JCSG+ suite. Each of the three sitting drops was tested with a different DNA substrate (16ds4ss, 21ds9ss5', 21ds4ss) in the presence of magnesium and AMP-PNP. AMP-PNP is a non-hydrolysable ATP ortholog. After two month of growth crystal grew in wells E12 and H3 of the -5mis-5M DNA mix, and in wells A5, E12, and H3 of 21ds9ss5' DNA mix. Crystals were small, but could be looped in cryoprotectant solutions. They were mounted and shot using the bending magnet x-ray source at the Canadian light source (CLS). They showed no diffraction beyond some ice ring patterns. Optimization was not pursued due to realization that the rSMARCAL1₂₈₇₋₈₅₀ protein was purified in higher yields, and had better stability as described previously.

5.2.2 rSMARCAL1₂₈₇₋₈₅₀

A summary of all rSMARCAL1₂₈₇₋₈₅₀ crystallization trials is presented in Table 5.1. The first crystallization trials with the rSMARCAL1₂₈₇₋₈₅₀ protein were set and incubated at room temperature for six crystallization suites. Additionally a dilution of protein concentration was tested on a JCSG+ suite. However, precipitation was observed in 80% to 90% of the wells.

To determine if the temperature was a factor affecting monodispersity of this protein, dynamic light scattering (DLC) was employed. A temperature gradient from 4°C to 25°C was tested, and DLC results show formation of aggregates as the temperature raised to 25°C (data not shown). From that point on crystal trays were set up at 4°C. The next crystallization trials were prepared in the cold room and tested six suites, but no crystals were seen. The next batch of protein was used to

set up 11 screens with the presence of AMP-PNP and magnesium. Two drops were set each with a different DNA substrate.

At this point another DLS was performed to determine lowest salt concentrations that can be used. We found that down to 150 mM NaCl there was no aggregates, however 100 mM salt showed aggregation, therefore the next trials were performed at 200 mM NaCl to be prudent. Furthermore, in these trials, we used sodium orthovanadate, which is an ATP analog. Sodium orthovanadate belongs to a class of phosphate analogs that have been previously used in crystallography to determine multiple structures of ATPases⁷⁵. These last crystallography trials tested five screens and had two drops, one with DNA the other without. This generated one hit, which was previously seen by Curtis Hodges, which had already prepared a optimization grid for this condition. This grid was used with the same conditions, however, the crystal did not reproduce.

Table 5.1: Summary of all crystallization trials

Protein Concentration	Temp	DNA substrate	ATP substrate	Suites
3.5 mg/mL	RT	No	No	Wizard I & II, Classics I & II, MPD, PEG
3.5, 3.0, 2.5 mg/mL	RT	No	No	JCSG+
3.0 mg/mL	4°C	No	No	Wizard I & II, Classics, MPD, PACT, PEG II
6.0 mg/mL	4°C	16ds9ss, 21ds9ss5'	AMP-PNP	Wizard I & II, PEG I & II, MPD, PACT, JCSG+, KeraFast, ProComplex, Natrix, Nucleix, PEG/Ion
3.0 mg/mL	4°C	No	Sodium Orthovanadate	Wizard I & II, PEG, Classics, JCSG+
3.0 mg/mL	4°C	fork DNA	Sodium Orthovanadate	Wizard I & II, PEG, Classics, JCSG+

All trials that include DNA also contained magnesium.

5.3 Discussion

A large focus was put on crystallographic trials of rSMARCAL1₂₈₇₋₈₅₀, which contains the second HARP and helicase domains. As described in Chapter 3, this construct was expressed in higher yields, it is more stable in crystallography buffer, and it showed no proteolytic degradation as opposed to other constructs. Chapter 4 has shown rSMARCAL1₂₈₇₋₈₅₀ has retained DNA binding and ATPase activity, thus determining it is functional. rSMARCAL1₂₈₇₋₈₅₀ is a similar construct to the human SMARCAL1 (325-890) determined to be the minimal enzymatic fragment or the human SMARCAL1 (325-954) that was modeled by SAXS studies⁵⁹.

It is unfortunate that crystallization trials were unsuccessful. However, there are further possibilities to produce a crystal of SMARCAL1. There are 41 known SMARCAL1 orthologs that can be tried, one of which is the bovine SMARCAL1 fragment that can already be produced in high yields and purity in a bacterial expression system⁴⁸. Further crystallography trials can be attempted, where ATP and DNA in absence of magnesium, due to requirement of magnesium for hydrolysis this could trap a protein bound to ATP and DNA⁶¹. Similarly, DNA alone, ATP alone, ADP and DNA, can be attempted to produce SMARCAL1 protein crystals.

The use of glycerol free buffers for SMARCAL1 crystallography can be reassessed. Chapter 3 shows there is a much greater stability of our purified rat proteins in the presence of glycerol, especially for the rSMARCAL1₂₁₀₋₈₁₇ protein. Low concentration of glycerol can be used to enhance crystallization for unstable proteins⁷⁶. The benefits upon protein stabilization may outweigh the harm to nucleation^{77,78}.

Chapter 6

Conclusions

6.1 Conclusions

We purified six rat SMARCAL1 proteins, where truncation mutants determined a stable minimal fragment. Characterization assays determined the bacterial expression of rat SMARCAL1 proteins produces functionally similar proteins to an insect cell expressed human SMARCAL1. The rSMARCAL1₂₈₇₋₈₅₀ protein was expressed in the highest yield and displayed no proteolytic degradation upon purification. DNA binding was similar to other proteins, however, the deletion of the first HARP domain may have some disregulatory effects on the ATPase activity. The rSMARCAL1₂₈₇₋₈₅₀ protein was therefore a hopeful target for crystallization trials, although we were unsuccessful in obtaining crystals.

Very low levels of dsDNA stimulated ATPase activity have previously been shown for SMARCAL1⁶⁹. However, we have determined that SMARCAL1 attains the same catalytic rates in presence of dsDNA as for a fork DNA substrate. Furthermore, we determined the SMARCAL1 affinity to dsDNA is increased (5-23 fold) over a fork DNA substrate. It would be interesting to produce a SMARCAL1 core helicase domains alone, and verify if fork DNA has a similar affinity to dsDNA. This would effectively determine if SMARCAL1 lost the additional recognition of the fork by HARP domains.

The involvement of SMARCAL1 in pathways of DNA damage response and cell cycle progression has lead to the hypothesis it may be exploited to sensitize tumor cells to cancer therapy^{79,80}. Aminoglycoside antibiotics can be inactivated by phosphorylation to generate phosphoaminoglycosides⁸¹. Phosphoaminoglycosides are a class of potent SMARCAL1 inhibitors, additionally they target the whole SWI2/SNF2 family⁸². The ATPase assay here optimized is well suited for high throughput screening of drug libraries and it can be a powerful tool in the discovery of new SMARCAL1 and SWI2/SNF2 inhibitors. The discovery of a specific SMARCAL1 inhibitor may also bring a greater understanding in SMARCAL1 cel-

lular functions, by allowing new experimental methods to be developed.

SMARCAL1 crystallization can also benefit from the use of phosphoaminoglycosides, which could trap SMARCAL1 in an inactive state. Crystallization could be attempted on a SMARCAL1 phosphomimetic mutation S652D effecting the linker region between helicase domains⁵⁵. However, further mutagenesis within the linker region would be required to completely inactivate SMARCAL1. This would generate an inactive protein where DNA and ATP can bind but no hydrolysis would ensue due to the loss of flexibility of the linker region⁵⁵. Bovine SMARCAL1 truncations have shown DNA can bind to SMARCAL1 in the absence of the second helicase domain. Thus, a truncation of the C-terminal deleting the second helicase domain could reduce the size and flexibility of SMARCAL1, making it more suitable for crystallization.

Bibliography

References

- [1] Watson, J. D. and Crick, F. H. C. (1953) THE STRUCTURE OF DNA. *Cold Spring Harbor Symposia on Quantitative Biology* 18, 123–131.
- [2] Kornberg, A. (1960) Biologic synthesis of deoxyribonucleic acid. *Science (New York, N.Y.)* 131, 1503–1508.
- [3] ABDEL-MONEM, M., DURWALD, H., and HOFFMANN-BERLING, H. (1976) Enzymic Unwinding of DNA. 2. Chain Separation by an ATP-Dependent DNA Unwinding Enzyme. *European Journal of Biochemistry* 65, 441–449.
- [4] Scott, J. F., Eisenberg, S., Bertsch, L. L., and Kornberg, A. (1977) A mechanism of duplex DNA replication revealed by enzymatic studies of phage phi X174: catalytic strand separation in advance of replication. *Proceedings of the National Academy of Sciences of the United States of America* 74, 193–197.
- [5] Hübscher, U. and Stalder, H. P. (1985) Mammalian DNA helicase. *Nucleic acids research* 13, 5471–5483.
- [6] Fairman-Williams, M. E., Guenther, U.-P., and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology* 20, 313–324.
- [7] Tanner, N. K., Cordin, O., Banroques, J., Doère, M., and Linder, P. (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Molecular cell* 11, 127–138.
- [8] Peterson, C. L. and Tamkun, J. W. (1995) The SWI-SNF complex: a chromatin remodeling machine? *Trends in biochemical sciences* 20, 143–146.
- [9] Pazin, M. J. and Kadonaga, J. T. (1997) SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein-DNA interactions? *Cell* 88, 737–740.
- [10] Reisman, D., Glaros, S., and Thompson, E. A. (2009) The SWI/SNF complex and cancer. *28*, 1653–1668.
- [11] Tang, L., Nogales, E., and Ciferri, C. (2010) Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Progress in Biophysics and Molecular Biology* 102, 122–128.
- [12] Hopfner, K.-P., Gerhold, C.-B., Lakomek, K., and Wollmann, P. (2012) Swi2/Snf2 remodelers: hybrid views on hybrid molecular machines. *Current Opinion in Structural Biology* 22, 225–233.
- [13] Hopfner, K.-P. and Michaelis, J. (2007) Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics. *Current Opinion in Structural Biology* 17, 87–95.

- [14] Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S., and Wigley, D. B. (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* 97, 75–84.
- [15] Flaus, A. (2006) Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic acids research* 34, 2887–2905.
- [16] Weinstock, G. M., McEntee, K., and Lehman, I. R. (1979) ATP-dependent renaturation of DNA catalyzed by the recA protein of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 76, 126–130.
- [17] Wu, Y. (2012) Unwinding and Rewinding: Double Faces of Helicase? *Journal of Nucleic Acids* 2012, 1–14.
- [18] Chen, J. Y. G. G. J., Ghosal, G., and Chen, J. (2012) The HARP-like Domain-Containing Protein AH2/ZRANB3 Binds to PCNA and Participates in Cellular Response to Replication Stress. *Molecular cell* 47, 410–421.
- [19] Ghosal, G., Yuan, J., and Chen, J. (2011) The HARP domain dictates the annealing helicase activity of HARP/SMARCAL1. *EMBO reports* 12, 574–580.
- [20] Rezazadeh, S. (2012) RecQ helicases; at the crossroad of genome replication, repair, and recombination. *Molecular biology reports* 39, 4527–4543.
- [21] Bernstein, K. A., Gangloff, S., and Rothstein, R. (2010) The RecQ DNA helicases in DNA repair. *Annual Review of Genetics* 44, 393–417.
- [22] Machwe, A., Karale, R., Xu, X., Liu, Y., and Orren, D. K. (2011) The Werner and Bloom Syndrome Proteins Help Resolve Replication Blockage by Converting (Regressed) Holliday Junctions to Functional Replication Forks. *Biochemistry* 50, 6774–6788.
- [23] Machwe, A., Xiao, L., Groden, J., and Orren, D. K. (2006) The Werner and Bloom Syndrome Proteins Catalyze Regression of a Model Replication Fork. *Biochemistry* 45, 13939–13946.
- [24] Bétous, R., Couch, F. B., Mason, A. C., Eichman, B. F., Manosas, M., and Cortez, D. (2013) Substrate-selective repair and restart of replication forks by DNA translocases. *Cell reports* 3, 1958–1969.
- [25] Singleton, M. R., Scaife, S., and Wigley, D. B. (2001) Structural Analysis of DNA Replication Fork Reversal by RecG. *Cell* 107, 79–89.
- [26] Gregg, A. V., McGlynn, P., Jaktaji, R. P., and Lloyd, R. G. (2002) Direct Rescue of Stalled DNA Replication Forks via the Combined Action of PriA and RecG Helicase Activities. *Molecular cell* 9, 241–251.

- [27] Yeeles, J. T., Poli, J., Marians, K. J., and Pasero, P. (2013) Rescuing Stalled or Damaged Replication Forks. *Cold Spring Harbor perspectives in biology* 5.
- [28] Manosas, M., Perumal, S. K., Croquette, V., and Benkovic, S. J. (2012) Direct Observation of Stalled Fork Restart via Fork Regression in the T4 Replication System. *Science (New York, N.Y.)* 338, 1217–1220.
- [29] Petermann, E. and Helleday, T. (2010) Pathways of mammalian replication fork restart. *Nature reviews. Molecular cell biology* 11, 683–687.
- [30] Weinert, T., Kaochar, S., Jones, H., Paek, A., and Clark, A. J. (2009) The replication fork’s five degrees of freedom, their failure and genome rearrangements. *Current Opinion in Cell Biology* 21, 778–784.
- [31] Driscoll, R. and Cimprich, K. A. (2009) HARPing on about the DNA damage response during replication. *Genes & development* 23, 2359–2365.
- [32] Blow, J. J., Ge, X. Q., and Jackson, D. A. (2011) How dormant origins promote complete genome replication. *Trends in biochemical sciences* 36, 405–414.
- [33] Yusufzai, T. and Kadonaga, J. T. (2010) Annealing helicase 2 (AH2), a DNA-rewinding motor with an HNH motif. *Proceedings of the National Academy of Sciences of the United States of America* 107, 20970–20973.
- [34] Ciccia, A., Nimonkar, A. V., Hu, Y., Hajdu, I., Achar, Y. J., Izhar, L., Petit, S. A., Adamson, B., Yoon, J. C., Kowalczykowski, S. C., Livingston, D. M., Haracska, L., and Elledge, S. J. (2012) Polyubiquitinated PCNA recruits the ZRANB3 translocase to maintain genomic integrity after replication stress. *Molecular cell* 47, 396–409.
- [35] Schimke, R. N., Horton, W. A., and King, C. R. (1971) Chondroitin-6-sulphaturia, defective cellular immunity, and nephrotic syndrome. *Lancet* 2, 1088–1089.
- [36] Spranger, J., Hinkel, G. K., Stöss, H., Thoenes, W., Wargowski, D., and Zepp, F. (1991) Schimke immuno-osseous dysplasia: a newly recognized multisystem disease. *The Journal of pediatrics* 119, 64–72.
- [37] Petty, E. M., Yanik, G. A., Hutchinson, R. J., Alter, B. P., Schmalstieg, F. C., Levine, J. E., Ginsburg, D., Robillard, J. E., and Castle, V. P. (2000) Successful bone marrow transplantation in a patient with Schimke immuno-osseous dysplasia. *The Journal of pediatrics* 137, 882–886.
- [38] Boerkoei, C. F., Takashima, H., John, J., Yan, J., Stankiewicz, P., Rosenbarker, L., André, J.-L., Bogdanovic, R., Burguet, A., Cockfield, S., Cordeiro, I., Fründ, S., Illies, F., Joseph, M., Kaitila, I., Lama, G., Loirat, C., McLeod,

- D. R., Milford, D. V., Petty, E. M., Rodrigo, F., Saraiva, J. M., Schmidt, B., Smith, G. C., Spranger, J., Stein, A., Thiele, H., Tizard, J., Weksberg, R., Lupski, J. R., and Stockton, D. W. (2002) Mutant chromatin remodeling protein SMARCAL1 causes Schimke immuno-osseous dysplasia. *Nature genetics* 30, 215–220.
- [39] Elizondo, L. I., Huang, C., Northrop, J. L., Deguchi, K., Clewing, J. M., Armstrong, D. L., and Boerkoel, C. F. (2006) Schimke immuno-osseous dysplasia: A cell autonomous disorder? *American Journal of Medical Genetics Part A* 140A, 340–348.
- [40] Baradaran-Heravi, A., Raams, A., Lubieniecka, J., Cho, K. S., DeHaai, K. A., Basiratnia, M., Mari, P.-O., Xue, Y., Rauth, M., Olney, A. H., Shago, M., Choi, K., Weksberg, R. A., Nowaczkyk, M. J. M., Wang, W., Jaspers, N. G. J., and Boerkoel, C. F. (2012) SMARCAL1 deficiency predisposes to non-Hodgkin lymphoma and hypersensitivity to genotoxic agents in vivo. *American Journal of Medical Genetics Part A* 158A, 2204–2213.
- [41] Carroll, C., Badu-Nkansah, A., Hunley, T., Baradaran-Heravi, A., Cortez, D., and Frangoul, H. (2013) Schimke immunoosseous dysplasia associated with undifferentiated carcinoma and a novel SMARCAL1mutation in a child. *Pediatric Blood & Cancer* pp. 000–000.
- [42] Clewing, J. M., Fryssira, H., Goodman, D., Smithson, S. F., Sloan, E. A., Lou, S., Huang, Y., Choi, K., Lücke, T., Alpay, H., André, J.-L., Asakura, Y., Biebuyck-Gouge, N., Bogdanovic, R., Bonneau, D., Cancrini, C., Cochat, P., Cockfield, S., Collard, L., Cordeiro, I., Cormier-Daire, V., Cransberg, K., Cutka, K., Deschenes, G., Ehrlich, J. H. H., Fründ, S., Georgaki, H., Guillén-Navarro, E., Hinkelmann, B., Kanariou, M., Kasap, B., Kilic, S. S., Lama, G., Lamfers, P., Loirat, C., Majore, S., Milford, D., Morin, D., Özdemir, N., Pontz, B. F., Proesmans, W., Psoni, S., Reichenbach, H., Reif, S., Rusu, C., Saraiva, J. M., Sakallioğlu, O., Schmidt, B., Shoemaker, L., Sigaudy, S., Smith, G., Sotsiou, F., Stajic, N., Stein, A., Stray-Pedersen, A., Taha, D., Taque, S., Tizard, J., Tsimaratos, M., Wong, N. A. C. S., and Boerkoel, C. F. (2007) Schimke immunoosseous dysplasia: suggestions of genetic diversity. *Human mutation* 28, 273–283.
- [43] Yue, Z., Xiong, S., Sun, L., Huang, W., Mo, Y., Huang, L., Jiang, X., Chen, S., Hu, B., and Wang, Y. (2010) Novel compound mutations of SMARCAL1 associated with severe Schimke immuno-osseous dysplasia in a Chinese patient. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 25, 1697–1702.
- [44] Elizondo, L. I., Cho, K. S., Zhang, W., Yan, J., Huang, C., Huang, Y., Choi, K., Sloan, E. A., Deguchi, K., Lou, S., Baradaran-Heravi, A., Takashima, H.,

- Lucke, T., Quiocho, F. A., and Boerkoel, C. F. (2009) Schimke immunoosseous dysplasia: SMARCAL1 loss-of-function and phenotypic correlation. *Journal of medical genetics* 46, 49–59.
- [45] Baradaran-Heravi, A., Cho, K. S., Tolhuis, B., Sanyal, M., Morozova, O., Morimoto, M., Elizondo, L. I., Bridgewater, D., Lubieniecka, J., Beirnes, K., Myung, C., Leung, D., Fam, H. K., Choi, K., Huang, Y., Dionis, K. Y., Zonana, J., Keller, K., Stenzel, P., Mayfield, C., Lucke, T., Bokenkamp, A., Marra, M. A., van Lohuizen, M., Lewis, D. B., Shaw, C., and Boerkoel, C. F. (2012) Penetrance of biallelic SMARCAL1 mutations is associated with environmental and genetic disturbances of gene expression. *Human Molecular Genetics* 21, 2572–2587.
- [46] Mesner, L. D., Sutherland, W. M., and Hockensmith, J. W. (1991) DNA-dependent adenosinetriphosphatase A is the eukaryotic analog of the bacteriophage T4 gene 44 protein: immunological identity of DNA replication-associated ATPases. *Biochemistry* 30, 11490–11494.
- [47] Mesner, L. D., Truman, P. A., and Hockensmith, J. W. (1993) DNA-dependent adenosinetriphosphatase A: Immunoaffinity purification and characterization of immunological reagents. *Biochemistry* 32, 7772–7778.
- [48] Muthuswami, R., Truman, P. A., Mesner, L. D., and Hockensmith, J. W. (2000) A eukaryotic SWI2/SNF2 domain, an exquisite detector of double-stranded to single-stranded DNA transition elements. *The Journal of biological chemistry* 275, 7648–7655.
- [49] Coleman, M. A., Eisen, J. A., and Mohrenweiser, H. W. (2000) Cloning and characterization of HARP/SMARCAL1: a prokaryotic HepA-related SNF2 helicase protein from human and mouse. *Genomics* 65, 274–282.
- [50] Ring, H. Z., Vameghi-Meyers, V., Wang, W., Crabtree, G. R., and Francke, U. (1998) Five SWI/SNF-Related, Matrix-Associated, Actin-Dependent Regulator of Chromatin (SMARC) Genes Are Dispersed in the Human Genome. *Genomics* 51, 140–143.
- [51] Lewis, L. K., Jenkins, M. E., and Mount, D. W. (1992) Isolation of DNA damage-inducible promoters in Escherichia coli: regulation of polB (dinA), dinG, and dinH by LexA repressor. *Journal of bacteriology* 174, 3377–3385.
- [52] McKinley, B. A. and Sukhodolets, M. V. (2007) Escherichia coli RNA polymerase-associated SWI/SNF protein RapA: evidence for RNA-directed binding and remodeling activity. *Nucleic acids research* 35, 7044–7060.
- [53] Byun, T. S. (2005) Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes & development* 19, 1040–1052.

- [54] Cortez, D. (2005) Unwind and slow down: checkpoint activation by helicase and polymerase uncoupling. *Genes & development* 19, 1007–1012.
- [55] Couch, F. B., Bansbach, C. E., Driscoll, R., Luzwick, J. W., Glick, G. G., Betous, R., Carroll, C. M., Jung, S. Y., Qin, J., Cimprich, K. A., and Cortez, D. (2013) ATR phosphorylates SMARCAL1 to prevent replication fork collapse. *Genes & development* 27, 1610–1623.
- [56] Bansbach, C. E., Betous, R., Lovejoy, C. A., Glick, G. G., and Cortez, D. (2009) The annealing helicase SMARCAL1 maintains genome integrity at stalled replication forks. *Genes & development* 23, 2405–2414.
- [57] Sonoda, E., Sasaki, M. S., Morrison, C., Yamaguchi-Iwai, Y., Takata, M., and Takeda, S. (1999) Sister chromatid exchanges are mediated by homologous recombination in vertebrate cells. *Molecular and cellular biology* 19, 5166–5169.
- [58] KUO, L. J. and YANG, L.-X. (2008) γ -H2AX-a novel biomarker for DNA double-strand breaks. *In Vivo* 22, 305–309.
- [59] Betous, R., Mason, A. C., Rambo, R. P., Bansbach, C. E., Badu-Nkansah, A., Sirbu, B. M., Eichman, B. F., and Cortez, D. (2012) SMARCAL1 catalyzes fork regression and Holliday junction migration to maintain genome stability during DNA replication. *Genes & development* 26, 151–162.
- [60] Nongkhlaw, M., Dutta, P., Hockensmith, J. W., Komath, S. S., and Muthuswami, R. (2009) Elucidating the mechanism of DNA-dependent ATP hydrolysis mediated by DNA-dependent ATPase A, a member of the SWI2/SNF2 protein family. *Nucleic acids research* 37, 3332–3341.
- [61] Nongkhlaw, M., Gupta, M., Komath, S. S., and Muthuswami, R. (2012) Motifs Q and I are required for ATP hydrolysis but not for ATP binding in SWI2/SNF2 proteins. *Biochemistry* 51, 3711–3722.
- [62] Oakley, G. G. and Patrick, S. M. (2010) Replication protein A: directing traffic at the intersection of replication and repair. *Frontiers in bioscience : a journal and virtual library* 15, 883–900.
- [63] Fanning, E. (2006) A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic acids research* 34, 4126–4137.
- [64] Ciccia, A., Bredemeyer, A. L., Sowa, M. E., Terret, M.-E., Jallepalli, P. V., Harper, J. W., and Elledge, S. J. (2009) The SIOD disorder protein SMARCAL1 is an RPA-interacting protein involved in replication fork restart. *Genes & development* 23, 2415–2425.
- [65] Yusufzai, T., Kong, X., Yokomori, K., and Kadonaga, J. T. (2009) The annealing helicase HARP is recruited to DNA repair sites via an interaction with RPA. *Genes & development* 23, 2400–2404.

- [66] Yuan, J., Ghosal, G., and Chen, J. (2009) The annealing helicase HARP protects stalled replication forks. *Genes & development* 23, 2394–2399.
- [67] Postow, L., Woo, E. M., Chait, B. T., and Funabiki, H. (2009) Identification of SMARCAL1 as a component of the DNA damage response. *The Journal of biological chemistry* 284, 35951–35961.
- [68] Hockensmith, J. W., Wahl, A. F., Kowalski, S., and Bambara, R. A. (1986) Purification of a calf thymus DNA-dependent adenosinetriphosphatase that prefers a primer-template junction effector. *Biochemistry* 25, 7812–7821.
- [69] Yusufzai, T. and Kadonaga, J. T. (2008) HARP is an ATP-driven annealing helicase. *Science (New York, N.Y.)* 322, 748–750.
- [70] Bryksin, A. V. and Matsumura, I. (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques* 48, 463–465.
- [71] Kiianitsa, K., Solinger, J. A., and Heyer, W.-D. (2003) NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Analytical Biochemistry* 321, 266–271.
- [72] Fernández, F. J. and Vega, M. C. (2013) Technologies to keep an eye on: alternative hosts for protein production in structural biology. *Current Opinion in Structural Biology* 23, 365–373.
- [73] Rai, M. and Padh, H. (2001) Expression systems for production of heterologous proteins. *CURRENT SCIENCE-BANGALORE-* 80, 1121–1128.
- [74] Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhe Paganon, S., Park, H.-w., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S. H., Rao, Z., Shi, Y., Terwilliger, T. C., Kim, C.-Y., Hung, L.-W., Waldo, G. S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J. L., Stevens, R. C., Lesley, S. A., Wilson, I. A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M. I., Eschenfeldt, W. H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S. K., Emtage, J. S., Sauder, J. M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S. C., Bonanno, J. B., Fiser, A., Swaminathan, S., Studier, F. W., Chance, M. R., Sali, A., Acton, T. B., Xiao, R., Zhao, L., Ma, L. C., Hunt, J. F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C. K., Wang, D., Wang, H., Jiang, M., Montelione, G. T., Stuart, D. I., Owens, R. J., Daenke, S., Schütz, A., Heinemann, U., Yokoyama, S., Büssow, K., and Gunsalus, K. C. (2008) Protein production and purification. *Nature Methods* 5, 135–146.

- [75] Davies, D. R. and Hol, W. G. J. (2004) The power of vanadate in crystallographic investigations of phosphoryl transfer enzymes. *FEBS letters* 577, 315–321.
- [76] Page, R. and Stevens, R. C. (2004) Crystallization data mining in structural genomics: using positive and negative results to optimize protein crystallization screens. *Methods* 34, 373–389.
- [77] Vera, L., Czarny, B., Georgiadis, D., Dive, V., and Stura, E. A. (2011) Practical Use of Glycerol in Protein Crystallization. *Crystal Growth & Design* 11, 2755–2762.
- [78] Sousa, R. (1995) Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. *Acta Crystallographica Section D Biological Crystallography* 51, 271–277.
- [79] Zhang, L., Fan, S., Liu, H., and Huang, C. (2012) Targeting SMARCAL1 as a novel strategy for cancer therapy. *Leukemia Research* 427, 232–235.
- [80] University of Virginia Patent Foundation. (2000) Mammalian DNA-dependent ATPase a polypeptides and fusions thereof (US Patent Office).
- [81] Muthuswami, R., Mesner, L. D., Wang, D., Hill, D. A., Imbalzano, A. N., and Hockensmith, J. W. (2000) Phosphoaminoglycosides Inhibit SWI2/SNF2 Family DNA-Dependent Molecular Motor Domains . *Biochemistry* 39, 4358–4365.
- [82] Dutta, P., Tanti, G. K., Sharma, S., Goswami, S. K., Komath, S. S., Mayo, M. W., Hockensmith, J. W., and Muthuswami, R. (2012) Global epigenetic changes induced by SWI2/SNF2 inhibitors characterize neomycin-resistant mammalian cells. *PloS one* 7, e49822.