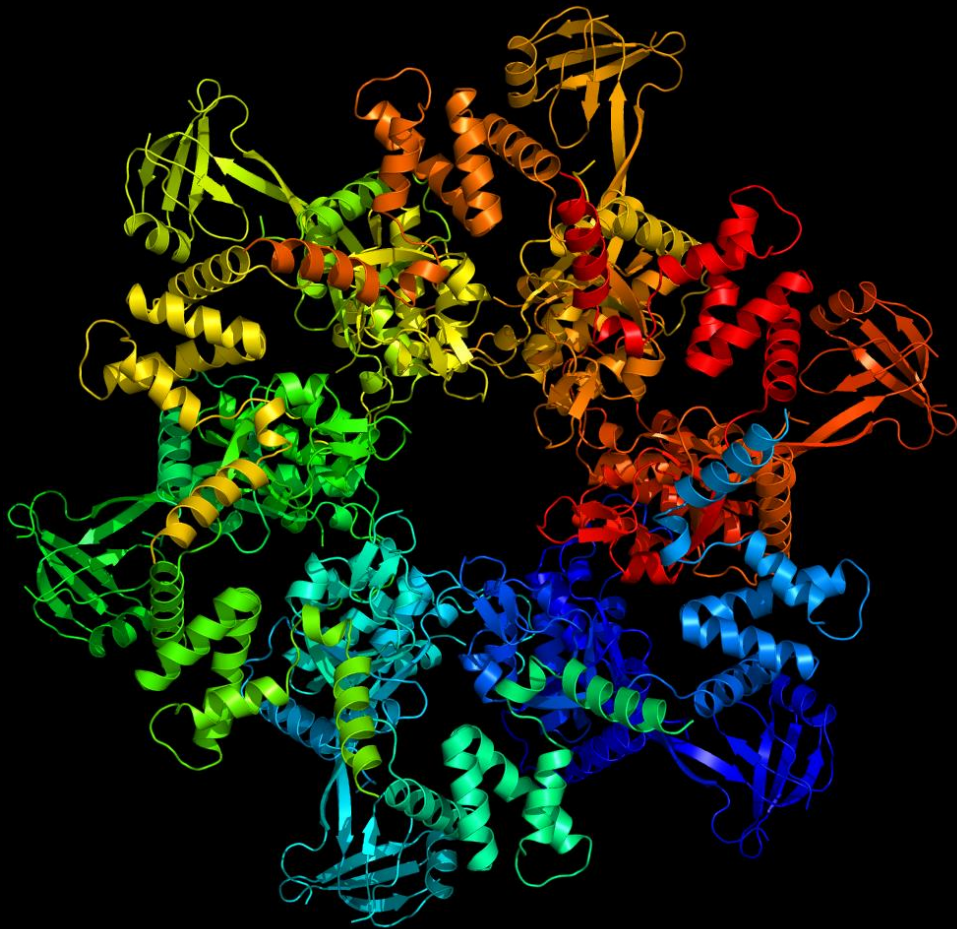


# Structural insights into the human multifunctional protein RuvBL2

Sara Teresa Neves da Silva



Dissertation presented to obtain the Ph.D degree in Structural Biochemistry

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,  
November, 2017



UNIVERSIDADE  
**NOVA**  
DE LISBOA



**Photo taken after the oral dissertation and discussion, on the 27<sup>th</sup> November 2017. From left to right: Sabine Gorynia, Édouard Bertrand, Pedro Matias (supervisor), Sara Silva, Tiago Bandeiras (co-supervisor), Sandra Macedo-Ribeiro, Professor Maria Arménia Carrondo (president of the jury) and Teresa Santos-Silva.**

# Structural insights into the human multifunctional protein RuvBL2

Sara Teresa Neves da Silva

Dissertation presented to obtain the Ph.D degree in Structural  
Biochemistry

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras, November, 2017





## Foreword

The work described in the present dissertation concerns the study of the human multi-functional protein RuvB-Like 2. The main focus of this work was the elucidation of the crystallographic structure of this potential drug target, complemented by lower resolution techniques, in order to assess different conformations and elucidate the possible mode of action. These structural studies constitute the third chapter of this dissertation. The second chapter is constituted by functional analyses, necessary in order to gain an understanding of the characteristics of the target protein, both at the functional level, but especially with the aim of improving stability for the structural analyses.

This dissertation is therefore divided in four chapters. The first chapter includes a general introduction to the RuvB-Like area, including published results from all areas of study of these proteins. The purpose of this introduction is to review and correlate the results produced by the several groups studying RuvB-Like proteins, putting them in the cellular context. The second chapter includes an introduction to DNA-binding proteins, followed by a description of the results on DNA-binding properties of *hsRuvBL1* and *hsRuvBL2*, and studies on the influence of tags on stability and oligomerization of *hsRuvBL2*. The third chapter focuses on structural analyses of *hsRuvBL2*, using X-ray crystallography, electron microscopy and small angle X-ray scattering. The final discussion aims to correlate the previously described results, and connect them with the state of the art, including data on the aggresome, produced with resource to protein produced during this work.

## Acknowledgments

The work achieved in this dissertation was only possible due to the contribution of many people, to whom I am indebted:

I deeply thank Pedro Matias, my supervisor, for the support and guidance throughout these years, whenever help was needed or a question arose. For the support, understanding and patience during the most challenging times, which was deeply felt and appreciated, and for believing in my capacities.

I deeply thank my co-supervisor, Tiago Bandejas, for the guidance and excellent teaching abilities. For the inspiration, provided during every scientific discussion, always with good mood and a joke, even during the (frequent) stressful times. For all the concern, patience and help, both visible and on the “backstage” during the most difficult times, which was always felt and appreciated.

Professor Maria Arménia, the Head of the Macromolecular Crystallography Unit, for always inspiring us to try new things and network with each other and with scientists outside the group, thus encouraging our scientific and personal development. For the concern and support throughout these years.

José Brito, for collecting the crystallographic data of *hsRuvBL2* and for the help provided in the determination of the structure, always with attention to detail. For always finding the time to clarify my doubts, even more than once, with patience, great teaching skills and humour.

All my colleagues in the Macromolecular Crystallography Unit and in the Structural Biology for Drug Discovery group, for contributing to a fun and

collaborative working environment, where I learned much about work and life. In particular to Filipe Rollo, Bernardo Caniço, Adriana Temporão and Diana Silva, who arrived at the group while I was writing, for their friendship and support during this difficult process.

João Carita, for the large-scale production of biomass and help with cell disruption whenever needed.

Ricardo Coelho, for help provided during the (many) problems with the crystallization robot.

Bruno Correia and Colin Mcvey, for help with the implementation of the electrophoretic mobility shift assays.

Rocío Arranz, Carlos Óscar Sorzano and Roberto Melero, for the support provided in the electron microscopy experiments. Rocío Arranz and Jaime Benito, and the other members of the group of Structure of Macromolecular Assemblies and the Biocomputing unit, for receiving me in their groups.

Mark Tully, James Douth and Robert Rambo, for support with the SAXS experiments.

Christine Ebel, for performing the analytical ultracentrifugation experiments and data analyses.

Phillipe Carpentier, for the support with the High Pressure cryocooling of crystals.

Hassan Berhali, for guidance during the controlled dehydration experiments at BM14.

Louise Bird, for the support with the high throughput cloning of c-Myc constructs.

Fundação para a Ciência e Tecnologia (FCT), Fundo Europeu de Desenvolvimento Regional (FEDER) (through COMPETE2020) and Instruct for funding.

A special thanks to my friends Diana Macedo, Pedro Silva and André Santos, the group formed in the beginning of our PhD classes from the geeks sitting in the front row, for the least productive, most fun working sessions.

Mariana, my best friend since we met when we were 3 years old, a big thank you for your support and for being so interested in my work and all things science. All my friends from outside the world of science, for their support and friendship.

A very special thank you to my family, my parents Belmira Teresa Pinguinhas and Manuel António Neves da Silva and my sister, Vera Teresa Neves da Silva, for your love, patience and support, which were a source of motivation.

## Thesis publications

The work described in this dissertation resulted in two publications:

Zaarur N, Xu X, Lestienne P, Meriin AB, McComb M, Costello CE, Newnam GP, Ganti R, Romanova NV, Shanmugasundaram M, **Silva STN**, Bandejas, TM, Matias PM, Lobachev KS, Lednev IK, Chernoff YO, Sherman MY; RuvbL1 and RuvbL2 enhance aggresome formation and disaggregate amyloid fibrils. EMBO J. 2015; 34: 2363–2382. doi:10.15252/embj

**Silva STN**, Brito JA, Arranz R, Sorzano CO, Ebel C, Douth J, Tully M, Rambo R, Carazo JM, Carrascosa J, Matias PM, Bandejas TM; Structure of human RuvB-Like 2 provides mechanism for coupling between ATP binding and mechanical action.  
*In preparation*

## Dissertation Abstract

RuvB-Like transcription factors, RuvBL1 and RuvBL2, function in cell cycle regulation and development. They have been attributed the functions of chaperone, transcription regulator and helicase, sometimes in an ATP-dependent fashion, but just how these functions are regulated in each protein is still a mystery, that many groups have been working to understand. There is already a vast, albeit scattered amount of knowledge gathered about RuvBL1 and RuvBL2 proteins. However, the functions of these important proteins still need to be placed into the context of the various signalling pathways of which they are a part. This may be a daunting task, since the functions attributed to these proteins seem to be used in various combinations depending on the complex in which they are included. To add further complexity, these proteins can work separately or in complex, and in a way that seems to be regulated either by their oligomeric state, binding partners and/or post-translational modifications. RuvBL1 and RuvBL2 have been found to be associated with the aetiology of a number of cancers and other diseases, such as heart hyperplasia and ciliopathies. As such, with this work, our group strived to provide a contribution to this area, a contribution we hope will be useful particularly in the health and pharmaceutical industry areas, since the structure of human RuvBL2 determined during this thesis is a widely recognized potential drug target.

The path to a better understanding of any protein is through the observation of its structure, and that was the main purpose of this work: to obtain the atomic-resolution structure of the missing player in the human RuvB-Like family, RuvB-Like 2 (*hsRuvBL2*). We present the atomic structure of human RuvBL2 with a level of completion that provides novel insight into its biology. The *hsRuvBL2* structure resolves the mobile domain II, which is responsible for protein-protein interactions and ATPase activity regulation. We demonstrate how ATP

binding may lead to domain II motion through interactions with the N-terminal loop and further show that inserted affinity tags affect *hsRuvBL2* oligomerization and stability in solution. A comparison with its homolog *hsRuvBL1* shows differences in surface charge distribution that may account for differences in regulation. Finally, single particle EM data reveals that single-stranded DNA can promote the oligomerization of monomeric *hsRuvBL2*. The structural information that was gathered in this work shows that *hsRuvBL2* presents specific motifs at the surface level that allow *hsRuvBL1* and *hsRuvBL2* to be distinguished in the context of a signalling pathway. Furthermore, this was the first **apo structure** obtained for a eukaryotic RuvBL2. Comparison of our structure with the structures of ADP-bound truncated *hsRuvBL2*, and fungal ADP-bound *ctRuvBL2* (*Chaetomium thermophilum*), suggests a putative mode of action for the coupling of ATP binding with a mechanical movement that could be the basis of ATP-dependent activities. Since crystallography provides a detailed, but somewhat static structural model, we used electron microscopy to gain some insight into the plasticity of the *hsRuvBL2* complex. Curious, although of still unknown biological relevance, is the observation that *hsRuvBL2* is also able to form heptamers.

We combined the obtained structural knowledge with DNA binding assays, since the DNA binding mode of RuvB-Like proteins is still a major focus of discussion, and largely unknown. Our results suggest that *hsRuvBL2* must be monomeric at the onset of activity in order to bind DNA, since we could not observe a DNA electrophoretic shift in the presence of hexameric *hsRuvBL2*. While the low concentrations necessary to obtain monomeric *hsRuvBL2* precluded the observation of DNA binding by electrophoretic mobility shift assay, electron microscopy allowed the observation that monomeric *hsRuvBL2* associates into a ring, possibly a hexamer, in the presence of single-stranded DNA. EMSA assays further showed that *hsRuvBL1* monomers are also able to bind ssDNA, and seem

to be able to confer to hexameric *hsRuvBL2* the ability to bind ssDNA, or alternatively to form bridges between DNA and hexameric *hsRuvBL2*.

An analysis of the oligomerisation states of *hsRuvBL2* with and without tags, shows that, as observed in yeast, tag presence and position interfere with the oligomerisation state. We also observed that they interfere with concentration-dependent oligomerisation state, such that C-terminally tagged *hsRuvBL2* is always a hexamer, while the N-terminally tagged form varies its oligomerisation state depending on protein concentration.

## **Resumo da Dissertação**

Os factores de transcrição da família RuvB-Like, RuvBL1 e RuvBL2, funcionam na regulação do ciclo celular e do desenvolvimento. A estas proteínas estão atribuídas as funções de chaperone, reguladores de transcrição e helicase, por vezes de forma dependente de ATP. No entanto, a forma como estas diferentes funções são reguladas é ainda um mistério, que diversos grupos tentam elucidar. Existe já um espólio vasto de informação sobre a RuvBL1 e RuvBL2. No entanto, as funções destas proteínas ainda têm de ser colocadas no contexto das diversas vias de sinalização das quais fazem parte. Esta pode ser uma tarefa árdua, já que as funções atribuídas a estas proteínas parecem ser usadas em diversas combinações, dependendo do complexo onde estão incluídas. A dificultar o processo, estas proteínas podem funcionar separadamente ou em complexo, e de uma forma que parece ser regulada pelo seu estado oligomérico, parceiros de ligação e/ou por modificações pós-translacionais. Tanto a RuvBL1 como a RuvBL2 estão associadas com a etiologia de vários cancros e outras doenças, como hiperplasia cardíaca e ciliopatias. Como tal, o nosso grupo pretendeu produzir uma pequena contribuição para a área de estudo destas proteínas, uma contribuição que esperamos que seja útil, em particular nas áreas de saúde e indústria, já que a estrutura da RuvBL2



humana determinada durante esta tese é um potencial alvo de compostos farmacológicos amplamente reconhecido.

O caminho para a compreensão do funcionamento de qualquer proteína passa pela análise da sua estrutura, e foi este o principal objectivo deste trabalho: a obtenção da estrutura de resolução atómica remanescente da família das RuvBLs, a RuvBL2 humana (*hsRuvBL2*). Aqui apresentamos a estrutura atómica da RuvBL2 humana suficientemente completa para providenciar nova informação quanto à sua biologia. A estrutura da *hsRuvBL2* resolve o domínio II, um domínio móvel responsável por interacções proteína-proteína e regulação da actividade ATPase. Neste trabalho tentamos demonstrar de que forma a ligação de ATP pode levar à movimentação do domínio II através de interacções com o *loop* N-terminal, e demonstramos ainda que as *tags* de afinidade afectam a oligomerização e estabilidade da *hsRuvBL2* em solução. Uma comparação com a homóloga *hsRuvBL1* demonstra diferenças na distribuição de cargas superficiais que podem justificar algumas diferenças de regulação entre as duas proteínas. Finalmente, dados obtidos por microscopia electrónica revelam que a presença de ADN de cadeia simples promove a oligomerização de *hsRuvBL2* monomérica. A informação estrutural recolhida neste trabalho demonstra que a *hsRuvBL2* apresenta motivos específicos na sua superfície que permitem que seja distinguida da *hsRuvBL1* no contexto de uma via de sinalização. Adicionalmente, esta foi a primeira estrutura *apo* obtida para uma RuvBL2 de um organismo eucariota. Uma comparação da estrutura aqui obtida com a estrutura da *hsRuvBL2* truncada ligada a ADP, e com a estrutura da RuvBL2 de *Chaetomium thermophilum*, também ligada a ADP, sugere um potencial modo de acção que conecta a entrada de ATP com um movimento mecânico do domínio II que poderá ser a base de actividades dependentes de ATP. Visto que a cristalografia de raios-X providencia um modelo estrutural detalhado, porém algo estático, recorreremos à microscopia electrónica para obter informação acerca da plasticidade do complexo *hsRuvBL2*. Observámos

que esta proteína é capaz de formar heptâmeros, um facto curioso mas de relevância biológica ainda indeterminada.

Combinámos o conhecimento estrutural obtido com ensaios de ligação ao ADN, já que o modo de ligação ao ADN das RuvBLs é ainda um foco de discussão, e pouco caracterizado. Os nossos resultados demonstram que a *hsRuvBL2* tem de estar no estado monomérico no início da reacção, por forma a estabelecer ligação com o ADN, visto que não observámos variação na distância percorrida pelo ADN em gel na presença de *hsRuvBL2* hexamérica, quando comparada com a distância percorrida pelo ADN isoladamente. Tendo em conta que as concentrações às quais a *hsRuvBL2* é monomérica são demasiado baixas para permitir a observação de ligação por ensaios de variação de mobilidade electroforética em gel (EMSA), utilizámos microscopia electrónica para observar que a *hsRuvBL2* monomérica se reorganiza em anéis na presença de ADN de cadeia simples. Os ensaios por EMSA mostraram ainda que **monómeros de *hsRuvBL1*** são também capazes de se ligar ao ADN de cadeia simples, e parecem conferir à *hsRuvBL2* hexamérica a capacidade de se ligar ao ADN, ou alternativamente podem actuar como pontes de ligação entre o ADN e o hexâmero de *hsRuvBL2*.

Uma análise das formas oligoméricas de *hsRuvBL2* com e sem *tags* de afinidade demonstrou que a presença e posição das *tags* interferem com o estado de oligomerização. Observámos também que há uma influência na variação de estado oligomérico conforme a concentração de proteína: a *hsRuvBL2* com *tag* C-terminal é sempre hexamérica, enquanto que a *hsRuvBL2* com *tag* na extremidade N-terminal adquire diferentes estados oligomérico conforme a concentração a que se encontra.

# Table of Contents

## Chapter 1 - Introduction

<b>1.1 A simplified view of the eukaryotic cell.....</b>	<b>4</b>
The crowded cell	4
Eukaryotic cell compartmentalization	5
DNA packing and organization	8
1.1.3 AAA+ PROTEINS	11
1.1.4 HELICASES	16
1.1.5 CHAPERONES	20
1.1.6 RUVB AND RUVB-LIKE PROTEINS	25
1.1.7 ROLES OF RUVBLs IN THE CELL AND CELL CYCLE	31
<b>1.2 RuvBL1 and RuvBL2 in higher-order complexes .....</b>	<b>36</b>
<b>1.3 RuvBLs in development and disease.....</b>	<b>47</b>
1.3.1 DEVELOPMENT	47
1.3.2 CANCER	49
1.3.3 HYPOXIA	52
REFERENCES	55

## Chapter 2 - Oligomerization and DNA-binding of RuvBL2; Preliminary study of binding partner c-Myc

INTRODUCTION TO ssDNA-BINDING PROTEINS	69
<b>2.1 hsRuvBL2 stability, oligomerization and DNA binding.....</b>	<b>72</b>
2.1.1 METHODS	72
Protein expression and purification	72
Differential scanning fluorimetry	73
Size exclusion chromatography coupled to multi-angle laser light scattering	75

Analytical ultracentrifugation	76
Small angle X-ray scattering	77
Electrophoretic mobility shift assays (EMSA)	78
Analysis of hsRuvBL2 binding to DNA by negative staining EM	79
2.1.2 RESULTS AND DISCUSSION	80
Purification tags and nucleotides affect hsRuvBL2 stability and oligomerization state	80
Stability of hsRuvBL2 oligomers varies with concentration	86
Insights into hsRuvBL1 and hsRuvBL2 binding to DNA	91

## **2.2 Preliminary study of RuvBL-interacting protein c-Myc..... 96**

2.2.1 METHODS	97
Construct design	97
Cloning procedures	99
Protein expression and purification	100
2.2.2 RESULTS AND DISCUSSION	102
REFERENCES	104

## **Chapter 3 - Structure of human RuvBL2**

### **3.1 Current structural knowledge on RuvB-Like proteins..... 108**

### **3.2 Structure of RuvBL2 at 2.8 Å by X-ray crystallography ..... 119**

3.2.1 METHODS	120
Protein production	120
Protein crystallization and crystal optimization	121
Data collection	127
Structure solution and refinement	128
3.2.2 RESULTS AND DISCUSSION	131
Purification of RuvBL2	131
Crystallization of RuvB-Like 2	132
Initial attempts to solve the structure from low resolution data	135
Data collection of crystals diffracting to 2.8 Ångstrom	138
Atomic structure of full-length RuvB-Like 2	139
Structural basis for coupling ATP binding to mechanical action	145

<b>3.3 Structure of RuvBL2 by electron microscopy.....</b>	<b>152</b>
3.3.1 METHODS	153
Protein production	153
Data collection - negative staining EM	155
Data processing and refinement of negative staining data	155
3.3.2 RESULTS AND DISCUSSION	156
Low resolution structure of RuvBL2 by negative staining EM	156
REFERENCES	160

***Chapter 4 - Discussion***

<b>4.1 Discussion.....</b>	<b>167</b>
<b>4.2 Concluding remarks and future perspectives.....</b>	<b>179</b>
REFERENCES	183

<b>Supplementary Information.....</b>	<b>185</b>
---------------------------------------	------------



# *Part 1*

---

# Chapter 1

---

## INTRODUCTION



<b>1.1 A SIMPLIFIED VIEW OF THE EUKARYOTIC CELL .....</b>	<b>4</b>
The crowded cell	4
Eukaryotic cell compartmentalization	5
DNA packing and organization	8
1.1.3 AAA <sup>+</sup> PROTEINS	11
1.1.4 HELICASES	16
1.1.5 CHAPERONES	20
1.1.6 RUVB AND RUVB-LIKE PROTEINS	25
1.1.7 ROLES OF RUVBLs IN THE CELL AND CELL CYCLE	31
<b>1.2 RUVBL1 AND RUVBL2 IN HIGHER-ORDER COMPLEXES.....</b>	<b>36</b>
<b>1.3 RUVBLs IN DEVELOPMENT AND DISEASE .....</b>	<b>47</b>
1.3.1 DEVELOPMENT	48
1.3.2 CANCER	49
1.3.3 HYPOXIA	52
<b>References .....</b>	<b>55</b>

This introduction aims to integrate relevant information presently available on RuvB-Like proteins with knowledge of the eukaryotic cell. The elucidation of the human RuvBL2 structure, determined in this work, will contribute to a better understanding of this protein family, which contains proteins that function both separately and in complex. Since they are deeply involved in such fundamental processes as development, chromatin remodelling and disease, including multiple cancer types, it is fundamental to understand their roles in the cell. In that way, one can follow the next step in regulating their activity with a better defined strategy in mind: the design of chemical compounds that will affect a particular activity, without interfering with their parallel workings inside the cell, a risk when targeting such a ubiquitous protein. Fortunately, this protein comprises discrete areas, targetable by multiple regulators, which may provide some clues as to where to begin, as long as it is possible to correlate the effect of targeting such areas to the outcome in the cell, and at the organism level.

## **1.1 A SIMPLIFIED VIEW OF THE EUKARYOTIC CELL**

### *The crowded cell*

The eukaryotic cell environment is based on an aqueous moiety, with a high concentration of macromolecules (50-400 mg/mL), leading to macromolecular crowding effects, such as volume exclusion of reactants and reduced diffusion coefficients of macromolecules<sup>1</sup>. Such an overcrowded environment precludes random movements and chance collisions as the basis for the protein-protein contacts needed for catalytic activities to occur. In fact, overcrowding promotes protein association and self-association of monomers<sup>2,3</sup>. Therefore, the formation of multi-subunit assemblies is facilitated by the

environment, and improves catalytic efficiency, through the formation of multi-subunit complexes composed of the enzymes that are part of a catalytic pathway. Many different proteins have been identified so far in such supramolecular assemblies, with varying degrees of binding affinities and complexity. RuvBL1 and RuvBL2 are two of those proteins that take part in various assemblies, either individually<sup>4-7</sup> or as a heteromeric complex<sup>8-13</sup>.

### *Eukaryotic cell compartmentalization*

The function of a protein is strongly dictated by its sub-cellular location and by interactions with binding partners, which may also be correlated with the deposition of post-translational modifications (epigenetic markers). Additionally, the cell proteome distribution varies during the cell cycle, and a few genes are only transcribed for specific cell types. Furthermore, about one-third of its proteins are found in multiple organelles, which suggests roles in diverse and multiple pathways, such as is the case for RuvB-Like proteins. This wide distribution justified a brief description of the eukaryotic cell, for visual aid (**Fig. 1.1**).

The eukaryotic cell is highly compartmentalized, which confers functional segregation. It contains a true **nucleus**, delimited by a nuclear membrane. The inner membrane serves as an anchoring site for chromatin, and the outer membrane is continuous with the endoplasmic reticulum. Nuclear pores distributed across the membrane allow the exchange of macromolecules between the nucleus and the cytoplasm. The nucleoplasm supports the chromosomes and the nucleoli, and it is the place where transcription and replication events take place. The nucleoli are structures not bound by membranes, where ribosomes are synthesized, processed and assembled, a complex process controlled by nucleolar

substructures such as the fibrillar centre. The nucleoli also comprise proteins involved in stress responses and cell cycle regulation.

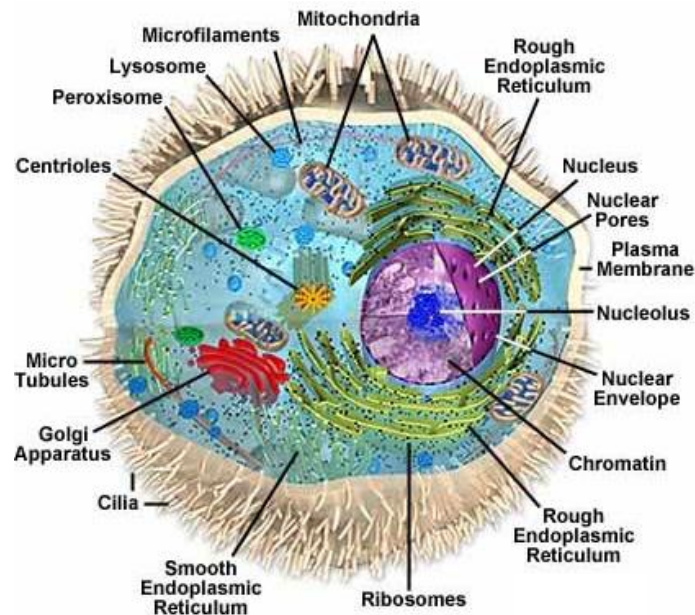


Figure 1.1 – Representation of the eukaryotic cell. Highlighted are the most relevant organelles and cellular structures. Depicted from <http://www.carolguze.com/text/102-7-eukaryoticcells.shtml>.

The **cytoskeleton** is composed of three types of fibres. The **microtubules** are the stiffest of the three, and are responsible for spindle formation during mitosis. **Actin filaments** are polarized, dynamic filaments with the ability to quickly assemble and disassemble, making them essential in the response to outside stimuli and in cell motility. As such, they are in direct contact with membrane-bound proteins and focal adhesions. The latter, also known as force-sensing and transducing complexes, are multi-subunit assemblies constructed by layers connecting actin filaments to the extracellular matrix. Functions of the different constituents of the focal adhesions start from receptor-matrix binding, linkage to actin and force transduction, intracellular signal transduction and actin

polymerization and regulation. Two examples of components of focal adhesions are **PI3Ks** (phosphatidylinositol kinases) and **Arp2/3** (actin related proteins 2 and 3). The **Arp2/3 complex** binds to the pointed ends of actin filaments, nucleating them, and facilitating the formation of actin-based protrusions, essential for motility. Finally, the **intermediate filaments** provide support to the cell and resistance to mechanical stress, and anchor the chromosomes to the inner part of the nuclear membrane, the lamina.

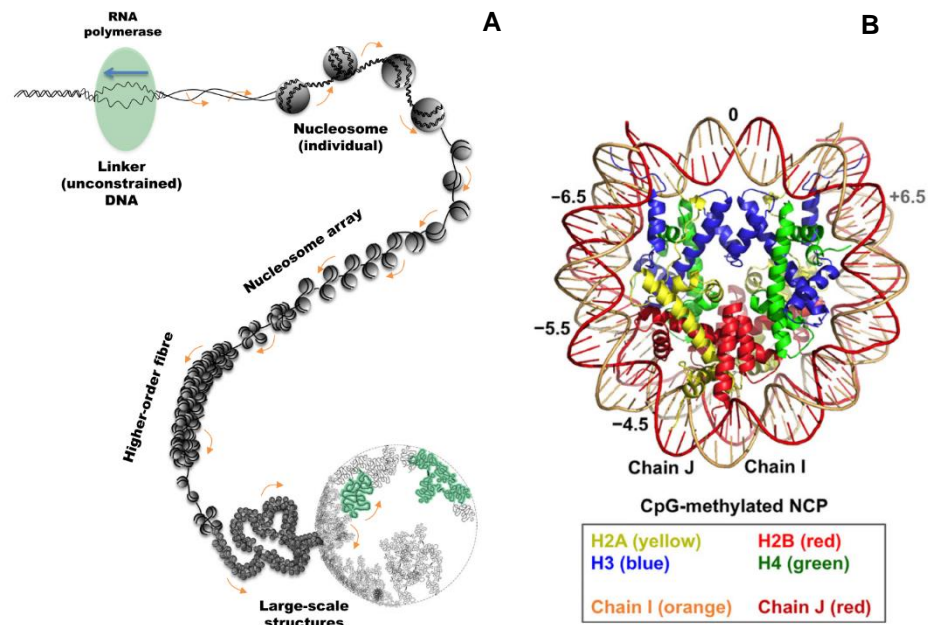
The **cytoplasm** comprises the cytosol, a semifluid liquid that contains mostly proteins, amino acids and ions, plus all non-nuclear organelles. The **endoplasmic reticulum (ER)** is a membrane-bound organelle, and the first step in the secretory pathway, together with the **Golgi apparatus (GA)**<sup>14</sup>. **Vesicles** are small, enclosed organelles that are formed to create a dedicated environment to perform a variety of specialized functions, such as specialized metabolic reactions, transport, degradation of biomolecules and secretion. The **centrosome** is an organelle responsible for microtubule organization in the cell. It is composed of two centrioles, embedded in the pericentriolar matrix, and it is the microtubule organising centre (MTOC) of the cell during G1 and G2 phases of the cell cycle (see section 1.1.1). During mitosis the centrosome will mature into the spindle pole, and may form cilia by elongating microtubules<sup>15</sup>. The **mitochondria** are involved in the production of energy in the form of ATP, are involved in cell death, signalling and differentiation, and are the only organelles to possess their own genome (in animals). Interestingly, it was recently shown that **RuvBL2** possesses a mitochondrial targeting signal in one of its isoforms (isoform 3, Accession: NP\_001308120.1), and indeed RuvBL2 and RuvBL1 were identified in mitochondria-enriched fractions. It has been suggested that, either by alternative splicing or alternative translation initiation, a form of RuvBL2 could be produced starting only at M46, thus activating this mitochondrial targeting signal. The

function of mitochondrial RuvBL2 isoform 3 is not yet clear, but it was shown to interact with the mitochondrial DNA polymerase gamma<sup>16</sup>.

Each cell is delimited by the **plasma membrane**, which provides a selective barrier with the outside environment. Proteins located in the membrane are responsible for cell to cell contacts, signal transduction, transport and adhesion to other cells or the extracellular matrix. Some proteins possess a signal peptide that promotes active transport out of the cell, and are thus **secreted** via the secretory pathway.

### *DNA packing and organization*

Human chromosome sizes range between 50,000,000 and 300,000,000 bp.



**Figure 1.2** – A) Structural organisation of genomic DNA. In order for the processing machinery to access the DNA, a section of the genome has to be unconstrained (uncoiled), for which the surrounding will become supercoiled, up to a limit. From<sup>17</sup>. B) Crystallographic structure of the nucleosome. Depicted from<sup>18</sup>.

In each cell, the DNA fibre is more than 2 meters long and can only be accommodated in the cell nucleus by iterative folding (**Fig. 1.2A**) into a condensed structure, the chromosome. Linear arrays of nucleosomes are coiled into a 30 nm fibre<sup>19</sup>, linked and stabilised by linker histone H1. The basic folding unit - the nucleosome - is composed of 146 base pairs supercoiled in two turns around a histone octamer, composed of two H2A-H2B heterodimers and a H3-H4 tetramer (**Fig. 1.2B**), forming the fundamental unit of chromatin<sup>8,18,20</sup>.

Histones not only help in DNA packing, but can also be interchanged between different forms that bind more or less tightly to DNA, aiding in transcription regulation, together with DNA sequence and remodelling machinery. Histone post-translational modifications (PTMs) mark locations of double strand breaks and replication fork stalling, for further processing by the cellular repair machinery. After a DNA strand break or replication fork stall, canonical forms of **H2A** are replaced at the site of the break by **H2AX**, which contains a conserved SQ(E,D)(I,L,F,Y) motif at the carboxyl terminus, a common PI3K kinase (PIKK) phosphorylation motif. The serine residue is phosphorylated immediately after a double strand break, by PIKKs DNA-PKcs (DNA-dependent protein kinase catalytic subunit) and ATM (ataxia-telangiectasia mutated), forming a number of  $\gamma$ -**H2AX** foci that approximates the number of DSBs. H2AX phosphorylation triggers a series of molecular events that activate DNA repair response, and proteins such as BRCA1, Nbs1, Rad50 and Rad51 have been found to form foci that co-localize with  $\gamma$ -H2AX foci<sup>21,22</sup>. The consequence is an extensive change in local chromatin composition, including acetylation, ubiquitination, potential **H2A.Z** deposition and eventually histone eviction, concomitant with end resection<sup>8</sup>. For example, the RuvBL-containing **INO80** chromatin remodelling complex is recruited by  $\gamma$ -H2AX to sites of replication fork stalling as cells enter the S-phase, and promotes efficient progression of replication by stabilising stalled replisomes. INO80 also functions in collaboration with **SRCAP** to replace

$\gamma$ -H2AX with the H2AZ variant in case of DSBs<sup>23</sup>. This facilitates nucleosome eviction prior to DNA strand repair, since H2AZ nucleosome structure and biochemical studies suggest that its association with DNA may be weaker than that of H2AX<sup>22</sup>. In yeast, **Htz1** (homolog of human H2AZ) promotes deposition of Gcn5 histone acetyltransferase, which acetylates histone H3 in response to UV stress, in Htz1-containing nucleosomes. Htz1 subsequently stabilises the interaction of Rad14 with damaged DNA, promoting nucleotide excision repair (NER) after UV irradiation<sup>24</sup>.

The mobility of a chromosomal locus varies with cell cycle stage and with chromatin status. DNA mobility, particularly at sites of double strand breaks (DSB), is not simply caused by Brownian motion, but is a controlled and regulated cellular process, promoted by enzymes. With about 53 different types of Snf2-type **ATPases** (enzymes that catalyse the hydrolysis of ATP to ADP with concomitant DNA/RNA remodelling) in humans, it is expected that each one will have different, even if maybe partially redundant, impacts on chromatin composition, mobility and structure which may, at least in the case of INO80, impact on chromatin dynamics. However, the activity of the DNA damage response (DDR) system and chromatin remodellers is not restricted to damaged sites, as checkpoint activation leads to an increase in chromatin mobility that is dependent on the INO80-dependent ATR (ataxia telangiectasia Rad3-related) kinase Mec1 and its downstream target kinase Rad53<sup>8</sup>. In case of difficult-to-repair DSBs and collapsed replication forks, DNA is relocalized from the nucleoplasm to the nuclear pore, an event that is dependent on the presence of histone H2A.Z, which implies an involvement of the RuvBL-containing **SWR1** complex. In fact, depending on the DSB cause, chromatin re-localization may occur towards the nuclear pore or the membrane protein Mps3. This is dependent on the cell cycle stage, such that association of persistent DSBs with Mps3 (which suppresses recombination) requires Rad51 and end resection, which occurs only



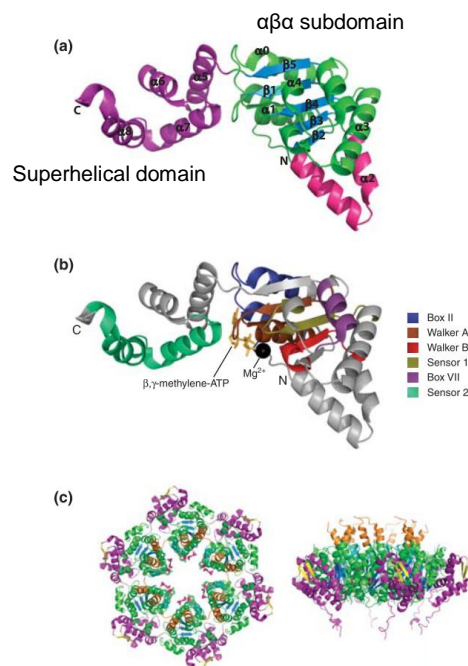
in the S phase, while DSB association with the nuclear pore (which favours non-canonical recombination) occurs also in G1 phase<sup>8</sup>.

### 1.1.3 AAA<sup>+</sup> PROTEINS

ATPases associated with various cellular activities (**AAA<sup>+</sup> ATPases**) are present in all domains of life, and eukaryotic genomes typically encode 50 – 100 different proteins belonging to this family<sup>25</sup>. The AAA<sup>+</sup> ATPases fall within the second major group of the P-loop NTPases, the **ASCE** (additional strand catalytic E, for the characteristic catalytic glutamate residue within the Walker B motif). Proteins of this superfamily use the energy obtained from ATP binding and hydrolysis to perform a variety of biochemical activities. Members of this superfamily include helicases, chaperones and regulatory components of proteolytic machines, which usually assemble into hexameric rings or helical structures (although pentameric, heptameric and octameric examples also exist)<sup>25</sup>. As such, this superfamily comprises members involved in the normal maintenance of the cell, but also in stress response. The defining feature of the numerous members of this group is an ATP-binding module, which associates into arrays to originate a functional complex. ATP binding and hydrolysis events at the interface of neighbouring subunits lead to conformational changes within the AAA<sup>+</sup> assembly that are responsible for the enzymatic activity of the complex<sup>26,27</sup>. The interface between protomers thus undergoes adjustments according to the bound nucleotide state, while maintaining integrity of the complex. Interestingly, it was observed that of all the AAA<sup>+</sup> ATPases analysed, RuvBL1 protomers had the largest surface contact area. In fact, the ATP binding pocket in RuvBL1 is so tightly occluded that it seemed to render impossible the exchange of the bound ADP, which could justify the observed low ATPase and

lack of helicase activities. It was suggested that a co-factor may be necessary to enable these activities, similarly to other members of this family<sup>25,28</sup>.

This large superfamily is characterized by the heterogeneity of its members' structures and functions. However, there are some common structural features: all possess a core  $\alpha\beta\alpha$  nucleotide-binding domain and an adjacent  $\alpha$ -helical domain, composed of two helical hairpins associated to form a left-handed superhelical structure (**Fig. 1.3**). The latter is a smaller domain, poorly conserved at the primary sequence level, but highly conserved structurally, such that it is a hallmark of the AAA+ ATPase superfamily. In these multimeric complexes, the ATP-binding site is located at the interface between neighbouring subunits: one protomer provides the main nucleotide binding pocket, composed of Walker A,



**Figure 1.3**– General structural features of the AAA+ ATPase core. A) The ATPase core is subdivided into an  $\alpha\beta\alpha$  subdomain and a superhelical subdomain. B) Distribution of the canonical AAA+ motifs. C) Top and side views of the HsIU from *Haemophilus influenzae*. The superhelical “lid” domain (purple), locates to the outside of the ring structure. Additional helices (orange) form the I-domain, characteristic of HsIU-family proteins. From<sup>29</sup>.

Walker B, sensor I and sensor II motifs; the adjacent protomer contributes a conserved arginine (box VII, Second Region of Homology – SRH – or SRC motif)<sup>30</sup>. These conserved motifs all map to the AAA+ core  $\alpha\beta\alpha$  nucleotide fold, except for **sensor II**, which is located in the helical subdomain, known as the AAA+ “lid”. An arginine “finger”, extending from one monomer to the adjacent one, completes the ATP binding pocket, and polarises the  $\gamma$  phosphate to facilitate hydrolysis. This important residue has also been shown to be responsible for oligomerization of hexameric ATPases, since its mutation abolished oligomerization, as well as ATP binding and hydrolysis and DNA translocation. Addition of wild-type monomers to the mutated ones resulted in dimer formation, since only part of the population was able to provide this residue to promote the interaction<sup>31,32</sup>. The **sensor I** motif is typically an Asn, but can also be another polar residue, such as Ser, Thr or Asp; this motif participates in hydrolysis by coordinating the attacking water nucleophile, together with the Walker B residues. The **Walker A** motif has a consensus sequence GXXXXGK[T/S] (X represents any amino acid), in which the Lysine present on the P-loop (phosphate-binding loop) contacts directly the  $\beta$  and  $\gamma$  phosphates of ATP, making it critical for ATP binding. The **Walker B** motif consensus sequence is hhhhDE (h represents a hydrophobic amino acid). In this motif, the glutamate polarises a water molecule for in-line attack of the  $\gamma$  phosphate during ATP hydrolysis, and its mutation to glutamine or alanine abolishes this activity<sup>25,31</sup>. These mutants can nevertheless be useful in order to obtain a “trapped” structure that may allow the observation of transient conformational states<sup>25,27,33,34</sup>. This glutamate residue has a pivotal role in translation of nucleotide state to enzyme activation state. Indeed, analysis of several ADP and ATP bound AAA+ ATPases shows that two major conformational changes occur between the ATP and ADP complexes. The first major movement occurs in the glutamate sidechain itself, which rotates by approximately 100° in the ATP complex relative to its position in

the ADP complex, and forms a hydrogen bond with an asparagine residue in an adjacent strand. This “glutamate switch” thus traps the ATP-bound complex in a higher energy and inactive form, in contrast to the ADP-bound “active” form. The second major movement occurs in an area farther from the active site, and may involve loops that interact directly with the Glu-Asn pair, providing communication between the ATPase and ligand-binding sites. The known exceptions to date are the HslUV/Clp family proteases and the RuvB-Like helicases, in which the glutamate contacts a conserved threonine or serine residue instead of the Asn residue. This switch pair formation requirement upon ATP binding is suggested to be a way of repressing ATPase activity until there is ligand binding (a mechanism which is reversed in the cases where the ligand represses instead of activating)<sup>31</sup>.

Although AAA+ ATPases require ATP to perform activities that other enzymes can perform without, their ATP turnover is frequently poor in comparison with other ATPases such as hexokinase or even DNA helicases (turnover numbers are typically 0.1 to a few per second for AAA+ enzymes, versus several hundred per second for the “simpler” metabolic ATPases). Furthermore, the ATPase activity is often positively or negatively regulated by a ligand, and sometimes the same ligand can have opposite effects on different proteins. For instance, DNA binding leads to an increase in the ATPase activity of Replication Factor C, and yet represses ATPase activity of the Origin Recognition Complex. ATP hydrolysis suppression is necessary in complex enzymatic systems, with complex assembly pathways and multiple intermediate steps. ATPase activity is enabled at the point when ATP turnover is necessary for completion of the reaction or recycling of components. In this way, ATP binds during complex assembly, but its hydrolysis is suppressed until the complex is fully assembled<sup>31</sup>.

Sequence and structure comparisons show that AAA+ ATPases went through significant divergent events both before and after the emergence of the last common ancestor between bacteria, archaea and eukarya. Members of this superfamily can be subdivided into groups, clades and families, where the clade is defined by the distinct structural elements located within and around the AAA+ core. The **five groups** described are: Classic AAA ATPases (classical AAA clade or extended AAA group); helicases and clamp loaders (HEC); protease, chelatase, transcriptional activators and transport (PACTT); ExeA (type II secretory pathway ATPase); and signal transduction ATPases with numerous domains (STAND). RuvB-Like/Rvb proteins belong to the classical AAA group, which also includes NSF, CDC48, Pex, Bcs, proteasomal ATPase, katanin, Vps4, FtsH and Clp Domain 1 (d1) families. The last two families, as well as RuvBL, diverge from the others<sup>25,29</sup>. However, topological analysis of the ATP binding pocket revealed RuvBL1 to have closer structural similarities with the HEC group (DnaA, Orc, CDC6 clade)<sup>25</sup>.

The great functional diversity of AAA+ proteins is a reflex of their diversity of accessory domains and co-factors, the diversity of their oligomeric assemblies and heterogeneity of residues that define key nucleotide-binding features. These variations define the specificity of AAA+ assemblies for different substrates and the mechanisms responsible for coupling conformational changes within the assembly. Interestingly, adaptor proteins have been identified which interact with the ATPase module to regulate its activity by increasing its target specificity, without necessarily compromising other activities of the ATPase. These adaptor proteins are an elegant way for the cell to strictly regulate the activity of AAA+ proteins by regulation of the adaptor proteins themselves. An example of an adaptor protein in higher eukaryotes is p47, which interacts with the AAA+ ATPase p97, regulating the p97-mediated fusion of Golgi and transitional endoplasmic reticulum membranes, as well as the growth of the

nuclear envelope. Interestingly, many adaptor proteins identified so far, despite their sequence variability, use the N-terminus of the ATPase as the interaction platform for binding and regulating the ATPase activity<sup>26</sup>.

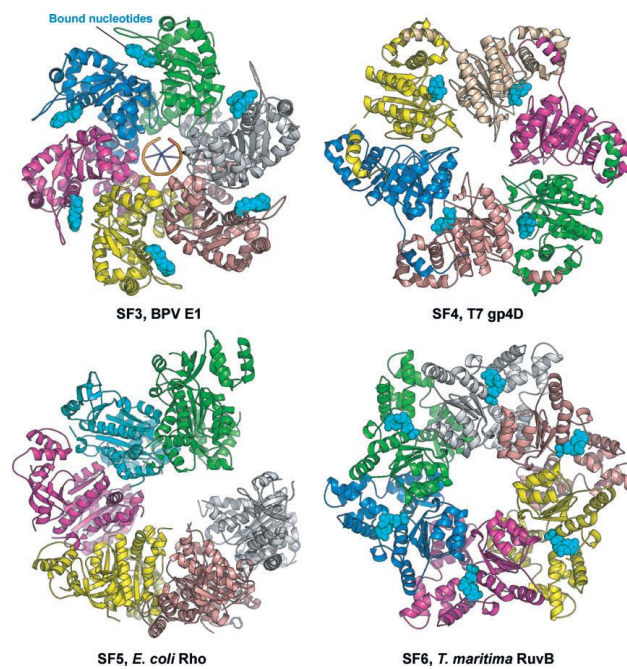
### 1.1.4 HELICASES

Nucleic acid helicases are molecular motor proteins that couple the chemical hydrolysis of a triphosphate nucleotide with the unwinding of the energetically stable DNA or RNA duplexes. This process allows access to single stranded templates for proteins involved in transcription, replication, recombination and repair machineries. The mode of unwinding of each helicase depends on the structure and type of nucleic acid substrate. Helicases can thus be classified as DNA-DNA, DNA-RNA or RNA-RNA dependent NTPases. The interaction of helicases with each nucleic acid substrate may be regulated by other proteins that modulate their access to the substrate. Such regulatory mechanisms ensure that *in vivo* each helicase will be active only in the pathway and at the moment when they are required<sup>35</sup>. The first helicase was discovered in 1976 in *Escherichia coli*, and two years later another was found in an eukaryotic organism, the lily flower. Since then, helicases have been found to be ubiquitous to all living organisms and many viruses<sup>36,37</sup>.

Most helicases share a few common biochemical features, namely nucleic acid binding, NTP binding and nucleic acid-stimulated hydrolysis of NTP. Helicases do not necessarily unwind nucleic acids, and those that do may or may not translocate along the strand. Most commonly, helicases bind preferentially to a ssDNA strand in a sequence-independent way, and from there they start to unwind the DNA double helix unidirectionally. A few exceptions preferentially bind dsDNA, such as RecBCD, simian virus 40 (SV40) large antigen and, notably, RuvB helicases<sup>36</sup>.

Helicase classification is made difficult by the complexity of existing systems, the variability of oligomerization states, nucleic acid specificity, translocation polarity, preference for single or double stranded DNA/RNA, and the addition of regulatory modules. The most recent classification system has been proposed by Singleton *et al* in 2007<sup>38</sup>, which divides helicases into 6 superfamilies according to their most prevalent functional motifs and characteristics. All superfamilies have common motifs that mark them as helicases: the Walker A (G(x)<sub>4</sub>GKT) and Walker B (DExD/H) motifs, and the “arginine finger”, that is involved in energy coupling. **Superfamily 1** includes only *bona fide* monomeric ssDNA helicases, i.e., proteins that translocate along single stranded nucleic acids and catalyse the separation of double strands<sup>39</sup>. **Superfamily 2** is the largest, encompassing 10 intensively studied families, including DEAD-box RNA helicases, the RecQ-like and the Swi/Snf family (where the complexes SWR1, INO80 and RAD54 are included). Swi/Snf complexes are involved in several processes in the cell that are correlated with chromatin remodelling, such as replication, DNA repair, RNA polymerase regulation, cell cycle progression and tumour progression. The mechanisms employed by these large supramolecular complexes involve mostly histone manipulation, i.e., altering the structure of nucleosomes in order to change DNA accessibility. Chromatin remodelling complexes are composed of multiple subunits, including one member of the Swi/Snf family, and other accessory proteins. The INO80 complex, for instance, not only includes the SF2 helicase Ino80, but also the RuvBL1/2 dodecameric complex, and 12 other proteins<sup>40,41</sup> (see section 1.3.2 for more information on chromatin remodelling complexes containing RuvBL proteins as one of the components). Superfamilies 3-6 include helicases that form hexamers and double hexamers: **SF3** is composed entirely of viral helicases<sup>42</sup>, and includes AAA<sup>+</sup> proteins with multiple enzymatic activities, a 3' to 5' processivity and the ability to form both hexamers and dodecamers<sup>38</sup>. SF3 helicases possess an

origin binding domain, which allows viruses to bypass the host-cell regulatory mechanisms, and a characteristic C motif<sup>43</sup> (**Fig. 1.4**); **SF4** includes only helicases with 5' to 3' polarity; **SF5** is composed of Rho5, which while being closely related to SF4 helicases, was attributed to a different superfamily on the basis of differences in sequence (**Fig. 1.4**). It functions as transcription terminator, unwinding RNA:DNA duplexes and releasing newly-formed RNA transcripts<sup>44</sup>. **SF6** comprises only DNA helicases, unlike SF1-5, which contain both DNA and RNA helicases<sup>36</sup>. As in the case of SF3, members of this family have many characteristics of AAA+ proteins (see section 1.1.3). However, SF6 members display also characteristic sensor 1 and 2 motifs (**Fig. 1.4**). Representative SF6 members include the mini chromosome maintenance (MCM) protein complex, which is the helicase component of the eukaryotic replicative holo-helicase CMG<sup>45</sup>, and prokaryotic RuvB (see section 1.1.6).



**Figure 1.4 - Representative helicases from Superfamilies 3 to 6.** Of notice is the general trend towards a ring-shaped structure. From <sup>38</sup>.



Although the prokaryotic RuvB has been included into SF6, RuvB-Like proteins have not been allocated to any particular superfamily. Their structural characteristics support their inclusion in SF6, considering the presence of SF6-specific motifs Walker A and B, and sensor 1 and 2, as well as some of the biochemical characteristics known so far, namely their classification as AAA+ ATPase with multiple cellular functions and ssDNA specificity. Like MCM, RuvBLs interact with binding partners through the outer part of the ATPase core, in the latter case through the accessory domain II module<sup>46</sup>. Curiously, the Rho helicase, the constituent member of SF5, was observed by EM to have an open “notched” conformation (**Fig. 1.4**), which is proposed to correspond to an open ring or absence of one subunit. This conformation was also observed by EM for RuvBL2 (unpublished result, see chapter 2, section 2.5.2). This diversity of characteristics among superfamilies creates a vast and complex number of systems, adapted to the particular needs of the cell.

Some proteins have been classified as helicases with basis on their sequence, displaying the characteristic helicase motifs, but do not perform the biochemical helicase activity. This is the case of SWI/SNF family, from SF2 (see section 1.2), which couples ATP binding and hydrolysis with chromatin remodelling through nucleosome rearrangements that alter chromatin accessibility<sup>47</sup>.

A genome-wide analysis of human helicases<sup>48</sup> identified 64 RNA helicases and 31 DNA helicases, including 5 RecQ members, KU70, 9 MCM (minichromosome maintenance) proteins, nucleolin, 2 chromodomain helicases, 2 DNA repair helicases, lymphoid-specific helicases, RuvB-Like 1 and 2, Pif1, Twinkle, BACH1, RecQ5 isoforms alpha, beta and gamma and regulator of telomere elongation helicase (RTEL1).

### 1.1.5 CHAPERONES

Most major processes that take place in the cell are performed by multi-component complexes, usually through energy-dependent conformational changes. Examples include complexes involved in chromatin remodelling, proteasomes and ribosomes. The complexity of such structures foretells the need for specific support in the assembly process. The help of “molecular matchmakers”, able to use the energy gained from ATP hydrolysis to induce conformational changes in one or both elements of a molecular pair, was initially described for protein-DNA complex formation<sup>49</sup>. However, since then this concept has been expanded to include other types of complexes, such as multi-protein complexes<sup>50</sup>. Chaperones can thus be ascribed the function of promoting the non-covalent assembly of other proteins or protein complexes. Chaperones may also interact with other proteins to prevent aggregation and/or promote proper folding, many times in an ATP-dependent manner. Many evolutionarily related chaperones belong to the AAA+ family, and use ATP-dependent conformational alterations in the ATPase core to perform their functions. Some AAA+ proteins still specialise in disassembly, remodelling or disaggregation, acting either in concert with a proteasomal machine for further target degradation, or by themselves<sup>51</sup>. Many of these proteins share a hexameric structure and a conserved protein-processing pore. However, it has been observed that they do not necessarily exert their disassembling/remodelling functions through translocation-dependent unfolding, but may do so through other ATP-dependent interactions with the target protein (e. g. the  $\sigma^{54}$ -activating enzymes)<sup>51</sup>.

RuvBL proteins are fundamentally involved in various mechanisms that implicate a chaperone function. They are involved in the assembly of **chromatin remodelling complexes** INO80, SWR/SRCAP and TIP60, but not as the catalytic

subunit. In all cases the RuvBL1/2 complex seems to have a scaffolding function<sup>12</sup>. However, as part of the larger SWR/SRCAP complexes, responsible for histone exchange, the RuvBL1/2 complex was found to have not only a scaffolding function, but also to be able to perform *in vitro* by itself the exchange of canonical histone H2A for the more labile H2A.Z<sup>52</sup>, another function added to the panoply of activities already identified for these versatile proteins.

As part of the **R2TP** complex, RuvBL1/2 participates in the assembly of box C/D and box H/ACA snoRNPs. R2TP is involved in the assembly of core factors in both snoRNPs, but recruits specific factors in each family. These factors chaperone the core components as they are assembled, regulate the assembly and disassembly of pre-snoRNP intermediates and regulate the activity of intermediate subcomplexes<sup>53,54</sup>. In eukaryotes and archaea, box C/D snoRNAs modify rRNAs and possibly mRNAs, and box H/ACA snoRNAs have motifs included in the RNA sequence of telomerase, a complex that works in the synthesis of telomeric DNA<sup>54</sup>. As a chaperone in snoRNPs assembly, the R2TP complex regulates ribosome biogenesis and consequently controls cell proliferation.

In mammalian cells, R2TP interacts through RuvBL1/2 with intermediate H/ACA snoRNP core components and their co-chaperones, namely SHQ1 and dyskerin/NAP57, dissociating them for subsequent assembly in the holo-complex. The dyskerin interaction is suggested to make use of the RuvBL1/2 central channel, which seems fit to thread the unstructured dyskerin C-terminus. This tail is not necessary for binding to RuvBL1/2, but it is necessary for separation of dyskerin from SHQ1, which further suggests a mechanism of action whereby the complex “grabs” both proteins and follows with a rotational movement of the barrel-like structure, which could be responsible for the forceful separation of both proteins. This possibility may be supported by the fact that point mutations in the Walker A and Walker B motifs of RuvBL2 lead to severe defects in snoRNA

accumulation<sup>55</sup>. This model seems to be applicable also to C/D core proteins, since some of their components possess similarly unstructured, highly charged tails, which are essential for yeast survival and for mammalian nucleolus maturation<sup>11,12</sup>. The biogenesis of box C/D snoRNPs is mediated by the R2TP/HSP90 complex and other assembly factors. These assembly factors are as of yet uncharacterized, although pull-down assays performed in yeast have identified TAF9, NOP17 and BCD1 as intermediate factors, which are not present in the mature snoRNP. Of these, NOP17 and BCD1 have been shown to interact directly with RuvBL1 and RuvBL2, respectively<sup>56</sup>. Interestingly, interaction of R2TP with box C/D snoRNPs was shown to decrease in poorly growing cells, as R2TP re-localizes from the nucleus to the cytoplasm, depending on the cell growth phase and nutrient condition<sup>57</sup>.

All core components of H/ACA snoRNPs also assemble with TERC (Telomerase RNA component) to form TERC-containing RNP, in a process involving the RuvBL1/2:dyskerin complex and dependent on the ATPase activity of RuvBL1. In addition to TERC, the **Telomerase** complex includes TERT (Telomerase reverse transcriptase) and the TERC-binding protein dyskerin. As part of the RuvBL1/2 complex, RuvBL1 was shown to interact directly both with TERT and with dyskerin. Telomerase holoenzyme adds DNA repeats to telomeres, nucleoprotein structures that protect the termini of chromosomes. In cancer cells, Telomerase activity is upregulated, conferring immortality by protecting chromosome ends indefinitely, contrary to the gradual loss of activity observed in normal cells. Both RuvBL1 and RuvBL2 were shown to be essential for Telomerase activity and for the accumulation of TERC and dyskerin, as depletion of RuvBL1, RuvBL2 or dyskerin leads to loss of TERC. Furthermore, the low-activity TERT:RuvBL1/2 complex seems to be highly abundant relative to the highly active TERT:TERC:dyskerin complex, which suggests that the former may also be a target for regulation, or that its assembly may require some time.

Interestingly, the TERT:RuvBL1/2 complex is especially abundant during S phase, a time-dependent association that may justify the cell cycle-dependence of Telomerase formation<sup>58</sup>. Additionally, since RuvBL1 and RuvBL2 are involved in maintenance of dyskerin levels, they could also be targets of interest in congenital dyskeratosis, a disease caused by low activity of dyskerin and decreased TERC levels, which results in low telomerase activity and much shorter telomere length<sup>59</sup>. It is further suggested that epigenetic modifications of RuvBL1 and RuvBL2 may be an interesting venue of study in the context of Telomerase activity regulation, since it has been observed that, e. g., SUMOylation of RuvBL1 and RuvBL2 modifies their transcriptional activity and protein-protein interaction<sup>59-61</sup>.

Still as components of R2TP, RuvBL1 and RuvBL2 are involved in the assembly and stability of **PIKKs**, via the concerted interaction of the R2TP:HSP90/Prefoldin-Like complex, the TTT complex and the PIKKs<sup>12</sup>.

The RuvBL1/2 complex was also recently shown to have chaperone functions in protein homeostasis, under stress conditions as yet unidentified (other than heat shock or proteasome inhibition, which did not induce RuvBL expression). In this role, RuvBL1/2 complex was shown to promote the formation of an organelle, the aggresome, which functions as storage compartment for aggregated proteins, and possibly also in their proteasomal or autophagic degradation<sup>62</sup>. This organelle is formed as an alternative to the ubiquitin-proteasome degradation pathway, when the accumulation of aggregates becomes too extensive<sup>63</sup>, and indeed overexpression of RuvBL1 and RuvBL2 protected yeast cells exposed to proteotoxic heat stress. Furthermore, depletion of either RuvBL1 or RuvBL2 suppressed aggresome formation in mammalian cells expressing aggregate-forming synphilin – a substrate of the aggresome. To help elucidate the role of RuvBL1/2 complex in this process, we collaborated with the group of Michael Sherman, by providing RuvBL1/2 full-length complex, which

was used for interaction assays with synphilin (since this was the total of our contribution to this study, and occupied only a small portion of the total duration of my PhD, the RuvBL1/2 complex purification protocol is only briefly described in the discussion section). It was thus assessed that RuvBL chaperone activity was independent of R2TP complex (see sections 1.1.7 and 1.2), and that neither RuvBL1 nor RuvBL2 had the ability to promote aggresome formation by themselves, but that the RuvBL1/2 complex formation was necessary for this function. This is consistent with their dodecameric, barrel-like structure, as well as their inner channel charge distribution, which as a mix of positive and negative charges from both types of monomers, supports a possible association with a polypeptide chain. Co-immunoprecipitation followed by mass spectrometry identified both RuvBL1 and RuvBL2 as interacting with synphilin. Interestingly, this interaction occurs between the ankyrin repeat domain of synphilin, which is the domain responsible for its aggregation, and K372 of RuvBL1, which is located on the surface of the ATPase core and in close proximity to the central channel. This interaction is critical for assembly of the aggresome. In line with this, while in naïve cells, RuvBLs are distributed throughout the cytoplasm and nucleus, after proteasome inhibition they were recruited to protein aggregates and then transported to the aggresome. Interestingly, ATPase activity was significantly stimulated also by other amyloid-forming proteins, namely insulin and casein fibrils, an increase more noticeable in the domain II-truncated complex. In fact, even denatured BSA potently enhanced this activity, in contrast to the native form, which had no effect. Finally, it was concluded that the main effect of RuvBLs is actually on disaggregation, rather than recruitment of aggregates to the aggresome, since disassembly of aggresomes is strongly suppressed by RuvBL1 depletion<sup>62</sup>.

### 1.1.6 RUVB AND RUVB-LIKE PROTEINS

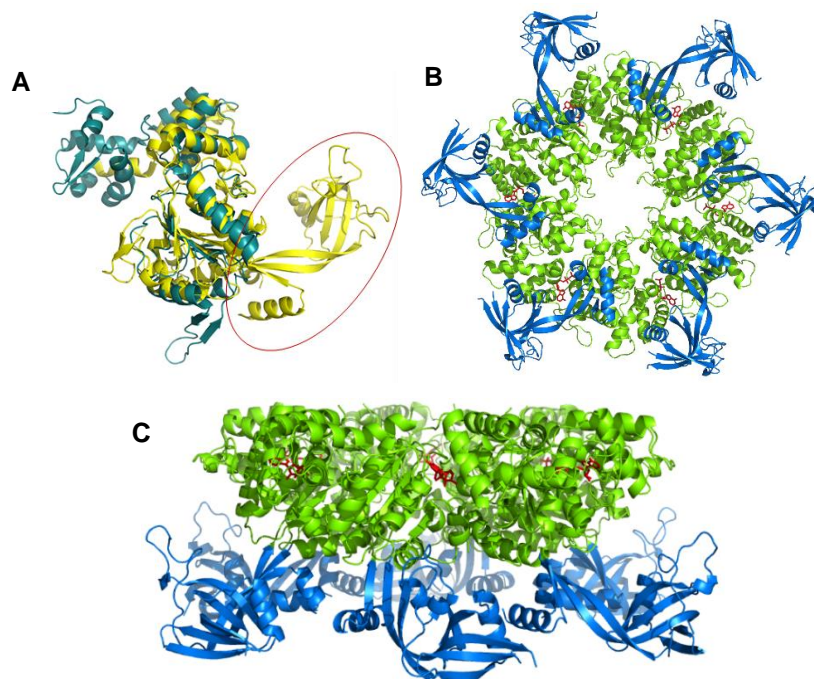
RuvBL1/Pontin/Rvb1/Tip49 was initially discovered in 1997 in rat, in association with the TATA-binding box protein (TBP), by Masato Kanemaki, from the group of Taka-Aki Tamura<sup>64,65</sup>. TBP is a general transcription factor that works at the core of multiprotein transcription factors, necessary for recruitment of all classes of RNA polymerases, which binds at the TATA box promoter element<sup>66</sup>. In 1998 RuvBL1 was found as part of a complex with RuvBL2/Reptin/Rvb2/Tip48, associated with the RNA polymerase II holoenzyme, and later in 1999 RuvBL2 was found as part of a complex with RuvBL1 alone in human cells<sup>67</sup>.

**Table 1.1 – Different RuvBL names and their origins.**

RuvBL1	RuvBL2	Name origin
TIP49	TIP48	<u>T</u> ATA-binding protein (TBP)- <u>i</u> nteracting <u>p</u> rotein <sup>68,69</sup>
TIP49a	TIP49b	<u>T</u> BP- <u>i</u> nteracting <u>p</u> rotein <sup>67</sup>
Pontin52	Reptin52	<u>R</u> epressing <u>P</u> ontin52 <sup>6</sup>
TAP54 $\alpha$	TAP54 $\beta$	<u>T</u> IP60- <u>a</u> ssociated <u>p</u> rotein <sup>70</sup>
TIH1	TIH2	<u>T</u> IP49a/b <u>h</u> omolog <sup>71</sup>
ECP54	ECP51	<u>e</u> rythrocyte <u>c</u> ytosolic <u>p</u> rotein <sup>72</sup>
NMP238	---	<u>n</u> uclear <u>m</u> atrix <u>p</u> rotein <sup>73</sup>
Rvb1	Rvb2	<u>R</u> uv <u>B</u> homolog <sup>74</sup>
p50	p47	protein <sup>75</sup>
RuvBL1	RuvBL2	Ruv (Resistance to UV) B-Like (RuvB homolog) <sup>76</sup>

RuvB-Like proteins are eukaryotic proteins with partial structural homology to bacterial RuvB (Resistance to UV) proteins (**Fig. 1.5**), the ATP-dependent motors of the RuvAB complex that drive branch migration of Holliday junctions formed during homologous recombination<sup>77</sup>. The RuvA, RuvB and

RuvC proteins are necessary for normal levels of cellular resistance to UV or ionizing radiation, or of mitomycin. Based on this partial homology, these proteins were immediately assumed to be ATP-dependent DNA helicases, and indeed they were shown to have helicase as well as ATPase activities *in vitro*<sup>67,78</sup>. Interestingly, RuvB-Like amino acid sequences are highly conserved not only in eukaryotes but also in *Archaea*, highlighting them as some of the most conserved nuclear dynamics-related proteins. Phylogenetic analyses further showed that the analysed *archaea* possess only one RuvBL copy, belonging to the RuvBL2 branch. Moreover, bacterial RuvB was shown to be closest to RuvBL2 than RuvBL1. Together, these facts suggest that RuvBL2 is the common ancestor of the RuvBL family, from which RuvBL1 would diverge after the emergence of eukaryotes.<sup>67</sup>



**Figure 1.5 – Main structural features of RuvB-Like proteins.** **A** - Structural alignment of human RuvBL2 (yellow) and bacterial RuvB (blue). The red circle highlights the extra domain II in RuvBL proteins. **B** – Bottom view of the RuvBL1 hexamer. The central ATPase core (green) is formed by domains 1 and 3, and displays a central channel. The extra domain II from each monomer (blue) protrudes outwards, producing 6 mobile arms. **C** – Side view of RuvBL1 hexamer. ATPase core in green, with bound nucleotide in red. Domains 2 in blue.

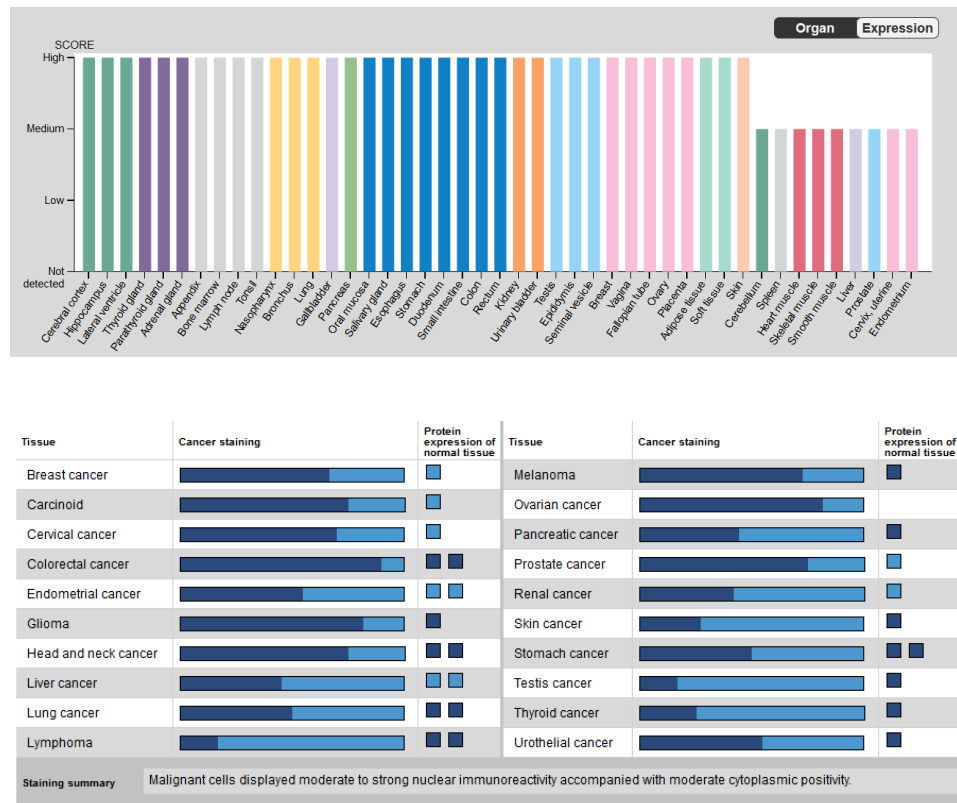


This raises the question for the need for two highly similar proteins, which often work together. It is possible that this need arose as organisms became more complex, in order to fine-tune more complex pathways, such as the Wnt/ $\beta$ -catenin signalling pathway, or pathways responsible for cell adherence (possibly during the transition to multicellularity), as can be observed in their antagonistic regulation of metastization processes by controlling the expression of metastasis Suppressor gene *KAI1*<sup>61,79</sup>.

RuvB hexamers contain only an ATPase core, and associate to form double rings through the mediation of RuvA octamers previously primed through interaction with a Holliday junction<sup>80</sup>. Compared to prokaryotic RuvB, RuvB-Like proteins comprise an extra domain II, located between domains I and III, encircled in **fig. 1.5A** (interestingly, domain II of RuvBL proteins locates spatially to roughly the same position occupied by RuvA in relation to RuvB<sup>80</sup>, such that a double RuvB ring acquires conformation similar to a RuvBL1/2 double ring<sup>81-83</sup>). Domains I and III form an ATPase core, responsible for nucleotide binding and hydrolysis (in green in **figs. 1.5B and 1.5C**) whereas domain II forms a mobile arm that protrudes from the more rigid core, such that in a ring the ATPase core faces a central channel and six mobile domains protrude outwards. RuvBLs are usually found in the cell as single or double hexameric rings, with a 1:1 proportion of each<sup>9,84</sup>. Whether each ring is constituted of a mixture of RuvBL1 and RuvBL2 or homomeric still under debate, since the resolution of the structures obtained by cryo-electron microscopy (cryo-EM) does not yet allow a clear distinction between RuvBL1 and RuvBL2. It is also not always clear whether the contact between rings occurs between the ATPase cores or through the domains II. However, it is now mostly accepted that the latter interaction is probably the most frequent, since both EM and crystallographic structures of dodecamers obtained to date show double hexamers interacting through the domains DII, as well as alternating RuvBL1 and RuvBL2 monomers in the

hexameric rings. Nevertheless, both proteins are able to form homo-hexamers in vitro<sup>28,85</sup>. Oligomer formation was shown to be dependent on the presence of the Walker A and Walker B motifs. However, these were not necessary for binding to other proteins through domain II<sup>7</sup>.

RuvBL1 is encoded in chromosome location 3q21, a region of frequent rearrangements in leukaemia and solid tumours<sup>86,87</sup>. RuvBL2 is encoded in location 19q13.33 of chromosome 19, and is physically linked with the breast



**Figure 1.6 - RuvBL2 protein expression data, identified by specific antibody CAB012432, from Abcam (product 36569). Top:** Expression levels in several organs. **Bottom:** expression levels in cancer vs. normal cells. **Dark blue:** high expression levels; **light blue:** medium expression levels. The bars indicate the proportion of high vs. normal levels of RuvBL2 identified in a total number of analysed samples, for each cancer type. From the protein atlas annotation project.

cancer-related CGB/LHB gene cluster<sup>88</sup>. This chromosome has more than double

the gene density in comparison to the genome-wide average<sup>89</sup>. RuvBL2 is ubiquitously expressed, and high protein levels have been identified in many organs and tissues (**Fig. 1.6**), but are especially high in the testis and thymus<sup>67</sup>, a pattern closely repeated by the eukaryotic recombination factor Rad51; mRNA levels are conversely highest in the testis but relatively low otherwise<sup>90</sup>.

A variety of functions and seemingly unrelated roles have been attributed to RuvBL2, which has been identified mostly in the nucleus, but also in the cytoplasm and associated with the cytoplasmic membrane. Together with RuvBL1, it has been implicated in regulation of gene transcription, chromatin remodelling, DNA damage sensing and repair and in the assembly of protein and ribonucleoprotein complexes (**Fig. 1.7**). Both exert their roles as part of signalling pathways that are involved in nutrient sensing, RNA metabolism, DNA damage repair and pathophysiology, such as cancer, ciliary biology and disease. In fact, they have been shown to be overexpressed in many tumours, and to be directly involved in tumour biology, hypoxia and metastization<sup>4,5,60,91-96</sup>. RuvBLs perform a variety of functions. Regarding mechanisms of transcriptional regulation, RuvBLs are reported to be required for the induction of interferon-stimulated genes in response to interferon alpha, independently of their ATPase activity. They also have a role in the regulation of nuclear hormone receptors, such that both proteins can bind to oestrogen and glucocorticoid receptors, with agonistic effects<sup>97</sup>. Conversely, RuvBLs antagonistically regulate transcription during heart development<sup>98</sup>, are differently methylated in hypoxic cells, and genome-wide analyses show that they don't share many common targets in hypoxia signalling pathways<sup>4,5</sup>. Further studies are required to elucidate how these two proteins, which function as part of the same complex, can also achieve transcriptional specificity.

The answer may lie partly in the regulation of their oligomerization state, such that when they are working alone, they recognize different

promoters/regulators of expression, as suggested by Rowe and colleagues in 2008<sup>99</sup>. Here, using ChIP/re-ChIP assays, they show that the addition of phorbol ester induces the recruitment of RuvBL1/Tip60 activator complex to the promoter of the anti-metastatic *KAI1* gene, in a way that mutually excludes the recruitment of RuvBL2/ $\beta$ -catenin complex. How their oligomerization state is regulated is still a matter of debate, but one of the mechanisms may rely upon interaction with binding partners, as has been demonstrated for the histone H3. Acetylation and methylation of histone H3 N-terminus regulates the activity and oligomerization state of RuvBL1 and RuvBL2. In fact, RuvBL2 association with the progesterone receptor gene promoter is related to the H3 marks of the surrounding histones<sup>100</sup>. Another mechanism which may direct the transcriptional specificity of RuvBL1 and 2 may be related with specific post-transcriptional modifications (PTMs). As an example, SUMO-conjugated RuvBL2 is recruited to the nucleus, where together with  $\beta$ -catenin, it binds to the promoter of the metastasis suppressor gene *KAI1*, leading to metastasis<sup>60,61</sup>.

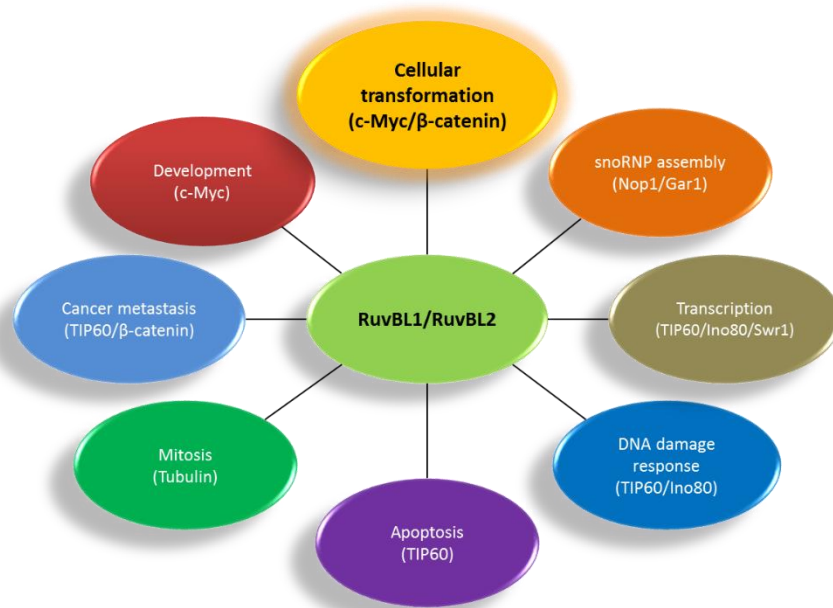
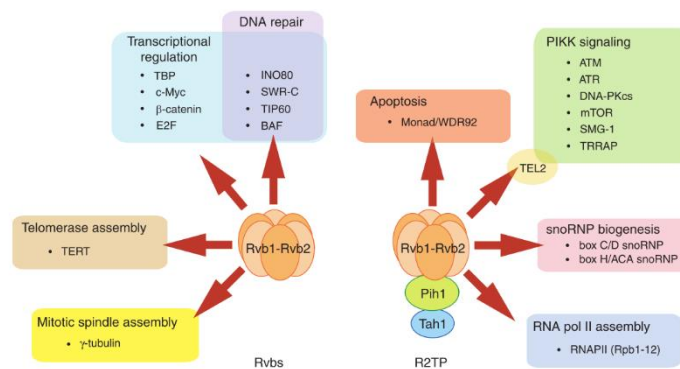


Figure 1.7 – RuvL1 and 2 are involved in several pathways in the cell. Adapted from<sup>101</sup>

### 1.1.7 ROLES OF RUVBLs IN THE CELL AND CELL CYCLE

RuvB-Like proteins are ubiquitous in the eukaryotic cell, and as such have a widespread influence in various pathways and cellular functions. The mechanisms that regulate their roles in each case are not completely understood. Their action is likely regulated by their oligomeric state and post-transcriptional modifications that will direct them towards interaction with specific partners or multi-subunit complexes, in response to the metabolic state of the cell, other external factors or even disease states<sup>4,60</sup>. A wide array of functions performed by RuvBLs is achieved through the activity of R2TP complex, and by their participation in other supramolecular assemblies, in a seemingly dynamic way, adapted to the needs of the cell (**Fig. 1.8**).



**Figure 1.8 – RuvBL complex functions mostly as part of larger supramolecular assemblies.** R2TP complex functions vs RuvBL complex functions. From<sup>57</sup>.

#### *Checkpoint pathways*

Checkpoint pathways are activated upon DNA damage, blocking progression of the cell cycle until the damage is repaired, before allowing the cycle to resume and transit to the next phase. The signalling cascade activates p53, leading to inactivation of cyclin dependent kinases, and preventing cell cycle progression from G1 to S (G1/S checkpoint), DNA replication (intra S checkpoint)

or G2 to mitosis (G2/M checkpoint)<sup>102</sup>. RuvBLs play roles in **checkpoint regulation**, usually as part of supramolecular assemblies, such as the INO80, SWR1 or TIP60 complexes.

In yeast, Mec1/ATR and Tel1/ATM phosphorylate subunits of the **INO80** complex (Ies4 and Ino80) during DNA damage response, affecting DNA damage checkpoint response, but not the DNA repair pathway. Checkpoint activation also leads to phosphorylation of subunits of the **SWR1** complex (Bdf1 and Swr1). Both complexes, which include RuvBL1 and RuvBL2 as components, thus participate in the process of checkpoint adaptation, by helping overcome an extended checkpoint arrest and re-enter the cell cycle with an unrepaired DSB. INO80 and SWR1 seem to have antagonistic activities in yeast, since cells deficient in Ino80 are unable to overcome checkpoint arrest, an effect which is counteracted by the additional deletion of Swr1<sup>8,103</sup>.

Human **TIP60** complex was observed to be recruited to chromatin after mitogenic stimulation, in an E2F-dependent manner, during late G1, resulting in acetylation of histone H3 and H4, and subsequent transition to S-phase<sup>104</sup>. TIP60 is also involved in activation of ATM, thus regulating the ATM-Chk2-Cdc25 and ATM-Chk2-p53 pathways. These are responsible for inhibition of DNA synthesis in response to DNA damage, and inhibition of Cdk activity, respectively<sup>103</sup>.

RuvBL2 was also found to form a complex with activating transcription factor 2 (ATF2), a member of the ATF-CREB family of transcription factors involved in cell cycle progression and differentiation<sup>105</sup>.

The subcellular distribution of RuvBLs is very cell-cycle-dependent, and undergoes particularly major changes during M phase. RuvBL1, which mostly localizes in the nucleus, was found to co-localize with  $\alpha$ - and  $\beta$ -tubulin in the mitotic spindle. In anaphase-to-telophase transition, RuvBL1 co-localizes with PLK1 (polo-like kinase 1)<sup>106</sup>, which interestingly has the ability to phosphorylate

RuvBL1 but not RuvBL2. Like RuvBL1, RuvBL2 was found to be excluded from the chromosomes in early mitosis, and a small population associated with the centrosome and mitotic spindle. However, unlike RuvBL1, it localizes to the midzone during telophase and to the midbody during cytokinesis. Despite the fact that RuvBL1 and RuvBL2 usually function together as a complex, the protomers separate during cytokinesis. Together with the fact that PLK1 associates with RuvBL1 *in vivo*, this suggests an unknown mechanism of differential regulation of RuvBLs during mitosis. Regardless, both RuvBLs are necessary for efficient chromosome alignment and segregation<sup>107-109</sup>.

#### *Transcription regulation*

As illustrated thus far, transcription of genes requires the coordinated action of many different factors, starting with the action of a) ATP-dependent chromatin remodelling complexes, which both mobilise nucleosomes and exchange histones from the DNA; and b) enzymes capable of adding or removing covalent modifications on histones, other proteins, or DNA, such as acetylation, methylation or ubiquitylation, which results in the recruitment of various protein complexes that recognize these modifications. RuvBLs are part of chromatin remodellers INO80 and SRCAP, which are involved in alterations in nucleosome mobility at the promoter region of a large subset of genes, and are also part of the TIP60 complex, responsible for nucleosomal H4/H2A HAT activity, as well as for acetylation of several non-histone proteins.

RuvBL1 and RuvBL2 may function together or independently, and have even been shown to work **antagonistically** in the regulation of transcription of several target genes, notably anti-metastatic KAI1<sup>61</sup>, and show antagonistic effects during early development in zebrafish embryos<sup>98</sup> (see sections on cancer and development). In *Drosophila*, RuvBL1 and RuvBL2 function antagonistically to regulate the expression patterns of Hox proteins: while RuvBL1 associates with

Trithorax group Brahma (Brm) complex, RuvBL2 associates with Polycomb group Polycomb Repressive Complex 1<sup>110</sup>, thus regulating expression of homeotic genes<sup>111,112</sup>.

When working in concert, RuvBL1 and RuvBL2 are usually found as members of larger supramolecular assemblies, although there is no indication so far that they may antagonistically regulate those complexes' activities in these cases. However, when demonstrated to work antagonistically, so far they have always been found to work in association with different complexes, such as Tip60 and  $\beta$ -catenin<sup>4,60,61,113</sup>.

#### *PIKK signaling*

RuvBL1 and RuvBL2 regulate all PIKK members. PIKKs are serine-threonine protein kinases that regulate DNA damage responses, nutrient-dependent signalling and nonsense-mediated mRNA decay (NMD). They phosphorylate proteins involved in cell cycle progression, DNA repair, apoptosis and cellular senescence<sup>12</sup>.

The PIKK family consists of 6 kinases. **ATM** is responsible for cellular response to DNA double strand breaks, by phosphorylating histone variant H2AX in response to DNA double strand break caused by environmental stresses such as UV light<sup>21</sup>. **ATR** phosphorylates histone H2AX in response to accidental DSBs caused by metabolic stresses<sup>21</sup>, and mutations in this protein lead to growth retardation and microcephaly. **DNA-PKcs** is involved in DSB repair by non-homologous end joining. **SMG1** (suppressor with morphological effect on genitalia 1) regulates nonsense-mediated decay of aberrant mRNA. **mTOR** (mammalian target of rapamycin) controls cell growth in response to environmental cues, mitogenic signals and nutrient availability. **TRRAP** (transformation/transcription domain-associated protein) is a regulator of gene expression via association with histone acetyltransferase complexes. All 6 PIKKs



require the HSP90 co-chaperone Tel2 (as part of the TTT complex) to associate to R2TP, through the PIH1D1 subunit<sup>12,114</sup>.

DNA damage response involves the concerted action of sensors, transducers and effectors. Sensors detect the damage, and activate PIKKs ATM, ATR and DNA-PKcs, which transmit the signal to the effector proteins. In fact, more than 900 phosphorylation sites have been identified that contain a consensus ATM or ATR phosphorylation motifs, in 700 proteins that are phosphorylated in response to ionizing radiation (IR)<sup>115</sup>. Depending on the severity of the damage, the effector proteins (which include transcription factor p53) activate the checkpoints which will arrest the cell cycle and either proceed to DNA repair or initiate the pathways which will lead to apoptosis or cell senescence. R2TP regulates DNA damage response of ATM, ATR and DNA-PKcs by affecting their stability, and by regulation of ATR activity, leading to decreased activation of p53<sup>114</sup>. Knockdown of RuvBL1 or RuvBL2 leads to decreased mRNA levels of ATM, ATR, DNA-PKcs, mTOR and TRRAP, but does not affect the abundance of other kinases. The RuvBL1/2 complex is also responsible for the phosphorylation and activation of direct downstream effectors of ATM, ATR, mTOR and SMG1, respectively Chk2, Chk1, p70 S6K and Upf1<sup>116</sup>, and interacts directly with the Fanconi Anemia complex, involved in DNA crosslink repair<sup>115,117</sup>. Human RuvBL2 was also found to be a potential target of ATM/ATR, as it includes putative phosphorylation sites<sup>118</sup>. The RuvBL1/2 complex further interacts with the URI prefoldin complex through Monad/WDR92. The URI complex is involved in nutrient sensing responses through regulation of gene responses and cell survival signalling downstream of mTOR<sup>118</sup>.

Protein kinases function as switches that coordinate the pathways by which a given signal transduction pathway is achieved, phosphorylating multiple components of particular pathways, sometimes even components of the same

multi-protein complex. The human genome encodes 518 kinases, and they are critical regulators of signal transduction pathways related to metabolism, transcription, cell cycle progression, cytoskeletal rearrangement and cell movement, apoptosis and differentiation. In eukaryotes, PIKKs initiate the cellular stress response in situations which may compromise genome integrity, mRNA translation or nutrient availability<sup>119</sup>. Overall, DNA damage signalling is known to regulate initiation of origins, fork stability, the intra-S phase checkpoint, and fork resumption after damage<sup>115</sup>.

## 1.2 RUVBL1 AND RUVBL2 IN HIGHER-ORDER COMPLEXES

As mentioned above, during transcription, replication and DNA repair, chromatin accessibility is dynamically altered by two main classes of chromatin remodellers: the first class includes ATP-dependent complexes that mobilize nucleosomes and alter accessibility to underlying DNA. The second class includes acetylases, deacetylases, methylases and demethylases, responsible for the addition or removal of covalent modifications on histone tails, thus altering the local compaction state of chromatin.

RuvBLs are frequently found associated with larger complexes as chaperones, or as activity regulators. Notably, the RuvBL1/2 complex is part of the R2TP complex, and is essential for INO80 complex activity due to its role in the incorporation of actin-related protein 5 (Arp5). Yeast RuvBLs also increase the stability of TIP60 complex, by increasing its histone acetyltransferase activity *in vitro* and protecting the catalytic subunit, Tip60, from heat inactivation. RuvBLs further contribute to TIP60 complex activation by preventing its p400-mediated inhibition, a role achieved through the interaction of RuvBLs with the SNF2 domain of p400<sup>120</sup>.

This section focuses on relevant RuvBL-containing complexes known to date. RuvBLs are commonly found associated with chromatin remodelling complexes, which highlights their importance at the transcription regulation level in various cellular settings, from development to disease and metabolic response to nutrient availability.

### *R2TP and snoRNPs*

The **R2TP complex**, highly conserved among eukaryotes, together with its co-chaperone HSP90, is involved in the biogenesis of **small nucleolar ribonucleoprotein** (snoRNP, required for biogenesis of ribosomal, small nuclear and transfer RNA) and **pre-ribosomal RNA processing**, and plays crucial roles in apoptosis, **PIKK signalling** and **RNA polymerase I, II and III assembly** (Fig. 1.9).

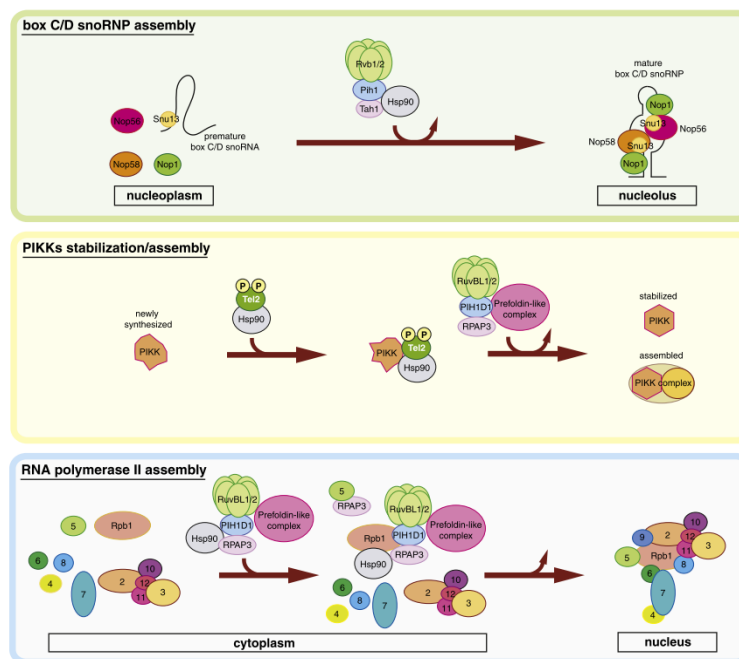
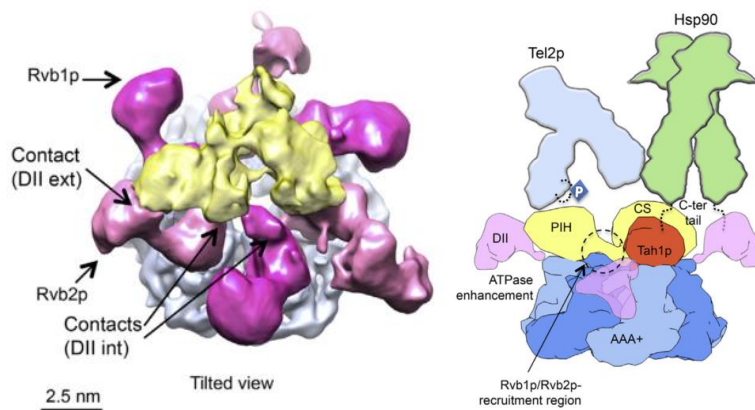


Figure 1.9 - R2TP complex functions. Adapted from<sup>116</sup>.

R2TP-HSP90 complex also takes part in the assembly of ribonucleoprotein particles, such as Telomerase and spliceosomal U4 snRNA.

The human/yeast R2TP complex comprises **RuvBL1/Rvb1**, **RuvBL2/Rvb2**, **RPAP3/Tah1/Spag** and **PIH1D1/Pih1**, of which **PIH1D1** targets R2TP to dyskerin (a central module in box H/ACA snoRNP), Rpb1 (a core subunit of RNA polymerase II) and Tel2 (a core subunit of the Tel2:Tti1:Tti2 complex, which is a central regulator of PIKK abundance and DNA damage response signalling pathways). The structure of the yeast homologue of R2TP has been recently determined by cryo-EM (**Fig. 1.10**)<sup>121</sup>. It shows that the complex is composed of one heterohexamer of RuvBL1/RuvBL2, with one Tah1p-Pih1p (RPAP3-PIH1D1) copy sitting at the bottom of the complex, bound to multiple DII domains. RPAP3 is the interaction bridge with the ubiquitous molecular chaperone HSP90, forming a complex essential in the assembly of box C/D snoRNPs<sup>53,114,116,122–124</sup>. The cryo-EM structure provides insights into the general mode of interaction between the RuvBL1-RuvBL2 and the Tah1-Pih1 subcomplexes. However, additional crystallographic information still needs to be gathered, so that the specific interactions of each complex can be characterised. Only then can some understanding be derived into the function and mode of regulation of each protein.

RuvBL2 binding to ATP, ADP and ATP $\gamma$ S (but not hydrolysis) led to R2TP disassembly and abolished binding to Nop58, a subunit of C/D snoRNPs. Since Walker A/B motifs were shown to be crucial for C/D snoRNPs, ATPase activity may play a role at a different stage of snoRNP assembly. RuvBL1/2 makes ATP-dependent contacts with R2TP components and box C/D core proteins, in an intermediary step during box C/D snoRNP assembly: they bind 15.5K when loaded with ATP and PIH1D1/RPAP3 otherwise. This intermediary assembly complex is further composed and stabilized by NOP58, NUFIP, ZNHIT3 and ZNHIT6 (BCD1), of which the latter also has ATP-dependent interactions with RuvBL1/2. Interestingly, many HIT-proteins seem to be specific interactors of RuvBL1/2, although some of these interactions are still to be characterized<sup>125</sup>.



**Figure 1.10 – Structure of the yeast R2TP complex.** **Left:** cryo-EM volume. For the RuvBL complex, the ATPase core is depicted in grey, and the DII domains are depicted in pink. The Tah1p-Pih1p complex is depicted in yellow. **Right:** Schematic representation of the R2TP complex and a possible mode of interaction with Hsp90 and Tel2p. Adapted from<sup>121</sup>.

Point mutations in Walker A/B of yeast RuvBL2 led to temperature-sensitive or lethal phenotypes. RuvBL2 depletion and mutation analyses revealed a crucial role in both C/D and H/ACA snoRNA accumulation, and depletion caused core subunits of each snoRNP family to accumulate in the nucleoplasm<sup>55</sup>. Finally, in situations of poor nutrient availability, R2TP is relocalized into the cytoplasm, decreasing interaction with snoRNP and consequently affecting pre-rRNA maturation<sup>57</sup>.

#### *RNA polymerase holoenzyme assembly*

There are three different RNA polymerases in eukaryotic cells: RNA pol I, which produces ribosomal RNA; RNA pol II, responsible for transcription of small nuclear RNAs and messenger RNAs; and RNA pol III, which synthesizes a range of small RNAs, including tRNAs. RNA pol II consists of 12 subunits, of which Rpb1 and Rpb2 form the active cleft.

R2TP and HSP90 together with the prefoldin-like complex, are involved in assembly of RNA polymerase II in the cytoplasm and subsequent transport to

the nucleus. Intermediate steps in the assembly of RNA pol II include interaction of HSP90 and its co-chaperone RPAP3/hSpagh/Tah1 with RNA pol II subunits Rpb1 (the largest RNA pol II subunit) and Rpb8. RPAP3, also interacts with the large subunits of RNA polymerases I and III, which suggests it may have a part in their assembly as well<sup>114,126</sup> (**Fig. 1.9**). The same interactions have been demonstrated in *Drosophila* as for yeast. In conclusion, R2TP is involved in the assembly of snoRNPs and all three RNA polymerases, and is thus deeply involved in the assembly of the machineries that produce tRNAs, mRNAs and ribosomes<sup>127</sup>.

*INO80, SRCAP/SWR1 and p400*

INO80 and SRCAP (Snf2-related CREBBP activator protein complex, known as Swi/Snf2-related [**SWR1**] in yeast) both belong to the Swi/Snf2-related INO80 family. The **INO80** complex works with other complexes to remodel chromatin at the promoter sites and sites of DNA double strand break (DSB), increasing accessibility to the DNA strand, by promoting sliding, positioning and eviction of nucleosomes, allowing the efficient recruitment of DNA transcription and repair machinery. In yeast, it was observed that after phosphorylation of H2A, histones H2B and H3 were lost, in a process that depended on INO80 ATP-dependent nucleosome sliding activity. This process rendered chromatin sensitive to nuclease activity, and allowed recruitment of RAD51 to DSBs<sup>128</sup>.

INO80 promotes both efficient progression and stabilization of the replication fork, as well as resumption of DNA synthesis under replication stress conditions, increasing tolerance to DNA damage. This complex is also involved in telomere maintenance, centrosome stability and chromosome segregation. The mechanisms of action involve eviction of H2AZ histone variant in favour of the canonical H2A, particularly at non-promoter sites. **SRCAP** catalyses the inverse reaction, depositing the H2AZ-H2B dimer in exchange for H2A-H2B<sup>129</sup>. Thus,

while sharing many similarities in subunit composition (**Table 1.2**), INO80 and SRCAP have opposing catalytic activities, and together contribute to H2A.Z genomic distribution. Besides the exchange activity shared with SRCAP, the INO80 complex is additionally able to mobilise other nucleosomes, regulating access to DNA, and thus contributing to transcription regulation of a large number of genes whose promoters are dependent on the Ino80 subunit (though not on RuvBL1 or RuvBL2)<sup>8,12,41,130–132</sup>. To sum up, INO80 and SRCAP seem to work together to exchange  $\gamma$ H2AX and H2AZ with free H2AX, thus dynamically renewing the substrates for phosphorylation and DNA repair responses<sup>23</sup>.

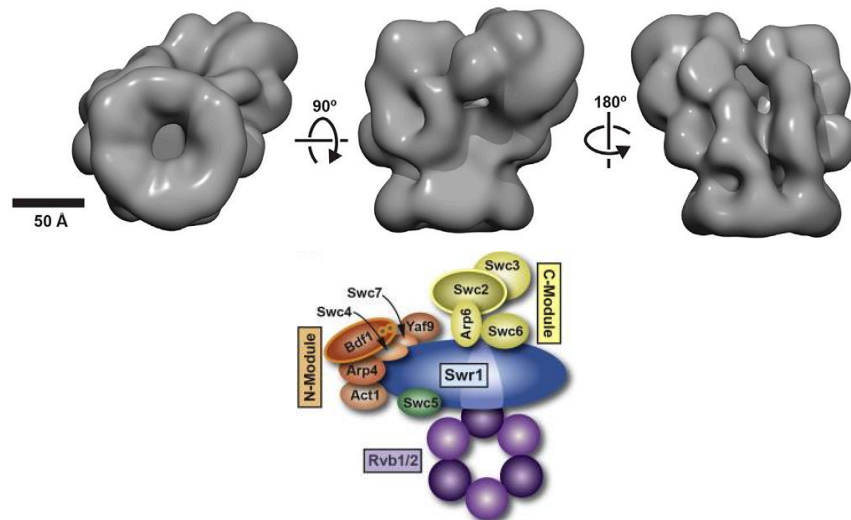
**Table 1.2 – Subunit composition of INO80 and SRCAP.** From<sup>133</sup>.

Subunit type	INO80 complex			SWR1 (yeast) and SRCAP (human) complexes	
	Saccharomyces cerevisiae	Homo sapiens	Drosophila melanogaster	Saccharomyces cerevisiae	Homo sapiens
ATPase	Ino80	INO80	INO80	Swr1	SRCAP
RuvB-like	Rvb1 and Rvb2	RUVBL1 and RUVBL2	Reptin and Pontin	Rvb1 and Rvb2	RUVBL1 and RUVBL2
Actin	Act1	$\beta$ -Actin	Actin	Act1	$\beta$ -Actin
Actin-related protein	Arp4, Arp5 and Arp8	BAF53A, ARP5 and ARP8	ARP5 and ARP8	Arp4 and Arp6	BAF53A and ARP6
YEATS protein	Taf14	Not known	Not known	Yaf9	GAS41
Non-conserved subunits*	Ies1, Ies2, Ies3, Ies4, Ies5, Ies6 and Nhp10	Amida, CCDC95, FLJ20309, IES2, IES6, MCRS1, NFRKB, UCH37 and YY1	Pleiohomeotic	Bdf1, Swc2, Swc3, Swc4, Swc5, Swc6 and Swc7	DMAP1, GAS41, tubulin, XPG, YL1 and ZnF-HIT1

\*Subunits in all other rows of this table are conserved. Act, actin; Arp, actin-related protein; BAF53A, BRG1-associated factor 53A; Bdf1, bromodomain factor; CCDC95, coiled-coil domain-

containing 95; DMAP1, DNA methyltransferase 1-associated protein 1; GAS41, glioma amplified sequence 41; Ies, INO80 subunit; MCRS1, microspherule protein 1; NFRKB, nuclear factor related to  $\kappa$ B-binding protein; Nhp10, non-histone protein 10; Rvb, RuvB-like; RUVBL, RuvB-like; SRCAP, SNF2-related CBP activator protein; Swc, SWR1 complex; Taf14, TBP-associated factor 14; UCH37, ubiquitin C-terminal hydrolase 37; XPG, xeroderma pigmentosa group G; Yaf9, yeast AF9; YEATS, Yaf9, ENL, AF9, Taf14, Sas5; YY1, yin yang 1; Znf-HIT1, zinc finger-His triad protein 1.

In each complex, the catalytic subunit that engages the nucleosomes is either the Ino80 or the Swr1 ATPase. In the **SWR1** complex the RuvBL1/2 subcomplex has a scaffolding function, assembling all functional subcomplexes in an ATP-dependent manner, where the insertion domains of RuvBLs (domain II) face the core of the SWR1 complex. Despite this orientation, observed by electron microscopy, crosslinking was found between residues of the ATPase core domain of the RuvBL complex and the Swr1, Swc2 and Arp4 subunits (**Fig. 1.11**). This



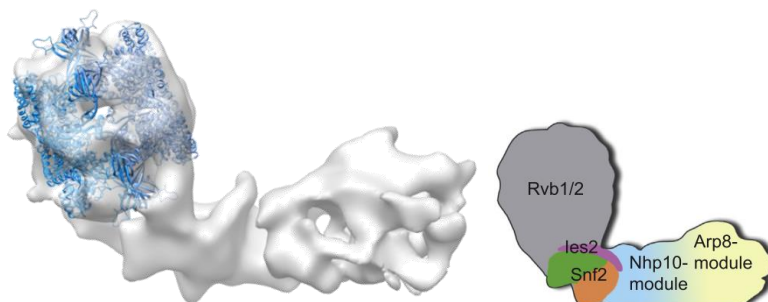
**Figure 1.11 – Cryo-EM structure of the SWR1 complex.** The RuvBL1/2 hexamer is located at the bottom, and seems to interact through the domain II arms, forming a scaffold (A). B) diagram representation of the subunits in the complex. From<sup>134</sup>

observation, together with the fact that the complex is composed of a discrete assembly of functional modules, suggests some flexibility within the complex. This led to the suggestion (supported by EM structures that also show high



flexibility of RuvBL1/2 complexes), that potential substrate- or nucleotide-dependent conformational changes of Swr1 or RuvBL1/2 could form the basis for histone dimer exchange activity<sup>9</sup>.

In **INO80**, the RuvBL1/2 has been observed to be present as a dodecameric double ring, located at the periphery of the elongated complex (**Fig. 1.12**), although some doubts remain as to the interpretation of the EM envelope, and the possibility remains that it may be a hexamer<sup>135</sup>. RuvBLs are essential for the complete assembly of the INO80 complex and, as in the SWR1 complex, are responsible for the recruitment of the catalytic Swr1-like subunit Ino80, through subunit Arp5<sup>12,84</sup>. Cross-linking coupled to mass spectrometry supports a ring structure composed of alternating RuvBL1 and RuvBL2 monomers, as observed in the existing RuvBL1/2 crystallographic structures<sup>46,84,85</sup>. Recent results suggest that the RuvBL1/2 complex may have a chaperoning function during the assembly of INO80: binding of an insert of the subunit Ino80 to a RuvBL1/2 hexamer leads to dodecamer formation, and maintains the natural conformational flexibility of the DII domains. Subsequent RuvBL1/2 binding to ATP leads to dodecamer disassembly, which is thought to be the final step in the cycle<sup>135</sup>.



**Figure 1.12** – Cryo-EM structure of the INO80 complex (A). The RuvBL1/2 dodecamer sits at the head of the complex (B). From<sup>46</sup>.

**SRCAP/SWR1** and **INO80**, and consequently RuvBL1 and RuvBL2, are involved in the regulation of chromatin mobility and association of DNA damage with perinuclear sites. These two chromatin remodellers are also relevant for

chromatin dynamics in terms of nucleosome composition, chromatin compaction and the precise locus position within the nuclear sphere. In summary, despite their many similarities in subunit composition, while SRCAP has been shown to incorporate the H2A.Z variant into nucleosomes, INO80 evicts unacetylated H2A.Z shortly after its incorporation at damaged sites<sup>136</sup>, and also remodels, spaces or removes entire nucleosomes.

An additional role for the INO80 complex is the interaction with the Cdc48 component of the Ubiquitin Proteasome System, during the degradation of RNA Polymerase II, as a result of genotoxic effects and stalling. It is also suggested that INO80 promotes the removal of ubiquitinated Rpb1 (the largest subunit of RNA pol II) from chromatin, thus keeping genome integrity<sup>137</sup>.

*TRRAP-TIP60 histone acetyltransferase*

**TIP60** (human nomenclature, homologous to yeast NuA4 [nucleosome acetyltransferase of histone 4], which does not possess RuvBLs) is a histone acetyltransferase (HAT), which means that it remodels chromatin by transferring an acetyl group to histones, therefore converting chromatin into the more relaxed and accessible euchromatin (thus rendering the DNA transcriptionally active). Like other Arp4-containing complexes (INO80 and SWR1), **NuA4** is recruited by  $\gamma$ H2AX<sup>8</sup>. This complex has also been shown to acetylate ATM protein kinase after DNA damage, thus activating ATM. TIP60 shares some subunits with the SWR1 complex and others with the NuA4 complex, suggesting that TIP60 may be a fusion of these two complexes<sup>12</sup>.

The acetyl-transferase activity of TIP60 is essential for accumulation of repair machinery at the site of DNA damage and removal of H2AX from chromatin<sup>12</sup>. In fact, histone acetylation (H3K56Ac) promotes deposition of the SRCAP/SWR1 complex, and consequent nucleosome exchange<sup>129,138,139</sup>, increasing chromatin dynamics<sup>140</sup>. HAT activity of TIP60 is also required for H4 acetylation

prior to phosphor-H2AX remodelling and dephosphorylation, during DNA damage response<sup>12</sup>. Interestingly, in mammalian cells exposed to ionizing radiation, it was shown that RuvBL1 silencing led to a 50% decrease in RAD51 foci formation, which was shown to be caused by impaired TRRAP-TIP60 function. INO80 and SRCAP/SWR1 activities did not seem to be affected, nor did the DNA damage signalling cascade effected by PIKKs, since the number of  $\gamma$ -H2AX foci was the same in RuvBL1-silenced and control cells<sup>141</sup>. Recently, the group of Claire Davies (2017) demonstrated that RuvBL1 is methylated by the PRMT5 arginine methyltransferase. This is shown to be critical for the acetyltransferase activity of TIP60, and leads to H4K16 acetylation. This facilitates 53BP1 displacement from DSBs, thus regulating homologous recombination<sup>142</sup>.

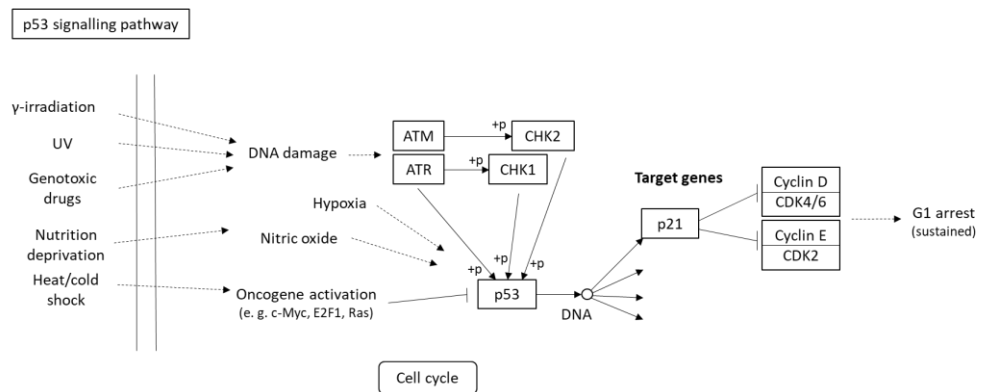
The TIP60 complex is mostly found associated with genes that have increased expression levels, where RNA Polymerase II is co-localized<sup>143</sup>. TIP60 is also responsible for p53 acetylation, necessary for its binding to promoters of pro-apoptotic genes<sup>12</sup>.

The **TRRAP** protein associates with several histone acetyltransferase complexes, such as SAGA, GCN5, PCAF and TIP60, working as a scaffold in their assembly, and as an intermediary for their regulation. Transcription regulator c-MYC binds to TRRAP in the HAT complexes, and induces acetylation of histones in target genes, in response to mitogenic signals. ChIP assays show that c-MYC recruits at least five subunits of this complex: TIP60, TRRAP, p400, BAF53, actin, RuvBL1 and RuvBL2<sup>103,144</sup>. The TIP60 complex is co-bound at about 50-60% of Myc/Max binding sites in mouse embryonic stem cells, and this seems to be directly involved in regulating the c-Myc-dependent transcriptional network<sup>143</sup>.

The activity of the mammalian TIP60 complex seems to be a combination of the activities of yeast NuA4 and SWR1 chromatin remodelling complexes. In fact, a heterogeneous population of TIP60 complexes has been found to date, suggesting a dynamic assembly, with a few varying subunits, among them

**p400**<sup>143</sup>. Like the human protein SRCAP, p400 is a human homologue of yeast Swr1 that is found in the human TIP60 complex and which also promotes the H2A.Z exchange into the promoter regions of p53 target genes. However, p400 and SRCAP seem to be mutually exclusive, and are usually found in different chromatin regions during specific cellular processes<sup>52</sup>, suggesting a plasticity in complex formation according to the needs of the cell.

RuvBL1 and RuvBL2 are common subunits of the TIP60 and SRCAP/p400 complexes, and were found in both small and large versions of the latter. The RuvBL1/2 complex was interestingly found to have the catalytic activity responsible for the H2A to H2A.Z exchange capabilities of the small SRCAP/p400 complex. This did not apply to the incorporation of H2A or H2A.X, suggesting H2A.Z-specific activity of the RuvBL1/2 complex. The mechanisms are as yet unknown, but these findings suggest yet another function for RuvBL complexes, in the targeted incorporation of H2A.Z histone variants, namely in the p21 promoter, at the p53 response element<sup>52</sup> (**Fig. 1.13**).



**Figure 1.13** – Part of the *Homo sapiens* p53 signalling pathway. Based on the KEGG pathway database<sup>145</sup>.

*Interaction with Yin Yang 1*

**Yin Yang 1** (YY1) is a transcription factor belonging to the Polycomb group which regulates a myriad of important cellular events. López-Perrote and colleagues found that the RuvBL1/2 complex interacts with YY1 during DNA damage repair, and that this interaction is important for the formation of Rad51 filaments foci, during homologous recombination. YY1 were shown to interact preferentially with RuvBL1, and Rad51 foci formation was shown to be dependent on RuvBL2 ATPase activity<sup>146</sup>.

### 1.3 RUVBLs IN DEVELOPMENT AND DISEASE

RuvBLs have been associated with several diseases, although little is still known about their mechanisms of action and regulation. The group of Bernhard Schermer has identified both proteins as interactors of NPHP1, a protein involved in the pathogenesis of nephronophthisis (cystic kidney disease), and of other **ciliopathy** proteins. Depletion of RuvBL1 in kidney tubular epithelial cells of a mouse model led to perinatal mortality due to severe cystic kidney disease, supporting their role in the molecular pathogenesis of cystic kidney disease and related ciliopathies<sup>97</sup>. In the zebrafish model, a mutation leading to RuvBL2 loss of function leads to kidney cyst formation and ventral body curvature, two phenotypes related to ciliary defects, and lead to Primary (genetic) Ciliary Dyskinesia. Further investigation showed that RuvBL2 interacts with cilia-related genes, and is essential for normal ciliary functions<sup>147</sup>. Curiously, *ruvbl2* transcription is up-regulated in *Chlamydomonas* during flagellum regeneration<sup>148</sup>.

### 1.3.1 DEVELOPMENT

RuvBLs are intimately linked to developmental processes, and their importance is felt at the onset of foetal development, as demonstrated by the fact that RuvBL1 and RuvBL2 null mice mutants are not viable. In fact, RuvBLs are essential for viability in all organisms analysed to date, including *Drosophila melanogaster*<sup>6</sup> and *Caenorhabditis elegans*. In zebrafish and *Xenopus*, knockdown with antisense morpholino oligonucleotides led to embryonic lethality<sup>149</sup>. Dutta and colleagues have also demonstrated that the ATPase activities of RuvBL1 and RuvBL2 were equally and independently essential for viability in *Saccharomyces cerevisiae*<sup>76</sup>; furthermore, a RuvBL1 plasmid could not rescue a RuvBL2-deficient strain of *S. cerevisiae*, further supporting that the two genes have non-redundant functions<sup>7</sup>. In early embryonic development, RuvBL2 interacts with Oct4, a key regulator for the development of the inner cell mass and for the self-renewal of embryonic stem cells (ESCs), to regulate pluripotency in ESCs and iPSCs (induced pluripotent stem cells)<sup>150,151</sup>.

Gain-of-function studies in *Xenopus* show that xRuvBL1 and xRuvBL2 overexpression lead to increased cell proliferation, and consequent convexity at the injection site. Contrariwise, knockdown led to reduced mitoses. Further studies showed that xRuvBL1 and xRuvBL2 act through interaction with c-Myc/Miz-1 to control cell proliferation during development. More specifically, the N-termini of both RuvBLs were shown to be essential for interaction with c-Myc, p21 repression and mitogenic function<sup>149</sup>.

In zebrafish embryos, RuvBL1 and RuvBL2 were found to have antagonistic roles in heart development, and a reduction in RuvBL1 leads to heart hyperplasia. Concordantly, the *liebeskummer (lik)* mutation in RuvBL2, which increases its ATPase activity and renders it DNA-independent, has the same effect. This mutation (which consists of an insertion of the amino acid residues

Phe, Cys and Arg within the OB fold) enhances the repressing function of the  $\beta$ -catenin/TCF-dependent *siamois* promoter. Thus, the balance of RuvBL1 and RuvBL2 regulates cardiac (and gut) growth in the early embryo<sup>98,152</sup>.

Some parasites such as *Plasmodium vivax*, *P. falciparum*, *P. knowlesi*, *Trypanosoma cruzi* and *Schistosoma mansoni* possess 3 RuvB-Like proteins. In the case of *P. falciparum*, a phylogenetic analysis revealed that *pfRuvBL1* and *pfRuvBL2* share more similarities to human and yeast RuvBL1, while *pfRuvBL3* was more similar to RuvBL2<sup>153</sup>. It is interesting to note this apparent exclusivity to human parasites. In the case of *P. falciparum*, RuvBL2 and 3 seem to interact at the intraerythrocytic mitosis, serving only to increase appreciably the helicase activity of RuvBL2<sup>154</sup>.

### 1.3.2 CANCER

It is clear at this point that chromatin remodellers play a vital role in maintaining the integrity of the genome.

RuvBL2 overexpression has been found to be a marker of poor prognosis, and it was found to be overexpressed in hepatocellular carcinoma (HCC, both RuvBL1 and RuvBL2), stomach, kidney, colon cancer, melanoma and bladder carcinoma, and involved in others. In HCC cells, RuvBL2 overexpression was not associated with changes in sequence in the cases analysed, and led to an increased presence in the cytoplasm. Silencing RuvBL2 in these cells led to an increase in the expression of several pro-apoptotic genes and reduced cell growth<sup>155</sup>. More recently, it was found that in HCC, RuvBL2 silencing led to a decrease in H2AX phosphorylation, in parallel with a decrease in protein expression of ATM and DNA-PKcs kinases (the major H2AX kinases), as well as a decrease in the number of BRCA1 and 53BP1 foci. It is suggested that RuvBL2 overexpression in HCC may be a factor of **resistance to treatment**, as it has been observed in subgroups

of **chemo-resistant breast and ovarian cancers**<sup>94</sup>. Later, it was determined that depletion of RuvBL1 led to similar effects, and in fact, the stability of one depended on the presence of the other, supporting a collaborative role, probably at the level of protein translation, when heteromers may likely be formed. Interestingly, these effects were observed in other cell types and species, suggesting that RuvBL1/2 co-depletion may be widespread<sup>93</sup>.

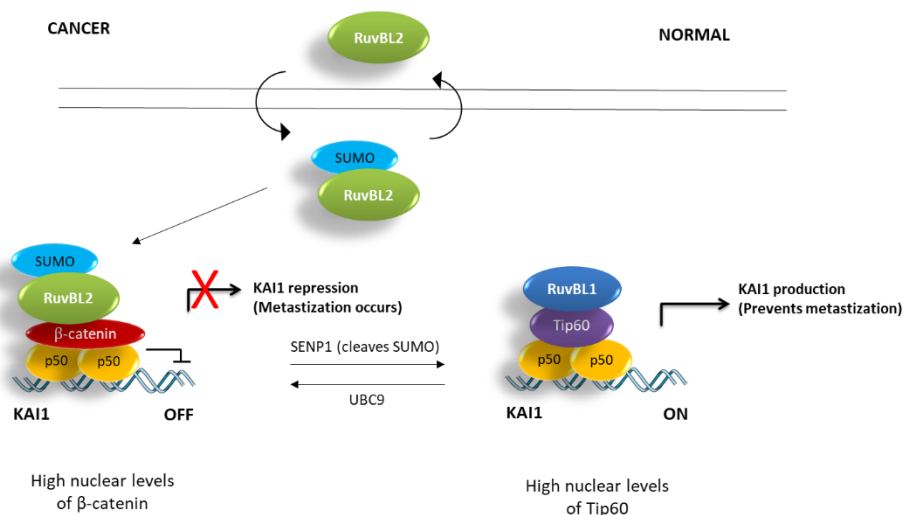
In E2f-dependent HCC, the E2f1 transcription factor recruits the RuvBL1/2 complex to E2f1 target genes, enhancing the transcriptional response. It is suggested that the mechanism involved relies on the ability of RuvBL1/2 complex to exchange H2A/H2AZ by itself<sup>52,156</sup>.

Additionally, RuvBL1 has been identified as a key modulator of apoptosis and oncogenesis through c-MYC and E2f1 interactions<sup>157</sup>, and was also shown to have a role in neoplastic transformation mediated by  $\beta$ -catenin (inhibition of RuvBL1 was linked to inhibition of histone acetylation of  $\beta$ -catenin target genes, which in turn led to an arrest in expression of  $\beta$ -catenin/TCF (T-cell factor) target genes)<sup>158</sup>. RuvBL2 was found to have an antagonistic activity<sup>6</sup>. In colon cancer cells that contain wild type p53, overexpression of EHF transcription factor leads to RuvBL1 upregulation, with a concomitant repression of p53 and consequent repression of apoptosis<sup>159</sup>.

RuvBL2 silencing also proved to have an **anti-proliferative** effect in cells from gastric and kidney cancer (RCC). In RCC, in fact, RuvBL2 displayed a behaviour similar to that in HCC, and further promoted cell migration and invasion, with RuvBL2 depletion leading to the same physiological effects as in HCC<sup>160</sup>. In gastric cancer cells, RuvBL2 was found to have a dual role in the transcription of *htert* gene: the RuvBL2/c-MYC complex-dependent association to the promoter, and regulation of hTERT mRNA, which decreased with silencing of both RuvBL1 and RuvBL2. Targeting of RuvBL2 is thus suggested as a therapeutic means to regulate Telomerase<sup>161</sup>. In colon cancer cells, RuvBL2 was



also found to be essential for *hTERT* full transcriptional activation, but dependent on the growth factor EGF<sup>162</sup>. In mammary HC11 cells, RuvBL1 and RuvBL2 were able to increase the oncogenic effect of c-MYC, by increasing its ability to repress the C/EBP $\delta$  promoter, an effect which was further increased when both proteins were coexpressed<sup>163</sup>. In HEK 293 cells, RuvBL1 was found to interact directly with c-MYC, through the MBII domain of the latter and the domain II of the former<sup>7</sup>. RuvBL2 was further found to act independently of the p53 pathway, by repressing the ARF (alternate reading frame, p14) tumour suppressor<sup>164</sup>. In paediatric acute myeloid leukaemia, chromosomal translocations frequently originate the fusion gene *MLL-AF9*. In immortalized human cord blood-derived hematopoietic progenitor cells, this led to an increased expression of RuvBL2, on which the cells depended for proliferation and survival<sup>165</sup>. In neuroblastoma cells, drug-induced apoptosis was shown to occur in a RuvBL2-dependent way. The end result was accumulation of acetylated H3, cell cycle arrest in G2/M phase and



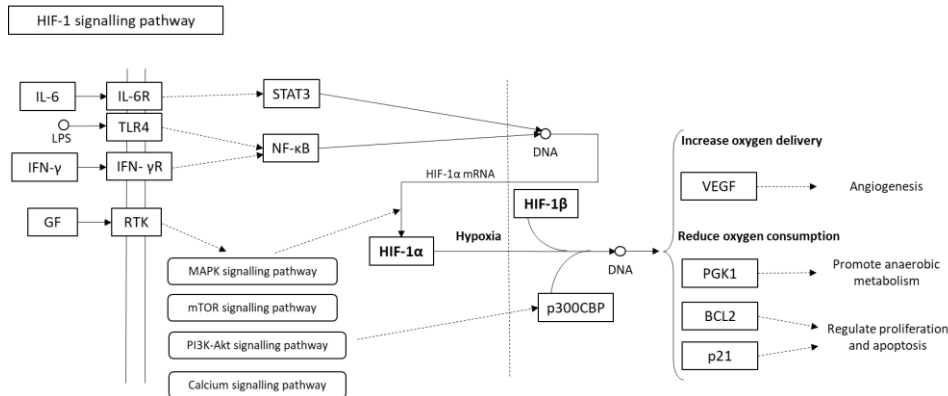
**Figure 1.14 – RuvBL1 (Pontin) and RuvBL2 (Reptin) work antagonistically to regulate expression of anti-metastatic protein KAI1. From<sup>61</sup>.**

activation of apoptosis cell signalling pathways<sup>166</sup>.

In prostate cancer cells, a subpopulation of RuvBL2 is SUMOylated, which results in its recruitment to the nucleus (**Fig. 1.14**). There, it forms a complex with the Wnt-signalling component  $\beta$ -catenin which represses the expression of anti-metastatic gene *KAI1*. This repressing function is abrogated with de-SUMOylation of RuvBL2, as well as by the expression of K456R mutant, which precludes SUMO-binding, and thus prevents nuclear re-localization. Hence, RuvBL2 SUMOylation seems to be an active control mechanism for the transcriptional repressive function of RuvBL2, by recruiting it to the nucleus. It follows that levels of Ubc9 and SENP1, the proteins responsible for RuvBL2 SUMOylation, seem to affect the *KAI1*-mediated metastatic potential, at least partly by altering the SUMOylation state of RuvBL2. De-SUMOylation of RuvBL2 in turn leads to recruitment of coactivator Tip60 to the promoter, activating the expression of *KAI1*. RuvBL1, as part of the TIP60 complex, acetylates histones at the promoter of the *KAI1* gene, leading to *KAI1* expression. It is thus suggested that RuvBL2 SUMOylation state may affect binding affinity of proteins for promoters<sup>12,60,167</sup>.

### 1.3.3 HYPOXIA

Hypoxia is a common feature in solid tumours due to decreased vasculature formation and deficient oxygen perfusion. Hypoxic cells have thus developed adaptations to hypoxic microenvironments, through expression of hypoxia-inducible transcription factor (HIF)-dependent genes. In particular, when under low oxygen concentrations, the oxygen-sensitive HIF-1 $\alpha$  subunit is stabilised and binds to oxygen-insensitive HIF-1 $\beta$  subunit. The HIF-1 $\alpha\beta$  heterodimer then binds to hypoxia-responsive elements located in enhancer



**Figure 1.15** - Part of the *Homo sapiens* HIF-1 signalling pathway. From the KEGG pathway database<sup>145</sup>.

regions, activating the transcription of target genes involved in metabolic pathways responsible for increasing adaptation to microaerobic environments, chemotherapeutic resistance, genomic instability, cell proliferation and metastasis<sup>4,168,169</sup> (**Fig. 1.15**).

Besides low oxygen levels, other factors may modulate HIF-1 activity, such as the presence of other oncogenes and tumour suppressors<sup>169</sup>. RuvBL1 and RuvBL2 have both been identified as two such factors, where RuvBL2 acts as a repressor of transcription and RuvBL1 as a co-activator, an example of the antagonistic roles played by these homologs.

During hypoxia, RuvBL1 is methylated by G9a and GLP methyltransferases, which promotes the formation of a complex with p300 and HIF-1 $\alpha$ . Concomitantly, binding of HIF-1 $\alpha$  $\beta$  complex to the promoter is hyperactivated, increasing the expression levels of proteins involved in cell survival and proliferation, tumour growth and invasive properties, such as Ets1 (ETS proto-oncogene 1, a transcription factor involved in stem cell development, cell senescence and death and tumorigenesis), KDM4B lysine demethylase and

IGFBP3 (insulin growth factor binding protein 3)<sup>4</sup>. RuvBL1 methylation is dynamic, and reversed as soon as cells return to normoxic levels.

RuvBL2 negatively regulates expression of HIF-1 target genes involved in metabolism, cell death and survival. The mechanism involved relies on the association of RuvBL2 with HIF-1 $\alpha$  in hypoxic conditions, in a methylation dependent manner: under hypoxic conditions, Lys67 of RuvBL2 is methylated by G9a methyltransferase, which leads to increased binding to HIF-1 $\alpha$ , negatively regulating expression of hypoxia target genes **VEGF** (Vascular Endothelial Growth Factor), **PGK1** (Phosphoglycerate Kinase 1) and **BNIP3** (BCL2 interacting Protein 3). *In vitro* results show that substitution with a RuvBL2 K67A mutant abolished this effect, and supported an interaction with a methylated K67-containing peptide. The RuvBL2-HIF-1 $\alpha$  complex binds to promoters concomitantly with a decrease in RNaseII recruitment. Reptin methylation thus accounts for a negative feedback loop, fine-tuning the expression of a subset of hypoxia target genes. At the cell level, RuvBL2 methylation led to a significant reduction in motility of a breast cancer cell line, implying it is involved in inhibition of cell migration. Knockdown of RuvBL2 or expression of a K67A mutant in the same cells under hypoxic conditions also led to a 2-fold increase in migratory potential, whereas expression of WT RuvBL2 reduced invasion. These results were recapitulated in a mouse model, and suggest that RuvBL2 hypoxia-induced methylation may affect tumour growth and invasive properties by inhibition of genes such as VEGF<sup>5</sup>.

## References

1. Schnell, S. & Turner, T. E. Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Prog. Biophys. Mol. Biol.* **85**, 235–60 (2004).
2. Rivas, G., Fernandez, J. A. & Minton, A. P. Direct observation of the self-association of dilute proteins in the presence of inert macromolecules at high concentration via tracer sedimentation equilibrium: Theory, experiment, and biological significance. *Biochemistry* **38**, 9379–9388 (1999).
3. Rivas, G., Fernández, J. a & Minton, a P. Direct observation of the enhancement of noncooperative protein self-assembly by macromolecular crowding: indefinite linear self-association of bacterial cell division protein FtsZ. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3150–3155 (2001).
4. Lee, J. S. *et al.* Hypoxia-induced methylation of a pontin chromatin remodeling factor. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13510–13515 (2011).
5. Lee, J. S. *et al.* Negative Regulation of Hypoxic Responses via Induced Reptin Methylation. *Mol. Cell* **39**, 71–85 (2010).
6. Bauer, A. *et al.* Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. *EMBO J.* **19**, 6121–30 (2000).
7. Wood, M. A., McMahon, S. B. & Cole, M. D. An ATPase/Helicase Complex Is an Essential Cofactor for Oncogenic Transformation by c-Myc. *Mol. Cell* **5**, 321–330 (2000).
8. Gerhold, C. B., Hauer, M. H. & Gasser, S. M. INO80-C and SWR-C: Guardians of the genome. *J. Mol. Biol.* **427**, 637–651 (2015).
9. Nguyen, V. Q. *et al.* Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. *Cell* **154**, 1220–31 (2013).
10. Jha, S., Gupta, A., Dar, A. & Dutta, A. RVBs are required for assembling a functional TIP60 complex. *Mol. Cell. Biol.* **33**, 1164–74 (2013).
11. Machado-Pinilla, R., Liger, D., Leulliot, N. & Meier, U. T. Mechanism of the AAA+ ATPases pontin and reptin in the biogenesis of H/ACA RNPs. *Rna* **18**, 1833–1845 (2012).
12. Nano, N. & Houry, W. A. Chaperone-like activity of the AAA+ proteins Rvb1 and Rvb2 in the assembly of various complexes. *Philos Trans R Soc L. B Biol Sci* **368**, 20110399 (2013).
13. McKeegan, K. S., Debieux, C. M. & Watkins, N. J. Evidence that the AAA+ proteins TIP48 and TIP49 bridge interactions between 15.5K and the related NOP56 and NOP58 proteins during box C/D snoRNP biogenesis. *Mol. Cell. Biol.* **29**, 4971–4981 (2009).
14. Gomez-Navarro, N. & Miller, E. Protein sorting at the ER–Golgi interface. *J. Cell Biol.* 1–10 (2016).
15. Vertii, A., Hehnly, H. & Doxsey, S. The Centrosome, a Multitalented

- Renaissance Organelle. *Cold Spring Harb. Perspect. Biol.* **8**, a025049 (2016).
16. Liyanage, S. U. *et al.* Characterizing the mitochondrial DNA polymerase gamma interactome by BioID identifies Ruvbl2 localizes to the mitochondria. *Mitochondrion* **32**, 31–35 (2016).
  17. Corless, S. & Gilbert, N. Effects of DNA supercoiling on chromatin architecture. *Biophys. Rev.* 245–258 (2016). doi:10.1007/s12551-016-0210-1
  18. Fujii, Y., Wakamori, M., Umehara, T. & Yokoyama, S. Crystal structure of human nucleosome core particle containing enzymatically introduced CpG methylation. *FEBS Open Bio* **6**, 1–17 (2016).
  19. Wu, C., McGeehan, J. E. & Travers, A. A metastable structure for the compact 30-nm chromatin fibre. *FEBS Lett.* **590**, 935–942 (2016).
  20. Chua, E. Y. D. *et al.* 3.9 Angstrom structure of the nucleosome core particle determined by phase-plate cryo-EM. **44**, 8013–8019 (2016).
  21. Rogakou, E. *et al.* Histone H2A variants H2AX and H2AZ. *Curr. Opin. Genet. Dev.* **12**, 162–169 (2002).
  22. Rudnizky, S. *et al.* H2A.Z controls the stability and mobility of nucleosomes to regulate expression of the LH genes. *Nat. Commun.* **7**, 12958 (2016).
  23. Seo, J. *et al.* Genome-wide reorganization of histone H2AX toward particular fragile sites on cell activation. *Nucleic Acids Res.* **42**, 1016–1025 (2014).
  24. Yu, Y., Deng, Y., Reed, S. H., Millar, C. B. & Waters, R. Histone variant Htz1 promotes histone H3 acetylation to enhance nucleotide excision repair in Htz1 nucleosomes. *Nucleic Acids Res.* **41**, 9006–9019 (2013).
  25. Wendler, P., Ciniawsky, S., Kock, M. & Kube, S. Structure and function of the AAA+ nucleotide binding pocket. *Biochim. Biophys. Acta* **1823**, 2–14 (2012).
  26. Dougan, D. A., Mogk, A., Zeth, K., Turgay, K. & Bukau, B. AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett.* **529**, 6–10 (2002).
  27. Erzberger, J. P. & Berger, J. M. Evolutionary Relationships and Structural Mechanisms of Aaa+ Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114 (2006).
  28. Matias, P. M., Gorynia, S., Donner, P. & Carrondo, M. A. Crystal structure of the human AAA+ protein RuvBL1. *J Biol Chem* **281**, 38918–38929 (2006).
  29. Snider, J., Thibault, G. & Houry, W. a. The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **9**, 216 (2008).
  30. Erzberger, J. P., Mott, M. L. & Berger, J. M. Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.* **13**, 676–683 (2006).
  31. Zhang, X. & Wigley, D. B. The ‘glutamate switch’ provides a link between ATPase activity and ligand binding in AAA+ proteins. *Nat. Struct. Mol.*

- Biol.* **15**, 1223–1227 (2008).
32. Zhao, Z. *et al.* An Arginine Finger Regulates the Sequential Action of Asymmetrical Hexameric ATPase in the Double-Stranded DNA Translocation Motor. *Mol. Cell. Biol.* **36**, 2514–2523 (2016).
  33. Saraste, M., Sibbald, P. R. & Wittinghofer, A. The P-loop - a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434 (1990).
  34. Rose, A. E., Brown, R. S. H. & Schlieker, C. Torsins: not your typical AAA+ ATPases. *Crit. Rev. Biochem. Mol. Biol.* **50**, 532–549 (2015).
  35. Wu, Y. Unwinding and rewinding: Double faces of helicase? *J. Nucleic Acids* **2012**, (2012).
  36. Jankowsky, E. & Margaret, E. An Introduction to RNA Helicases: Superfamilies, Families, and Major Themes. *RSC Biomol. Sci.* **19**, RNA He, (2010).
  37. Tuteja, N. & Tuteja, R. Prokaryotic and eukaryotic DNA helicases: Essential molecular motor proteins for cellular machinery. *Eur. J. Biochem.* **271**, 1835–1848 (2004).
  38. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
  39. Gilhooly, N. S., Gwynn, E. J. & Dillingham, M. S. Superfamily 1 helicases. *Front. Biosci. (Schol. Ed.)* **5**, 206–16 (2013).
  40. Byrd, A. K. & Raney, K. D. Superfamily 2 helicases. *Front. Biosci. (Landmark Ed.)* **17**, 2070–88 (2012).
  41. Conaway, R. C. & Conaway, J. W. The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends Biochem. Sci.* **34**, 71–7 (2009).
  42. Iyer, L. M., Aravind, L. & Koonin, E. V. Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. **75**, 11720–11734 (2001).
  43. Gorhalenya, A. E., Koonin, E. V & Wolf, Y. I. A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett.* **262**, 145–148 (1990).
  44. Banerjee, S., Chalissery, J., Bandey, I. & Sen, R. Rho-dependent transcription termination: more questions than answers. *J. Microbiol.* **44**, 11–22 (2006).
  45. Abid Ali, F. *et al.* Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nat. Commun.* **7**, 10708 (2016).
  46. Lakomek, K., Stoehr, G., Tosi, A., Schmailzl, M. & Hopfner, K. P. Structural basis for dodecameric assembly states and conformational plasticity of the full-length AAA+ ATPases Rvb1 . Rvb2. *Structure* **23**, 483–495 (2015).

47. Flaus, A. & Owen-Hughes, T. Mechanisms for ATP-dependent chromatin remodeling: the means to the end. *FEBS J.* **278**, 3579–3595 (2011).
48. Umate, P., Tuteja, N. & Tuteja, R. Genome-wide comprehensive analysis of human helicases. *Commun. Integr. Biol.* **4**, 1–20 (2011).
49. Sancar, A. & Hearst, J. E. Molecular Matchmakers. *Science (80-. )*. **259**, 1415–1420 (1993).
50. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
51. Sauer, R. T. & Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **80**, 587–612 (2011).
52. Choi, J., Heo, K. & An, W. Cooperative action of TIP48 and TIP49 in H2A.Z exchange catalyzed by acetylation of nucleosomal H2A. *Nucleic Acids Res.* **37**, 5993–6007 (2009).
53. Kakihara, Y. & Saeki, M. The R2TP chaperone complex: its involvement in snoRNP assembly and tumorigenesis. *Biomol Concepts* **5**, 513–520 (2014).
54. Massenet, S., Bertrand, E. & Verheggen, C. Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol.* **6286**, 00–00 (2016).
55. King, T. H., Decatur, W. a, Bertrand, E., Maxwell, E. S. & Fournier, M. J. A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. *Mol. Cell. Biol.* **21**, 7731–7746 (2001).
56. McKeegan, K. S., Debieux, C. M., Boulon, S., Bertrand, E. & Watkins, N. J. A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Mol. Cell. Biol.* **27**, 6782–93 (2007).
57. Kakihara, Y., Makhnevych, T., Zhao, L., Tang, W. & Houry, W. a. Nutritional status modulates box C/D snoRNP biogenesis by regulated subcellular relocalization of the R2TP complex. *Genome Biol.* **15**, 404 (2014).
58. Venteicher, A. S., Meng, Z., Mason, P. J., Veenstra, T. D. & Artandi, S. E. Identification of ATPases Pontin and Reptin as Telomerase Components Essential for Holoenzyme Assembly. *Cell* **132**, 945–957 (2008).
59. Baek, S. H. When ATPases pontin and reptin met telomerase. *Dev. Cell* **14**, 459–61 (2008).
60. Kim, J. H. *et al.* Roles of sumoylation of a reptin chromatin-remodelling complex in cancer metastasis. *Nat. Cell Biol.* **8**, 631–639 (2006).
61. Hee Baek, S. A novel link between SUMO modification and cancer metastasis. *Cell Cycle* **5**, 1492–1495 (2006).
62. Zaarur, N. *et al.* RuvbL1 and RuvbL2 enhance aggresome formation and disaggregate amyloid fibrils. *EMBO J* **34**, 2363–2382 (2015).
63. Olzmann, J. a, Li, L. & Chin, L. S. Aggresome formation and neurodegenerative diseases: therapeutic implications. *Curr. Med. Chem.* **15**, 47–60 (2008).



64. Kanemaki, M. *et al.* Molecular cloning of a rat 49-kDa TBP-interacting protein (TIP49) that is highly homologous to the bacterial RuvB. *Biochem. Biophys. Res. Commun.* **235**, 64–8 (1997).
65. Makino, Y., Kanemaki, M., Kurokawa, Y., Koji, T. & Tamura, T. A. A rat RuvB-like protein, TIP49a, is a germ cell-enriched novel DNA helicase. *J. Biol. Chem.* **274**, 15329–15335 (1999).
66. He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 1–21 (2016).
67. Kanemaki, M. *et al.* TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *J. Biol. Chem.* **274**, 22437–22444 (1999).
68. Kurokawa, Y., Kanemaki, M., Makino, Y. & Tamura, T. A. A notable example of an evolutionary conserved gene: studies on a putative DNA helicase TIP49. *DNA Seq* **10**, 37–42 (1999).
69. Wood, M. A., McMahon, S. B. & Cole, M. D. An ATPase/Helicase Complex Is an Essential Cofactor for Oncogenic Transformation by c-Myc. *Mol. Cell* **5**, 321–330 (2000).
70. Ikura, T. *et al.* Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* **102**, 463–473 (2000).
71. Lim, C. R. *et al.* The *Saccharomyces cerevisiae* RuvB-like protein, Tih2p, is required for cell cycle progression and RNA polymerase II-directed transcription. *J. Biol. Chem.* **275**, 22409–22417 (2000).
72. Salzer, U., Kubicek, M. & Prohaska, R. Isolation, molecular characterization, and tissue-specific expression of ECP-51 and ECP-54 (TIP49), two homologous, interacting erythroid cytosolic proteins. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1446**, 365–370 (1999).
73. Holzmann, K. *et al.* Identification and characterization of the ubiquitously occurring nuclear matrix protein NMP 238. *Biochem. Biophys. Res. Commun.* **252**, 39–45 (1998).
74. Jónsson, Z. O. *et al.* Rvb1p and Rvb2p are Essential Components of a Chromatin Remodeling Complex that Regulates Transcription of over 5% of Yeast Genes. *J. Biol. Chem.* **276**, 16279–16288 (2001).
75. Kikuchi, N. *et al.* Molecular shape and ATP binding activity of rat p50, a putative mammalian homologue of RuvB DNA helicase. *J. Biochem.* **125**, 487–94 (1999).
76. Qiu, X. B. *et al.* An eukaryotic RuvB-like protein (RUVBL1) essential for growth. *J. Biol. Chem.* **273**, 27786–27793 (1998).
77. Dickman, M. J. *et al.* The RuvABC resolvosome. *Eur. J. Biochem.* **269**, 5492–5501 (2002).
78. Papin, C. *et al.* 3'- to 5' DNA unwinding by TIP49b proteins. *Febs J* **277**, 2705–2714 (2010).
79. Yang, X. *et al.* KAI1 protein is down-regulated during the progression of

- human breast cancer. *Clin. Cancer Res.* **6**, 3424–3429 (2000).
80. Yamada, K. *et al.* Crystal structure of the RuvA-RuvB complex: A structural basis for the holliday junction migrating motor machinery. *Mol. Cell* **10**, 671–681 (2002).
  81. López-Perrote, A., Muñoz-Hernández, H., Gil, D. & Llorca, O. Conformational transitions regulate the exposure of a DNA-binding domain in the RuvBL1-RuvBL2 complex. *Nucleic Acids Res.* **40**, 11086–11099 (2012).
  82. Silva-Martin, N. *et al.* The Combination of X-Ray Crystallography and Cryo-Electron Microscopy Provides Insight into the Overall Architecture of the Dodecameric Rvb1/Rvb2 Complex. *PLoS One* **11**, e0146457 (2016).
  83. Lakomek, K., Stoehr, G., Tosi, A., Schmailzl, M. & Hopfner, K.-P. Structural Basis for Dodecameric Assembly States and Conformational Plasticity of the Full-Length AAA+ ATPases Rvb1/Rvb2. *Structure* **23**, 1–13 (2015).
  84. Tosi, A. *et al.* Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* **154**, 1207–19 (2013).
  85. Gorynia, S. *et al.* Structural and functional insights into a dodecameric molecular machine - The RuvBL1/RuvBL2 complex. *J. Struct. Biol.* **176**, 279–291 (2011).
  86. Rynditch, A., Pekarsky, Y., Schnittger, S. & Gardiner, K. Leukemia breakpoint region in 3q21 is gene rich. **193**, 49–57 (1997).
  87. Kashuba, V. *et al.* Not I linking/jumping clones of human chromosome 3: mapping of the TFRC, RAB7 and HAUSP genes to regions rearranged in leukemia and deleted in solid tumors. *FEBS Lett.* **419**, 181–185 (1997).
  88. Parfait, B. *et al.* Human TIP49b/RUVBL2 gene: genomic structure, expression pattern, physical link to the human CGB/LHB gene cluster on chromosome 19q13.3. *Ann. Génétique* **43**, 69–74 (2000).
  89. Grimwood, J. *et al.* The DNA sequence and biology of human chromosome 19. *Nature* **428**, 529–535 (2004).
  90. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (80-. )*. **347**, 1260419–1260419 (2015).
  91. Huber, O. *et al.* Pontin and reptin, two related ATPases with multiple roles in cancer. *Cancer Res* **68**, 6873–6876 (2008).
  92. Grigoletto, A., Lestienne, P. & Rosenbaum, J. The multifaceted proteins Reptin and Pontin as major players in cancer. *Biochim. Biophys. Acta - Rev. Cancer* **1815**, 147–157 (2011).
  93. Haurie, V. *et al.* Adenosine triphosphatase Pontin is overexpressed in hepatocellular carcinoma and coregulated with Reptin through a new posttranslational mechanism. *Hepatology* **50**, 1871–1883 (2009).
  94. Raymond, A.-A. *et al.* Reptin regulates DNA double strand breaks repair in human hepatocellular carcinoma. *PLoS One* **10**, e0123333 (2015).

95. Rosenbaum, J. *et al.* The Emergence of the Conserved AAA+ ATPases Pontin and Reptin on the Signaling Landscape. *Sci. Signal.* **6**, mr1-mr1 (2013).
96. Ménard, L. *et al.* In vivo silencing of Reptin blocks the progression of human hepatocellular carcinoma in xenografts and is associated with replicative senescence. *J. Hepatol.* **52**, 681–689 (2010).
97. Matias, P. M. *et al.* The AAA+ proteins Pontin and Reptin enter adult age: from understanding their basic biology to the identification of selective inhibitors. *Front. Mol. Biosci.* **2**, 17 (2015).
98. Rottbauer, W. *et al.* Reptin and Pontin Antagonistically Regulate Heart Growth in Zebrafish Embryos. *Cell* **111**, 661–672 (2002).
99. Rowe, A., Weiske, J., Kramer, T. S., Huber, O. & Jackson, P. Phorbol Ester Enhances KAI1 Transcription by Recruiting Tip60/Pontin Complexes. *Neoplasia* **10**, 1421–32 (2008).
100. Queval, R., Papin, C., Dalvai, M., Bystricky, K. & Humbert, O. Reptin and Pontin oligomerization and activity are modulated through histone H3 N-terminal tail interaction. *J Biol Chem* **289**, 33999–34012 (2014).
101. Jha, S. & Dutta, A. RVB1/RVB2: running rings around molecular biology. *Mol Cell* **34**, 521–533 (2009).
102. Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K. & Linn, S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu.Rev.Biochem.* **73**, 39–85 (2004).
103. Jha, S. & Dutta, A. RVB1/RVB2: running rings around molecular biology. *Mol Cell* **34**, 521–533 (2009).
104. Taubert, S. *et al.* E2F-dependent histone acetylation and recruitment of the Tip60 acetyltransferase complex to chromatin in late G1. *Mol. Cell. Biol.* **24**, 4546–4556 (2004).
105. Cho, S. *et al.* TIP49b , a Regulator of Activating Transcription Factor 2 Response to Stress and DNA Damage. *Society* **21**, 8398–8413 (2001).
106. Bandejas, T. M. *et al.* Structure of wild-type Plk-1 kinase domain in complex with a selective DARPIn. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 339–353 (2008).
107. Gartner, W. *et al.* The ATP-dependent helicase RUVBL1/TIP49a associates with tubulin during mitosis. *Cell Motil. Cytoskeleton* **56**, 79–93 (2003).
108. Gentili, C. *et al.* Chromosome Missegregation Associated with RUVBL1 Deficiency. *PLoS One* **10**, e0133576 (2015).
109. Sigala, B., Edwards, M., Puri, T. & Tsaneva, I. R. Relocalization of human chromatin remodeling cofactor TIP48 in mitosis. *Exp. Cell Res.* **310**, 357–369 (2005).
110. Diop, S. B. *et al.* Reptin and Pontin function antagonistically with PcG and TrxG complexes to mediate Hox gene control. *EMBO Rep.* **9**, 260–266 (2008).

111. Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* **1**, e1500447–e1500447 (2015).
112. Connelly, K. & Dykhuizen, E. Compositional and Functional Diversity of Canonical PRC1 Complexes in Mammals. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1860**, 233–245 (2016).
113. Lee, J. S. *et al.* Negative regulation of hypoxic responses via induced Reptin methylation. *Mol. Cell* **39**, 71–85 (2010).
114. Von Morgen, P., Horejsi, Z. & Macurek, L. Substrate recognition and function of the R2TP complex in response to cellular stress. *Front. Genet.* **5**, 1–12 (2015).
115. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science (80-. )*. **316**, 1160–1166 (2007).
116. Kakihara, Y. & Houry, W. A. The R2TP complex: discovery and functions. *Biochim. Biophys. Acta* **1823**, 101–7 (2012).
117. Rajendra, E., Garaycochea, J. I., Patel, K. J. & Passmore, L. A. Abundance of the Fanconi anaemia core complex is regulated by the RuvBL1 and RuvBL2 AAA+ ATPases. *Nucleic Acids Res* **42**, 13736–13748 (2014).
118. Izumi, N., Yamashita, A. & Ohno, S. Integrated regulation of PIKK-mediated stress responses by AAA+ proteins RUVBL1 and RUVBL2. *Nucleus* **3**, 29–43 (2012).
119. Bakkenist, C. J. & Kastan, M. B. Initiating cellular stress responses. *Cell* **118**, 9–17 (2004).
120. Jha, S., Gupta, A., Dar, A. & Dutta, A. RVBs are required for assembling a functional TIP60 complex. *Mol. Cell. Biol.* **33**, 1164–74 (2013).
121. Rivera-Calzada, A. *et al.* The Structure of the R2TP Complex Defines a Platform for Recruiting Diverse Client Proteins to the HSP90 Molecular Chaperone System. *Structure* (2017). doi:10.1016/j.str.2017.05.016
122. Hurov, K. E., Cotta-Ramusino, C. & Elledge, S. J. A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. *Genes Dev.* **24**, 1939–1950 (2010).
123. Vaughan, C. K. Hsp90 picks PIKKs via R2TP and Tel2. *Structure* **22**, 799–800 (2014).
124. Horejsi, Z. *et al.* CK2 phospho-dependent binding of R2TP complex to TEL2 is essential for mTOR and SMG1 stability. *Mol. Cell* **39**, 839–850 (2010).
125. Verheggen, C., Pradet-balade, B. & Bertrand, E. SnoRNPs , ZNHIT proteins and the R2TP pathway. **6**, 1–2
126. Boulon, S. *et al.* HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Mol. Cell* **39**, 912–924 (2010).

127. Benbahouche, N. E. H. *et al.* Drosophila spag is the homolog of RNA polymerase II-associated protein 3 (RPAP3) and recruits the heat shock proteins 70 and 90 (Hsp70 and Hsp90) during the assembly of cellular machineries. *J. Biol. Chem.* **289**, 6236–6247 (2014).
128. Tsukuda, T., Fleming, A. B., Nickoloff, J. A. & Osley, M. A. Chromatin remodelling at a DNA double-strand break site in *Saccharomyces cerevisiae*. *Nature* **438**, 379–383 (2005).
129. Henderson, L. E. *et al.* A Histone Acetylation Switch Regulates H2A.Z Deposition by the SWR-C Remodeling Enzyme. 195–200 (2013).
130. Bao, Y. & Shen, X. SnapShot: Chromatin remodeling: INO80 and SWR1. *Cell* **144**, 158–158.e2 (2011).
131. Shimada, K. *et al.* Ino80 Chromatin Remodeling Complex Promotes Recovery of Stalled Replication Forks. *Curr. Biol.* **18**, 566–575 (2008).
132. Gospodinov, a. *et al.* Mammalian Ino80 Mediates Double-Strand Break Repair through Its Role in DNA End Strand Resection. *Mol. Cell. Biol.* **31**, 4735–4745 (2011).
133. Morrison, A. J. & Shen, X. Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nat Rev Mol Cell Biol* **10**, 373–384 (2009).
134. Nguyen, V. Q. *et al.* Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. *Cell* **154**, 1220–1231 (2013).
135. Zhou, C. Y. *et al.* Regulation of Rvb1/Rvb2 by a Domain within the INO80 Chromatin Remodeling Complex Implicates the Yeast Rvbs as Protein Assembly Chaperones. *Cell Rep.* **19**, 2033–2044 (2017).
136. Alatwi, H. E. & Downs, J. A. Removal of H2A.Z by INO80 promotes homologous recombination. *EMBO Rep.* **16**, 986–994 (2015).
137. Lafon, A. *et al.* INO80 Chromatin Remodeler Facilitates Release of RNA Polymerase II from Chromatin for Ubiquitin-Mediated Proteasomal Degradation. *Mol. Cell* **60**, 784–796 (2015).
138. Wang, F., Ranjan, A., Wei, D. & Wu, C. Comment on ‘ A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme ’. *Science (80-. )*. **353**, 358–b (2016).
139. Wang b, F., Ranjan, A., Wei, D. & Wu, C. Comment on ‘ A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme ’. *Science (80-. )*. **353**, 358–b (2016).
140. Ikura, T. *et al.* DNA Damage-Dependent Acetylation and Ubiquitination of H2AX Enhances Chromatin Dynamics. *Mol. Cell. Biol.* **27**, 7028–7040 (2007).
141. Gospodinov, A., Tsaneva, I. & Anachkova, B. RAD51 foci formation in response to DNA damage is modulated by TIP49. *Int. J. Biochem. Cell Biol.* **41**, 925–933 (2009).
142. Clarke, T. L. *et al.* PRMT5-Dependent Methylation of the TIP60 Coactivator RUVBL1 Is a Key Regulator of Homologous Recombination. *Mol. Cell* **65**, 900–916.e7 (2017).

143. Ravens, S., Yu, C., Ye, T., Stierle, M. & Tora, L. Tip60 complex binds to active Pol II promoters and a subset of enhancers and co-regulates the c-Myc network in mouse embryonic stem cells. *Epigenetics Chromatin* **8**, 45 (2015).
144. Frank, S. R. *et al.* MYC recruits the TIP60 histone acetyltransferase complex to chromatin. *EMBO Rep.* **4**, 575–80 (2003).
145. Doyon, Y. & Côté, J. The highly conserved and multifunctional NuA4 HAT complex. *Curr. Opin. Genet. Dev.* **14**, 147–154 (2004).
146. Lopez-Perrote, A. *et al.* Structure of Yin Yang 1 oligomers that cooperate with RuvBL1-RuvBL2 ATPases. *J Biol Chem* **289**, 22614–22629 (2014).
147. Zhao, L. *et al.* Reptin/Ruvbl2 is a Lrrc6/Seahorse interactor essential for cilia motility. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12697–702 (2013).
148. Stolc, V., Samanta, M. P., Tongprasit, W. & Marshall, W. F. Genome-wide transcriptional analysis of flagellar regeneration in *Chlamydomonas reinhardtii* identifies orthologs of ciliary disease genes. *Proc Natl Acad Sci U S A* **102**, 3703–3707 (2005).
149. Etard, C., Gradl, D., Kunz, M., Eilers, M. & Wedlich, D. Pontin and Reptin regulate cell proliferation in early *Xenopus* embryos in collaboration with c-Myc and Miz-1. *Mech. Dev.* **122**, 545–556 (2005).
150. Do, E. *et al.* Reptin Regulates Pluripotency of Embryonic Stem Cells and Somatic Cell Reprogramming Through Oct4-Dependent Mechanism. *Stem Cells* **0**, 3126–3136 (2014).
151. Boo, K. *et al.* Pontin functions as an essential coactivator for Oct4-dependent lincRNA expression in mouse embryonic stem cells. *Nat. Commun.* **6**, 6810 (2015).
152. Petukhov, M. *et al.* Large-scale conformational flexibility determines the properties of AAA+ TIP49 ATPases. *Structure* **20**, 1321–1331 (2012).
153. Ahmad, M., Afrin, F. & Tuteja, R. Identification of R2TP complex of *Leishmania donovani* and *Plasmodium falciparum* using genome wide in-silico analysis. *Commun. Integr. Biol.* **6**, (2013).
154. Ahmad, M. & Tuteja, R. *Plasmodium falciparum* RuvB2 translocates in 5'-3' direction, relocalizes during schizont stage and its enzymatic activities are up regulated by RuvB3 of the same complex. *Biochim. Biophys. Acta* **1834**, 2795–2811 (2013).
155. Rousseau, B. *et al.* Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma. *Hepatology* **46**, 1108–18 (2007).
156. Tarangelo, A. *et al.* Recruitment of Pontin/Reptin by E2f1 amplifies E2f transcriptional response during cancer progression. *Nat. Commun.* **6**, 10028 (2015).
157. Dugan, K. A., Wood, M. A. & Cole, M. D. TIP49, but not TRRAP, modulates c-Myc and E2F1 dependent apoptosis. *Oncogene* **21**, 5835–43 (2002).

- (2002).
158. Feng Y, Lee N, F. E. TIP49 regulates B-catenin-mediated neoplastic transformation and T-cell factor target gene induction via effects on chromatin remodeling. *Cancer Res* **63:8726–34**, 8726–8734 (2003).
  159. Taniue, K., Oda, T., Hayashi, T., Okuno, M. & Akiyama, T. A member of the ETS family, EHF, and the ATPase RUVBL1 inhibit p53-mediated apoptosis. *EMBO Rep.* **12**, 682–9 (2011).
  160. Ren, J. *et al.* Overexpression of Reptin in renal cell carcinoma contributes to tumor malignancies and its inhibition triggers senescence of cancer cells. *Urol. Oncol. Semin. Orig. Investig.* **31**, 1358–1366 (2013).
  161. Li, W. *et al.* Reptin is required for the transcription of telomerase reverse transcriptase and over-expressed in gastric cancer. *Mol. Cancer* **9**, 132 (2010).
  162. Flavin, P. *et al.* RuvBl2 cooperates with Ets2 to transcriptionally regulate hTERT in colon cancer. *FEBS Lett.* **585**, 2537–44 (2011).
  163. Si, J., Yu, X., Zhang, Y. & DeWille, J. W. Myc interacts with Max and Miz1 to repress C/EBPdelta promoter activity and gene expression. *Mol. Cancer* **9**, 92 (2010).
  164. Xie, C., Wang, W., Yang, F., Wu, M. & Mei, Y. RUVBL2 is a novel repressor of ARF transcription. *FEBS Lett.* **586**, 435–41 (2012).
  165. Osaki, H. *et al.* The AAA+ ATPase RUVBL2 is a critical mediator of MLL-AF9 oncogenesis. *Leukemia* **27**, 1461–1468 (2013).
  166. Zhan, Q. *et al.* RuvBL2 Is Involved in Histone Deacetylase Inhibitor PCI-24781-Induced Cell Death in SK-N-DZ Neuroblastoma Cells. *PLoS One* **8**, (2013).
  167. Kim, J. H. *et al.* Transcriptional regulation of a metastasis suppressor gene by Tip60 and beta-catenin complexes. *Nature* **434**, 921–926 (2005).
  168. Brown, J. M. in *Methods in enzymology* **435**, 297–321 (2007).
  169. Chan, D. A., Krieg, A. J., Turcotte, S. & Giaccia, A. J. HIF gene expression in cancer therapy. *Methods Enzymol.* **435**, 323–45 (2007).

# *Part 2*

---



# *Chapter 2*

---

**OLIGOMERIZATION AND  
DNA-BINDING OF RUVBL2.  
PRELIMINARY STUDY OF  
BINDING PARTNER C-MYC.**

INTRODUCTION TO ssDNA-BINDING PROTEINS	69
<b>2.1 <i>hsRuvBL2</i> stability, oligomerization and DNA binding .....</b>	<b>72</b>
2.1.1 METHODS	72
Protein expression and purification	72
Differential scanning fluorimetry	73
Size exclusion chromatography coupled to multi-angle laser light scattering	75
Analytical ultracentrifugation	76
Small angle X-ray scattering	77
Electrophoretic mobility shift assays (EMSA)	78
Analysis of <i>hsRuvBL2</i> binding to DNA by negative staining EM	79
2.1.2 RESULTS AND DISCUSSION	80
Purification tags and nucleotides affect <i>hsRuvBL2</i> stability and oligomerization state	80
Stability of <i>hsRuvBL2</i> oligomers varies with concentration	86
Insights into <i>hsRuvBL1</i> and <i>hsRuvBL2</i> binding to DNA	91
<b>2.2 Preliminary study of RuvBL-interacting protein c-Myc.....</b>	<b>96</b>
2.2.1 METHODS	97
Construct design	97
Cloning procedures	99
Protein expression and purification	100
2.2.2 RESULTS AND DISCUSSION	102
<b>References.....</b>	<b>104</b>

## **INTRODUCTION TO ssDNA-BINDING PROTEINS**

In order to copy or repair DNA strand breaks, the double helix must be opened to expose the two complementary strands. The need to process single stranded DNA (ssDNA) led to the onset of a specialized group of proteins. To understand how they act, and to be able to predict whether a protein may bind ss or dsDNA, the characteristics of the DNA substrate must be known. The single DNA strand is composed of repeating nucleotide units. Each nucleotide is in turn composed of a phosphate backbone, a sugar and a nucleoside base. ssDNA-binding proteins (SSBPs) make use of all of these features to recognise and bind to ssDNA. Since one of the non-bridging oxygens in each phosphate group is negatively charged, the DNA single strand can be considered a negatively charged polymer, with positive charges from some exposed bases. Thus, SSBPs frequently line their DNA-binding surfaces with the positively charged amino acid residues Lysine and Arginine<sup>1</sup>. In fact, since negative charges attract proteins, and the dsDNA is composed of two strands bound via the bases, such that the phosphate backbones protrude outwards, most DNA in the cell is associated with proteins, such as histones<sup>2</sup>.

SSBPs are known to form electrostatic interactions via the positively-charged guanidinium group of their arginine residues with the negatively charged delocalised  $\pi$ -cloud of the aromatic nucleoside base. SSBPs can also make stacking interactions between the aromatic planar side chains of tyrosine and phenylalanine residues and the planar nucleoside bases in ssDNA. This additional strategy provides an alternative binding surface that can be used in conjunction with binding to the backbone phosphodiester group. SSBPs may further interact in a non-sequence specific way with the ssDNA via hydrogen bonds from the amino acid side chains, or even the amide or carbonyl groups. More rarely, the sugar

moiety is also sometimes involved in interactions with SSBPs. DNA and RNA sugar groups differ by an extra 2' hydroxyl group in the sugar rings of the RNA bases. This allows for steric exclusion of 2' hydroxyls in ssDNA-binding proteins, or affinity adjustment in proteins that can bind both, but make use of the 2' hydroxyl group to increase their affinity towards RNA<sup>1</sup>. The single DNA strand is much more labile and susceptible to chemical and nucleolytic attacks than dsDNA, and many SSBPs have been identified as part of larger complexes with other genome maintenance proteins. These facts suggest that DNA metabolic processes are likely to occur upon a ssDNA/SSB nucleoprotein filament, instead of on naked ssDNA<sup>3</sup>.

One of the main features that allows surface recognition of ds vs. ssDNA is their flexibility: dsDNA is constrained by base pairing, which promotes the formation of a double helix with a rigid spacing of 0.34 nm between each phosphate, and 10 base pairs per turn (in B-form DNA). On the other hand, in ssDNA the internucleotide spacing can vary quite widely<sup>1</sup>.

The OB-fold, so named after its initially identified oligonucleotide/oligosaccharide binding properties, is commonly found in SSBPs. According to the SCOP (Structural Classification of Proteins) database, the OB-fold is present in sixteen different superfamilies. This fold consists of two three-stranded, anti-parallel  $\beta$ -sheets, with an  $\alpha$ -helix frequently found packed between strands 3 and 4, forming a cleft. Despite their rather constant 'Greek key' structure and ubiquity, the primary sequence of the OB-fold is not well conserved in SSBPs, and varies between 70 and 150 amino acids. The tertiary structure can vary quite widely in the length of each element, particularly in loop length. OB-folds are frequently found as recognition domains in larger macromolecular complexes. Although the majority of OB-folds is related to ssDNA recognition, they have also been described at the interface of protein-protein interactions<sup>3,4</sup>. Finally, the association constant of an individual OB-fold for ssDNA is relatively low (around 70

$10^5$ - $10^6$   $M^{-1}s^{-1}$ ), but their modular nature allows them to work together in order to raise the affinity<sup>1</sup>.

SSBPs do not unwind dsDNA. Rather, they bind and stabilise the single DNA strand as it is produced by the action of a helicase, ssDNA bubbles or transiently available 5' or 3' ends. Their function is to protect the exposed single DNA strand prior to the action of the subsequent enzyme in the pathway, which may be a DNA polymerase or another protein involved in DNA recombination and repair<sup>1</sup>.

The primary sequence of RuvBL2 provides some clues as to its helicase nature, namely by its sequence similarity to bacterial RuvB helicase. Furthermore, it has been described to bind ssDNA and unwind upstream duplexes *in vitro*, under two conditions: that the protein should be in the **monomeric state at the onset of activity**<sup>5</sup>; and that the DNA single stranded portion should be **at least 30 nucleotides long**<sup>5</sup>. ATP hydrolysis was found to increase in the presence of DNA, and helicase activity was shown to be dependent on the presence of ATP, suggesting that ATP binding is at least necessary for this processive activity to occur<sup>5</sup>. Since ATP binding should only occur when at least two monomers associate to form a dimer – and thus a complete nucleotide binding pocket, with the addition of the *trans*-arginine finger, this suggests that DNA single strands may act as scaffolds for RuvBL2 oligomerization. Interaction with histone tails has also been shown to promote formation of RuvBL2 rings. Although DNA-binding and helicase activities have also been reported for RuvBL1, they are not as well described as for RuvBL2<sup>6-9</sup>. However, while RuvBL2 has 3' → 5' processivity, RuvBL1 has occasionally been described to work in the opposite direction, although few advances have been made to establish the details of RuvBL1 helicase activity<sup>7</sup>. However, the notion that DNA processivity increases for the heteromeric complex, suggesting a synergistic or complementary mode of action, appears to be consensual<sup>6,9</sup>. The ablation of domain 2 from both the individual proteins and the

RuvBL1/RuvBL2 complex also greatly increases ATPase and helicase activities, hinting at a regulatory function for this domain<sup>6</sup>. Interestingly, some groups have shown that, in the RuvBL1/RuvBL2 dodecameric complex (which occurs naturally when expressed without the use of tags), not only do the rings interact through the domains 2, but they do so in a flexible way. As such, two conformations were observed: a compact one, in which the external regions of the D2 domains are intimately connected, and an extended one, in which these domains rotate and stretch, pushing the rings apart, and putatively exposing the DNA-binding region of domain 2 (see table 3.1)<sup>10</sup>.

## 2.1 *hsRuvBL2* stability, oligomerization and DNA binding

### 2.1.1 METHODS

#### *Protein expression and purification*

The codon-optimized sequence of *hsRuvBL2* with a C-terminus-His<sub>6</sub> tag including a 3C protease cleavage site was obtained from Genscript. The vector *pET49b\_ruvbl2\_Cter\_His* was transformed into *E. coli* B834 and gene expression was induced by the addition of 100 µM IPTG at 30°C for 19 h. The dry cells were resuspended in **lysis buffer** (50 mM phosphate buffer pH 7, 500 mM NaCl, 50 mM imidazole, 100 µM ADP, 1 mM MgCl<sub>2</sub>) supplemented with EDTA-free protease inhibitor tablet (Roche) and benzonase (Novagen), thoroughly homogenised using an ULTRA-TURRAX (IKA T18 basic) and passed twice through a Z basic cell disruptor (Constant Systems) at 15000 psi. The cell lysate was separated by centrifugation at 35000 *g* for 35 min, filtered through a 0.22 µm filter, injected

through a 5 mL **HisTrap** HP column (GE Healthcare) and eluted with a gradient of **buffer B** (50 mM phosphate buffer pH 7, 500 mM NaCl, 1 M imidazole, 100  $\mu$ M ADP). Fractions containing *hsRuvBL2* were pooled and **desalted** to remove imidazole by exchanging into a buffer containing 50 mM Na/KPO<sub>4</sub> pH 7.0, 500 mM NaCl, 100  $\mu$ M ADP and 1 mM MgCl<sub>2</sub>. A fraction of **tagged *hsRuvBL2*** was kept aside at this point, for analyses of the tagged form of *hsRuvBL2*. The remaining fraction was incubated overnight at 4°C with HRV3C protease in a 1:100 ratio. The mix of untagged *hsRuvBL2*, free tags, uncleaved *hsRuvBL2* and 3C protease was applied to GST and HisTrap columns set in tandem, and the flow through of **untagged *hsRuvBL2*** was collected. Finally, both tagged and untagged protein were in turn injected on a **HiLoad 16/60 Superdex 200** pg column (GE Healthcare), previously equilibrated with 50 mM phosphate buffer pH 7, 500 mM NaCl, 100  $\mu$ M ADP, 1 mM MgCl<sub>2</sub>. Thermal shift assays (TSA) were performed at several stages, to determine the best sample buffer and to assess stability of *hsRuvBL2* in the presence and absence of different nucleotides. Prior to the final concentration step, ADP was added to a concentration of 4 mM.

### *Differential scanning fluorimetry*

The protein melting temperature ( $T_m$ ) determination was performed by monitoring protein unfolding with the fluoroprobe SYPRO Orange dye (Molecular Probes), which although completely quenched in an aqueous environment, emits fluorescence upon binding to the exposed hydrophobic patches during protein unfolding. This increase in fluorescence can be measured as a function of temperature. The assays were performed in an iCycle iQ5 Real Time PCR Detection System (Bio-Rad), equipped with a charge-coupled device (CCD) camera and a Cy3 filter with excitation and emission wavelengths of 490 and 575 nm, respectively. This equipment can detect the fluorescence changes in 96-well plates

simultaneously (low profile plate, Bio-Rad) and thus can be used for parallel differential scanning fluorimetry (DSF) assays. The 96-well plates were sealed with optical quality sealing tape (Bio-Rad) and centrifuged at 2500 g for 30 seconds immediately prior to the assay, to remove possible air bubbles. The plates were subsequently heated from 20 to 90°C with stepwise increments of 0.5°C with a 60-second equilibration time, followed by the fluorescence read out. In a typical assay with a total volume of 20  $\mu$ l, a protein concentration of 0.125 mg/ml and a 5-fold dye concentration (stock is 5000 fold) were used to guarantee the best signal to noise ratio. The assay was prepared by adding protein to the mix of dye-buffer solution.

An initial sample **buffer screening** was performed using a screen of 96 buffers with varying buffer molecules, pH and salt concentrations. For buffer screenings, the dye dilution buffer used was HEPES pH 7.5 at 50 mM, since HEPES has the lowest pH variation with temperature. The assay was prepared by adding 2  $\mu$ l of protein and dye mix to 18  $\mu$ l of new buffer (100 mM), in this way testing the protein behaviour in a set of 96 different buffers. When the **effect of nucleotides** was tested, all assays were prepared in previously determined optimal protein buffer (50 mM Na/K phosphate, 500 mM NaCl, 4 mM MgCl<sub>2</sub>). In this case, prior to the assays the protein was incubated with nucleotides with a molar excess higher than 10-fold (4 mM of nucleotide to ca. 300  $\mu$ M *hsRuvBL2*). All assays were performed with pure protein, from the peak corresponding to a hexameric oligomer, collected after the last size exclusion purification step. Fluorescence intensities versus temperature were used to calculate the protein melting temperature ( $T_m$ ) by determining the first derivative ( $d(Rfu)/dT$ ) to extract the minima, which corresponds to the exact transition as the inflection point of the melting curve. The higher the  $T_m$ , the more stable the protein.



### *Size exclusion chromatography coupled to multi-angle laser light scattering*

SEC-MALLS was performed by Christine Ebel at the Institut de Biologie Structurale (IBS) in Grenoble. This technique was used to determine, for each sample, the hydrodynamic radius and molecular weight, in order to analyse the influence of tags and nucleotides in the oligomerization state of *hsRuvBL2*.

The protein sample was thawed on ice for 1 hour, centrifuged at  $96500 \times g$  just prior to the experiment and injected into a WTC050N5 (Wyatt) column equilibrated at  $20^{\circ}\text{C}$  in sample buffer (50 mM Na/K  $\text{PO}_4$  pH 7, 500 mM NaCl, 1 mM  $\text{MgCl}_2$ , 500  $\mu\text{M}$  ATP or ADP). Coupled to the chromatographic system were: a multi-wavelength absorbance detector (Shimadzu SPD-M20A), which measures the absorbance of the eluate between 190 and 700 nm; a static light scattering detector (WYATT mini DAWN TREOS), which measures scattering at three different angles from a sample illuminated by a 658 nm laser beam and makes it possible to determine the molecular weight (MW) from the intensity of the scattered light; a dynamic light scattering (Quasi Elastic Light Scattering – QELS) detector (WYATT DynaPro Nanostar), which measures the fluctuation of the scattered light intensity at  $90^{\circ}$ , from which the hydrodynamic radius ( $R_h$ ) can be calculated. **MALLS** analyses the time averaged (1s) scattered light intensity  $I$ , and allows for the determination of MW (and radius of gyration  $R_g$  if greater than 20 nm). **QELS** analyses the fluctuations of the scattered light intensity,  $I$ , as a function of time, and allows the determination of the diffusion coefficient  $Dt$ , and thus hydrodynamic radius  $R_h$ . The analysis was performed using the ASTRA software version 5.4.3.18.

*hsRuvBL2* without tags and with tags on the N- and C-terminus were analyzed in buffer with ADP. Additionally, untagged *hsRuvBL2* was analyzed after dialysis to buffer with ATP.

### ***Analytical ultracentrifugation***

Analytical ultracentrifugation (AUC) analysis of sedimentation coefficients of *hsRuvBL2* was also performed in collaboration with Christine Ebel, from the Institut de Biologie Structurale, in Grenoble. AUC is one of the most precise analytical techniques for the calculation of the oligomerization state of a complex. The aim was to determine differences in the oligomerization state of the different *hsRuvBL2* samples in a buffer containing ADP: untagged and with tags on the C- or N-terminus.

AUC is a powerful technique for the quantitative characterization of macromolecular associations in solution. **Sedimentation velocity** measures the rate at which the boundaries of molecules move during mass redistribution, as a consequence of exposure to high centrifugal fields. An **equilibrium sedimentation** experiment is performed to determine the concentration distribution after equilibrium is reached<sup>11</sup>. **Sedimentation velocity (SV)** was used in this work to determine the proportion of different oligomeric forms for each sample, as well as their molecular weight, and to observe whether the *hsRuvBL2* complex undergoes concentration-dependent dissociation.

Sedimentation velocity experiments were performed at 35000 revs per min and 20°C, in a XL-I analytical ultracentrifuge using an Anti-50 rotor (Beckman Coulter, USA), with 3 and 12 mm path length Ti double-sector centrepieces equipped with sapphire windows (Nanolytics GmbH, Germany), loaded with 110 and 420 µL, respectively, depending on protein concentration, sample and reference solvent. Acquisitions were made using interference optics. Due to the

presence of 4 mM ADP, absorbance detection could not be used (for a 3 mm path length,  $A_{280}$  is higher than 1.5). *HsRuvBL2* samples were stored at  $-80^{\circ}\text{C}$ , in 50 mM Na/K phosphate, 500 mM NaCl, 1 mM  $\text{MgCl}_2$ , and 4 mM ADP. Prior to the experiments, samples were thawed and diluted in the same buffer. The reference channels were filled with the buffer without ADP.

The analyses were done with the Sedfit software<sup>12</sup>, v14.1. Partial specific volumes,  $\bar{v}$ , were calculated from composition with Sedfit. The program SEDNTERP (<http://sednterp.unh.edu/>), was used for the analysis, with buffer density,  $\rho = 1.025 \text{ g mL}^{-1}$  and viscosity,  $\eta = 1.06 \text{ cP}$ .

Sedimentation velocity profiles were analysed in terms of continuous distribution  $c(s)$  of sedimentation coefficients,  $s$ <sup>13</sup>. Peak integration provides estimates of  $s$  and of the signals. The linear fit  $s = s_0(1 - k'sc)$ , where  $c$  is the protein concentration, was used to estimate the sedimentation coefficient at infinite dilution,  $s_0$ .

For homogeneous samples, the non-interacting species analysis provides independent estimates of  $s$  and of the diffusion coefficient,  $D$ , which was used in the Svedberg equation  $s_0/D = M(1 - \rho\bar{v})/RT$ , to provide an experimental value for the molar mass,  $M$ , with  $R$  the gas constant and  $T$  the temperature.

The value of  $s_0$  was also interpreted considering  $D = RT/N_A 6\pi\eta R_H$ , leading to a modification of the Svedberg equation:  $s_0 = M(1 - \rho\bar{v}) / (N_A 6\pi\eta R_H)$ , where  $N_A$  is the Avogadro's number and  $R_H = f/f_{\min} R_{\min}$  is the hydrodynamic radius, with  $f/f_{\min}$  the frictional coefficient and  $R_{\min}$  the radius of the anhydrous volume.

### ***Small angle X-ray scattering***

SAS (small angle scattering) techniques detect the scattering of X-ray photons - SAXS (or neutrons - SANS), which occurs at very small angles from the

incident beam. SAS is useful for retrieving information from “disordered” systems, where the arrangement of molecules is random.

SAXS was used to assess the oligomerization state of the samples in solution, as well as to detect variations in shape, such as different compaction states of the constituent domains. This technique provides only a low resolution envelope, so the main purpose in this work was to compare the results in solution with the structures obtained with cryo-EM and X-ray crystallography. The results obtained in solution should be closer to the protein behaviour *in vivo*, since there are no mobility constraints. This is thus a good complementary technique to gain additional insight into the size, shape and mobility of the constituent domains of the protein/complex.

HsRuvBL2 was analysed at beamline B21 of the Diamond Light Source (Didcot, UK), both in solution using a sample changer and through in-line analysis of the sample applied to a size-exclusion chromatography column. Analysis **in solution** was performed by merging datasets from serial dilutions of a *hsRuvBL2* fraction from the centre of the S200 peak (in the final purification stage), from 0.6 to 4.5 mg/mL. **In-line analysis** was performed by injection of a fraction at 5.5 mg/mL, also of a sample from the middle of the S200 peak, in a Shodex 403kw column. Measurements were performed at several points during peak elution. All samples were filtered through a 1 MDa filter prior to analysis to eliminate larger aggregates.

### ***Electrophoretic mobility shift assays (EMSA)***

Electrophoretic mobility shift assays were performed to assess *hsRuvBLs* ability to bind DNA in native conditions. Since it has already been established that RuvBL2 binds only to single stranded DNA<sup>14</sup>, M13mp18 was used in these assays. M13mp18 is a large molecule (2.236 MDa, 7249 bases), and as such allows binding

of several RuvBL molecules, thus producing a weight variation large enough to be identified in the gel. Prior to the agarose gel separation, the protein was incubated with ssDNA in a reaction mixture with a total volume of 20  $\mu\text{L}$ , of which 0.5  $\mu\text{L}$  (1.12 nM) DNA, variable amounts of protein (50 or 25  $\mu\text{M}$ ) and completed with reaction buffer (25 mM HEPES/KOH pH 8, 2.5 mM  $\text{Mg}(\text{CH}_3\text{COO})_2$ , 100 mM KCl, 0.2 mM DTT)<sup>14</sup> supplemented with 4 mM ATP and 2 mM  $\text{MgCl}_2$ . Negative controls were performed by adding either only DNA or only protein to the reaction buffer. The reaction occurred for 1 h at 23°C, after which 1  $\mu\text{L}$  of 50% glycerol was added, the reaction was loaded in a 0.6% agarose gel in TBE 1x and run for 2h30m at 80 V.

The DNA bands in the gel were observed by staining with SYBR Gold DNA stain (S11494, ThermoFisher), which binds to ssDNA with a high quantum yield. Fluorescence was detected with a Fuji TLA-5100, using a 473 nm excitation wavelength (LPB filter, 700 V).

### ***Analysis of hsRuvBL2 binding to DNA by negative staining EM***

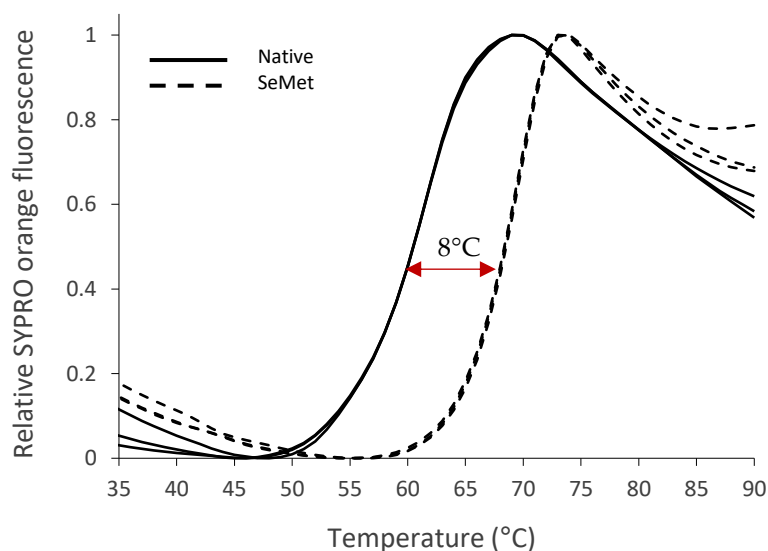
Immediately prior to EM data collection (protocol described in Chapter 3), *hsRuvBL2* was incubated with M13mp18 circular ssDNA, in conditions previously observed to lead to binding to DNA<sup>14</sup>. DNA binding had not been observed for RuvBL2 hexamers, either in published data by other groups, or in our hands; however, after observing EM micrographs that showed hexamer dissociation of *hsRuvBL2* into its individual components after a prolonged dilution time (more than 1 day), we hypothesized that the rings would be able to re-form around DNA, based on the known mechanisms of action of other helicases. To prove this hypothesis, *hsRuvBL2* was incubated with M13mp18 DNA after hexamer disruption. A control incubation was prepared without DNA. After 30 minutes of incubation at room temperature (about 20°C) in buffer with ATP, both samples (with and without DNA) were adhered to the Rhodium/Copper EM grids (pre-

treated with a carbon coating and rendered hydrophilic through an electrical glow discharge), washed with Tris-containing buffer, and incubated with uranyl acetate.

## 2.1.2 RESULTS AND DISCUSSION

### *Purification tags and nucleotides affect hsRuvBL2 stability and oligomerization state*

In order to study *hsRuvBL2*, the problem of low stability in solution had to be overcome. To tackle this, differential scanning fluorimetry assays were

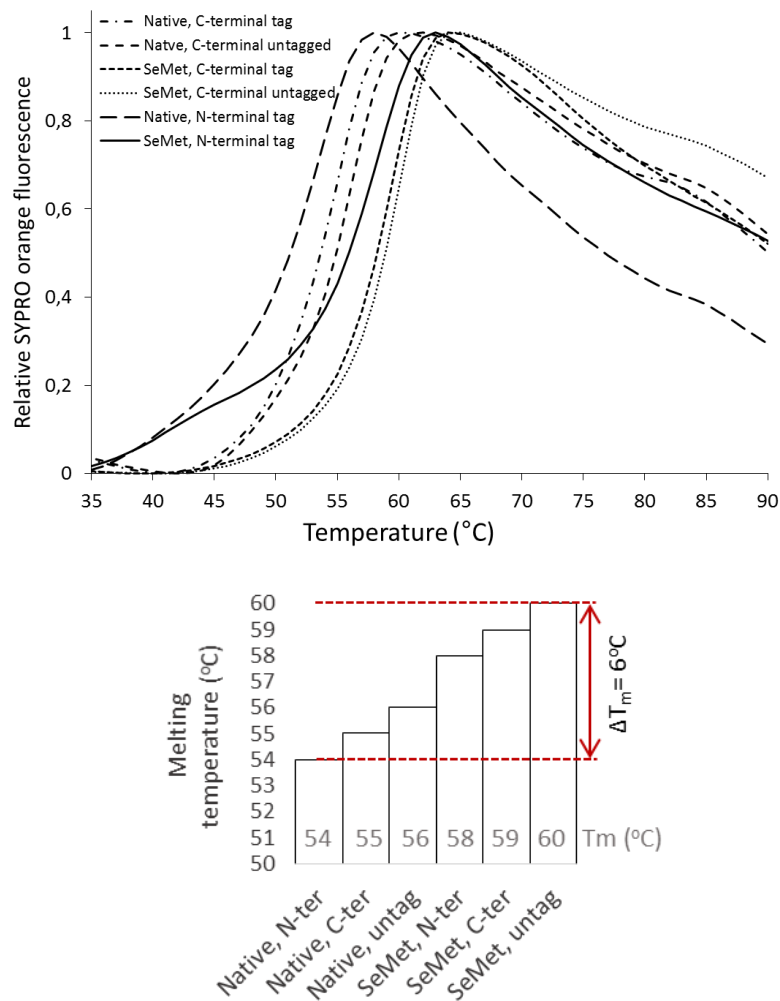


**Figure 2.1 – Fluorescence intensity vs. temperature for the thermal unfolding of the *hsRuvBL1ΔDII/hsRuvBL2ΔDII* complex in the native (continuous line) and selenomethionine substituted (dotted line) forms. These initial assays show an increase in thermal stability of 8°C between the native and the selenomethionine-substituted truncated RuvBL1/2 complex.**

performed, which allowed for the quantitative analysis of *hsRuvBL2* stability in

solution under various conditions. We first started by analysing two samples of the DII-truncated RuvBL1/RuvBL2 complex previously studied by our group: ***hsRuvBL1* $\Delta$ DII/*RuvBL2* $\Delta$ DII, in the native and selenomethionine-substituted (selenomet) forms (Fig. 2.1)**. This initial analysis clearly indicated that **selenomethionine substitution led to an increase in the thermal stability of the complex**, with a difference of 8°C in the melting temperature. These results prompted us to produce selenomethionine-substituted *hsRuvBL2*, with a view of increasing stability for crystallization purposes. Additionally, since tagged *hsRuvBL2* has been used as a tool for *in vivo* assays, to enable the identification of binding partners (such as c-myc<sup>15</sup>), we further inquired whether tags also had an influence on RuvBL2 behaviour, knowing that affinity tags placed on the N-terminus of *scRuvBL1* (from *Saccharomyces cerevisiae*) and *scRuvBL2* were shown to induce alternative oligomeric states in the yeast Rvb1:Rvb2 complex<sup>16</sup>. Specifically, Cheung and colleagues found that Histidine-tagged *scRvb1* and *scRvb2* form a dodecameric complex, which is reverted to a hexamer upon tag removal. Thus, to address the issue of whether the location of the affinity tags would also affect the assembly and stability of human *hsRuvBL2*, we expressed recombinant *hsRuvBL2* in *E. coli*, with affinity (FLAG and His) tags either at the N- or C-terminus, both in the native and selenomet form. It became clear during purification that the tags placed at the N-terminus greatly destabilized *hsRuvBL2*, as this sample, in the native form, was prone to aggregation during concentration steps. We analysed the thermal stability of the native and selenomet *hsRuvBL2* by DSF, both with the affinity tags on the C- and on the N-terminus, and further produced untagged *hsRuvBL2* by cleaving the tags on the C-terminus. DSF assays (Fig. 2.2) show that thermal stability of *hsRuvBL2* is lowest for the N-terminal tagged forms (54 and 58°C, for the native and selenomet forms, respectively), followed by the C-terminal tagged forms (55 and 59°C). Stability was maximized after cleaving the C-terminal tags (56 and 60°C).

Together, these results suggest that the insertion of affinity tags may be



**Figure 2.2** – *HsRuvBL2* thermal stability in solution is affected by tags, as observed by differential scanning fluorimetry assays. An N-terminal tag decreases stability the most, followed by a C-terminal tag. The untagged *hsRuvBL2* (tag removed from the C-terminus) is the most stable. All constructs are further stabilized by selenomethionine substitution. However, this increased stability did not result in the successful crystallization of the SeMet variant (see chapter 3).

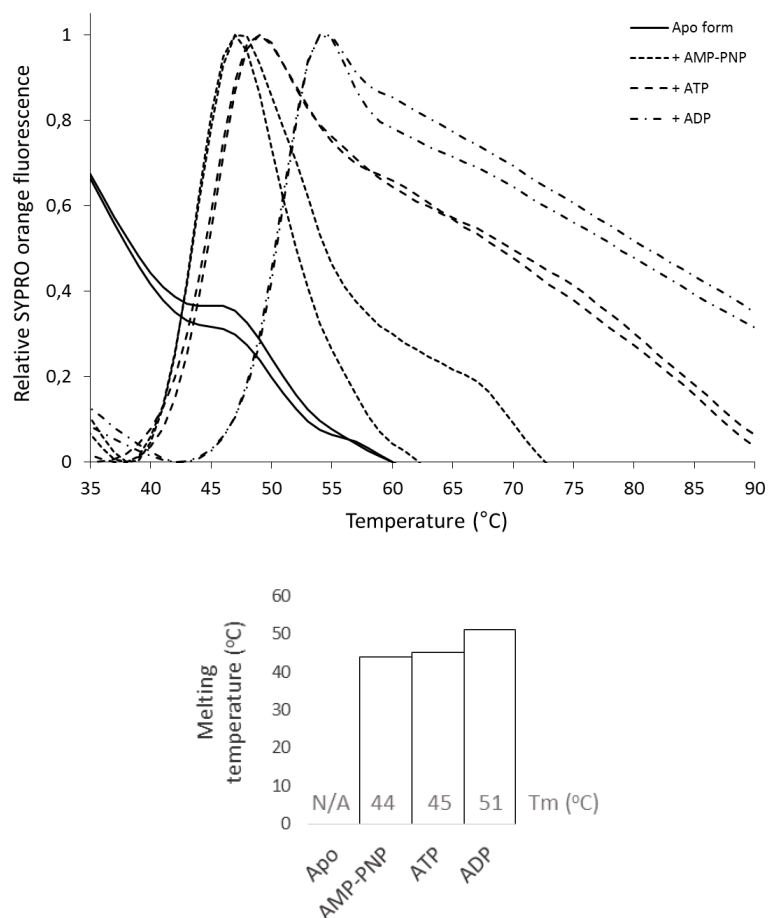
detrimental to *hsRuvBL2* folding and stability. Such an influence may be better



understood in the context of the overall structure of *hsRuvBL2*, which is discussed in the next chapter. Briefly, the N-terminus forms a long (ca. 40 residues) unstructured loop in the absence of a nucleotide in the binding pocket. In structures where *RuvBL2* is bound to a nucleotide, this loop becomes organized, partly through interaction with the bound nucleotide. These observations suggest that the N-terminus may have a function related to ATP binding, and that the addition of affinity tags may interfere with this function. The C-terminal tags are less destabilizing. The C-terminus of human *RuvBL2* is organized into an antenna-like helix, which protrudes outward from the top of the hexameric complex. Compared to that of *hsRuvBL1*, the C-terminal helix of *hsRuvBL2* is longer, and in yeast, this motif may be involved in interactions with binding partners, such as *scPih1p*<sup>17</sup>.

Altogether, these results suggest that the use of tags for *in vivo* studies may be disadvantageous, since it can compromise protein behaviour in different conditions, as well as generate non-physiological complexes. On the other hand, given the need for the use of tags in some studies, the use of both tagged forms of *RuvBL2* should be analysed during complex formation studies *in vivo*, bearing in mind that the influence of tags in oligomerisation plasticity may produce results different from the *in vivo* reality.

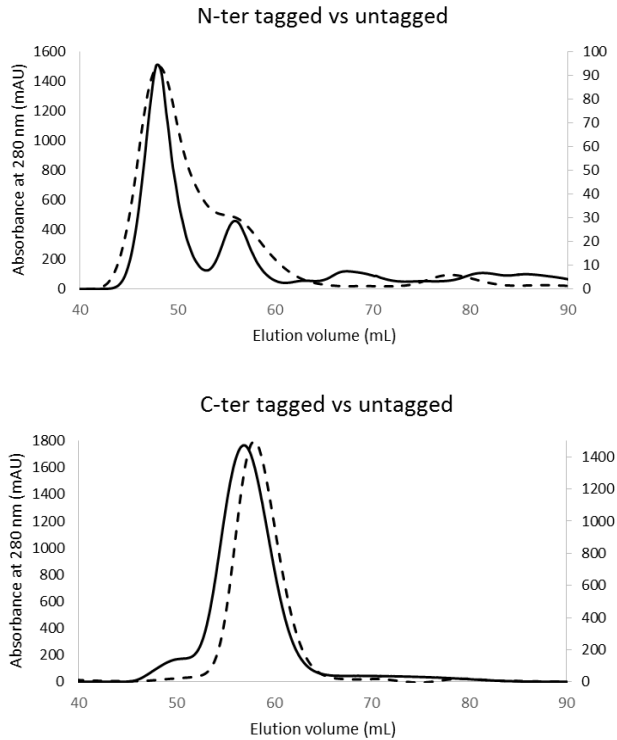
Thermal stability of *hsRuvBL2* was further tested in the presence and absence of nucleotides (**Fig. 2.3**), since ATP is the natural substrate of *RuvBL2*. The melting curve of apo *hsRuvBL2* (full line) suggests a lack of structural integrity, denoted by the high initial fluorescence and lack of a defined melting transition. A DSF curve with such high initial fluorescence indicates the presence of exposed hydrophobic patches at the onset of the assay (20°C), and thus that the sample is not properly folded. This precludes the expected unfolding pathway, and hence the absence of a well-defined melting transition. Conversely, *hsRuvBL2* pre-incubated with nucleotides produced curves with lower initial fluorescence, and well defined melting transitions. These curve characteristics indicate that in the



**Figure 2.3 – Thermal stability of *hsRuvBL2* in the absence and presence of nucleotides, determined by differential scanning fluorimetry assays.** Apo *hsRuvBL2* (native untagged construct) shown as a full line versus thermal stability upon incubation with nucleotides (dotted lines). Pre-incubation of *hsRuvBL2* with nucleotides leads to curves with a well-defined melting curve transition, and lower initial fluorescence, which is indicative of a proper 3-dimensional fold.

presence of nucleotides there are no exposed hydrophobic patches in *hsRuvBL2* at the beginning of the assay, and thus that the protein is properly folded. ADP provided the highest thermal stabilization effect to *hsRuvBL2* in solution, and thus produced an optimal system for the subsequent analyses.

Having determined that *hsRuvBL2* in the presence of ADP provides the most stable sample, we pursued the goal of clarifying whether affinity tags have any influence in human RuvBL2 oligomerization. For this purpose we used size-exclusion chromatography. We produced *hsRuvBL2* expressed with affinity tags



**Figure 2.4 – SEC profiles of *hsRuvBL2* with tags on the N- and C-terminus, before and after tag removal.** Oligomeric behaviour seems to be maintained after tag removal. **Top:** When *hsRuvBL2* is expressed with a tag on the N-terminus, it forms both hexamers and dodecamers, both with (solid line) and without (dashed line) tag. However, when the tag is placed on the C-terminus (**bottom**), *hsRuvBL2* forms mostly hexamers (tagged *hsRuvBL2* - solid line), an oligomeric state also maintained after tag removal (dashed line).

either on the C- or N-terminus. We also analysed both samples after tag removal. Analytical SEC results showed that *hsRuvBL2* expressed with an N-terminal tag assembles into a mixture of dodecamers and hexamers (**Fig. 2.4 top**), as had been

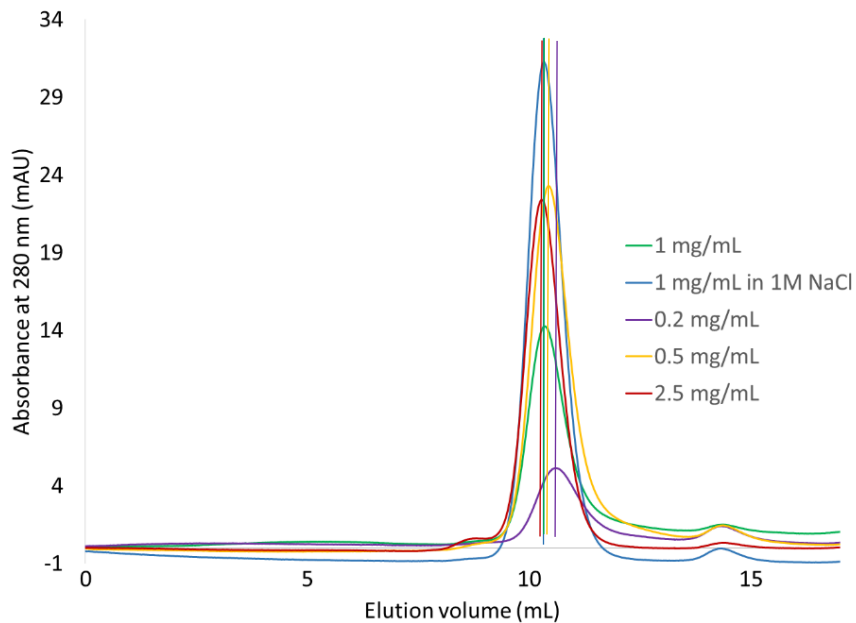
previously observed by our group by SAXS<sup>6</sup>. We further observed that these oligomeric forms are still assembled after the tag is removed. On the other hand, when the tags were inserted in the C-terminus, *hsRuvBL2* assembled mostly as a hexamer, again without modifications after tag removal (**Fig. 2.4 bottom**). These results suggest that, in *hsRuvBL2*, the oligomeric behaviour of *hsRuvBL2* may be determined during protein biogenesis, and that this behaviour is maintained after tag cleavage.

### ***Stability of *hsRuvBL2* oligomers varies with concentration***

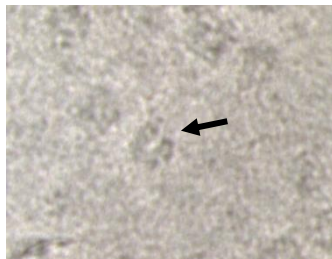
Having observed that tags influence the oligomers formed by *hsRuvBL2*, we further asked whether the tag position may also affect the oligomeric plasticity depending on *hsRuvBL2* concentration. For this, we analysed *hsRuvBL2* with tags on the C-terminus, on the N-terminus, and an untagged form, obtained by tag cleavage from the C-terminus.

Initial oligomerization state analyses at different concentrations were performed using classical analytical SEC studies using a Superdex 200 10/300 column, in a buffer containing 500 mM and 1 M NaCl (**Fig. 2.5**). At these high ionic strengths, the observed oligomer is consistently a hexamer. As the protein concentration is lowered ca. 10-fold from 2.5 mg/mL to 0.2 mg/mL, the apparent molecular weight slightly decreases. Additionally, we have observed by electron microscopy that at low concentrations the hexameric rings start to acquire an open conformation, which may lead to a lower apparent molecular weight (**Fig. 2.6**). We further attempted to obtain more accurate values for the molecular weight of the observed oligomer using SEC-MALLS. However, this analysis was not successful, as no peaks were observed for the tagged samples, suggesting adherence to the guard column. The analysis of untagged hexameric *hsRuvBL2* in presence of ATP and ADP provided an equivalent molecular weight, roughly corresponding to a

hexamer. However, these results are inconclusive regarding the influence of nucleotides on oligomeric state, since it cannot be demonstrated that there was nucleotide exchange.

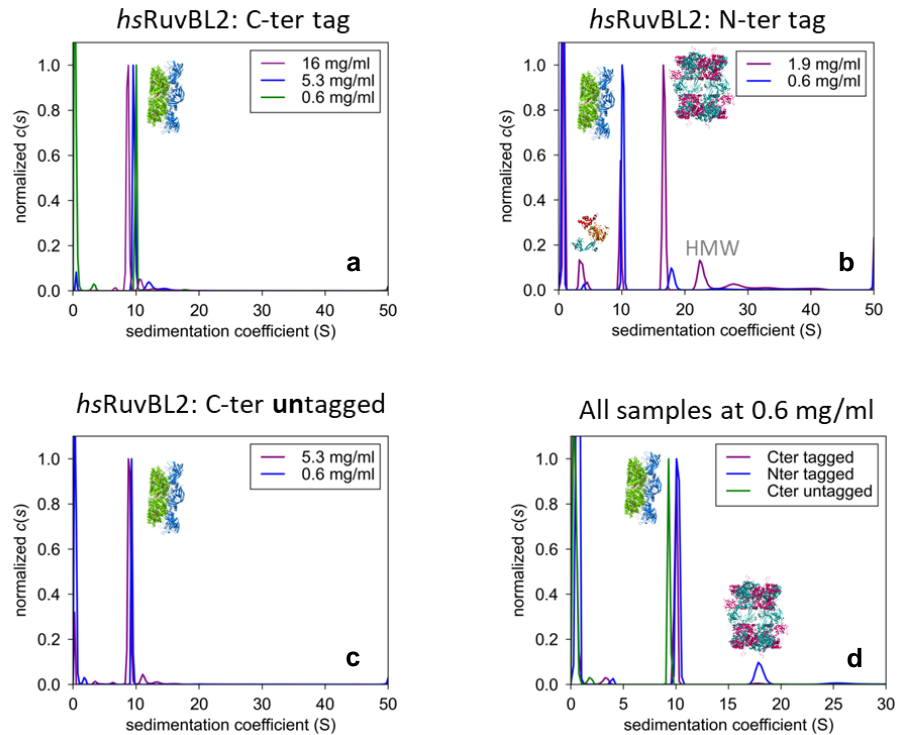


**Figure 2.5 – Oligomerization state of untagged *hsRuvBL2* at different protein concentrations, in buffer with 500 mM NaCl and 1 M NaCl.** At the tested protein and salt concentrations, the observed oligomeric species corresponds to a hexamer. However, as protein concentration decreases, there is a slight shift towards a lower apparent molecular weight. This may indicate partial subunit loss or a change in conformation, such as partial ring opening.



**Figure 2.6 – Open ring conformations were observed in *hsRuvBL2* samples after they had been diluted to about 100  $\mu\text{g}/\text{mL}$  for a few hours.** The arrow indicates an open (notched) ring. Micrograph of a sample observed by negative staining EM.

Analytical ultracentrifugation analyses (**Fig. 2.7**) were performed to compare sedimentation velocity (SV) profiles. These were analysed as a continuous distribution of sedimentation coefficients,  $c(s)$ , which allowed the analysis of homogeneity and concentration effects. Using this technique, we observed that **C-terminally tagged *hsRuvBL2*** forms a mostly homogeneous complex at all



**Figure 2.7 – Analytical ultracentrifugation profiles of *hsRuvBL2* constructs.** Tags affect the oligomerisation equilibrium of *hsRuvBL2*: while *hsRuvBL2* with C-terminal affinity tags is hexameric at all concentrations tested, N-terminally tagged *hsRuvBL2* assembles into higher MW oligomers as protein concentration is increased.

concentrations tested, with a major oligomer at about 10 S, which corresponds to a hexamer – ca. 90% of the signal (**Fig. 2.7a and c**).

The **N-terminal tagged** *hsRuvBL2* at low concentrations (0.6 mg/mL) forms a mixture of hexamers (77%) and dodecamers (14%), while at higher concentrations (1.9 mg/mL), the hexamer population decreases to 17% concomitant with an increase in dodecamer population to 45% (**Fig. 2.7b**). The expected value for a globular compact species with the same mass of the *hsRuvBL2* dodecamer is 18.3 S, and the observed value was  $17 \pm 1$  S. It can thus be concluded that for this construct there is an equilibrium between hexamer and dodecamer formation, with a higher percentage of dodecamer being assembled as concentrations increase. In addition, larger oligomers/aggregates were detected between 18 and 45 S (6% at 0.6 mg/mL and 27% at 1.9 mg/mL), which are not detected in significant amounts for the other construct (**Fig. 2.7**). It is possible, considering the absence of evidence in the literature for the formation of such high molecular weight species in RuvBLs, that the N-terminal tag promotes the formation of aberrant, or non-physiological, complexes.

Sedimentation (s) and diffusion (D) coefficients were obtained from data acquired for C-terminal **tagged and untagged** *hsRuvBL2*. The mean of the D values were used to calculate masses in solution, providing molecular weights of **288 and 354 KDa** (for comparison, the corresponding theoretical molecular weights are 323 and 307 KDa), close to the expected value for an **association state of  $6 \pm 1$** . The sedimentation coefficient was combined with the theoretical molar mass of the hexamer, to obtain an estimated **hydrodynamic radius (RH) of  $6.8 \pm 0.2$  nm**, and frictional ratio of  $1.5 \pm 0.06$ , corresponding to **slightly elongated shapes** (the typical value for a globular compact shape is 1.25).

In all constructs there is also a minor population of free monomer, at 3.5 S (1-5%). Curiously, in the N-terminally tagged construct, this population increases concomitantly with the increase in protein concentration (**Fig. 2.7b**).

In conclusion, AUC analyses show that expression of *hsRuvBL2* **with tags on the C-terminus** produces a homogeneously hexameric sample, before and after

tag removal, in the conditions used. Despite the fact that this construct was also shown to be the most stable by TSA, the presence of tags may negatively affect some experiments, as observed by the absence of curves when SEC/MALLS was attempted using these samples, which may be caused by protein adhesion to the guard column. The presence of tags in this position also precluded the formation of crystals, which form after about 6 h when using untagged *hsRuvBL2*. On the other hand, **N-terminally tagged *hsRuvBL2*** is heterogeneous, with a mixed population of hexamers and dodecamers, larger aggregates and dissociated species.

Combining the observations from this section and the previous one, it becomes apparent that the position of affinity tags affects both the stability and oligomer formation in *hsRuvBL2* in response to variations in protein concentration. Furthermore, the oligomeric behaviour in each case is maintained after tag removal, which suggests that either oligomeric behaviour is determined during protein biogenesis, or that the two residues that remain after tag cleavage (Gly and Pro) may still have some influence (which seems less likely).

This difference in behaviour of the differently tagged *hsRuvBL2* constructs suggests that tag placement may interfere with other analyses, namely interactions with other partners and with specific activities. As demonstrated by TSA, the N-terminally tagged construct of *hsRuvBL2* is slightly more unstable, suggesting that tag placement in the N-terminus may be more detrimental. However, both the N-terminal loop and the C-terminal helix have putative functions, namely in interaction with other proteins<sup>18,19</sup>, hence it might be better for interaction assays to analyse the results obtained using both constructs. Additionally, equilibrium dynamics is affected by tag position: C-terminally tagged *hsRuvBL2* is consistently hexameric, while N-terminally tagged *hsRuvBL2* forms oligomers in a concentration-dependent manner.



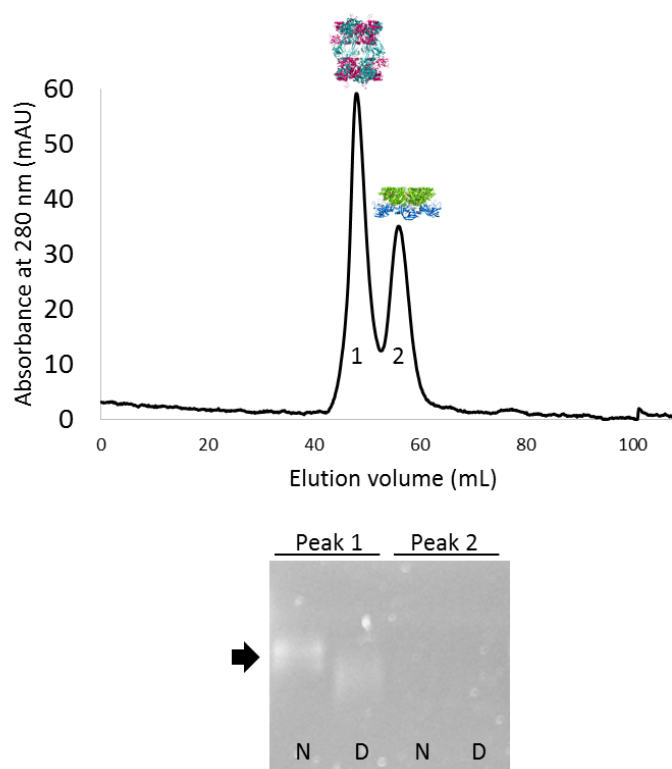
SEC/MALLS was performed to assess the MW of *hsRuvBL2* in buffers with ADP and ATP. The latter was obtained by dialysis of the ADP-purified sample to a buffer with ATP. The observed peaks correspond to a MW of 303 KDa and 296 KDa, of *hsRuvBL2* in buffer with ADP and ATP, respectively. These values are within the error margins for a hexamer in both cases. Whether there was nucleotide exchange upon dialysis cannot be determined; however, from observation of the existing structures and activities<sup>6,14,20</sup>, nucleotide exchange is precluded in the hexameric form, due to the very occluded entrance to the nucleotide binding site. Only in the DII-truncated complex was nucleotide exchange observed. The observed hydrodynamic radii ( $7.0 \pm 0.1$  and  $7.15 \pm 0.1$  nm) are also within the error margins, as compared to the values obtained by AUC. Since we cannot be sure of nucleotide exchange in the binding pocket, no conclusion can be drawn from these results regarding the influence of nucleotides. Notwithstanding, they support the results obtained by AUC.

### ***Insights into *hsRuvBL1* and *hsRuvBL2* binding to DNA***

RuvBL2 binding to DNA is independent of nucleotide sequence, and restricted to single-stranded DNA<sup>6,14</sup>. Papin and colleagues have also clearly shown that only monomeric RuvBL2 (yeast and human) can bind the polynucleotide chain, which suggests a mechanism of action whereby RuvBL2 oligomerizes around a single chain overhang, and only then starts to perform the upstream unwinding of the remaining double helix, in an ATP-dependent process<sup>14</sup>.

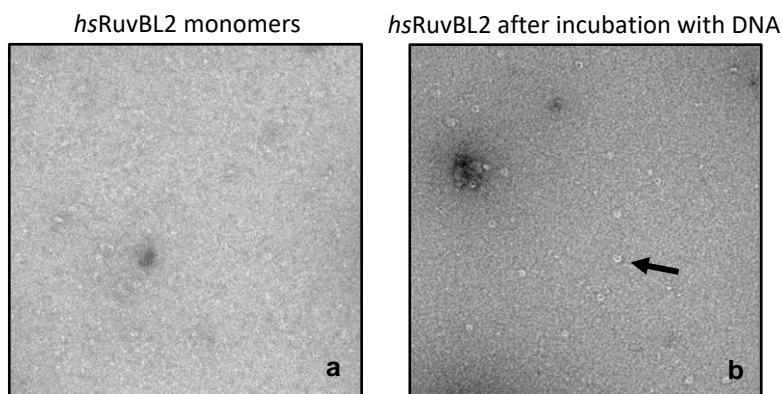
Here, we aimed to expand the current knowledge on the DNA-binding mode of *hsRuvBL2*, and gain further insight into *hsRuvBL1* and *hsRuvBL2* binding to DNA. In our hands, *hsRuvBL2* binding to DNA was first observed while purifying the N-terminally tagged, selenomethionine-substituted *hsRuvBL2*. The elution peaks (Absorbance at 280 nm) obtained by size-exclusion chromatography

were applied into an agarose gel and stained with SYBR green (Fig. 2.8). A coomassie-stained acrylamide gel confirmed the presence of protein in the peak corresponding to the putative dodecamer, albeit in much lower amounts than in the hexameric peak. The fact that the peak corresponding to the dodecamer is the more intense is due to the contribution of DNA in the sample. DNA presence was identified exclusively in the peak corresponding to the dodecameric complex, which prompts the question of whether DNA promotes the formation of dodecamers or if previously formed dodecamers have a tendency to bind DNA.



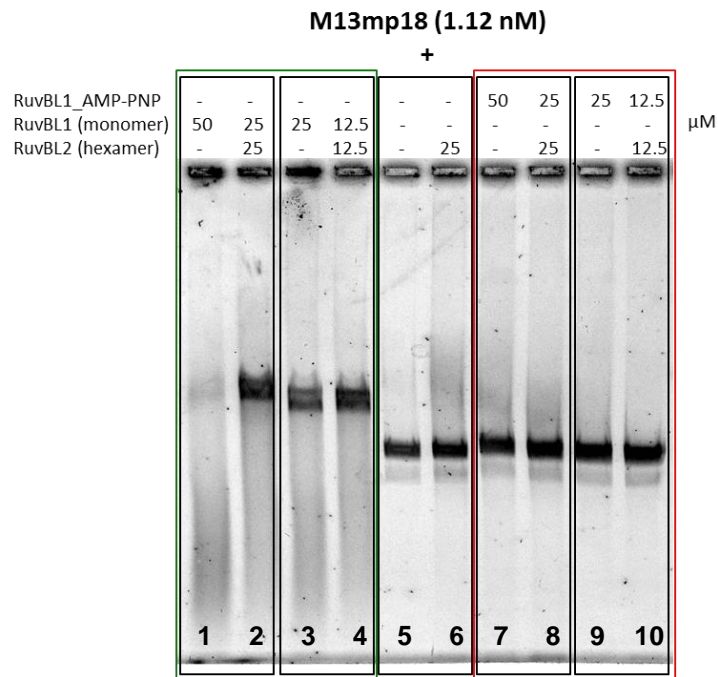
**Figure 2.8 – Initial observations of DNA bound to *hsRuvBL2* (N-terminally tagged).** Both the “dodecamer” and the “hexamer” peaks of a size-exclusion chromatography purification step were observed in an agarose gel. SYBR green staining indicates the presence of DNA exclusively in the “dodecamer” peak. N - Native samples; D – Heat denatured samples.

Papin and colleagues have observed that RuvBL2 is required to be in the monomeric form at the onset of DNA binding, and also that RuvBL2 activity and oligomerization are modulated by interaction with binding partners, such as histones<sup>5,21</sup>. These observations suggest that DNA binding may also be a driver of oligomerization, a hypothesis we addressed using single-particle electron microscopy. During the EM experiments, we observed the dissociation of *hsRuvBL2* hexamers into monomers (**Fig. 2.9a**) after dilution to the very low concentration of about 100  $\mu\text{g/mL}$ , a concentration which limits the techniques that allow the observation of the structural changes taking place upon interaction between *hsRuvBL2* and DNA. When monomeric *hsRuvBL2* was incubated with single-stranded DNA, the previously disassembled rings were re-formed (**Fig. 2.9b**). Although the circular M13mp18 DNA strand could not be observed at the resolution obtained by negative - staining EM, it can be assumed that, if the rings assembled around the DNA chain, the latter would remain associated with the protein complex.



**Figure 2.9 – Monomeric *hsRuvBL2* assembles into rings in the presence of ssDNA.** a) At a concentration of 100  $\mu\text{g/mL}$ , *hsRuvBL2* slowly dissociates into monomers (and possibly dimers). This dissociation can be observed by EM, since the monomers are too small to be clearly identified. Conversely, hexamers (and possibly heptamers) are large enough to be observed by EM. Thus, it was possible to observe their formation as a consequence of incubation with M13mp18 ssDNA (b).

Alternatively, DNA binding may occur through the external region of domain 2, as suggested for *hsRuvBL1*<sup>20</sup>, since both proteins display an OB-fold within this region. However, if binding occurred only through domain 2, it probably would also be observed in the hexameric forms. Furthermore, helicase activity – for which prior binding to the DNA strand is required – was shown in complexes where the external region of domain 2 was deleted<sup>22</sup>.



**Figure 2.10 – Electrophoretic mobility shift assay with hexameric *hsRuvBL2*, monomeric *hsRuvBL1* and M13mp18 ssDNA shows an interplay between the two proteins that enables hexameric *hsRuvBL2* to bind DNA. *HsRuvBL2* is not able to bind ssDNA when in the hexameric form (lane 6). However, when co-incubated with ssDNA and monomeric *hsRuvBL1* (lanes 2 and 4), a shift in the DNA running distance is observed which indicates binding of both *hsRuvBL1* and *hsRuvBL2* to DNA (as compared to the shifts in lanes 1 and 3, in which the **total** molar amount of protein is the same as in lane 2). Parallel experiments done with AMP-PNP (lanes 5 and 6) support the ATP-dependence of binding to DNA. Agarose gel stained with SYBR gold DNA stain.**

We confirmed the lack of binding of hexameric *hsRuvBL2* to M13mp18 by EMSA (**Fig. 2.10, lanes 5 and 6**), as the shift displayed by DNA alone (**lane 5**) or in presence of hexameric *hsRuvBL2* (**lane 6**) was the same in both lanes. Interestingly, hexameric *hsRuvBL2* was able to bind DNA when co-incubated with monomeric *hsRuvBL1* (**lanes 2 and 4**): in these lanes, the shift in DNA running distance corresponds to a DNA which is bound to the total number of *hsRuvBL1* and *hsRuvBL2* molecules in the reaction. Lane 3 contains 25  $\mu\text{M}$  of *hsRuvBL1*, and lane 4 contains 12.5  $\mu\text{M}$  each of *hsRuvBL1* and *hsRuvBL2*. Binding to DNA is shown to be dependent on ATP hydrolysis, since prior co-incubation of monomeric *hsRuvBL1* with the non-hydrolysable ATP analogue AMP-PNP precludes the binding of any *hsRuvBL* to the DNA (**lanes 7 - 10**), and as such the running distance of DNA in the gel is the same as the one observed for free M13mp18.

The question remains whether *hsRuvBL2* returns to a monomeric state or open ring conformation upon DNA release. Our observation of open (notched) rings by EM (**Fig. 2.6**), shows that *hsRuvBL2* may acquire that conformation, and it may be one way through which the rings are released from DNA. A second possibility would be a total ring fragmentation, in a way similar to what happens with the CMG helicase at the end of DNA replication<sup>23</sup>.

Finally, despite presenting an OB-fold, it is possible that the functions of the external region of domain 2 may be more related to its indirect interaction with the nucleotide in the binding pocket (through a mechanism explored in the third chapter), than to a direct interaction with DNA. It is also worth noting that while the OB-fold is mostly present in nucleic acid-binding proteins, they can also occur in larger proteins as recognition domains, and have even been found at protein-protein interfaces<sup>4</sup>. Thus, although an interaction of the domain II with DNA was observed for *hsRuvBL1*<sup>20</sup>, it is possible that the OB-fold in *hsRuvBL2* may also be important for its interaction with other proteins in a supramolecular assembly, such as within the Ino80 complex, where cross-linking/mass spectrometry analysis

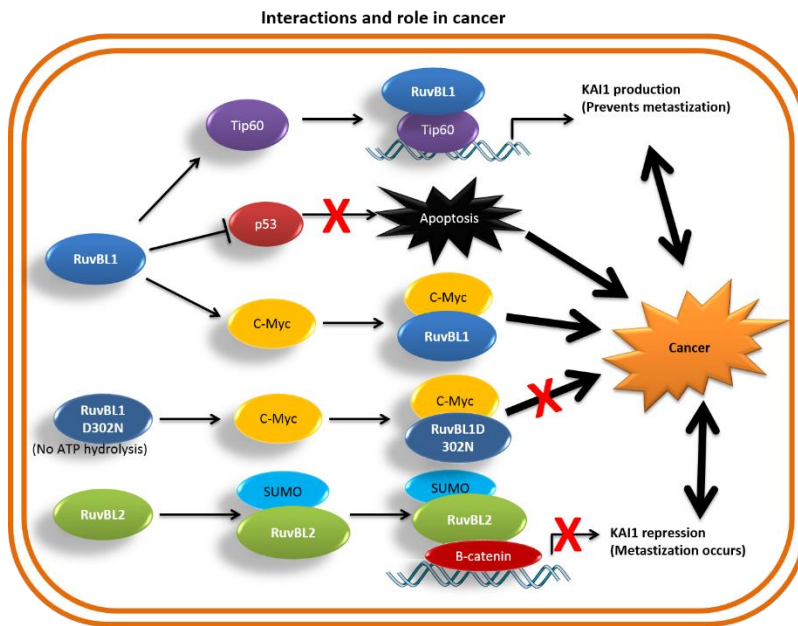
shows a strong interaction profile between the OB-folds of *ctRuvBLs* and Ies2<sup>24</sup>. Interestingly, EM micrographs showed *hsRuvBL2* hexamers to be frequently found in direct lateral contact (not shown), forming continuous strings, which is also suggestive of a tendency of domain 2 to act as a mediator in protein-protein interactions.

## 2.2 Preliminary study of RuvBL-interacting protein c-Myc

c-Myc is an oncogenic transcription regulator, for which a full-length structure has never been determined. To date, the only known structure is that of the C-terminal domain, responsible for interaction with DNA, in complex with the C-terminal domain of Max protein. The N-terminal portion of c-Myc, the Transcription Activation Domain (TAD), is responsible for interaction and regulation of other proteins, including RuvBL1. It has been known for some time that the interaction between RuvBL1 and c-Myc occurs between Myc Box II (MBII) and the domain DII of RuvBL1. This interaction, when de-regulated, most often leads to oncogenesis<sup>25</sup>. **Fig. 2.11** depicts some oncogenic pathways in which RuvBLs are involved. Considering this, and the importance of c-Myc as a pharmaceutical target, our proposal was accepted by the Oxford Protein Production Facility for the cloning, high-throughput expression and solubility screening of an assembly of constructs of human c-Myc (*hsc-Myc*, henceforth referred to as c-Myc). The main objective was to find a combination of factors leading to the attainment of soluble protein constructs, with emphasis on the segment responsible for the interaction with RuvBL1. On the long term, the expression of c-Myc MBII, together with the existing knowledge on the expression, purification and crystallization of *hsRuvBL1* and 2, will enable us to follow the next rational step: the study of the interaction between *hsRuvBL1*, *hsRuvBL2* and c-Myc. The main goal of that study will be to

understand how the interaction between c-Myc and *hsRuvBLs* occurs on a structural level. This knowledge may lead to the development of a strategy to modulate this interaction.

The present work was developed at the Research Complex at Harwell (part of the Oxford Protein Production Facility), under the supervision of Dr. Louise Bird.



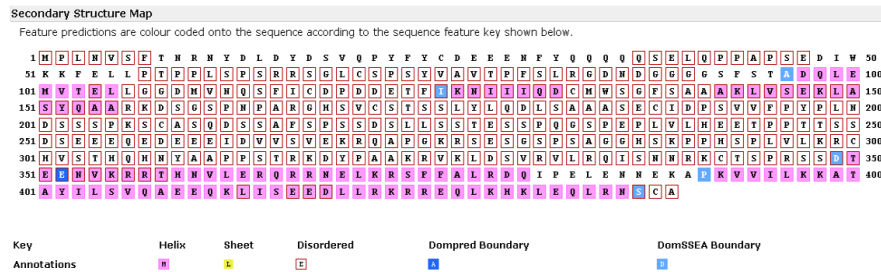
**Figure 2.11 – Involvement of *hsRuvBL1* and 2 in cancer pathways.** The *hsRuvBL1*-c-Myc cancer progression pathway is dependent on the ATPase activity of RuvBL1.

## 2.2.1 METHODS

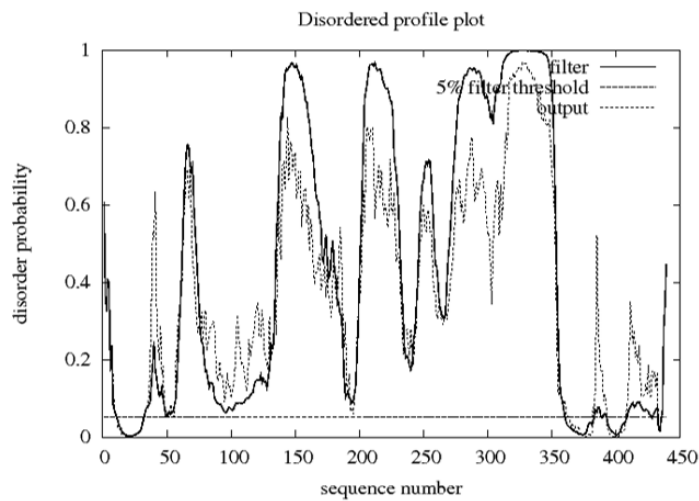
### *Construct design*

c-Myc is an intrinsically disordered protein, according to Disorpred and PsiPred<sup>26</sup> (Fig. 2.12). Wood and colleagues determined, through specific deletions,

that c-Myc binds to RuvBL1 through amino acids 118-152 of c-Myc<sup>25</sup>. Thus, constructs were selected in order to include this region, while also trying to maintain secondary structure-containing domains and avoiding the more disordered regions, which tend to decrease solubility and crystallization. **Table 2.1** enumerates the constructs and vectors selected for expression. 47 constructs were



A



B

**Figure 2.12 – c-MYC secondary structure prediction.** A) PsiPred secondary structure prediction shows most of c-Myc as being disordered. B) Disordered indicates the regions with highest probability of being disordered.

designed, based on their functional significance and disorder probability. One positive control (GFP) was included, for visual assessment of the success of the expression tests.



Tags were chosen on the basis of their records as solubility enhancing. All vectors include also a 6xHis tag and a cleavage site for HRV 3C protease. The latter feature is indispensable, for the presence of tags, being very large and mobile elements, may hinder crystallization of the target protein.

**Table 2.1. Constructs tested at the OPPF.** MBI – Myc Box I; MBII – Myc Box II; TAD – Transcription Activation Domain. SUMO – Small Ubiquitin-like Modifier; GST – Glutathione-S-transferase; HALO – Halotag (Promega); MBP – Maltose Binding Protein; TF – Trigger Factor; NusA – E. coli NusA; GFP - Green Fluorescent Protein (positive control for expression).

Tags (vector)	Constructs							
	1-439 (FL)	15-158 (MBI + II)	1-160 (MBI + II)	1-262 (TAD)	36-160 (MBI + II)	91-158 (MBII)	96-155 (MBII)	48-158 (MBII)
SUMO (pOPINS3C)	X	X	X	X	X	X	X	X
GST (pOPINJ)	GFP	X	X	X	X	X	X	X
HALO (pOPINHALO)	X	X	X	X	X	X	X	X
MBP (pOPINM)	X	X	X	X	X	X	X	X
TF (pOPINTF)	X	X	X	X	X	X	X	X
NusA (pOPINNUSA)	X	X	X	X	X	X	X	X

### *Cloning procedures*

Template DNA (pET49-b\_c-Myc) was obtained from GenScript, codon-optimized for expression in *E. coli*. PCR products of our target constructs were obtained using Phusion Flash polymerase, from Thermo. PCR products were cleaned using magnetic beads (Azincourt AMPure XP, from Beckman Coulter), after treatment with DpnI restriction enzyme to remove the methylated template DNA. Cloning was achieved using In-Fusion technology, from Clontech. For this, the linearized vectors were mixed with our PCR products, and the In-Fusion polymerase creates regions of homology at the ends. After 30 minutes at 42°C, the reaction was stopped by adding EDTA and the ligation was transformed in highly

competent OmniMAX cells, from Invitrogen. Cells were plated in 24 well plates, in LB mixed with 50 µg/mL carbenicillin, X-gal and IPTG. Individual white colonies were picked and grown overnight in Power Broth mixed with 50 µg/mL carbenicillin. Glycerol stocks and minipreps were prepared for each clone, using the Theonyx Liquid Performer Robot and the Wizard SV96 kit from Promega. A PCR was performed to confirm the presence of the tagged insert, using as forward primer pOPINF (similar to T7 but larger, since annealing temperature of the T7 primer is too low), and the reverse primer specific for each insert. Insert and tag sizes were verified by agarose gel using SYBR Safe gel stain. The high efficiency observed in our PCRs is most likely a consequence of the high quality of the template, since it was optimized for codon usage in *E. coli* and also in order to abolish any secondary structures that the template might form, which would diminish the efficiency of primer binding.

### ***Protein expression and purification***

Each vector was transformed in *E. coli* Lemo21 and Rosetta 2 competent cells. Lemo21 cells are adequate for the expression of membrane, toxic or difficult soluble proteins, and by adding L-rhamnose to the growth medium, levels of expression can be tightly tuned by modulating the level of lysozyme (lysY), the inhibitor of T7 RNA polymerase. Rosetta 2 cells contain the pRARE2 plasmid, which allows for the translation of 7 codons mostly used by eukaryotes, by supplying the respective tRNAs. However, since the gene contained in our expression vector had been codon optimized, this strain worked as simple BL21 strain. One colony of each was grown overnight at 600 rpm, in 600µL of Power Broth supplemented with 30 µg/mL chloramphenicol and 50 µg/mL carbenicillin.

150 µL of the overnight culture was inoculated into 3 mL (5%) of Power Broth or auto-induction medium supplemented with carbenicillin and

chloramphenicol. For Lemo21 cells, media were additionally supplemented with 250  $\mu$ M Rhamnose. (Levels of expression may be optimized later by varying the concentration of Rhamnose - no Rhamnose will produce the same effect as if we were inducing in a BL21 strain). Growth occurred at 37°C, 600 rpm. For cells growing in Power Broth medium, induction was achieved by adding 1 mM IPTG after 5 – 6 h of growth. After the addition of IPTG, growth proceeded at 20°C overnight. Cells in auto-induction medium were transferred to 25°C after 5 hours, and left to grow overnight. The Rosetta strain cells took longer to achieve induction OD, which is very likely due to the toxicity caused by basal levels of protein expression. A positive control was used for immediate visual assessment of induction efficiency, by inducing expression of GFP under the same conditions as the constructs.

Purification of 1 mL induced cells was performed on the Qiagen Bio Robot 8000 or Theonyx Robot, using GE His Mag sepharose magnetic beads. The cell pellet was first lysed, by freezing at -80°C for 20 minutes and then adding lysis buffer (Na Phosphate pH 8, 10 mM Imidazole, DNase I, Lysozyme and a small percentage of detergent – this small percentage is abolished when physically disrupting cells, on a large-scale). The protein was purified from the supernatant after centrifuging 30 minutes at 5000  $\times$  g with the magnetic beads, washed with buffer containing 20 mM Imidazole and eluted with 60  $\mu$ L of buffer containing 250 mM Imidazole. 10  $\mu$ L of each sample were mixed with 10  $\mu$ L 2x loading buffer and 10  $\mu$ L of this mix were applied in a 24 well 10% BT gel. The gel run was performed at 200 V for 39 minutes and stained directly using Generon Quick Coomassie stain.

## 2.2.2 RESULTS AND DISCUSSION

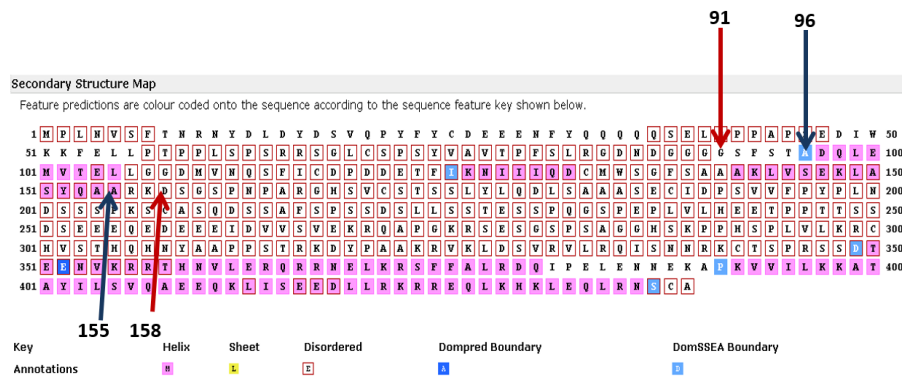
Of the 192 tested conditions, 5 produced promising results, both in PB and auto-induction medium, all of which using the Lemo21 cells. In many cases, cells failed to grow even before induction, mainly in Rosetta, which indicates toxicity caused by basal levels of protein expression. Purified protein also suffered proteolysis in the vast majority of cases, which resulted in the production of only the tags, or smaller portions of the expressed construct.

**Table 2.2. Results of c-Myc constructs expression tests.** Green squares indicate successful expression and presence of non-degraded protein after the final purification tests. The darker squares indicate higher levels of protein at the end of the purification step. The two darker constructs were successfully produced in PB and auto-induction medium.

Tags (vector)	Constructs							
	1-439 (FL)	15-158 (MBI + II)	1-160 (MBI + II)	1-262 (TAD)	36-160 (MBI + II)	91-158 (MBII)	96-155 (MBII)	48-158 (MBII)
SUMO (pOPINS3C)	X	X	X	X	X	X	X	X
GST (pOPINJ)	GFP	X	X	X	X	X	X	X
HALO (pOPINHALO)	X	X	X	X	X	X	X	X
MBP (pOPINM)	X	X	X	X	X	X	X	X
TF (pOPINTF)	X	X	X	X	X	X	X	X
NusA (pOPINUSA)	X	X	X	X	X	X	X	X

The best results were obtained for c-Myc constructs 91 – 158 and 96 – 155 (Table 2.2), which are the minimal constructs that still include MBII. These start on a region predicted as not unstructured and end after a helix, which may contribute to their increased stability (Fig. 2.13). Several solubility tags were tested. The tag that produced the most stable protein was TF (Trigger Factor), which is a prokaryotic ribosome-associated chaperone protein.

The optimization of the MBII purification protocol is ongoing. So far, thermal shift assays have not yielded a conclusive result for an improved purification buffer, as initial fluorescence is high and overall signal intensity very low, due most likely to its intrinsically disordered nature and small size (6 KDa). Additionally, preliminary NMR analyses indicate a disordered tridimensional structure, and low stability, at a concentration of 1 mg/mL. As a consequence of its lack of an ordered structure, MBII has yielded no crystals to date (either with or without tag, although in the former case, it may be due to the high solubility provided by the tag), and co-crystallization with *hsRuvBL1* has likewise failed to provide crystals to date. Future optimization procedures will include co-expression with *hsRuvBL1*, both of the MBII domain, and the full-length c-Myc.



**Figure 2.13 – Successfully expressed constructs of c-Myc.** HT purification of c-MYC constructs was most successful with the two constructs indicated, possibly because they contain helices at the boundaries, and are relatively small constructs.

## References

1. Shamoo, Y. Single-stranded DNA-binding Proteins. *Encycl. Life Sci.* 1–7 (2002).
2. Lodish, H., Berk, A. & Zipursky, S. in *Molecular Cell Biology* Section 4.1 (W. H. Freeman, 2000).
3. Walker, J. M. *Single-Stranded DNA binding proteins.* (2012).
4. Theobald, D. L., Mitton-Fry, R. M. & Wuttke, D. S. Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 115–133 (2003).
5. Papin, C. *et al.* 3'- to 5' DNA unwinding by TIP49b proteins. *Febs J* **277**, 2705–2714 (2010).
6. Gorynia, S. *et al.* Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex. *J Struct Biol* **176**, 279–291 (2011).
7. Kanemaki, M. *et al.* TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuVB-like DNA helicase, TIP49a. *J. Biol. Chem.* **274**, 22437–22444 (1999).
8. Makino, Y., Kanemaki, M., Kurokawa, Y., Koji, T. & Tamura, T. A. A rat RuvB-like protein, TIP49a, is a germ cell-enriched novel DNA helicase. *J. Biol. Chem.* **274**, 15329–15335 (1999).
9. Gribun, A., Cheung, K. L. Y., Huen, J., Ortega, J. & Houry, W. A. Yeast Rvb1 and Rvb2 are ATP-Dependent DNA Helicases that Form a Heterohexameric Complex. *J. Mol. Biol.* **376**, 1320–1333 (2008).
10. López-Perrote, A., Muñoz-Hernández, H., Gil, D. & Llorca, O. Conformational transitions regulate the exposure of a DNA-binding domain in the RuvBL1-RuvBL2 complex. *Nucleic Acids Res.* **40**, 11086–11099 (2012).
11. Cole, J. L., Lary, J. W., Moody, T. & Laue, T. M. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. **84**, 143–179 (2008).
12. Brown, P. H. & Schuck, P. A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. *Comput. Phys. Commun.* **178**, 105–120 (2008).
13. Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **78**, 1606–1619 (2000).
14. Papin, C. *et al.* 3'- to 5' DNA unwinding by TIP49b proteins. *FEBS J.* **277**, 2705–2714 (2010).
15. Wood, M. A., McMahon, S. B. & Cole, M. D. An ATPase/Helicase Complex Is an Essential Cofactor for Oncogenic Transformation by c-Myc. *Mol. Cell* **5**, 321–330 (2000).

16. Cheung, K. L. Y., Huen, J., Kakihara, Y., Houry, W. A. & Ortega, J. Alternative oligomeric states of the yeast Rvb1/Rvb2 complex induced by histidine tags. *J. Mol. Biol.* **404**, 478–92 (2010).
17. Rivera-Calzada, A. *et al.* The Structure of the R2TP Complex Defines a Platform for Recruiting Diverse Client Proteins to the HSP90 Molecular Chaperone System. *Structure* (2017). doi:10.1016/j.str.2017.05.016
18. Etard, C., Gradl, D., Kunz, M., Eilers, M. & Wedlich, D. Pontin and Reptin regulate cell proliferation in early *Xenopus* embryos in collaboration with c-Myc and Miz-1. *Mech. Dev.* **122**, 545–556 (2005).
19. Lakomek, K., Stoehr, G., Tosi, A., Schmailzl, M. & Hopfner, K.-P. Structural Basis for Dodecameric Assembly States and Conformational Plasticity of the Full-Length AAA+ ATPases Rvb1/Rvb2. *Structure* **23**, 1–13 (2015).
20. Matias, P. M., Gorynia, S., Donner, P. & Carrondo, M. A. Crystal structure of the human AAA+ protein RuvBL1. *J Biol Chem* **281**, 38918–38929 (2006).
21. Queval, R., Papin, C., Dalvai, M., Bystricky, K. & Humbert, O. Reptin and Pontin oligomerization and activity are modulated through histone H3 N-terminal tail interaction. *J. Biol. Chem.* **289**, 33999–4012 (2014).
22. Gorynia, S. *et al.* in *Methods in Enzymology* 55–62 (2012). doi:10.1007/978-94-007-2530-0
23. Maric, M., Maculins, T., De Piccoli, G. & Labib, K. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science* **346**, 1253596 (2014).
24. Tosi, A. *et al.* Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* **154**, 1207–19 (2013).
25. Wood, M. A., McMahon, S. B. & Cole, M. D. An ATPase/Helicase Complex Is an Essential Cofactor for Oncogenic Transformation by c-Myc. *Mol. Cell* **5**, 321–330 (2000).
26. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

# *Chapter 3*

---

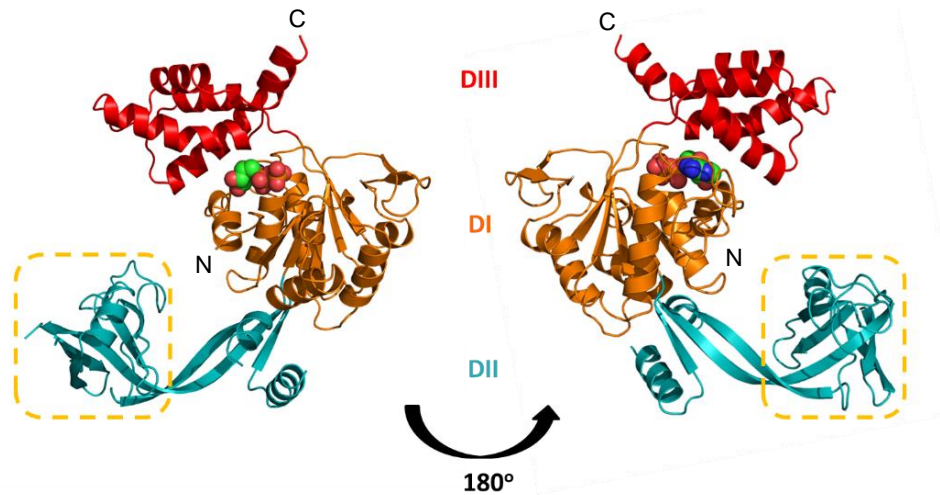
STRUCTURE OF HUMAN RUVBL2



<b>3.1 Current structural knowledge on RuvB-Like proteins .....</b>	<b>108</b>
<b>3.2 Structure of RuvBL2 at 2.8 Å by X-ray crystallography .....</b>	<b>120</b>
3.2.1 METHODS	120
Protein production	120
Protein crystallization and crystal optimization	121
Data collection	127
Structure solution and refinement	128
3.2.2 RESULTS AND DISCUSSION	131
Purification of RuvBL2	131
Crystallization of RuvB-Like 2	132
Initial attempts to solve the structure from low resolution data	135
Data collection of crystals diffracting to 2.8 Ångstrom	138
Atomic structure of full-length RuvB-Like 2	139
Structural basis for coupling ATP binding to mechanical action	145
<b>3.3 Structure of RuvBL2 by electron microscopy.....</b>	<b>152</b>
3.3.1 METHODS	153
Protein production	153
Data collection - negative staining EM	155
Data processing and refinement of negative staining data	155
3.3.2 RESULTS AND DISCUSSION	156
Low resolution structure of RuvBL2 by negative staining EM	156
<b>References .....</b>	<b>160</b>

### 3.1 Current structural knowledge on RuvB-Like proteins

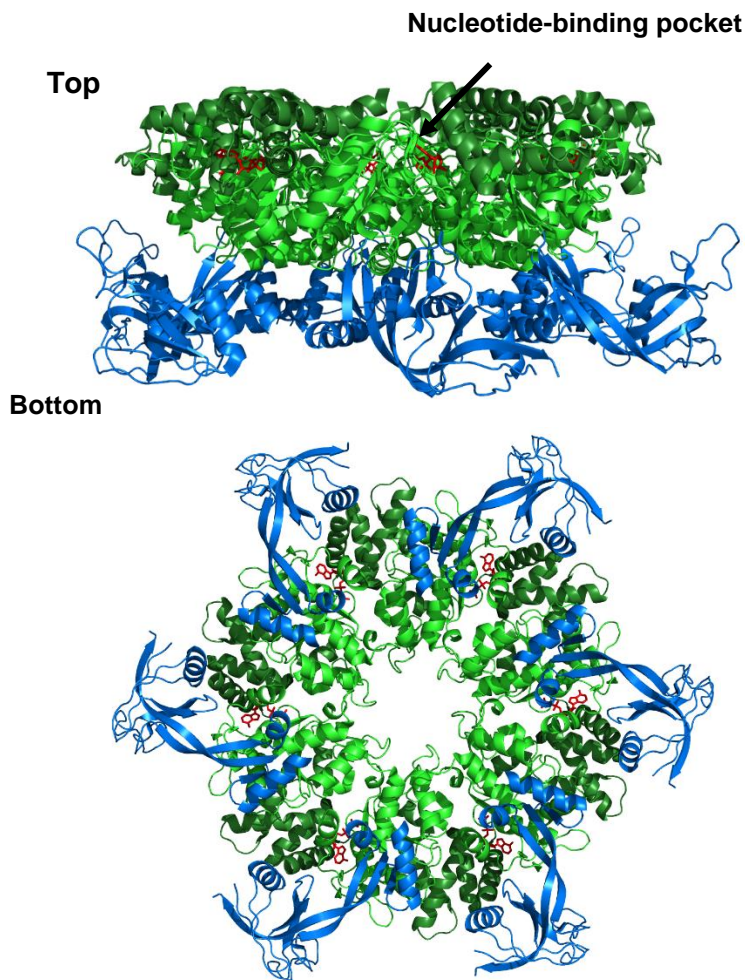
RuvB-Like proteins are widespread in Archaea and eukaryotes. To date, structural studies have been mostly based on electron microscopy (EM) and X-ray crystallography, with a few SAXS analyses to correlate with the overall protein shape in solution. The first structure of the human RuvBL1/RuvBL2 complex was published in 2007<sup>1</sup>, followed by the yeast Rvb1/Rvb2 complex in 2008<sup>2</sup>, both by negative staining EM. In 2005, RuvBL1 crystals had already been produced<sup>3</sup>, and in 2006 the first crystallographic structure of RuvBL1 alone was published<sup>4</sup>.



**Figure 3.1 – Cartoon model of the human RuvBL1 monomer.** Domain I is depicted in orange, domain II in cyan and domain III in red. ADP atoms are shown as spheres. The OB-fold is located at the outermost part of domain II (highlighted).

RuvBL1 and RuvBL2 are homologous proteins that share 43% sequence identity and 65% sequence similarity, and are composed of three domains, I, II and III (**Fig. 3.1**). The Rossmann-like  $\alpha\beta\alpha$  fold (**domain I**) and the canonical all- $\alpha$  subdomain (**domain III**) form the ATPase core. The smaller domain III forms a “lid” located near the P-loop (phosphate-binding loop), which covers the

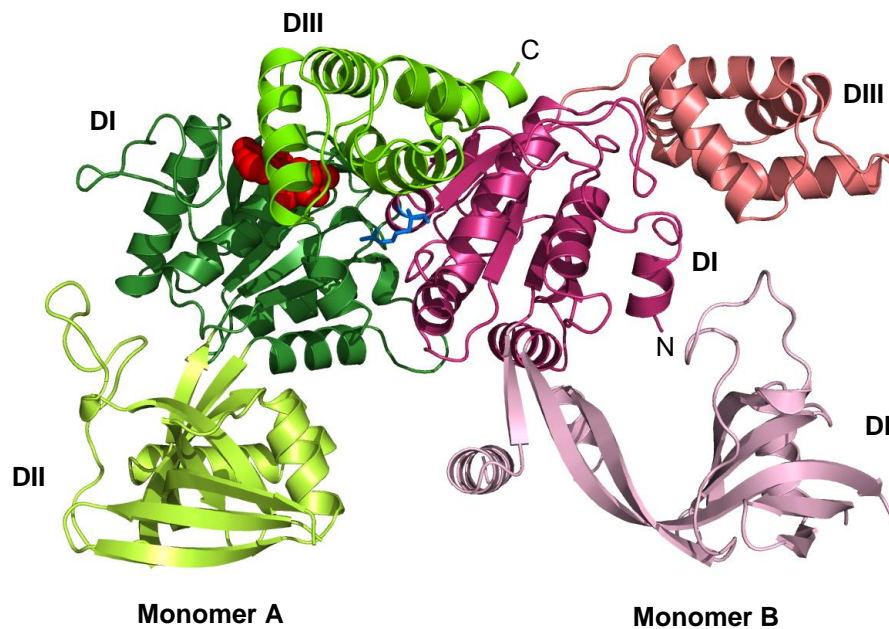
nucleotide-binding pocket at the interface of domain I and domain III. Between the canonical Walker A and Walker B motifs of domain I there is an insertion domain



**Figure 3.2 – Cartoon model of the *hsRuvBL1* hexamer (PDB ID 2c9o).** The ATPase core is depicted in green (light for domain I and dark for domain III), and the domain II in blue. ADP is represented as red sticks. **Top:** side view; **bottom:** bottom view.

(**domain II**), that bears no significant sequence similarity to other known domains, except for residues 131-227 (*RuvBL1*), which organise into an OB-fold. Domain II is unique to *RuvB*-Like proteins. Like all AAA+ ATPases, *RuvBLs* assemble into a

hexameric torus-shaped complex (Fig. 3.2). The interfaces form a complete nucleotide binding pocket, which includes a *trans*-arginine finger from the adjacent monomer (Fig. 3.3).



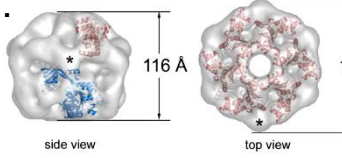
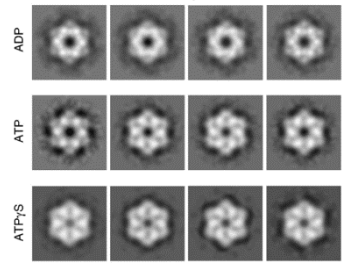
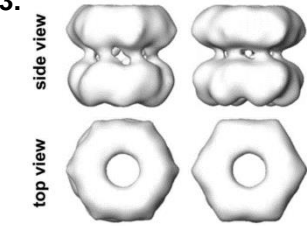
**Figure 3.3 – Interface between two monomers of RuvBL1.** The interfaces are formed by contacts between domains I and III. The two adjacent monomers are represented respectively in shades of green and in shades of pink, each shade defining a domain. The ATP molecule from the left monomer is depicted in red spheres. The *trans*-Arginine finger from the right-side monomer, that completes the nucleotide binding pocket, is depicted in blue sticks.

Structural studies show that, when mixed, RuvBLs form complexes composed of alternating RuvBL1 and RuvBL2 monomers, assembling as hexamers<sup>5,6</sup> or dodecamers<sup>2,5-15</sup>, both in humans, yeast and *Chaetomium thermophilum*. However, only structures of RuvBL1/2 dodecamers have been reported (Table 3.1). These proteins are also able to form homomeric complexes, of

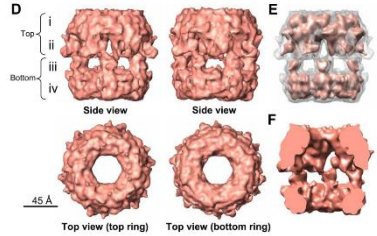
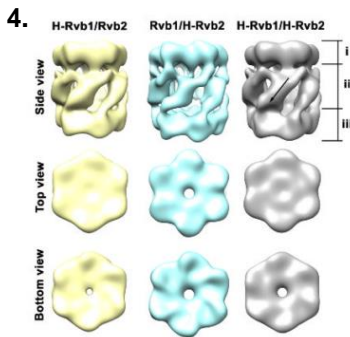
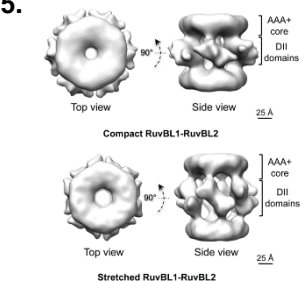
which thus far only the crystallographic structure of hexameric RuvBL1 has been determined (Fig. 3.2)<sup>4,16</sup>.

**Table 3.1 – Structures of RuvBL1/RuvBL2 complex determined to date by electron microscopy.**

Some differences between the different studies are described, in what refers to expression, tags used and type of complex obtained. Differences in the purification buffers used may also have an important contribution to the observed structures. When the type of complex obtained is not described as hetero or homododecamer/hexamer, it is due to the fact that the resolution attained in that work did not allow for that distinction.

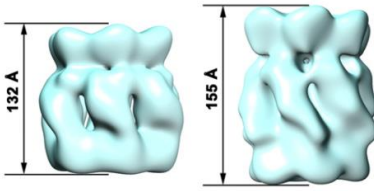
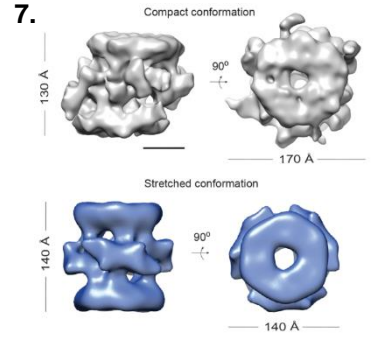
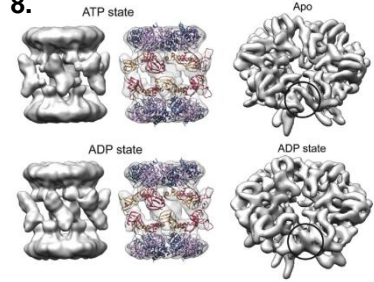
EM structure	Complex production	Remarks
<p><b>1.</b></p>  <p>side view 116 Å top view 158 Å</p> <p>Puri et al 2007</p>	<p>Separately expressed in <i>E. coli</i>. Complex assembly by <b>co-incubation</b> of partially purified RuvBL1 and RuvBL2_His6.</p>	<p>Human. Forms asymmetric dodecamers with equimolar RuvBL1 and RuvBL2 amounts. RuvBL1/RuvBL2-His6.</p>
<p><b>2.</b></p> <p>2D class averages obtained from reference-free image classification</p>  <p>ADP ATP ATP-S</p> <p>Gribun et al 2008</p>	<p>Separately expressed in <i>E. coli</i> with N-terminal tags. Tags were cleaved prior to <b>co-incubation</b>.</p>	<p>Yeast. Forms heterohexamer, 6-fold symmetric. Mixed in presence of ADP.</p>
<p><b>3.</b></p>  <p>side view top view EDTA ATP</p> <p>Torreira et al 2008</p>	<p><b>Co-expression</b> in recombinant <b>baculovirus</b> (insect cells).</p>	<p>Yeast. Rvb1/His-Rvb2. Negative staining-EM.</p>

**Table 3.1 – Structures of RuvBL1/RuvBL2 complex determined to date by electron microscopy.** Some differences between the different studies are described, in what refers to expression, tags used and type of complex obtained. Differences in the purification buffers used may also have an important contribution to the observed structures. When the type of complex obtained is not described as hetero or homododecamer/hexamer, it is due to the fact that the resolution attained in that work did not allow for that distinction.

EM structure	Complex production	Remarks
 <p><b>Torreira et al 2008</b></p>	<p>Yeast. Rvb1/His-Rvb2. Forms asymmetric dodecamers with equimolar amounts of RuvBL1 and RuvBL2. Cryo-EM.</p>	
 <p><b>Cheung et al 2010</b></p>	<p>Purified <b>endogenous</b> Rvb1/Rvb2 from <i>Saccharomyces cerevisiae</i>.</p>	<p>Yeast. N-terminal tagged Rvb1 and Rvb2 form dodecamers, and the untagged proteins form hexamers. Expression system and mode of complex assembly do not affect oligomeric state.</p>
 <p><b>López-Perrote et al 2012</b></p>	<p><b>Co-transformation of His-RuvBL1 and RuvBL2 in <i>E. coli</i>.</b></p>	<p>Human. Untagged RuvBL1/RuvBL2 (negative staining) and His-RuvBL1/RuvBL2 (cryo-EM). Dodecameric assembly maintained after tag removal. Compact and stretched conformations independent of nucleotide present.</p>

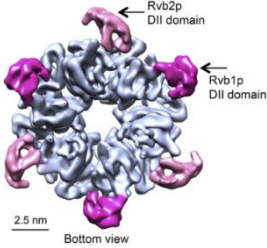


**Table 3.1 – Structures of RuvBL1/RuvBL2 complex determined to date by electron microscopy.** Some differences between the different studies are described, in what refers to expression, tags used and type of complex obtained. Differences in the purification buffers used may also have an important contribution to the observed structures. When the type of complex obtained is not described as hetero or homododecamer/hexamer, it is due to the fact that the resolution attained in that work did not allow for that distinction.

EM structure	Complex production	Remarks
<p><b>6.</b></p>  <p><b>Jeganathan et al 2015</b></p>	<p><b>Co-expression</b> of His-Rvb1 and Rvb2 in <i>E. coli</i>.</p>	<p>Yeast. Structure of untagged complex by negative staining EM. 8% of complex is dodecamer, with four levels of compaction.</p>
<p><b>7.</b></p>  <p><b>Silva-Martin et al 2016</b></p>	<p><b>Expressed separately</b> in <i>E. coli</i> and subsequently <b>co-incubated</b> as monomers.</p>	<p><i>Chaetomium thermophilum</i> (fungus). Cryo-EM of untagged dodecamer. Stretched and compact conformations. Asymmetric features.</p>
<p><b>8.</b></p>  <p><b>Ewens et al 2016</b></p>	<p><b>Co-expression</b> of His-Rvb1 and Rvb2 in <i>E. coli</i>.</p>	<p>Yeast. Cryo-EM of untagged Rvb1/Rvb2. Nucleotide-dependent bent and straight conformations. Rotation of one hexamer relative to the other in a degree dependent on nucleotide.</p>

**Table 3.1 – Structures of RuvBL1/RuvBL2 complex determined to date by electron microscopy.**

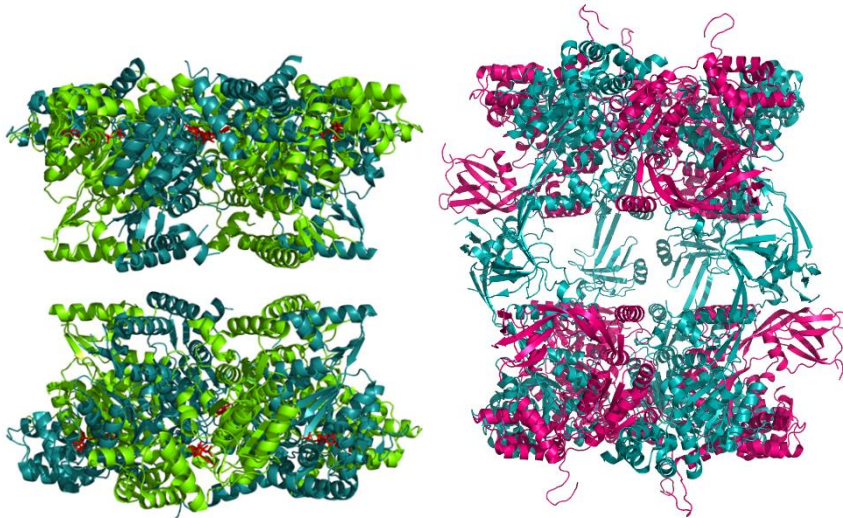
Some differences between the different studies are described, in what refers to expression, tags used and type of complex obtained. Differences in the purification buffers used may also have an important contribution to the observed structures. When the type of complex obtained is not described as hetero or homododecamer/hexamer, it is due to the fact that the resolution attained in that work did not allow for that distinction.

EM structure	Complex production	Remarks
<p>9.</p>  <p>2.5 nm Bottom view</p> <p>Rivera-Calzada et al 2017</p>	<p>Co-expression of Myc-RuvBL1 and His-RuvBL2 in <i>E. coli</i>.</p>	<p>Yeast. Cryo-EM of tagged heterohexamers.</p>

In the human proteins, the abscission of domain II led to preferential dodecamer formation (Fig. 3.4)<sup>12</sup>, while in yeast the dodecameric complex was shown to be formed due to the purification tags since, in their absence, hexamers preferentially formed<sup>6</sup>. Although this information should be taken into consideration (particularly when using tags for *in vivo* studies), it probably does not impair the validity of the structural information gathered from structures thus obtained, in particular considering that double ring formations have been observed to occur naturally<sup>17</sup>. The question of whether a dodecamer of homohexameric rings of RuvBL1 and RuvBL2 can exist is still unanswered. Some low resolution EM structures show an asymmetric dodecamer<sup>1,18</sup>, where one of the halves of the dodecamer is wider than the other (position 1 of table 3.1). Nevertheless, the possibility should be considered that this may be a transient conformational state, which has never been observed to date in the structures of double heterohexameric rings. Asymmetry caused by conformational states induced by nucleotide binding may also be considered. However, in structures where conformational differences



due to nucleotides have been observed, the diameter of both rings remained the same (position 7 of **table 3.1**). Alternatively, the asymmetric dodecamers could be composed of two distinct types of hexamers, or might be formed instead through contacts between domains II of the top ring and the ATPase core of the bottom ring (i.e., top-to-bottom, instead of the so far observed bottom-to-bottom).



**Figure 3.4** – Cartoon representations of the RuvBL1/RuvBL2 dodecamer. **Left:** human domain II-truncated complex (PDB ID 2xsz). RuvBL1 monomers depicted in light green and RuvBL2 monomers in blue-green. ADP represented in red sticks. **Right:** *Chaetomium thermophilum* complex. RuvBL1 monomers depicted in blue and RuvBL2 monomers depicted in pink. Side views.

The highest resolution structures show that the interactions between hexamers occur between the domain II of RuvBL1 and RuvBL2<sup>7-9,12</sup> (**Fig. 3.4**, right). These domains are linked to the ATPase core by a flexible  $\beta$ -sheet link, which provides flexibility to the cage-like complex. The complex is thus able to acquire different conformations, depending on the nucleotide present in the binding pocket, and possibly on bound DNA or other proteins. The connection between ATP hydrolysis and mechanical movement remains a key missing link to our understanding of the biological functions of this complex, and a topic we aimed to address in this work. In the yeast dodecameric complex, ATP and ADP binding

lead, respectively, to the opposing movement of both hexamers<sup>14</sup> by 29 and 32 degrees, where the hinges are formed by the connections between the  $\alpha/\beta$  tips of the domains II (OB folds) of RuvBL1 and RuvBL2<sup>8</sup>. So far there are no comparative 3D structures of apo vs. nucleotide-bound hexamers, but it has been suggested that the motion of domains II in a hexamer would be the same as in the dodecamer<sup>8</sup>. This motion leads to a compaction of the dodecamer height, from 145 Å to 130 Å, observed in a number of studies<sup>10,19</sup> (**Table 3.1**).

Interestingly, it has been observed in several instances that nucleotide binding leads to a rearrangement and stabilization of the N-terminal loop<sup>7,20,21</sup>, in addition to the already described movement of domain II. Molecular dynamics simulations of RuvBL hexamers and monomers in solution show that the major movements are associated with domain II and the C-terminal helix (domain III). In all cases, domain II acquires a wide variety of conformations, and its mobility is further enhanced by the addition of 3 residues (Phe-Cys-Arg) within the OB-fold, in the lik mutant<sup>22,23</sup>. Interestingly, the addition of these residues also increases the mobility of amino acid residues at the interface between domains I and III, where the nucleotide binding pocket is located, suggesting a connection between the motion of domain II and nucleotide exchange, supported by the results obtained with the domain II-truncated complex, which show an increase in ATPase activity<sup>12,24</sup>.

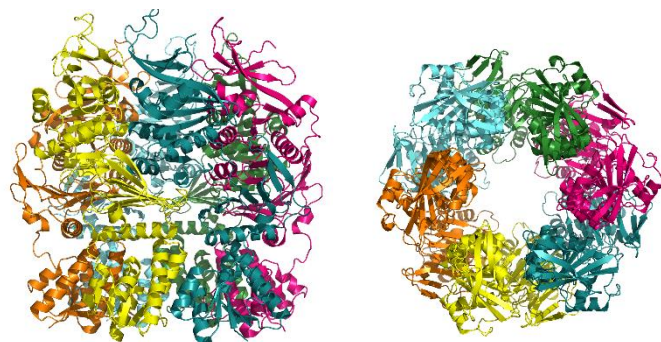
The double rings observed by EM enclose a large central cavity. It has been suggested that this central channel could be used to pass DNA strands; both diameter and charge suggest that it would most likely be ssDNA, although a structure of a RuvBL-DNA complex has yet to be produced. The existing structures show side openings, suggestive of a possible extrusion mechanism similar to that of the archaean MCM<sup>25</sup> or the simian virus 40 LTag<sup>26</sup>. On the other hand, the motion caused in the dodecamer by ATP and ADP binding exposes the OB-fold<sup>8</sup>, which could wrap DNA in a way reminiscent of a film reel. This type of function would

fit well in complexes such as the INO80, where the dodecamer sits at the DNA-binding head, with the equatorial line between hexamers perpendicular to the “neck” of complex, exposing the external part of the domains II<sup>17</sup> (see chapter 1).

The issue of the physiologically relevant forms is still under debate, the reason for this being that ATPase and helicase activities have been observed in the heterohexamer, the truncated heterododecamer and in the RuvBL2 monomer but not in the RuvBL2 hexamer, and different results have been observed for the full-length heterododecamer<sup>2,18,21,27</sup>. However, this may not be the best way to assess the biological significance of the oligomer, since: a) oligomerization and activity are known to be regulated by cofactors present in the cell, such as the Histone H3 N-terminal tail<sup>16</sup>; b) some functions, such as chaperone, may not require ATPase activity; and c) activity data has frequently been determined with undefined oligomeric forms<sup>16</sup>.

Alternatively to working together, RuvBL1 and RuvBL2 also have independent activities<sup>27-29</sup>, and roles in the cell<sup>30,31</sup>. The oligomeric form in which they enact those roles is still a largely unexplored question. When working independently, their activities and association to binding partners are known to be regulated by markers such as SUMO or methyl groups<sup>30-34</sup>.

Despite the many structural and biochemical publications, some fundamental questions remain unanswered, thwarted by the mobile nature of these proteins and the large size of their complex, which hinder the crystallization efforts. What is the mechanism and biological function of DNA binding? How are the different activities of RuvBLs regulated? How do they perform their activities when in complex vs in the monomeric form? How is the hydrolysis of ATP translated into movement with functional meaning? So far, structural studies have focused on the human and fungal proteins, both using X-ray crystallography and electron microscopy. With the exponential improvement the latter technique has experienced in recent years, more structures of the full-length dodecameric RuvBL



NW Score	Identities	Positives	Gaps
-87	94/546(17%)	175/546(32%)	124/546(22%)
SiR2	1 MNDMLDKSALLFGVSKYLEKGIITGNVLIHKSLLAELERESNDGLVSAE		50
hsR2	1 -----MATV		4
SiR2	51 IALDEVKRLKDI--ERILVNFIVG---DDSK-----		79
hsR2	5 TATTKVPEIRDVTRIERIGAHSHIRGLDDALEPRQASQGMVGLAARR		54
SiR2	80 --GEANLSREYCLEKGCIIVTADETQKKICDAMGIQYNFLQPLKQGLSF		127
hsR2	55 AAGVLEMIREKGIAGRAVLIAGQPGTGKTAIAMGM---AQLGPDTPF		100
SiR2	128 ESFFDDETMSLHI--KEDTVPRAKKGKPGNKKFVNLSDKPMPLSTDVRMIAN		176
hsR2	101 TAIAGSEIFSLMSKTEALTQAFR-----RSIGV		129
SiR2	177 EIINAVRLIKGSFVEIERRGSLIIQLGNRVVITRPPLSDGWEITITREV		226
hsR2	130 RIKEETEIIIEGVEVIQ-----IDRPATGTGSKVGLTLK		164
SiR2	227 VRKRELYNLDERLIKRLKLE---ERAEGIIAGAPG---MGATTFAQAL		268
hsR2	165 TTEMETIYDLGTMIESLTKDKVQAGDVITIDKATGKISKLG-RSFTR--		211
SiR2	269 AEVYMRGLKIVKTIIESPRDMHLPPEITQYSKNYAEIGELHDI-----		310
hsR2	212 ARDYDAMGSQTKFVQCP--DGEL----QKRKEVVHIVSLHEIDVINSRTQ		255
SiR2	311 ---LLSRPDYTVYDEMKNDEDFKLYVDLRLAGVGMV--GVVHATSPIDA		355
hsR2	256 GFLALFSGDTGEIKSEVREQINAKV-AEWREEGKAEIIPGVLF---IDE		300
SiR2	356 IHRFVNRVDIGTIPNILDIIIFINSGNVSKVYVLEMTVKVPAGLKEADLA		405
hsR2	301 VH---MLDI-----ESFSFLNRA-----LESDMA		321
SiR2	406 RPVVEIKDLATGNTEYEIIVFGEQTMIVPVNRGIT-----MSNMEFKI		448
hsR2	322 -FV-----LIMATNRGITRIRGTSVQSPHGIFI		348
SiR2	449 SKIVNNIIPNATVKY--EDGEYVIVIP--KEEIGKYNRKLQRLKREKKN		495
hsR2	349 D-LLDRLLIVSTTPYSEKDTKQLLRIRCEEEDVEMSEDAYTVLTRIGLET		397
SiR2	496 NIKIKIKLSD-----		505
hsR2	398 SLRVAIQLITAASLVCRKRKGTVEVQDDIKRVVSYFLDESRSRSTQYMKYQ		447
SiR2	506 -----	505	
hsR2	448 DAFLFNLKGETMDS	463	

**Figure 3.5 – Crystallographic model of an *archaean* RuvB-Like Holiday junction helicase from *Sulfolobus islandicus*. Top:** Cartoon model of *Sulfolobus islandicus* RuvB-Like helicase (side view on the left and top view on the right). **Bottom:** Alignment of *S. islandicus* RuvBL (PDB ID 5F4H) with human RuvBL2 (PDB ID 5N7R). Alignment with EMBOSS Needle<sup>35</sup>.

complexes are being published, at increasingly higher resolutions. The supramolecular associations of RuvBLs with other proteins in high molecular weight complexes has been reviewed before (see chapter 1). Only two EM structures of such complexes have been determined to date; however, this is a fast-growing field, and we should expect the next few years to bring larger multi-complex structures.

Recently, the crystallographic structure of an archaean RuvB-Like Holiday junction helicase has been determined to 2.7 Å resolution (**Fig. 3.5, top**, no publication available). However, despite being annotated as a RuvB-Like protein, the barrel-like complex does not have the outwards protruding domains characteristic of the (so far) known structures of eukaryotic RuvBLs. The sequence alignment (**Fig. 3.5, bottom**) also shows very limited conservation to the human RuvBL2 (17% identity and 32% similarity, calculated with BLAST, Needleman-Wunsch alignment), and does not have the same domain distributions as *hsRuvBLs*.

The three-dimensional structures of the human and fungal RuvBL1/RuvBL2 complexes are very similar to that of helicases, with which they also share function. However, they also bear similarities to known chaperones, such as HSP60/HSP10, and this function is perhaps the most commonly found so far in multi-complex assemblies, where RuvBLs are present as a dodecameric complex. The difficulty in discerning the true function of these proteins comes from the fact that they have multiple activities, and it is challenging to attribute a certain topological area to a specific function, especially when they may overlap. This may be the case of the OB-fold, which binds DNA - at least in RuvBL1<sup>4</sup>, and proteins in both RuvBL1 and RuvBL2<sup>17,36</sup>.

## 3.2 Crystallographic structure of *hsRuvBL2* at 2.8 Å

### 3.2.1 METHODS

#### *Protein production*

*hsRuvBL2* was produced for crystallization purposes using the same purification protocol described in the previous chapter. During optimization of protein stability and crystallization, different constructs of *hsRuvBL2* were tested in order to improve the diffraction quality of the crystals: native and selenomethionine-substituted, with tags on the C- or N-terminus, and after tag removal from the C-terminus (the yield of untagged protein after tag removal from the N-terminus was too low for crystallization purposes). The native, N-terminus tagged form of *hsRuvBL2* was the only construct not amenable to concentration to 10 mg/mL or crystallization. Good quality-diffracting crystals were obtained from native, untagged *hsRuvBL2*.

The following steps were added to the purification protocol in order to increase protein purity and homogeneity for crystallization purposes. After the tag cleavage and removal step, *hsRuvBL2* was diluted in buffer to a final NaCl concentration of 50 mM and 500 µM ADP, injected into a **Resource Q** column and eluted with a gradient of buffer with 1 M NaCl. The resulting peaks were separately applied in the final S200 HiLoad 16/60 size-exclusion column, previously equilibrated in a buffer containing 20 mM phosphate pH 8, 200 mM NaCl, 3 mM MgCl<sub>2</sub> and 500 µM ADP. *hsRuvBL2* eluted as a hexamer at a concentration of 10 mg/mL.

*hsRuvBL2* was dialysed to buffers at different pH values prior to crystallization. Crystals diffracting to 2.8 Å were obtained using protein dialysed to pH 6.

### ***Protein crystallization and crystal optimization***

Crystallization drops were set up with *hsRuvBL2* tagged on the C- and N-terminus, and with untagged *hsRuvBL2*. Native and selenomethionine-substituted forms of the constructs were tested.

In order to find conditions promoting *hsRuvBL2* crystallisation, initial crystallization screens were set up, with drops of 0.2 nL total volume in 96 well-plates, using a Honeybee Cartesian Dispensing System, from Genomic Solutions. The crystallization screens performed are listed in **table 3.2**. Initial crystal hits obtained from the commercial screens were tentatively and continually optimized by varying combinations of the following factors: temperature; sitting/hanging-drop; micro batch (under oil); capillary diffusion; streak-seeding; addition of compounds (3 screens of 96 different compounds – “additives” - were tested for the most promising crystallization conditions); ionic liquids (one screen of 24 different ILs); precipitant concentration; buffer concentration; crystallization buffer pH; protein buffer pH; salt concentration; co-crystallization with nucleotides; proportion of protein/crystallization buffer (v/v) in the drop; addition of cryoprotectants; use of irregular surfaces to promote seeding. When crystal size was above 100  $\mu\text{m}$ , the diffraction quality was tested at room temperature, either at the *in-house* source (Bruker AXS Proteum X8 diffractometer with a Pt135 CCD detector, coupled to a Microstar-I rotating anode X-ray generator with Montel mirrors) or at a synchrotron beamline. This allows the assessment of the crystal diffraction quality prior to addition of cryoprotectants. Additionally, *in situ* diffraction measurements were performed at a synchrotron beamline, with crystals still in the drops. For this, we used EMBL-designed plates with an X-ray permissive base. At the ESRF beamline BM-14, the plate was fitted in a support equivalent to the goniometer used for mounting loops with single crystals.

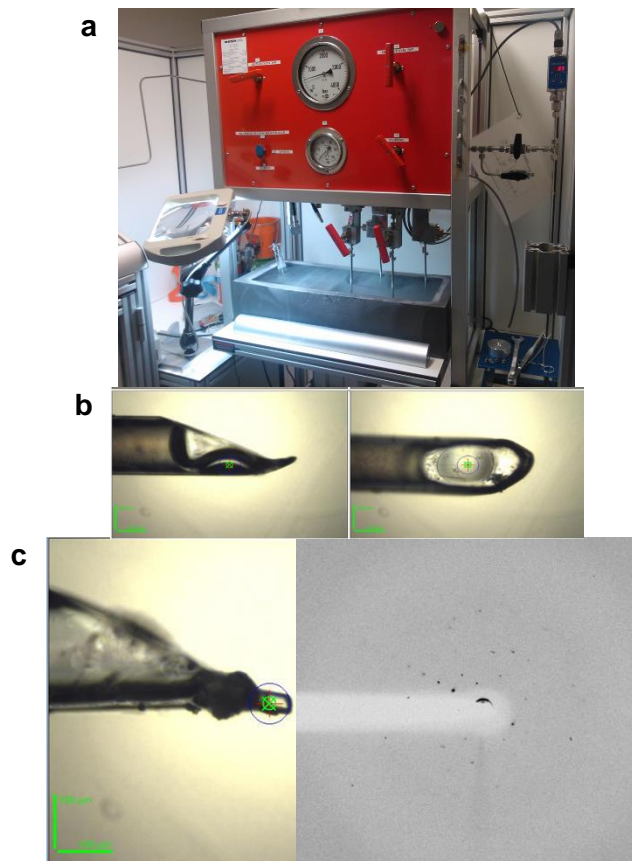
**Table 3.2 – List of crystallization screens performed with *hsRuvBL2*.** Most were used more than once for each construct, as changes in protein purification protocol may lead to a final sample with different characteristics. The additive screens were performed for each condition that produced promising crystals. The OPTI screens were designed in house by Ricardo Coelho, and test concentrations of four precipitants: PEG, NaCl, MPD and 2-propanol. These screens were also used only in the reservoir, by making drops with the crystallization condition. Commercial screens are from Qiagen, Jena Biosciences (JBS), Molecular Dimensions (MD) or Hampton Research (HR).

<b>Crystallization screens</b>	<b>Additive screens</b>
Structure 1&2 (MD)	Additive screen (Hampton)
Index (HR)	Silver bullets 1 and 2 (HR)
Pact Premier (MD)	Ionic liquids (HR)
JCSG+ (MD)	
Morpheus (MD)	<b>Other screens</b>
Midas (MD)	Crystal dehydration kit (JBS)
Membgold (MD)	Methylation kit (JBS)
MembFac (MD)	
Nucleix (Qiagen)	
Nuc-Pro (JBS)	
Natrix (HR)	
Salt Rx (HR)	
Proplex (MD)	
Footprint + MacroSol (MD)	
Ammonium sulphate (Qiagen)	
OPTI 1 and 2 (handmade)	
PGA (MD)	
MPD (Qiagen)	
Cations (Qiagen)	
Anions (Qiagen)	
Cryos (Qiagen)	
Grid screen (sodium malonate, NaCl, sodium/potassium phosphate, ammonium sulfate) (HR)	

Since the crystal must be removed from the drop and flash-frozen to minimize radiation damage during data collection, different cryoprotectant solutions and freezing protocols were tested, followed by an analysis of the resolution of the crystal, as well as anisotropy. Several cryoprotectants were tested for each



crystallization condition: PEG200 and 400, Fomblin Y, Li<sub>2</sub>SO<sub>4</sub>, MPD, dioxane, ethylene glycol, glycerol, LiCl, TMAO, NaCl, sucrose, trehalose, xylitol, paraffin, paratone, LV cryo oil and malonate. Alternatively, in an attempt to avoid addition of cryoprotectants which might disrupt crystal packing, two additional strategies were performed with the most promising crystals: **high-pressure cryocooling** and **dehydration**.



**Figure 3.6 – High-pressure cryocooling.** a) High pressure cryocooling chamber, with connections to the liquid nitrogen pool below. b) Many crystals collapsed after exposure to high pressure levels. c) High pressure did not increase resolution, although diffraction spots became more defined.

Using **high-pressure cryocooling (HPC)**, crystals were harvested in specially designed loops (**Fig. 3.6b**), loaded into a high-pressure chamber where they were pressurized with Krypton gas at 2000 bar, and immediately cooled to liquid nitrogen temperature (**Fig. 3.6a**). The crystals were then transferred to regular cryocaps under liquid nitrogen, for data collection. HPC was performed in a pressure chamber coupled to a liquid nitrogen bath, handmade by Phillippe Charpentier (ESRF, Grenoble), and the crystals were measured at ESRF beamline BM14. During high pressure cryocooling, the water in the channels of the crystal lattice turns into high-density amorphous ice, instead of the more disruptive low-density amorphous ice. This precludes the use of infiltratory chemical cryoprotectants<sup>37-39</sup>.

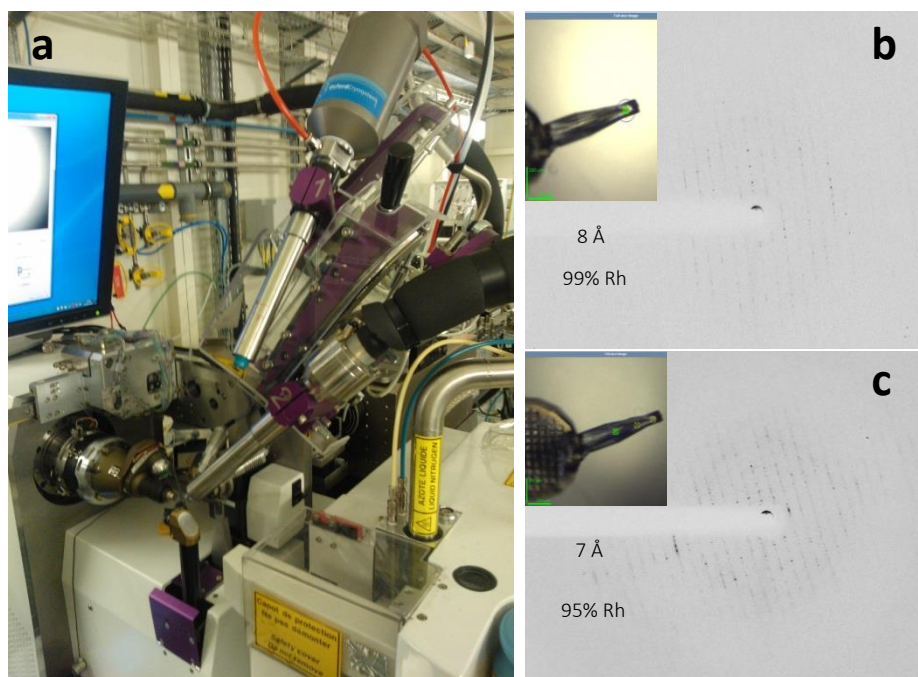
**Crystal dehydration** takes into consideration the dynamic nature of crystals. The molecules that compose the crystal lattice have a degree of flexibility. Upon controlled dehydration of the surrounding environment of the crystal, there is a loss of water molecules that causes a lattice rearrangement. While this rearrangement is unpredictable and may decrease the quality of diffraction, in some cases, a tighter packing may lead to an increase in internal order<sup>40</sup>.

Dehydrated crystals also dispense the addition of cryoprotectants. In this work, crystal dehydration was performed with two strategies: using **saturated salt solutions** or under **controlled relative humidity** conditions.

**Dehydration with salt solutions** was performed by harvesting the crystals with mesh loops mounted on GB-B3S goniometer bases, and immediately covering them with MicroRT capillaries (Mitegen), which had been previously filled with a saturated salt solution. 12 solutions from the Crystal Dehydration and Salvage kit (Jena Bioscience) were tested, and the crystals were equilibrated for 2 minutes up to 2 hours, and then cryocooled in liquid nitrogen.

**Controlled dehydration** was achieved at a synchrotron source, using a humidity control device (HC1) head, mounted at the BM14 beamline of the ESRF

(Fig. 3.7). The crystal was harvested using mounted mesh LithoLoops, and mounted on the goniometer. The relative humidity value (Rh) was first determined for the reservoir solution, and set up with the humidity controller nozzle for each crystal. The diffraction pattern of the crystal was analysed as humidity was decreased in 1-2% steps. Once the diffraction pattern attained the most satisfactory level achievable by that particular crystal, the HC1 head was quickly exchanged for the liquid nitrogen nozzle, the crystal was cryocooled and stored to collect a full dataset in a more intense beamline (ESRF ID29).



**Figure 3.7 – Controlled crystal dehydration was performed at BM14. a)** Liquid nitrogen (1) and humidity control head (2) setup at BM14. **b)** and **c)** diffraction pattern collected at the indicated level of relative humidity (Rh).

In some cases, crystal annealing was performed at the synchrotron source, by briefly blocking the flow of cold nitrogen gas to allow the crystal to thaw and re-freeze. This may lead to a reordering of the crystal lattice, which is afterwards tested by observing the resulting diffraction pattern.

Crosslinking was performed by setting up simultaneously a glutaraldehyde hanging drop in the same well of a *hsRuvBL2* sitting drop, such that by evaporation a mild crosslinking would occur. However, this led to protein aggregation in the drop. Lysine methylation was also attempted, in order to try to obtain a *hsRuvBL2* form that could crystallize in a more ordered way, but this random addition of methyl groups led to a complete aggregation of *hsRuvBL2*. This may be due to the high amount of Lysines at the surface of *hsRuvBL2*. The N-terminal residue is also a methylation target, which may have contributed to a decrease in stability.

### **Production of *hsRuvBL2* crystals diffracting to 2.8 Å**

Homogeneity of the *hsRuvBL2* sample was assessed by size-exclusion chromatography, SDS PAGE gels, and by observing the sample in micrographs obtained by negative staining electron microscopy. This allowed us to observe whether the sample formed aggregates or included a large population of open rings, and directed the optimisation of the different purification steps to produce a more homogeneous sample, constituted mainly of closed rings. The observation of EM micrographs also showed that the rings started to acquire an open conformation and to become disrupted after a few freeze/thaw cycles. Finally, the purified *hsRuvBL2* was dialysed to pH 6 immediately prior to crystallization. This pH is closer to the theoretical pI of *hsRuvBL2* of 5.66, calculated with the Crystallization pH predictor<sup>41</sup>.

Several commercial crystallization screens were tested initially (**table 3.2**). Of all the positive hits obtained, only a few were amenable to scale-up and optimisation. The crystallization drops were set up at 293 K, with a proportion of 0.7 µL of protein to 0.3 µL of reservoir solution.

The majority of the tested cryoprotectants negatively affected diffraction quality, with the exception of glycerol and ethylene glycol (25% v/v).

## ***Data collection***

### **An overview of the strategies used for crystal analysis and data collection**

In over 350 crystals that were tested, some techniques were used in order to optimize the collection of good quality data. The major problems observed with the crystals of *hsRuvBL2* throughout this work were diffraction to very low resolution (no diffraction at all for many crystallization conditions), multiplicity and anisotropy.

Initial assessments of crystal diffraction quality were performed with the **in-house X-ray diffraction system**, for crystals larger than 100  $\mu\text{m}$ . Subsequent diffraction experiments were performed at high brilliance synchrotron radiation facilities (ALBA, Soleil, Petra III, Diamond and ESRF).

Crystal screening *in situ* was performed, so as to discard the possibility that crystal removal from the drops was responsible for a decrease in diffraction quality due to a deformation of the crystal lattice.

When the crystal size permitted, a **grid analysis** of the different areas of the crystal was performed prior to collection, in order to search for discrete areas of higher diffraction power by hitting the crystal briefly with the X-ray beam in one direction. The area that produced the most promising diffraction pattern was marked for data collection, which was performed in the cases where the first beam exposure did not cause radiation damage.

Since we obtained mostly elongated prism-shaped crystals, **helical** collections were performed in cases when diffraction was of constant quality along the crystal - isotropic, so as to minimize radiation damage.

Ice coating of the loop surface was cleaned with a liquid nitrogen jet.

Diffraction data from a crystal diffracting to 2.8 Å was collected at 293.15 K at Proxima-I beamline, at the Soleil synchrotron source (St. Aubin, Paris, France), using a CCD detector (ADSC QUANTUM 315r).

### *Structure solution and refinement*

Data were indexed and integrated with XDS<sup>42</sup>, and the space group was determined with POINTLESS<sup>43</sup> and scaled with AIMLESS<sup>44</sup>, all within the autoPROC data processing pipeline. At this stage, a test set comprising about 5% of the measured reflections of the data set was flagged for cross-validation calculations ( $R_{\text{free}}$  flag).

Four datasets were collected from the same crystal, in a total of 347 images. Since each image is 1° wide, this means that a 347° “wedge” was collected in total, yielding a data set with high multiplicity in space group P6. The total amount of collected images was used in a first “brute force” approach, where the high multiplicity was favoured in detriment of good statistics, in order to obtain an initial electron density map. The four datasets were processed individually and scaled together.

The main data collection and processing statistics are listed in **Table 3.3**. Matthews coefficient calculations<sup>45</sup> indicated the presence of one molecule per asymmetric unit.

**Table 3.3 - Data collection and refinement statistics**

*Values in parenthesis refer to the highest resolution shell*

<b>Data collection</b>	
Beamline	Proxima-1 (SOLEIL – Paris – France)
Wavelength (Å)	0.9762
Space group	P 6 (168)

**Table 3.3 - Data collection and refinement statistics***Values in parenthesis refer to the highest resolution shell*

<b>Data collection</b>		
PDB entry	-	5N7R
	347°-wedge	40°-wedge
Cell dimensions		
$a, b, c$ (Å)	122.97, 122.97, 60.84	122.5, 122.5, 60.65
$\alpha, \beta, \gamma$ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	106.5 – 2.80 (2.81 – 2.80)	40.25 – 2.87 (2.88 – 2.87)
$R_{\text{merge}}$ (%)	38.6 (125.3)	10.5 (47.4)
$R_{\text{p.i.m.}}$ (%)	9.2 (44.1)	7.5 (34.0)
$\langle I/\sigma(I) \rangle$	14.4 (2.6)	10.2 (2.7)
Completeness (%)	100 (100)	93.2 (93.8)
Redundancy	16.4 (9.0)	2.7 (2.7)
$CC_{1/2}$ (%)	96.0 (74.3)	98.8 (88.5)
Wilson $B$ factor (Å <sup>2</sup> )	53.9	61.4
<b>Refinement</b>		
Resolution (Å)		40.25-2.87 (2.88 – 2.87)
No. Reflections		11413
$R_{\text{work}}/R_{\text{free}}$ (%)		17.6 / 25.6
No. Atoms		
Protein		2946
Water		59
$B$ factors (Å <sup>2</sup> )		
Average		53.1
Protein		53.3
Water		42.4
r.m.s. deviations from ideal values		
Bond lengths (Å)		0.01
Bond angles (°)		1.13
Ramachandran plot		
Favoured		96

**Table 3.3 - Data collection and refinement statistics***Values in parenthesis refer to the highest resolution shell*

<b>Data collection</b>	
Outliers	0.3
Rotamer outliers (%)	1.6
Clashscore	0.0
MolProbity score	0.92

The 3D structure of *hsRuvBL2* was solved by the molecular replacement method using Phaser<sup>46</sup>, with *hsRuvBL1* domains I and III (cropped from the PDB ID 2C9O) as the phasing model. Since domain II corresponds to a very large portion of the total protein chain, the phasing power of the ATPase core was not enough to immediately provide electron density for that part of the chain, except for an alpha helix and part of the chain leading from the core to the mobile domain. For the iterative **refinement** of the electron density map, electron density was initially obtained with Buster<sup>47</sup>, and **automated model building** performed with Buccaneer<sup>48,49</sup>, in iterative cycles. The chain of domain II was built from the slowly emerging electron density, until the phases could no longer be improved with these programs. For the final refinement round with Buster, the input structure factor data were obtained from a 40° wedge of diffraction images, chosen for their better integration statistics, thus ensuring a higher accuracy of the final model. **Manual model building/fitting** was done in Coot<sup>50</sup> against  $2|F_o| - |F_c|$  and  $|F_o| - |F_c|$  electron density maps. Final molecule **validation** was performed with MolProbity<sup>51</sup>.

### Electrostatic surface calculations

The software CHARMM was used to calculate the surface charges distribution in *hsRuvBL1* and *hsRuvBL2*. For this, .pdb files of the hexamers were



generated, minus the water molecules and with added hydrogens. These files were input into the PDB2PQR software to generate .pqr files, which contain the charges and radii of the atoms. Topological visualization of the electrostatic potential was produced in Pymol with the plug-in APBS<sup>52</sup>.

### 3.2.2 RESULTS AND DISCUSSION

#### *Purification of hsRuvBL2*

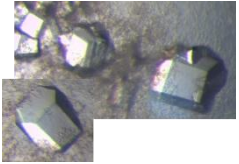

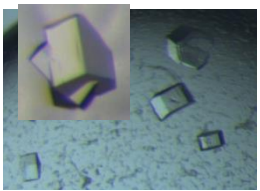
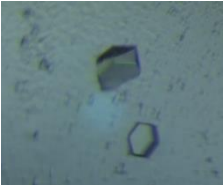
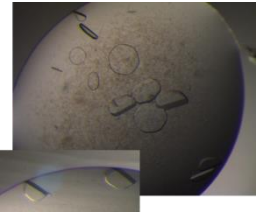
*hsRuvBL2* was first applied to a HisTrap and eluted with a 45% step of elution buffer (equivalent to about 500 mM imidazole) (**SI Fig. 1**). Immediately afterwards, the sample was exchanged into a buffer devoid of imidazole (**SI Fig. 2**), and incubated overnight at 4°C with HRV 3C protease, in the presence of 1 mM TCEP. The mix of cleaved His<sub>6</sub> tag, untagged *hsRuvBL2*, *hsRuvBL2*-His<sub>6</sub> and protease was applied to a HisTrap column in tandem with a GSTrap. The collected flowthrough contained only untagged *hsRuvBL2*, and the chromatogram (**SI Fig. 3**) indicates a high yield of cleaved sample. Untagged *RuvBL2* was injected into a Resource Q column and eluted with a gradient of elution buffer containing 1M NaCl (**SI Fig. 4**). Four peaks were separately collected, with conductivity levels 15.2, 20.2, 27.7 and 36 mS/cm, respectively. All four peaks were applied to the subsequent size-exclusion column, and all samples corresponded to hexameric *hsRuvBL2* (as calculated from a standard curve). **SI Fig. 5** corresponds to the elution curve of the sample collected in **SI Fig. 4**, eluted at 27.7 mS/cm conductivity, which produced the crystals that provided the *hsRuvBL2* structure. The analytical SDS-PAGE gel with samples from each purification step is depicted in **SI Fig. 6**.

### *Crystallization of hsRuvBL2*

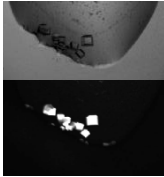
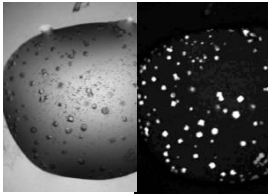
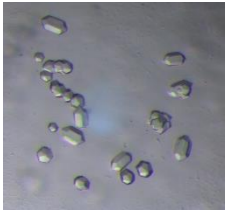

All constructs of *hsRuvBL2* produced crystals. The selenomethionine-substituted sample, initially expressed for the purpose of structure solution by the MAD (multi-wavelength anomalous dispersion) method, was actually more stable in solution, and allowed the purification of the N-terminally tagged construct. However, the native C-terminally tagged *hsRuvBL2* was already quite stable, and SeMet substitution rendered that construct so stable that it precluded the formation of crystals. **Table 3.4** describes the *hsRuvBL2* crystal forms obtained, and their respective conditions. All crystalline forms were extensively optimized, and a few of them subjected to post-crystallization treatments in order to increase resolution. Crystals that were deemed too small for diffraction analysis are not included in the table.

Of all the positive hits obtained, only a few were amenable to scale-up and optimisation. Condition B8 of the JCSG plus screen (0.2 M MgCl<sub>2</sub>, 0.1 M Tris pH 7, 10% PEG 8000), which produced small thin needles, provided the most promising results (**Fig. 3.8**). This hit was reproduced on a larger scale using sitting drop vapour diffusion on a 24-well Linbro plate, with 500 µL reservoir solution and 1 µL drop size. Needle size was increased by varying the crystallization condition components and concentrations, to obtain a final solution composed of 2-3% PEG 3350 and 250 mM MgCl<sub>2</sub>. The crystallization drops were set up at 293 K, with a proportion of 0.7 µL of protein to 0.3 µL of mother liquor. The cryoprotectant solution used consisted of the mother liquor supplemented with 1% PEG 3350, 25% glycerol and 1 mM ADP.

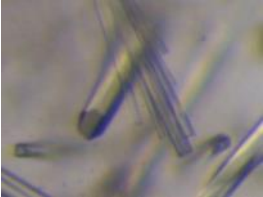
**Table 3.4 – The most representative crystals obtained with the different *hsRuvBL2* constructs.**  
Small non-optimisable crystallization hits are not represented. Untagged *hsRuvBL2* was expressed with tag on the C-terminus.

Crystal habit	<i>hsRuvBL2</i> construct	Crystallization conditions	Typical crystal size ( $\mu\text{m}$ )	Diffraction limit ( $\text{\AA}$ )
	N-ter tag, SeMet	50 mM cacodylate pH 6.5, 18 mM $\text{CaCl}_2$ , 2.5 mM spermine, 1.7% isopropanol 20°C	110x100	10
	N-ter tag, SeMet	50 mM cacodylate pH 7, 2.25 mM spermine, 36 mM $\text{MgCl}_2$ 20°C	Small	27
	N-ter tag, SeMet	20 mM $\text{MgCl}_2$ , 0.1 M MES pH 6, 3% isopropanol 20°C	Large	15 (multiple)
	N-ter tag, SeMet	50 mM cacodylate pH 6, 20 mM $\text{MgCl}_2$ , 1 mM spermine, 6% ethanol, 25% MPD 20°C	200x100	Very low diffraction
	N-ter tag, SeMet	100 mM Bis-Tris pH 5.5, 3.2 M NaCl 30°C	200x20	6.5 (multiple, anisotropic)

**Table 3.4 – The most representative crystals obtained with the different *hsRuvBL2* constructs.**  
Small non-optimisable crystallization hits are not represented. Untagged *hsRuvBL2* was expressed with tag on the C-terminus.

Crystal habit	<i>hsRuvBL2</i> construct	Crystallization conditions	Typical crystal size ( $\mu\text{m}$ )	Diffraction limit ( $\text{\AA}$ )
	N-ter tag, SeMet	5 mM MgCl <sub>2</sub> , 2 mM CaCl <sub>2</sub> , 50 mM Cacodylate pH 6, 15% 2-propanol, 1 mM spermine 20°C	50x50	No diffraction
	N-ter tag, SeMet	50 mM Cacodylate pH 7, 1 mM spermidine, 5 mM MgCl <sub>2</sub> , 10% tert-butanol 20°C	small	---
	Untagged, native	150 mM Mg Acetate, 100 mM Ammonium Acetate, 5% PEG 4000 20°C	60x20	No diffraction
	Untagged, SeMet	10% PEG 4000, 150 mM KCl 20°C	Plates ca. 5 thick	No diffraction

**Table 3.4 – The most representative crystals obtained with the different *hsRuvBL2* constructs.** Small non-optimisable crystallization hits are not represented. Untagged *hsRuvBL2* was expressed with tag on the C-terminus.

Crystal habit	<i>hsRuvBL2</i> construct	Crystallization conditions	Typical crystal size ( $\mu\text{m}$ )	Diffraction limit ( $\text{\AA}$ )
	Untagged, native	2% PEG 3350, 250 mM $\text{MgCl}_2$ , 30°C	200x30	2.8 (frequently anisotropic and multiple)

### *Initial attempts to solve the structure from low resolution data*

In previous structural studies with *hsRuvBL2*<sup>12,22</sup>, domain II was always truncated to increase the probability of crystallization. The full-length structure of RuvBL2 from a thermophilic fungus (*Chaetomium thermophilum*) was resolved only very recently, as part of the *ctRuvBL1/ctRuvBL2* complex<sup>7,20</sup>, illustrating well the difficulties inherent to obtaining diffracting crystals of this protein, mainly due to the high mobility of domain II. Despite these difficulties, we managed the first successful attempt at the crystallization of full-length *hsRuvBL2*, which produced crystals diffracting to 3.4  $\text{\AA}$  resolution.

The dataset was used to determine a preliminary structure of *hsRuvBL2*, in spite of the diffraction pattern being anisotropic and showing 20% twinning, as determined from the H-test<sup>53</sup>. Structure determination was performed by Molecular Replacement with PHASER<sup>46</sup>, as part of the CCP4 suite<sup>54</sup>, using *hsRuvBL1* (PDB ID 2c9o) as search model.

The electron density maps from the initial MR calculations using the full *hsRuvBL1* monomer as a single ensemble in PHASER did not show any electron

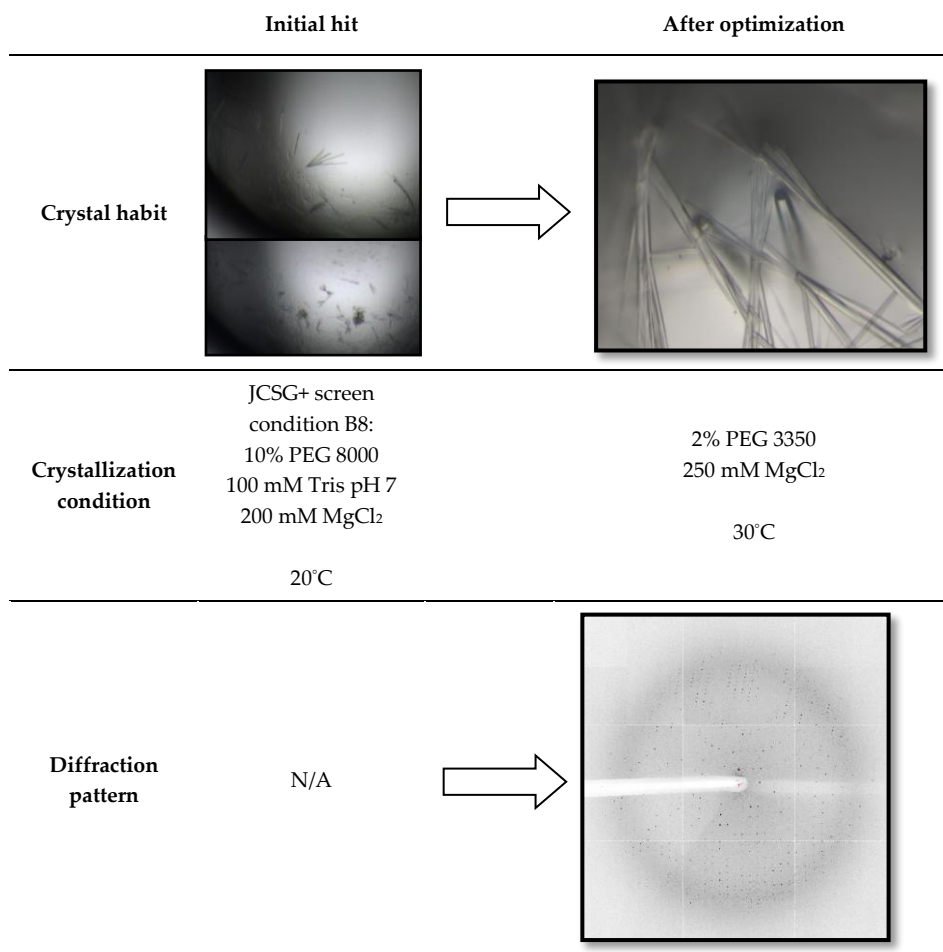


Figure 3.8 – Crystals of *hsRuvBL2*: the initial crystallization screening hit and the crystals after optimization.

density for domain II. To check the possibility that domain II might have adopted a different orientation in the *hsRuvBL2* monomer, the *hsRuvBL1* coordinates were divided into two separate ensembles: one contained domains I and III (the ATPase core), and the other comprised domain II alone. Furthermore, the connecting residues between both ensembles, Arg130 and Thr241 were removed to limit the possibility of solutions being rejected due to clashing between both ensembles.

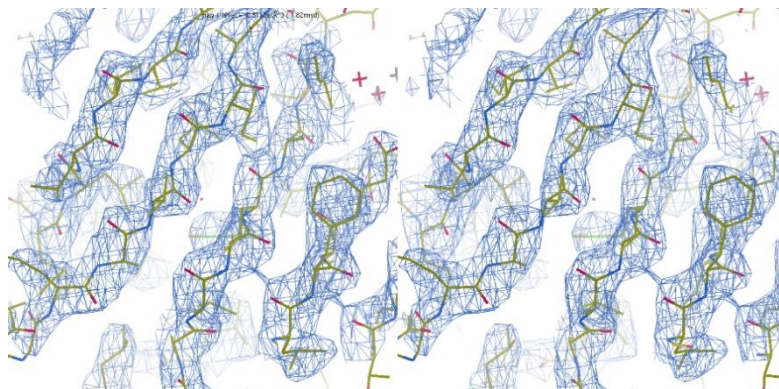
These calculations gave a clear solution for the ATPase core but failed again for the domain II ensemble. This suggested that the domain II structure in *hsRuvBL2* is sufficiently different from that in *hsRuvBL1*, precluding its use as search model in MR calculations, or may be an artefact of the low data resolution and quality, since domain II represents a much smaller fraction of the total scattering than the ATPase core. Nevertheless, the structure obtained for the ATPase core of *hsRuvBL2* provided insights into the crystal packing of full-length *hsRuvBL2*. The molecules associate to form hexamers, which stack along the short crystal axis *c* and form layers parallel to the *ab* crystal plane. Contacts between hexamers in the same layer appear to be weakly mediated by domains II. Combining these observations with the previously mentioned simulations by Petukhov and colleagues<sup>22</sup>, which show a very high mobility of domain II in *hsRuvBL2*, it is reasonable to admit that a better stacking of the domains II would be necessary to decrease anisotropy, increase resolution, and improve electron density map quality of this protein domain.

In order to improve the stacking of molecules, two approaches seemed to slightly decrease anisotropy and enhance reflection definition, if not resolution: co-crystallization with myo-inositol and dehydration. A small co-crystal of *hsRuvBL2* with myo-inositol was subjected to controlled dehydration from 98% to 90% relative humidity, using the HC-1 device at ESRF BM-14 (Grenoble, France), followed by a rapid exchange to the nitrogen gas nozzle (Oxford Cryosystems) for flash cooling of the dehydrated crystal. Thus treated, the crystal diffracted to 4.4 Å and its diffraction pattern could be indexed in space group P3. Upon analysis of that dataset, a blob of positive electron density could be seen in the space between the presumed locations of domains II in neighbouring hexamers which seemed able to accommodate one molecule of myo-inositol. This suggests the possibility that more ordered crystals may be obtained by promoting interactions that stabilise the highly mobile domain II. Also noteworthy was the absence of electron density in the nucleotide binding pocket, despite the addition of ADP prior to

crystallization. Although ADP was shown to increase *hsRuvBL2* melting temperature ( $\Delta T_m = + 6^\circ\text{C}$ ), and to be essential during protein purification, we cannot exclude that although ADP binds at the active site, it may be somehow released during crystallization, unlike what we observed for *hsRuvBL1*.

### *Data collection from crystals diffracting to 2.8 Å*

Prior to data collection, all crystals were transferred to a cryoprotectant solution composed of the mother liquor supplemented with 25% glycerol. Most crystals diffracted only to low resolution (around 4 Å) and were mostly anisotropic, even in *in situ* measurements at room temperature. Diffraction data from the best diffracting *hsRuvBL2* crystal were collected to 2.8 Å resolution. The initial analysis of the reciprocal lattice from two perpendicular diffraction images (performed with EDNA software<sup>55</sup>) suggested a space group, characterized the unit cell and provided a strategy for subsequent data collection (including optimal beam flux, so as to minimize radiation damage while maximizing the intensity of reflections, and thus amount of acquired data).



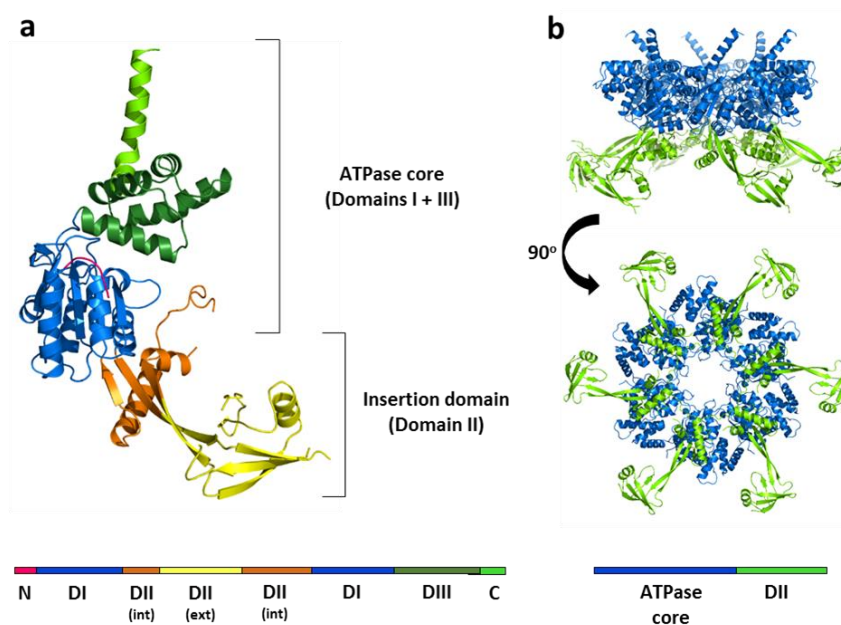
**Figure 3.9 – Stereo image of the electron density map of *hsRuvBL2*.** Image created with Coot<sup>50</sup>.



**Fig. 3.9** depicts a stereo view of the electron density calculated for the *hsRuvBL2* structure, included for quality assessment.

### *Atomic structure of full-length human RuvBL2*

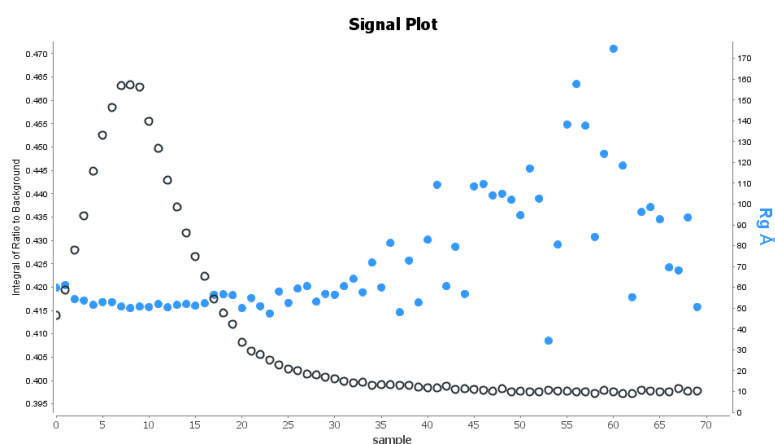
We solved the crystal structure of the human RuvB-Like 2 (*hsRuvBL2*) in the apo form, at 2.8 Å (**Table 3.4 and Fig. 3.10**). The structure shows a hexameric arrangement of monomers similar to other structures in the RuvB-Like family<sup>4,7,12,20</sup>.



**Figure 3.10** – Cartoon representations of the overall structure of *hsRuvBL2* monomer and hexamer. (a) The full-length apo *hsRuvBL2* monomer. Domains I, II and III are coloured blue, orange/yellow and green, respectively. The visible part of the N-terminal loop and the C-terminal helix are also identified in pink and light green, respectively. A linear schematic representation of the domains of RuvBL2, using the same colour, is shown below, highlighting the internal and external portions of domain II. (b) Side and bottom views of the RuvBL2 hexamer, highlighting the AAA+ core and domain II in blue and green, respectively.

The distribution of  $\alpha$ -helices and  $\beta$ -sheets is represented as a cartoon, obtained with

Pro-origami<sup>56</sup>, in **SI Fig. 7**. The hexameric arrangement of *hsRuvBL2* in solution was confirmed by SAXS coupled to size exclusion chromatography, which unequivocally showed the formation of a complex with a mean radius of gyration of *ca.* 52 Å throughout the elution peak (**Fig. 3.11**). The ring-shaped *hsRuvBL2* hexamer comprises an ATPase core that includes domains I and III, as well as the internal region of domain II from each protomer, and six outward-facing mobile units that comprise the external region of domains II. Protruding from the ATPase core are the six antennae-like  $\alpha$ -helices from the C-terminus. The *hsRuvBL2* hexamer has a central channel 24 Å wide on its narrowest part and although a double-stranded B-DNA molecule could be tightly fitted (not shown; fitting done with PDB entry molecule 1BNA, not considering deviations from B-form caused by binding), it has been clearly demonstrated biochemically that *hsRuvBL2* can only bind single-stranded DNA<sup>27</sup>.



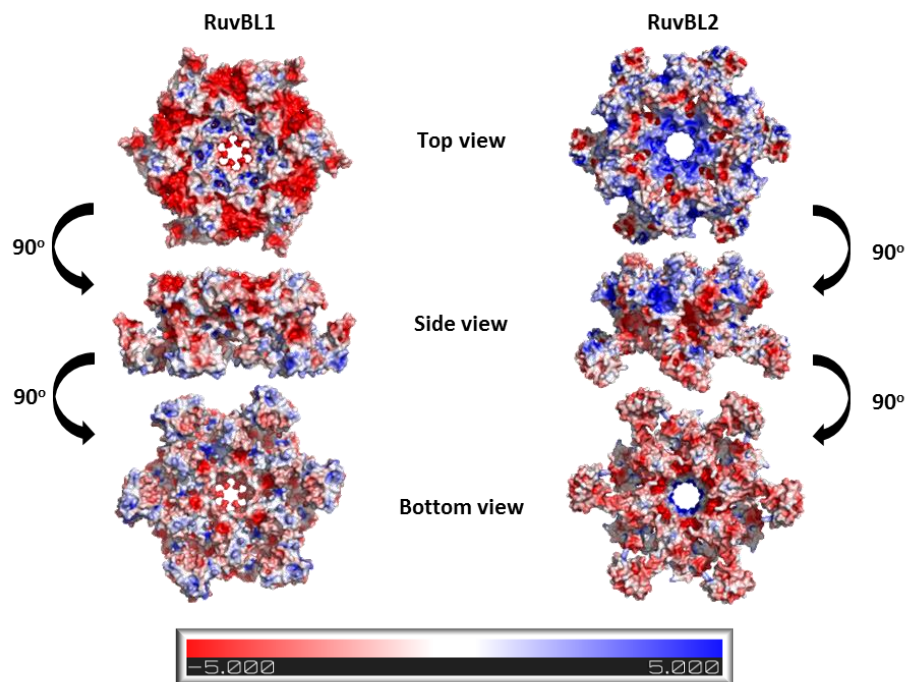
**Figure 3.11** – Small angle X-ray scattering coupled to size exclusion chromatography shows that *hsRuvBL2* has a constant radius of gyration (Rg) throughout the elution peak that corresponds roughly to a hexamer. UV absorbance in open circles, Rg in closed blue circles.

To the best of our knowledge, similar studies have not been made for RuvBL1. However, the smaller size of the inner channel of RuvBL1 (20 Å) and the

marked differences in the surface electrostatic charge distribution (**Fig. 3.12** and ref.<sup>21</sup> for comparison with the truncated dodecameric complex) suggest the possibility of a different mechanism. Since their activities have been observed to increase when working together as a heteromeric complex, it is also possible that their individual characteristics complement each other, with a consequent increase in DNA processing efficiency. The *hsRuvBL2* hexamer is 149 Å wide (compared to 147 Å in *hsRuvBL1*, which crystallized in a slightly more compact conformation, with the domains DII folded backwards into the ATPase core – see **figure 3.2**) and the ATPase core is 51 Å high (compared to 50 Å in *RuvBL1*<sup>4</sup> and 51 Å in the truncated human dodecameric complex<sup>21</sup>). The *ctRuvBL1*:*RuvBL2* dodecamer from *Chaetomium thermophilum* has a central channel with similar dimensions (25 Å) and is 118 Å wide<sup>79</sup>. The latter is a consequence of the compact conformation of *ctRuvBL2* in this complex. The interface between protomers in the *hsRuvBL2* hexamer, analysed with LigPlot<sup>57</sup> occurs mainly through hydrophobic interactions, but also through covalent bonds formed between residues: Asp352-Arg400, Arg273-Ser262, Ser310-Phe109/Ser106, Arg124-Phe261, Tyr340-Thr333, Arg314-Glu112, Gln78-Met443, Asn329-Tyr442, Leu354-Gln404, Ile356-Leu434, His344-Ser439 and Ser339-Arg334.

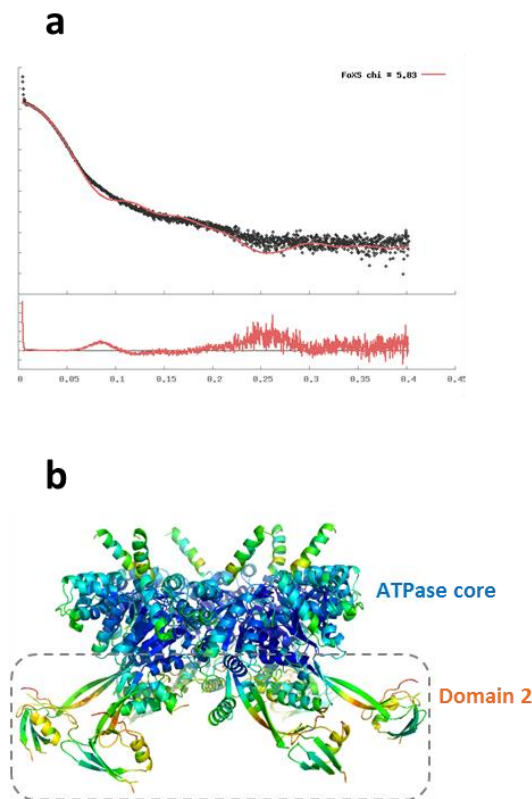
*hsRuvBL1* and *hsRuvBL2* are homologous proteins with 43% sequence identity and 65% sequence similarity that work together as part of large chromatin remodelling complexes. However, they also work antagonistically in many situations. It has been suggested that the antagonistic activities of *RuvBL1* and *RuvBL2* are the consequence of interactions with different, more or less specific partners<sup>58</sup>. To assess the structural basis for these interaction specificities, we analysed the structures of *hsRuvBL1* and *hsRuvBL2* for differing characteristics at the surface of both complexes. A comparison of the surface charge distribution of *hsRuvBL1* and *hsRuvBL2* hexamers (**Fig. 3.12**) shows marked differences in electrostatic potential distribution, mainly on the inner surface of the central

channel and on domain II, which suggests different mechanisms/affinities for binding the DNA strand and other proteins. *hsRuvBL1* and *hsRuvBL2* have different oligomerization dynamics, since hexamer formation occurs at a much higher protein concentration for *hsRuvBL1* than for *hsRuvBL2* (ref.<sup>21</sup> and this work). The sum of these observations suggests considerable differences in specificity that go beyond overall shape dissimilarity.



**Figure 3.12** – Surface charge distribution in the *hsRuvBL1* and *hsRuvBL2* hexamers. Differences in the distribution of amino acid residues at the surface of the rings may underlie their antagonistic activities. Colour by electrostatic potential on solvent accessible surface. Scale bar at the bottom, in kT/e.

Molecular dynamics studies highlighted a propensity of the external region of domain II in RuvBLs to acquire a variety of conformations in solution<sup>22</sup>. Comparison of the Guinier plot from an empirical SAXS profile of *hsRuvBL2* (**Fig. 3.13**, black) with a theoretical plot calculated from the crystallographic coordinates

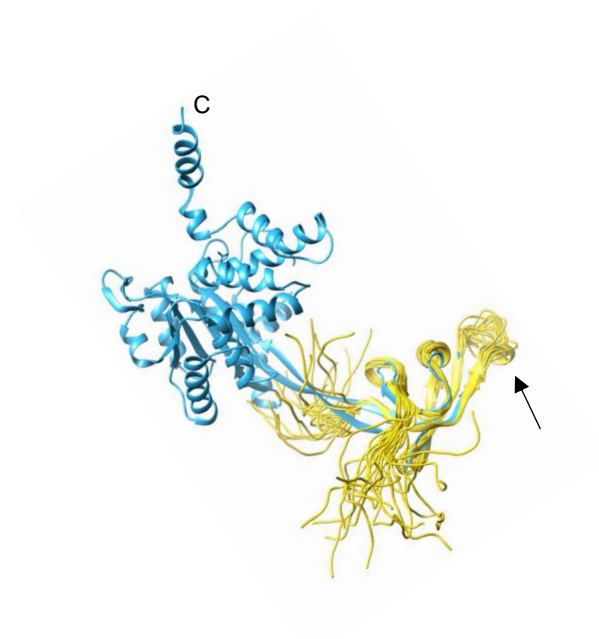


**Figure 3.13 – Mobility of the domain II of *hsRuvBL2*.** **a.** Guinier plot obtained empirically (black), superimposed with a theoretical plot calculated from the crystallographic coordinates (red). **b.** Side view of the *hsRuvBL2* hexamer. Atoms are coloured according to B-factor: Blue – lowest; orange - highest.

(red) shows that, in solution, the *hsRuvBL2*  $d_{max}$  is similar to that observed in the crystallographic structure: 140 – 150 Å in solution versus 149 Å measured between the extended tips of domain II. Comparison between the two profiles also shows an overall similar shape. However, while the crystallographic structure translates into a Guinier plot clearly divided into two well defined regions (which can be interpreted as the ATPase core and the domain II extension), the sample in solution produced an average curve consistent with a more homogeneous structure. Since  $d_{max}$  is similar, these results support a high mobility of the protruding domains in

the hexamer, as proposed by Petukhov and colleagues<sup>24</sup>, lending an appearance of homogeneity of structure to the plot obtained empirically.

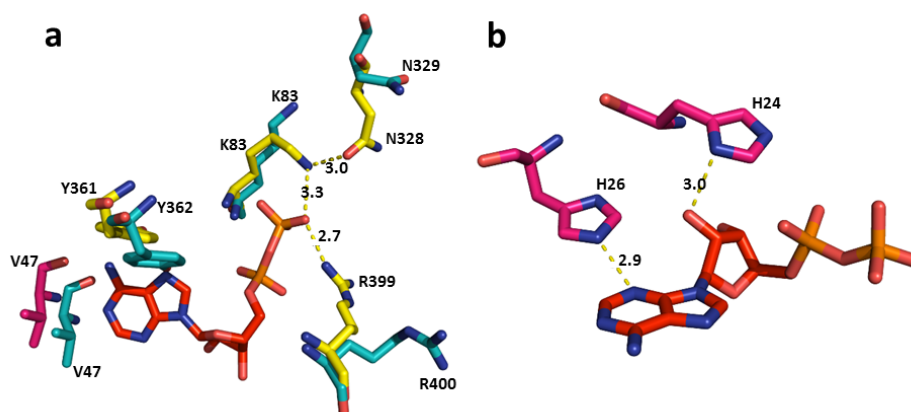
In addition to the overall mobility of domain II, there are three loop fragments that are too mobile to produce an electron density: D148 - S156, I202 - Q226 and R253 - Q255. The first two, larger, segments are naturally located on the external part of the domain. Using an available structure of domain II from *hsRuvBL2* obtained by NMR (PDB ID 2cqa), the structural information on this domain could be complemented by superimposing the two (Fig. 3.14). It becomes apparent that the lack of electronic density is a result of an intrinsic high loop mobility that is not restricted by intermolecular contacts in the crystal structure.



**Figure 3.14 – Superimposition of the crystallographic structure of *hsRuvBL2* (blue, PDB ID 5N7R) with the NMR model ensemble obtained for the external part of domain II (yellow, PDB ID 2CQA). The loop on the tip of domain II (black arrow) is not ordered in the crystal structure, but can be observed as highly mobile in the NMR structure. Structure alignment performed with Chimera<sup>59</sup>.**

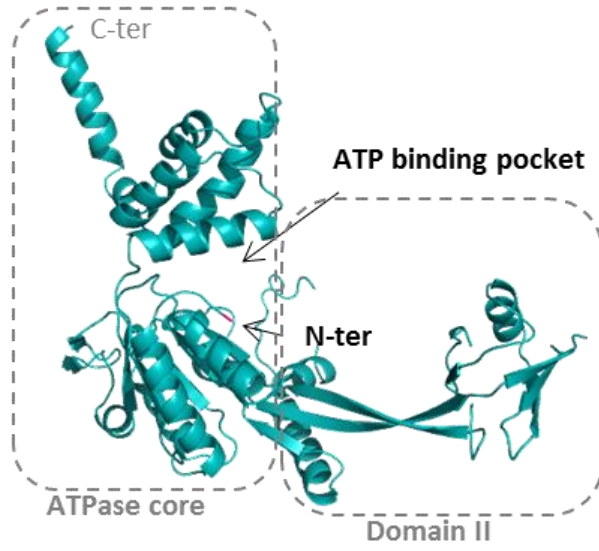
### Structural basis for coupling ATP binding to mechanical action

It is a sensible assumption that the basis of ATP-dependent activities in RuvB-Like proteins lies upon the mechanical consequences of ATP binding and hydrolysis. In order to elucidate what these consequences might be in the case of *hsRuvBL2*, we compared the structure of human apo *hsRuvBL2* (this work) with the ADP-bound form of RuvBL2 from *C. thermophilum*<sup>9</sup> (*ctRuvBL2*, PDB ID 4WW4), which has 68% identity and 85% similarity with *hsRuvBL2*, as well as a conserved disposition of the key residues that constitute the ATP binding pocket (Fig. 3.15). In *ctRuvBL2*, nucleotide binding leads to a tightening of the binding pocket near the phosphate tail. The  $\beta$ -phosphate of the ADP molecule interacts with neighbouring residues R399, K83 and N328 of *ctRuvBL2* (which correspond

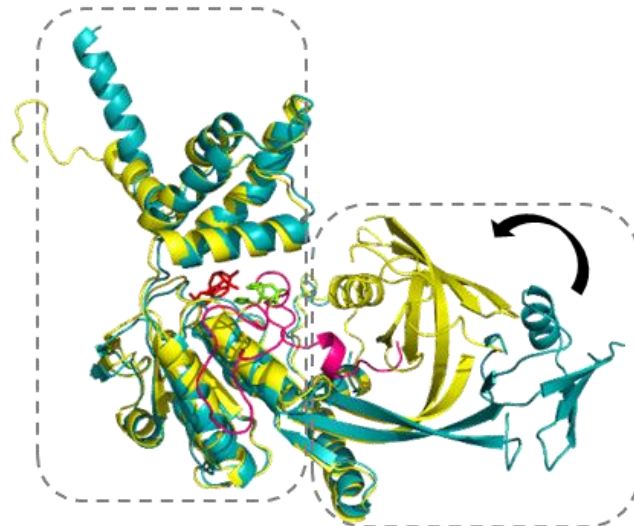


**Figure 3.15 – The nucleotide binding pocket and the N-terminus Histidines involved in the motion of the N-terminal loop towards the nucleotide. (a)** Superimposition of the nucleotide binding pocket of *C. thermophilum* (yellow, 4WW4) with the human homologous residues (blue, 5N7R). The ADP (depicted in red) is from the fungal structure. Atoms are coloured as follows: N – dark blue; O – salmon; P – dark yellow. V47 from *C. thermophilum* is depicted in pink since it is part of the N-terminal loop. **(b)** Position of Histidines 24 and 26 in relation to ADP in the binding pocket of *ctRuvBL2*. The interatomic distances shown are in Å.

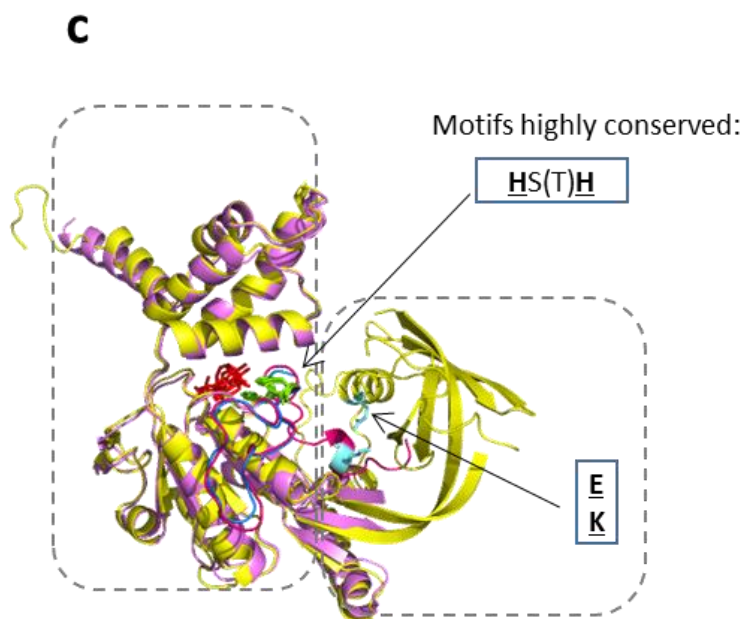
**a**



**b**







**Figure 3.16 – Structural comparison between existing structures of RuvBL2 (human RuvBL2, human RuvBL2 $\Delta$ DII and *C. thermophilum* RuvBL2) suggests a link between nucleotide binding and domain II motion. (a) The apo *hsRuvBL2* (blue, this work, PDB ID 5N7R) is associated with a stretched, or at least loose conformation of the domain II. (b) The apo *hsRuvBL2* is superimposed with the ADP-bound *ctRuvBL2* from *C. thermophilum* (yellow, PDB ID 4WW4). When compared with the apo *hsRuvBL2*, the ADP-bound *ctRuvBL2* displays a positioning of domain II proximal to the ATPase core and a more ordered N-terminal loop (pink). ADP is depicted in red sticks. (c) The ATP-bound, domain II-truncated *hsRuvBL2* (light pink, PDB ID 2XSZ) is superimposed with the full-length *ctRuvBL2* (4WW4). It becomes apparent that the absence of the external part of domain II has an influence in the organization of the N-terminal loop. In both nucleotide-bound forms, the N-terminal loop interacts with the nucleotide through two conserved histidines; however, in the truncated *hsRuvBL2*, the N-terminal loop (blue) remains disordered from the first residue up to the place of interaction with the nucleotide. The histidines that interact with the nucleotide are depicted in green and the motif is highlighted; residues involved in electrostatic interactions between the N-terminal loop and domain II are depicted in light blue and highlighted as well.**

to R400, K83 and N329 of *hsRuvBL2*) (**Fig. 3.16a**). Concomitantly, nucleotide entrance into the binding pocket causes a displacement of Y361 and V47 (Y362 and V47 in the human protein) due to hydrophobic interactions with the adenine ring, driving the initial rearrangement of the N-terminal loop, into a more ordered position in close proximity with the nucleotide binding pocket. We propose that these interactions initiate the coupling between nucleotide binding and mechanical movement of domain II, depicted in **Fig. 3.16**. This figure shows that, while apo *hsRuvBL2* has no electron density up to residue 47 (**Fig. 3.16a**, shown in pink), ADP-bound *ctRuvBL2* could be modelled for most of the N-terminal loop (**Fig. 3.16b**), supporting a connection between ADP binding and N-terminus reorganization, as previously suggested<sup>7</sup>. Furthermore, in ADP-bound *ctRuvBL2*, domain II is more tightly packed against the ATPase core, a conformation in which this domain interacts closely with the proximal part of the N-terminus. The apo-

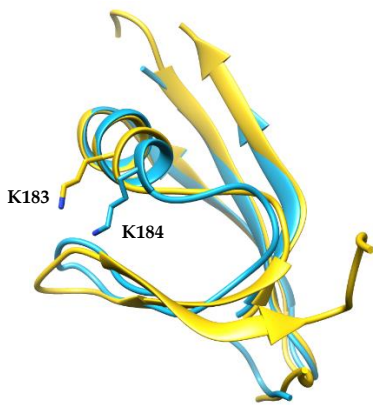


**Figure 3.17 – Model of domain II-truncated *hsRuvBL2* monomer (PDB ID 3uk6)<sup>24</sup>.** Of the twelve molecules in the asymmetric unit, most have the modelled N-terminal loop (pink). Only one does not have an ordered N-terminal loop, and two have a longer loop before the histidines, stabilised by interactions with the internal portion of domain II and the adjacent protomer. ATP depicted in red sticks (arrow).

*hsRuvBL2* structure was further compared with the ATP-bound form in the *hsRuvBL1* $\Delta$ DII:*RuvBL2* $\Delta$ DII complex<sup>12</sup> (**Fig. 3.16c**, PDB ID 2XSZ), showing that in the absence of an interaction between the N-terminus proximal segment and the

external region of domain II, the N-terminal loop is stabilized only from the point where it interacts with the bound nucleotide, through Histidines 40 and 42 (H24 and H26 in *ctRuvBL2*, not visible in apo *hsRuvBL2*), and the proximal part remains unstructured. This was also observed in the structure of domain II-truncated *hsRuvBL2* bound to ADP<sup>24</sup> (Fig. 3.17). These observations support the proposal of the following sequential events: in ATP-dependent activities, the nucleotide first binds to the nucleotide-binding pocket, eliciting a rearrangement of the surrounding residues, which causes the most distal part of the N-terminus to move.

The predicted movements that would occur in a RuvBL2 monomer, during its transition from the apo to the nucleotide-bound form include the movement of the N-terminal loop as it is tethered by interactions with the nucleotide through histidines 24 and 26 in *ctRuvBL2* (Fig. 3.16b – the structured N-terminus from *ctRuvBL2* is depicted in pink, with the tethered histidines in green). This initial interaction then draws the remaining part of the N-terminus into the close vicinity of the external region of domain II. Here, the conserved residues Glu11, Lys13 and Glu14 from the N-terminus, and Lys183 and Ile201 from domain II (Glu12, Arg14, Asp15 and Lys184 in *hsRuvBL2*) are responsible for the electrostatic interactions that sustain the positioning of domain II in close proximity to the ATPase core. Thus, our apo-*hsRuvBL2* structure provides the missing link that supports the mechanism proposal whereby the N-terminal loop provides an interface between the external region of domain II and the ATPase core. Interestingly, key residues involved in the proposed mechanism are conserved between the human and fungal forms (Fig. 3.18). Recent work by Zhou and colleagues (2017) with the yeast Rvb1/Rvb2 complex and an Ino80 insert (Ino80INS) supports this connection between nucleotide binding and movement of domain II. In their work, they observe that a dodecamer previously assembled due to Ino80INS binding to domain II, can be **collapsed into hexamers by ATP binding**, concomitantly with a loss of binding to Ino80INS. The collapse of dodecamers into hexamers is correlated



**Figure 3.18 – Alignments of RuvBL2 from *Homo sapiens* and *Chaetomium thermophilum*.** **Top:** alignment of protein sequences. The highlighted residues are involved in the reorganisation of the N-terminal loop. Secondary structure elements (SSE) of *hsRuvBL2* are depicted on top of the alignment. Alignment made with Clustal Omega<sup>60</sup> and SSE depiction obtained with ESPript 3.0<sup>61</sup>. **Bottom:** Structure alignment of the OB-folds from the two structures (human 5N7R, cyan and *C. thermophilum* 4WW4, yellow) shows that the residues involved in the stabilisation of domain II have the same spatial distribution. The lysine residues depicted as sticks have a role in the proposed mechanism, and are conserved between the two organisms (as indicated by a triangle in the top panel). Structure alignment obtained with Chimera<sup>59</sup>.

to the changes in dodecamer conformation observed by Ewens and colleagues (2016) upon nucleotide addition<sup>62</sup>, which causes a significant reconfiguration of domains DII. These changes in configuration may also be responsible for the loss of affinity to Ino80INS.

On the other hand, nucleotide binding to RuvBL1 does not seem to elicit the same structural changes as for RuvBL2. In fact, despite the similar initial interaction of the two conserved Histidines with the nucleotide, both in human and in *C. thermophilum* RuvBL1 the N-terminal loop is directed towards interactions with domain 1: Arg15 interacts with Pro96 from *ctRuvBL1*, and Lys11 with Pro95 from *hsRuvBL1*. Most of the conserved residues highlighted in the proposed mechanism are also conserved in RuvBL1, which suggests that additional residues may have a significant influence in the outcome of nucleotide binding. Possible differences between action mechanisms of RuvBL1 and RuvBL2 may also be connected to the observed differences in stability in consequence to tag placement in the N-terminal loop: while *hsRuvBL1* could be expressed and purified with no signs of instability, *hsRuvBL2* had a tendency to aggregate at higher concentrations. Curiously, the opposite effect was observed upon tag placement in the C-terminal helix: while *hsRuvBL2* retains some stability, *hsRuvBL1* formed inclusion bodies during expression tests. The structural reasons behind these different consequences

of tag placement on the C-termini may be related to the fact that in *hsRuvBL2*, this helix is longer than that of *hsRuvBL1*, while in the latter, the helix is shorter and is wholly involved in contacts with the adjacent monomer in the homohexamer.

### 3.3 Structure of *hsRuvBL2* by electron microscopy

**Single-particle electron microscopy** (EM) is a technique used to determine the structure of individual molecules about 250-300 KDa in size, using information gathered from electrons transmitted through the dispersed molecules in the sample (Transmission Electron Microscopy). It can be divided into negative-staining EM and cryo-EM. **Negative staining EM** provides structures of lower resolution (up to 15 Å), since contrast is provided by the use of a heavy atom salt, but it is easier to setup and process, as well as cheaper to perform, and thus is used for optimisation of sample preparation conditions. **Cryo-EM** has seen extraordinary improvements over the last few years as a technique capable of providing the atomic structures of molecules. The sample is directly cryopreserved in the sample buffer, and thus contrast is provided mainly through the phase differences caused by the interactions of the electron beam with the sample as compared to the surrounding buffer. It is much more expensive and challenging to setup and process, and thus it is only used after all conditions have been optimised using negative staining EM.

The study of the structure of *hsRuvBL2* was complemented by the use of electron microscopy in several ways. The *hsRuvBL2* complex consists of six modules with mobile domains that most likely acquire different conformations in solution<sup>24</sup>, which in principle would not be observed by X-ray crystallography. Furthermore, EM allows the identification of all types of particles present in the sample (provided their number is high enough to generate sufficient signal intensity vs noise upon averaging). This is particularly useful to compensate for

other techniques, such as AUC, for which the associated error may not allow discerning large complexes with a difference of one monomer in their composition.

### 3.3.1 METHODS

#### *Protein production*

hsRuvBL2 was obtained to a high degree of purity and homogeneity using the previously described protocol. Prior to the electron microscopy experiments, hsRuvBL2 samples were dialysed to a buffer containing Tris pH 8.0 instead of phosphate buffer, since the latter reacts with the heavy metal salt (uranyl acetate) used for negative staining, producing salt crystals. Samples were kept at 4°C, since freeze/thaw cycles promoted dissociation of oligomers, and analysed as soon as possible after purification.

#### *Sample preparation – negative staining*

Sample concentration in the grid was optimised to ca. 20 µg/mL; however, since sample concentration can be quite variable due to operator or equipment failures, concentrations were tested empirically by performing different dilutions with Tris-containing buffer. The method used in this work to perform negative staining is called the Flotation Method. The procedure involves floating the EM grid on a droplet of sample, to permit adsorption of the specimen. The grid is then blotted on filter paper, immediately transferred onto a nearby uranyl acetate drop for another 2 minutes, blotted again and it is then ready for use. Rhodium/Copper EM grids were pre-treated with a carbon coating, rendering the surface uniform. The hydrophobic surface of the grid was rendered hydrophilic through an electrical glow discharge.

Two forms of *hsRuvBL2* were studied: untagged and with tag on the N-terminus. The untagged protein was adhered to the EM grid for 2 minutes, blotted and stained with uranyl acetate for another 2 minutes.

N-terminally tagged *hsRuvBL2* forms dodecamers in the conditions used in this work. However, these are labile complexes, and easily separate into its constituent hexamers upon dilution, as demonstrated by analytical ultracentrifugation. In order to prevent dodecamer dissociation, glutaraldehyde was used in a very mild cross-linking procedure termed GraFix<sup>63</sup>, by which the sample was applied on top of a sucrose gradient, which was also a glutaraldehyde gradient. The final gradient parameters were as follows: 10 – 25% sucrose, 20 mM HEPES pH 8.0, 250 mM NaCl, 2 mM MgCl<sub>2</sub>, 1 mM ADP. The heaviest solution also contained 0.2% glutaraldehyde. Gradients were performed by applying the top (light) solution first, and then injecting the bottom (heavy) solution using a thin needle, at the bottom of an ultra-clear 11 x 60 mm tube, and dispensing slowly. Finally, both solutions were mixed to achieve a continuous gradient, with Gradient master (Biocomp). 40 µL of N-terminally tagged native *hsRuvBL2* (0.3 mg/mL, equivalent to *ca.* 180 pmol), was then gently applied on top and ultra-centrifuged at 37000 rpm for 13.5 h, in a SW60 Ti rotor (Beckman), at 4°C. Fractions (200 µL) were collected from top to bottom with a micropipette, and aliquoted on ice. No glycine was added to inactivate the glutaraldehyde, since the samples were either immediately desalted into a buffer with Tris (compatible with EM reagents and less harmful to the sample than HEPES) or later on diluted in this buffer.

Time for sample adhesion to the grid was increased for the N-terminally tagged *hsRuvBL2* sample, due to the very low concentration obtained after the GraFix procedure, to *ca.* 15 h at 4°C, inside a closed Petri dish. After this incubation period, the grid was washed and stained for 2 minutes. Unfortunately, despite the fact that some molecules resembling dodecamers could be observed, the number of particles per micrograph was too low for data processing.



### ***Data collection - negative staining EM***

Micrographs were obtained using a Tecnai F20 or a JEOL JEM-1010 TEM, with exposure time per picture of 0.2 - 0.9 s and a dose of 20 - 25 e<sup>-</sup>/Å<sup>2</sup>s, and collected using a CCD camera. Between 200 and 400 micrographs were taken for each sample, depending on the observed concentration of molecules per micrograph.

### ***Data processing and refinement of negative staining data***

In Transmission Electron Microscopy, in this particular case **single particle EM**, the objective is to produce a three-dimensional model (volume) of the macromolecule from 2-dimensional projections, by finding the relations between them. To achieve this, the collected data (projections, also called “particles”) are individually picked, and similar views are averaged to increase the signal-to-noise ratio. An initial density estimate is built by, e.g., applying the common lines method, which consists of performing the Fourier transform of the projections, finding the common lines between them, and building a 3D model in Fourier space, which is then back transformed in to real space. Different projection angles are generated and compared to the empirical projections, generating a cross correlation (similarity) coefficient. The projections with the highest cross correlation coefficients are assumed to have the same Euler angles as the empirical projections<sup>64-66</sup>. The best matching position for each particle image is determined, in an iterative refinement process, thus improving the model.

The hexameric form of *hsRuvBL2*, has a molecular weight of about 300 KDa, which lies on the lower limit of size that can be studied by cryo-EM (at the time of the experiment). This makes the identification of particles in micrographs very challenging, and it is essential that good contrast be obtained. It also renders essential an initial analysis of the molecules by negative staining, prior to the cryo-

EM studies. This has allowed for preliminary knowledge to be gained on the behaviour of the molecule, especially its shape and size on the micrograph.

Since we could observe both top and side views of the *hsRuvBL2* hexamers on the carbon coated grids, low resolution 3D reconstructions were performed. The Scipion program package (developed by the Biocomputación Unit team at CNB), was used to import the micrographs. Before particle picking, the contrast transfer function was estimated with CTFFind<sup>67</sup> and corrected using the Xmipp package. Particles were hand-picked with Xmipp<sup>68</sup>, by choosing all specimens that may correspond to individual particles, in all orientations found.

Selected particles were divided into 2D classes with c2d<sup>69</sup>. Each class is an average of a set of images with the same orientation/conformation. Similar 2D classes were used to produce an initial volume. All particles belonging to similar 2D classes were chosen to produce a 3D class, using as template the previously obtained initial volume. The 3D class was then refined to the final structure.

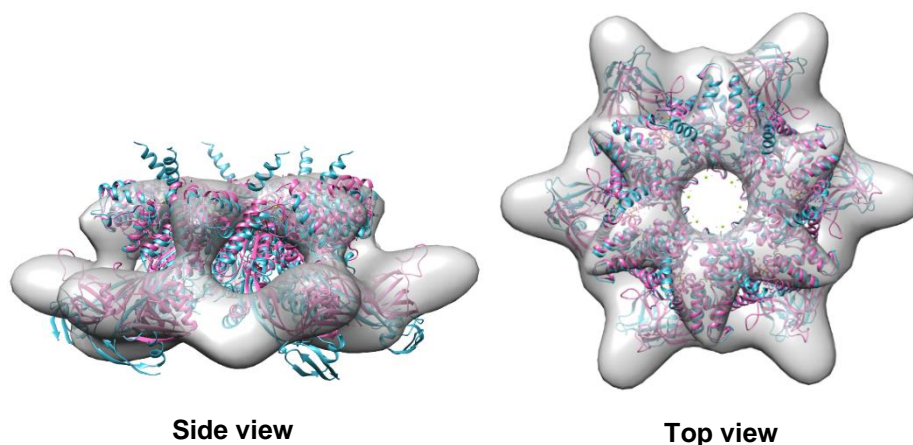
### 3.3.2 RESULTS AND DISCUSSION

The analysis of *hsRuvBL2* by negative staining EM has allowed us to greatly improve sample quality, by direct observation of the integrity of the *hsRuvBL2* rings on the micrographs, in parallel with improvements of the purification protocol. This had great impact on the quality of crystals obtained for X-ray diffraction studies.

#### *Low resolution structure of hsRuvBL2 by negative staining EM*

The fact that hexamers have high symmetry greatly contributes to the achievable quality of the final structure. A low resolution volume of *hsRuvBL2* was obtained (Fig. 3.19, grey surface), which fits nicely with the crystallographic model

of *hsRuvBL1* (pink) as expected, due to the high degree of identity between the two proteins. The fit was not as good for the crystallographic structure of *hsRuvBL2* (blue), mainly because the domains II are in a more extended conformation.

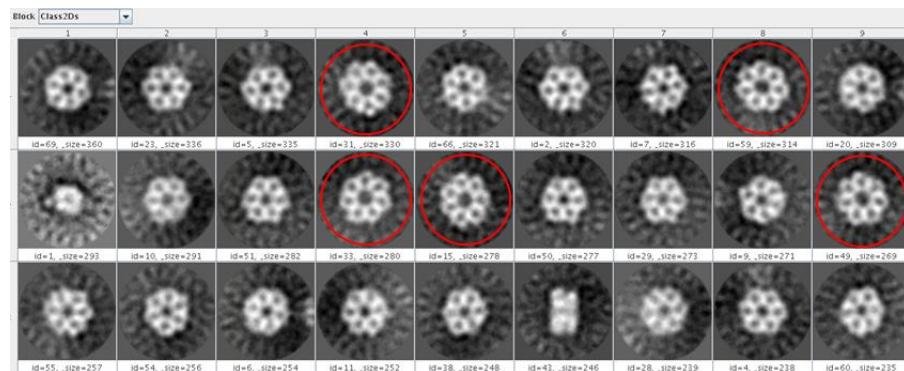


**Figure 3.19** – Low resolution envelope of the human *hsRuvBL2* hexamer, obtained by negative staining electron microscopy (gray), superimposed with the crystallographic structure of *hsRuvBL1* (pink) and *hsRuvBL2* (blue). **Left:** side view. The C-terminal helices visible in the crystallographic structure protrude from the EM envelope. **Right:** top view. The alignment was made by the more stable ATPase core, since the rmsd values between the cores of *hsRuvBL* molecules are generally lower than for the domain II. Domains II from the crystallographic structure of *hsRuvBL2* do not fit the volume, since the EM sample seems to be in a more compact conformation, more similar to that of the crystallographic structure of *hsRuvBL1*. Hence, the latter seems to be a better fit with the EM volume. Figure produced with Chimera<sup>59</sup>.

Additionally, the “antennae-like” C-terminal  $\alpha$ -helices of *hsRuvBL2* positioned on top of the ring do not appear in the EM volume of *hsRuvBL2*. The technique used (negative staining) may provide an explanation for these observations: this technique implies that the molecules are deposited on the EM grid, and afterwards a layer of a heavy atom solution (uranyl acetate) is deposited on top of the molecules. This is known to lead to deformation of the sample, and possibly flattening, which may justify the absence of an envelope portion

corresponding to these C-terminal  $\alpha$ -helices and a more horizontal disposition of the mobile domains.

Surprisingly, after averaging all particles picked, we observed that some 2D classes unequivocally belonged to heptamers (**Fig. 3.20**). This form of the protein corresponds to about 10% of all picked particles, and may be due to different reasons. It is possible that it corresponds to an error in assembly of the protein, with no biological significance. However, if this were the case, the formation of other random forms, such as pentamers or octamers would also be expected (unless precluded by steric constraints due to the arc formed by the ring).



**Figure 3.20** – *hsRuvBL2* is able to form heptamers. 10% of the total picked particles corresponded to this oligomeric form, highlighted in red.

It may also be the case that *hsRuvBL2* can, in some conditions, assemble as heptamers, and since the currently used conditions are not optimal for heptamers formation, only a small proportion was formed, due to it being putatively pre-designed for that. Refinement of observed heptamers was not possible due to the low number of particles obtained.

The *hsRuvBL2* rings acquired a very frequent open conformation about 5 hours after the samples were extensively diluted. The structure became open in a U or V shape, with the hinges seemingly between dimers or trimers (**Fig. 2.6**, see chapter 2). If the dilution of the sample was kept for extended lengths of time

(days), the rings completely disaggregated into monomers. This seems to suggest that, under certain conditions, it may be possible for *hsRuvBL2* to acquire an open conformation, a characteristic also observed in the Rho helicase<sup>70</sup>. The physiological significance of this finding is not yet understood, but it may be suggested that, upon hexamer formation, it can open up, in order, for instance, to bind DNA or release it. If a parallel is made between what is observed in this work for *hsRuvBL2*, and what is known of Rho helicase, it can be suggested that, either starting from a monomeric form, as observed by Papin and colleagues<sup>71</sup>, or from an open notch conformation, as observed for Rho<sup>72</sup>, *hsRuvBL2* re-forms a closed ring after binding to a polynucleotide chain (see chapter 2). This conformational transition would activate all the nucleotide-binding sites, which are only completed by the contribution of two adjacent monomers.

## References

1. Puri, T., Wendler, P., Sigala, B., Saibil, H. & Tsaneva, I. R. Dodecameric Structure and ATPase Activity of the Human TIP48/TIP49 Complex. *J. Mol. Biol.* **366**, 179–192 (2007).
2. Gribun, A., Cheung, K. L. Y., Huen, J., Ortega, J. & Houry, W. A. Yeast Rvb1 and Rvb2 are ATP-Dependent DNA Helicases that Form a Heterohexameric Complex. *J. Mol. Biol.* **376**, 1320–1333 (2008).
3. Gorynia, S. *et al.* Expression, purification, crystallization and preliminary X-ray analysis of the human RuvB-like protein RuvBL1. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **62**, 61–66 (2006).
4. Matias, P. M., Gorynia, S., Donner, P. & Carrondo, M. A. Crystal structure of the human AAA+ protein RuvBL1. *J Biol Chem* **281**, 38918–38929 (2006).
5. Cheung, K. L. Y., Huen, J., Houry, W. A. & Ortega, J. Comparison of the multiple oligomeric structures observed for the Rvb1 and Rvb2 proteins This paper is one of a selection of papers published in this special issue entitled 8th International Conference on AAA Proteins and has undergone the Journal's usual . *Biochem. Cell Biol.* **88**, 77–88 (2010).
6. Cheung, K. L. Y., Huen, J., Kakihara, Y., Houry, W. A. & Ortega, J. Alternative oligomeric states of the yeast Rvb1/Rvb2 complex induced by histidine tags. *J. Mol. Biol.* **404**, 478–92 (2010).
7. Silva-Martin, N. *et al.* The Combination of X-Ray Crystallography and Cryo-Electron Microscopy Provides Insight into the Overall Architecture of the Dodecameric Rvb1/Rvb2 Complex. *PLoS One* **11**, e0146457 (2016).
8. Ewens, C. A. *et al.* Architecture and Nucleotide-Dependent Conformational Changes of the Rvb1-Rvb2 AAA+ Complex Revealed by Cryoelectron Microscopy. *Structure* **24**, 657–666 (2016).
9. Lakomek, K., Stoehr, G., Tosi, A., Schmailzl, M. & Hopfner, K.-P. Structural Basis for Dodecameric Assembly States and Conformational Plasticity of the Full-Length AAA+ ATPases Rvb1/Rvb2. *Structure* **23**, 1–13 (2015).
10. Jeganathan, A. *et al.* Yeast rvb1 and rvb2 proteins oligomerize as a conformationally variable dodecamer with low frequency. *J. Mol. Biol.* **427**, 1875–86 (2015).
11. López-Perrote, A., Muñoz-Hernández, H., Gil, D. & Llorca, O. Conformational transitions regulate the exposure of a DNA-binding domain in the RuvBL1-RuvBL2 complex. *Nucleic Acids Res.* **40**, 11086–11099 (2012).
12. Gorynia, S. *et al.* Structural and functional insights into a dodecameric molecular machine - The RuvBL1/RuvBL2 complex. *J. Struct. Biol.* **176**, 279–291 (2011).
13. Niewiarowski, A. *et al.* Oligomeric assembly and interactions within the

- human RuvB-like RuvBL1 and RuvBL2 complexes. *Biochem. J.* **429**, 113–125 (2010).
14. Torreira, E. *et al.* Architecture of the pontin/reptin complex, essential in the assembly of several macromolecular complexes. *Structure* **16**, 1511–20 (2008).
  15. Puri, T., Wendler, P., Sigala, B., Saibil, H. & Tsaneva, I. R. Dodecameric structure and ATPase activity of the human TIP48/TIP49 complex. *J. Mol. Biol.* **366**, 179–92 (2007).
  16. Queval, R., Papin, C., Dalvai, M., Bystricky, K. & Humbert, O. Reptin and Pontin oligomerization and activity are modulated through histone H3 N-terminal tail interaction. *J. Biol. Chem.* **289**, 33999–4012 (2014).
  17. Tosi, A. *et al.* Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* **154**, 1207–19 (2013).
  18. Torreira, E. *et al.* Architecture of the Pontin/Reptin Complex, Essential in the Assembly of Several Macromolecular Complexes. *Structure* **16**, 1511–1520 (2008).
  19. López-Perrote, A., Muñoz-Hernández, H., Gil, D. & Llorca, O. Conformational transitions regulate the exposure of a DNA-binding domain in the RuvBL1-RuvBL2 complex. *Nucleic Acids Res.* **40**, 11086–99 (2012).
  20. Lakomek, K., Stoehr, G., Tosi, A., Schmailzl, M. & Hopfner, K. P. Structural basis for dodecameric assembly states and conformational plasticity of the full-length AAA+ ATPases Rvb1 . Rvb2. *Structure* **23**, 483–495 (2015).
  21. Gorynia, S. *et al.* Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex. *J Struct Biol* **176**, 279–291 (2011).
  22. Petukhov, M. *et al.* Large-scale conformational flexibility determines the properties of AAA+ TIP49 ATPases. *Structure* **20**, 1321–1331 (2012).
  23. Rottbauer, W. *et al.* Reptin and Pontin Antagonistically Regulate Heart Growth in Zebrafish Embryos. *Cell* **111**, 661–672 (2002).
  24. Petukhov, M. *et al.* Large-scale conformational flexibility determines the properties of AAA+ TIP49 ATPases. *Structure* **20**, 1321–1331 (2012).
  25. Brewster, A. S. & Chen, X. S. Insights into MCM Functional Mechanism: Lessons Learned from the archaeal MCM complex. *Crit Rev Biochem Mol Biol* **45**, 243–256 (2011).
  26. Li, D. *et al.* Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature* **423**, 512–518 (2003).
  27. Papin, C. *et al.* 3'- to 5' DNA unwinding by TIP49b proteins. *Febs J* **277**, 2705–2714 (2010).
  28. Kanemaki, M. *et al.* TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuVB-like DNA helicase, TIP49a. *J. Biol. Chem.* **274**, 22437–22444 (1999).

29. Makino, Y., Kanemaki, M., Kurokawa, Y., Koji, T. & Tamura, T. A. A rat RuvB-like protein, TIP49a, is a germ cell-enriched novel DNA helicase. *J. Biol. Chem.* **274**, 15329–15335 (1999).
30. Hee Baek, S. A novel link between SUMO modification and cancer metastasis. *Cell Cycle* **5**, 1492–1495 (2006).
31. Kim, J. H. *et al.* Roles of sumoylation of a reptin chromatin-remodelling complex in cancer metastasis. *Nat. Cell Biol.* **8**, 631–639 (2006).
32. Lee, J. S. *et al.* Negative regulation of hypoxic responses via induced Reptin methylation. *Mol. Cell* **39**, 71–85 (2010).
33. Lee, J. S. *et al.* Negative Regulation of Hypoxic Responses via Induced Reptin Methylation. *Mol. Cell* **39**, 71–85 (2010).
34. Lee, J. S. *et al.* Hypoxia-induced methylation of a pontin chromatin remodeling factor. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13510–13515 (2011).
35. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276–277 (2000).
36. Wood, M. A., McMahon, S. B. & Cole, M. D. An ATPase/Helicase Complex Is an Essential Cofactor for Oncogenic Transformation by c-Myc. *Mol. Cell* **5**, 321–330 (2000).
37. Linden, P. Van Der *et al.* Towards a high-throughput system for high-pressure cooling of cryoprotectant-free biological crystals. *J. Appl. Crystallogr.* **47**, 584–592 (2014).
38. Kim, C. U., Wierman, J. L., Gillilan, R. & Sol, M. A high-pressure cryocooling method for protein crystals and biological samples with reduced background X-ray scatter. *J. Appl. Crystallogr.* **46**, 234–241 (2012).
39. Fourme, R., Girard, E., Dhaussy, A. & Medjoubi, K. A new paradigm for macromolecular crystallography beamlines derived from high-pressure methodology and results. *J. Synchrotron Radiat.* **18**, 31–36 (2010).
40. Sanchez-weatherby, J. & Felisaz, F. Improving diffraction by humidity control : a novel device compatible with X-ray beamlines. *Acta Crystallogr D Biol Crystallogr* **65**, 1237–1246 (2009).
41. Kantardjiev, K. A. & Rupp, B. Protein isoelectric point as a predictor for increased crystallization screening efficiency. *Bioinformatics* **20**, 2162–2168 (2004).
42. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
43. Evans, P. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D* **67**, 282–292 (2010).
44. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
45. Kantardjiev, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865–1871 (2003).
46. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**,



- 658–674 (2007).
47. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2210–2221 (2004).
  48. Cowtan, K. Fast Fourier feature recognition. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **57**, 1435–1444 (2001).
  49. Cowtan, K. General quadratic functions in real and reciprocal space and their application to likelihood phasing. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 1612–1621 (2000).
  50. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
  51. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
  52. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, 665–667 (2004).
  53. Yeates, T. O. Simple statistics for intensity data from twinned specimens. *Acta Crystallogr. Sect. A* **44**, 142–144 (1988).
  54. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).
  55. Incardona, M. F. *et al.* EDNA: A framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Radiat.* **16**, 872–879 (2009).
  56. Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C. & Stuckey, P. J. Automatic generation of protein structure cartoons with pro-origami. *Bioinformatics* **27**, 3315–3316 (2011).
  57. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
  58. Diop, S. B. *et al.* Reptin and Pontin function antagonistically with PcG and TrxG complexes to mediate Hox gene control. *EMBO Rep.* **9**, 260–266 (2008).
  59. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E. Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* **7**, 339 (2006).
  60. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. (2011). doi:10.1038/msb.2011.75
  61. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. **42**, 320–324 (2014).
  62. Ewens, C. A. *et al.* Architecture and Nucleotide-Dependent Conformational Changes of the Rvb1-Rvb2 AAA+ Complex Revealed by Cryoelectron Microscopy. *Structure* **24**, 657–666 (2016).

63. Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
64. Brubaker, M. A. & Fleet, D. J. Building Proteins in a Day: Efficient 3D Molecular Reconstruction. 1–10 (2015). at [papers3://publication/uuid/A793B9DF-7BD9-43B8-A102-02B6877D25C1](https://papers3://publication/uuid/A793B9DF-7BD9-43B8-A102-02B6877D25C1)
65. Carazo, J. M., Sorzano, C. O. S., Otón, J., Marabini, R. & Vargas, J. Three-dimensional reconstruction methods in Single Particle Analysis from transmission electron microscopy data. *Arch. Biochem. Biophys.* **581**, 39–48 (2015).
66. Unit, B. Scipion Introductory Tutorial. 1–23 (2015).
67. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
68. Sorzano, C. O. S. *et al.* XMIPP: A new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* **148**, 194–204 (2004).
69. Sorzano, C. O. S. *et al.* A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Struct. Biol.* **171**, 197–206 (2010).
70. Jankowsky, E. & Margaret, E. An Introduction to RNA Helicases: Superfamilies, Families, and Major Themes. *RSC Biomol. Sci.* **19**, RNA He, (2010).
71. Papin, C. *et al.* 3'- to 5' DNA unwinding by TIP49b proteins. *FEBS J.* **277**, 2705–2714 (2010).
72. Kaplan, D. L. & O'Donnell, M. Rho factor: Transcription termination in four steps. *Curr. Biol.* **13**, 714–716 (2003).

# *Chapter 4*

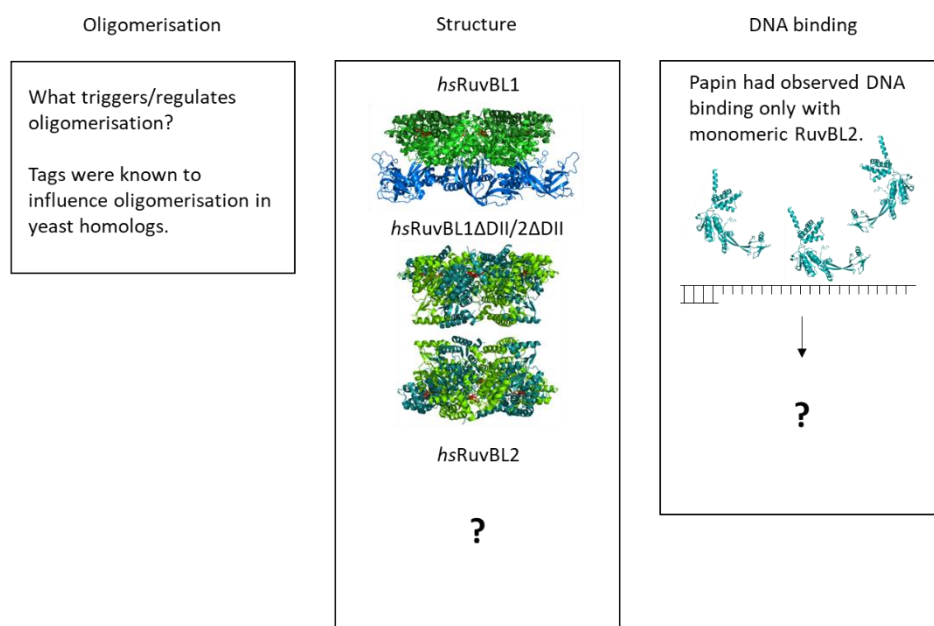
---

DISCUSSION

<b>4.1 Discussion .....</b>	<b>167</b>
<b>4.2 Concluding remarks and future perspectives .....</b>	<b>179</b>
<b>References.....</b>	<b>183</b>

## 4.1 Discussion

At the onset of this work, the atomic-level structures available of proteins from the RuvBL family included the full-length structures of the *Chaetomium thermophilum* RuvBL1/RuvBL2 complex<sup>1,2</sup>, the human RuvBL1/RuvBL2 truncated complex (lacking the regulatory domain II)<sup>3</sup>, the full-length human RuvBL1 complex<sup>4</sup> and a partial structure of human RuvBL2, which also lacked the important regulatory domain II. Further, the pool of knowledge on the oligomeric behaviour of this protein in solution was practically non-existent, and the DNA binding behaviour still contained considerable gaps, particularly on the changes that occur upon binding. Considering the known involvement of *hsRuvBL2* in a myriad of cancers and other pathologies, we felt it necessary to contribute to this area of knowledge (main gaps summarised in **fig. 4.1**), by focusing our efforts in the full-length *hsRuvBL2* structure elucidation. We further aimed to complement



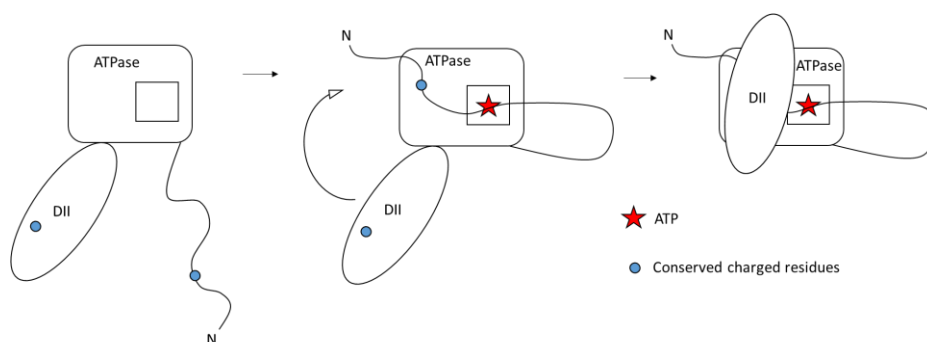
**Figure 4.1** – State of the art at the onset of the work of the main areas addressed in this thesis.

the high resolution structural information with low resolution information, using electron microscopy and small angle X-ray scattering. Information on protein behaviour in solution was obtained using size-exclusion chromatography, analytical ultracentrifugation and SAXS. We further aimed to gain further understanding on the mode of DNA binding of both human RuvBL proteins using electrophoretic mobility shift assays, and used electron microscopy to observe oligomeric changes of *hsRuvBL2* in the presence of DNA.

ATPase assays did not show any activity as compared with a positive control, even in the presence of ssDNA. This is probably because *hsRuvBL2* was in the hexameric form, which has been shown to be inactive for ATP binding and hydrolysis. The use of monomeric protein was however precluded by the concentration required to overcome the relatively low sensitivity of the colorimetric assay used. We will attempt to overcome this limitation in future studies by looking at other methods for ATP hydrolysis assays with higher sensitivity, preferably high-throughput (HT) methods, in order to analyse the behaviour of RuvBL2 in the presence of chemical compounds.

The main objective of this work was the structural study of the multi-function human RuvB-Like 2 protein, RuvBL2. This protein is involved in several fundamental pathways, from transcription regulation and cell-cycle control to chaperone in the biogenesis of snoRNPs, with implications on several complexes, such as ribosomes and the Telomerase complex. Due to the high mobility of the RuvBL2 oligomer, the attainment of a crystallographic structure required a continuous and extensive optimization of the purification and crystallization protocols. The structure described here is topologically similar to others in the same family<sup>1,2,4</sup>, albeit with some distinguishing features. In particular, the comparison of the surface electrostatic charges of *hsRuvBL2* and *hsRuvBL1* suggests that the differences in the regulation of their activities may be related to interactions with

distinct binding partners, and possibly the addition of post-translational modifications at different sites. Additionally, the structure of apo *hsRuvBL2*, when compared to nucleotide-bound structures, suggests that upon nucleotide binding, the N-terminus undergoes reorganization through coordination with two conserved histidines (H24 and H26 in *ctRuvBL2*, H40 and H42 in the *hsRuvBL2*). As a consequence of this movement, the N-terminus becomes suitably positioned to act as an interface between the ATPase core and domain II, through electrostatic contacts mediated by a conserved lysine (Fig. 4.2).



**Figure 4.2 – Schematics of the N-terminus reorganization.** Upon ATP (red star) entrance to the binding pocket (square), the N-terminal segment is re-organized, and forms an interface for interaction with domain II (DII). The latter domain thus moves into the close vicinity of the ATPase core, putatively stabilised by conserved charged residues.

The *hsRuvBL1* N-terminus also contains the conserved histidines that bind to ATP in the binding pocket. However, the fact that the nucleotide-bound structures of RuvBL1 do not display the same compact conformation for the domain 2 extension hints at a distinct mechanism of action and possibly regulation. The proposed mechanism of action for RuvBL2, which relies on the mobility of the N-terminus, would partly justify the, observed by us, disruptive effect on protein stability of an added purification tag at this extremity. These observations, together with the data obtained by analytical ultracentrifugation on oligomeric behaviour

according to tag placement, also suggest that in *hsRuvBL2*, the N- and C-termini may have an influence in oligomerization. The opposite effect was observed for *hsRuvBL1*, which in our hands could only be expressed in soluble form with the tag on the N-terminus. This further supports different mechanisms of action for the two RuvBLs, despite their sequence and structural similarity. The proposed mechanism may also provide a rationale for understanding the effects of mutations that interfere with the activity of RuvBL2, such as the **lik (FCR) mutation**, in which Phe-Cys-Arg residues are inserted within the OB fold (located in domain II), between G190 and D191 (**Fig. 4.3**). This mutant has an **increased and DNA-independent ATPase activity**, as well as an altered pattern of oligomerization, as compared to the wild-type protein<sup>5,6</sup>. In light of the mechanism of action proposed in this work, it is likely that the inserted residues in the *lik* mutant of RuvBL2 lead to the enhanced ATPase activity by increasing the bulk size of the OB-fold. The additional hydrophobic residue may further promote a preferentially compact



**Figure 4.3 – *hsRuvBL2* monomer (cyan).** The Phe-Cys-Arg insertion present in the *lik* mutant occurs between the G190 and D191 residues (orange sticks), within the OB-fold. This insertion increases the bulk size and changes electrostatic and hydrophobicity properties of the DII domain, which increases ATPase activity.



conformation. Together, these effects may lead to an increased ATP turnover by decreasing the amplitude of domain 2 motion in relation to the ATPase core, thereby increasing its frequency. Alternatively, a preferentially compact conformation may translate into a more accessible nucleotide binding site of the oligomer, thus increasing the rate of nucleotide entry and exchange. These suggestions on the effect of this mutation are however very speculative, and require the analysis of more structural intermediates of action, both to clarify the proposed mechanism and the putative influence of functional mutations.

The *hsRuvBL2* crystal structure presented in this study provides insight into the mechanisms whereby this protein, alone or in combination with *hsRuvBL1*, may contribute to the regulation of diverse cellular functions. Distinct structural features of RuvBLs are the basis of different, and sometimes antagonistic roles in, e. g., the regulation of development. Other aspects that may regulate their activities differently are their different rates of dissociation into monomers (and how this modulates their roles), ATPase and DNA binding kinetics (data already published on this regarding RuvBL2, for monomers and hexamers)<sup>7</sup>. Although it must be considered that in the cell these rates may be affected by cell crowding and specific regulators depending on the cell cycle phase, localization etc., the kinetic values observed *in vitro* may provide some indication of their predispositions. Interestingly, the group of Claire Davies has observed that RuvBL1 R205 is methylated by PRMT5, in the context of TIP60 regulation of histone acetylation, **but not** the RuvBL2 equivalent R206 (**Fig. 4.4**). In their publication (in co-authorship with our group), they base their justification for this fact on the observation that the location of these residues within domain II can adopt widely different orientations with respect to the ATPase core, and while so far all reported RuvBL1 structures show R205 to be exposed to the solvent (and therefore accessible to methylation), all RuvBL2 structures to date showed R206 to be occluded by the compact conformation of domain II<sup>8</sup>. The present study shows, however, that

RuvBL2 may acquire an open conformation of domain II, in which the discussed residues fall in similar topological areas, both solvent-accessible, as depicted in **Fig. 4.4**. Therefore, the fact that this arginine is methylated in RuvBL1 but not in RuvBL2 may be justified by either an increased prevalence of a compact conformation in RuvBL2, or the involvement of other regulatory factors.

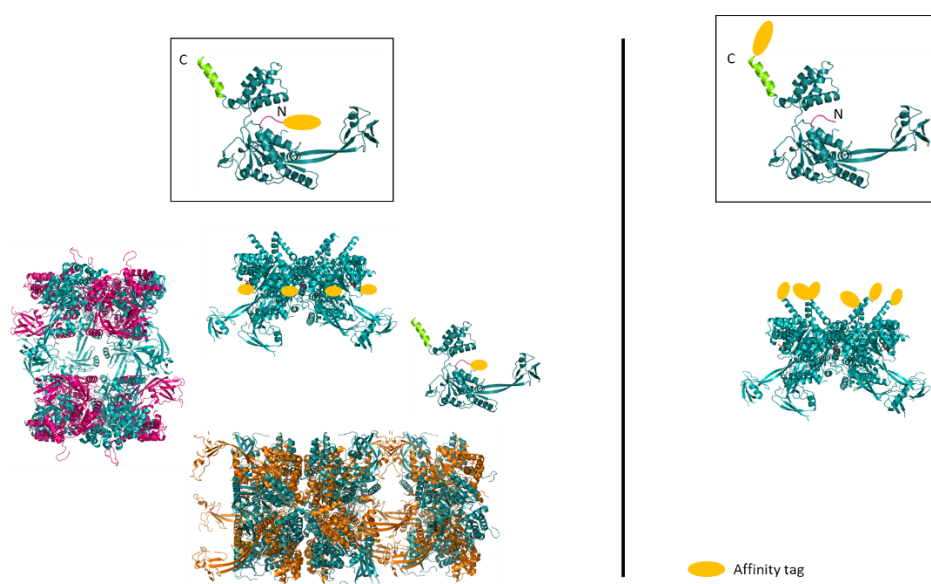


**Figure 4.4** – Superimposition of *hsRuvBL1* (pink) and *hsRuvBL2* (cyan). R205 from RuvBL1 is depicted in green sticks, and K201 from RuvBL2 in yellow sticks. The latter is depicted for spatial location of the area where R206 is roughly located, since this residue falls within a loop for which there is no built model. These two residues seem to be located in similar topological locations, which would be exposed when domain II is in an “open” conformation.

In order to gain insight into the flexibility of the *hsRuvBL2* oligomer, a negative staining electron microscopy analysis was performed. The obtained low-resolution structure is similar to others of the RuvBL family, and together with a SAXS analysis, supports a wide conformational flexibility of the outwards protruding domains II, as expected. Surprisingly, the EM particle averages also showed the presence of heptamers in the sample, in addition to the expected

hexamers. A cryo-EM analysis is currently being performed in order to obtain a higher resolution structure of this complex in conditions closer to the native state in solution.

The structural analysis was complemented with a study of the oligomerisation behaviour in solution, by analytical ultracentrifugation. This analysis aimed at defining the influence of tags in the human RuvBL2 oligomer, since such an interference had been described for the yeast complex<sup>9</sup>. Indeed, the oligomerisation state of human RuvBL2, as well as its stability, is also affected by tags. As with the yeast Rvb1 and Rvb2, *hsRuvBL2* forms dodecamers when expressed with affinity tags on the N-terminus (Fig. 4.5, left). Further, SEC and AUC analyses show that this construct also forms hexamers, monomers and high molecular weight species, with the proportion of each depending on protein



**Figure 4.5 – Influence of affinity tags in oligomer formation of *hsRuvBL2*.** Left: *hsRuvBL2* expressed with tags on the N-terminus (pink) will associate into various oligomeric forms in solution. Right: *hsRuvBL2* expressed with tags on the C-terminus (green) associates into hexamers (with a small percentage of heptamers, identified by electron microscopy).

concentration. The formation of different molecular weight oligomers seems to be maintained after N-terminal tag removal, as demonstrated by SEC analyses. When the affinity tags were placed on the C-terminus of *hsRuvBL2* however, only hexamers were formed at all concentrations tested, even after tag removal (**Fig. 4.5, right**). These observations suggest that oligomerization plasticity may be determined during protein biogenesis.

Finally, an analysis of the DNA-binding properties of *hsRuvBL2* was performed, by electrophoretic mobility shift assay. As previously observed in other studies<sup>7</sup>, hexameric *hsRuvBL2* did not bind ssDNA in our hands. However, when the same form was co-incubated with monomeric *hsRuvBL1*, we observed binding of both *hsRuvBL1* and *hsRuvBL2*. This binding ability was prevented by the prior co-incubation of *hsRuvBL1* monomers with AMP-PNP, which suggests that ATP hydrolysis may be necessary for ssDNA binding. Furthermore, we observed by electron microscopy that monomeric *hsRuvBL2* assembles into ring-shaped complexes in the presence of ssDNA. It is possible that *hsRuvBL2* oligomerizes around the DNA molecule, but this hypothesis still requires confirmation.

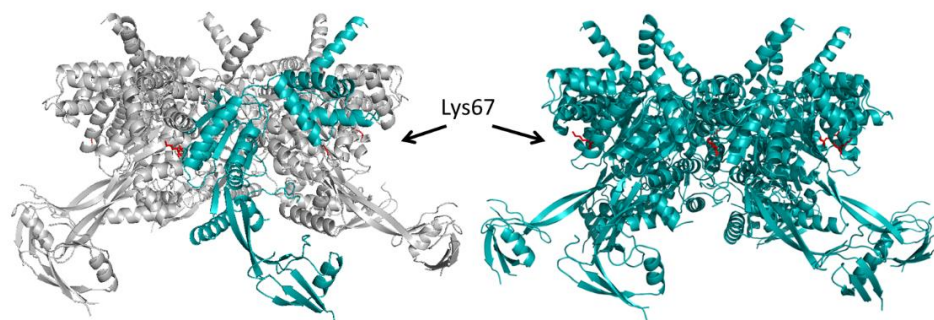
A collaboration with the group of Michael Sherman (Boston University School of Medicine) resulted in the publication of results connecting the *hsRuvBL1/RuvBL2* complex and the **disaggregation of amyloid fibrils**<sup>10</sup>. In short, this group identified *hsRuvBL1* and *hsRuvBL2* as being associated to aggresome formation through a full-genome siRNA screen in mammalian cells. The aggresome is an organelle formed to transport polypeptide aggregates in the cytosol to the centrosome. This system comes into play in aged and diseased cells, which do not have functioning chaperones or an active ubiquitin-proteasome system. Deletion of either *RuvBL1* or *RuvBL2* abolished aggresome formation in mammalian cells. Additionally, a crosslinking-MS/MS analysis using aggresome

substrate synphilin-1, identified RuvBL1 as a direct binding partner. The crosslinked residues were K663 of synphilin-1 and K372 of RuvBL1, the latter located at the top of the ATPase core. Furthermore, it was shown that interaction of RuvBL1 with synphilin-1 targeted this protein to the aggresome. Further binding assays show that RuvBL1 interacts with other misfolded peptides, such as Amyloid- $\beta$  and insulin fibres, with the consequent stimulation of the ATPase activity of the RuvBL1/2 complex, particularly of the domain II-truncated form.

Our contribution to this work was in the form of pure *hsRuvBL1/2* complex, both full-length and domain II-truncated<sup>3</sup>. Briefly, for the production of the full-length complex, each tagged protein was separately expressed in *E. coli*, and the pellets of both expressions were re-suspended together before cell disruption and purification, in the presence of ADP. The final step was a size exclusion S200 26/60 HiLoad column, which produced three peaks, corresponding to complexes with the apparent molecular weights 659, 546 and 502 KDa. These indicate molecules around the size of a dodecamer, and suggest either partial monomer loss or differences in shape that cause different elution profiles.

In cancer patients, *hsRuvBL2* overexpression is considered a mark of poor prognosis<sup>11-15</sup>. Previous works have shown that, in hypoxic conditions, *hsRuvBL2* is methylated at K67 (**Fig. 4.6**), which leads to downregulation of pro-apoptotic *BNIP3*, pro-angiogenic *PGK1*, and *VEGF*. Combining these observations, it is conceivable that, when a tumour reaches hypoxic state, overexpression of *hsRuvBL2*, and its consequent methylation *en masse*, may contribute to a large-scale downregulation of this subset of hypoxia target genes<sup>16</sup>. The consequences could be an involvement in the already described increased resistance of hypoxic tumours to chemo and radiotherapy treatments<sup>17,18</sup>. The crystallographic structure of *hsRuvBL2* can thus contribute to the development of small compounds or antibodies to either target the surface of *hsRuvBL2*-K67Me or aimed at the

disruption of the HIF-1 $\alpha$ :*hsRuvBL2*-K67Me complex. Either way, an understanding of this interaction and its effects at the cellular (and organism) level is necessary prior to developing a strategy to modulate it. On the other hand, the structure of *hsRuvBL2* may be of use in the development of a small compound targeted at the nucleotide-binding pocket, with the objective of decreasing hyperactivity, in the case, e. g., of the *lik* mutant or other mutations that result in heart hyperplasia or other malformations during development.



**Figure 4.6 – Cartoon model of the *hsRuvBL2* hexamer, highlighting the position of lysine 67.** **Left:** one protomer is represented in cyan and the remaining five in grey. **Right:** all protomers represented in cyan. Lysine 67 is depicted in red sticks. The structure of *hsRuvBL2* provides structural support to the biochemical results of post-translational modifications to specific residues. *hsRuvBL2* is methylated in Lysine 67 in hypoxic conditions by the methyltransferase G9a. Methylated *hsRuvBL2* binds to the promoters of a subset of hypoxia-responsive genes and negatively regulates transcription of these genes to modulate cellular responses to hypoxia. By combining biochemical and structural approaches, a rational strategy can be defined for the development of a targeted regulatory molecule.

Further studies on possible functions of RuvBLs and distinct modifications to their surface residues may provide a framework for the development of compounds that can regulate the specific activities of RuvBL1 and RuvBL2, depending on the tissue, cell stage or stress conditions of the target cells. By targeting specific modifications of RuvBLs that result from particular metabolic

cellular states, it might be possible to acquire the degree of specificity desired when aiming for the treatment of a specific cell subset.

RuvBLs have been identified as taking part in many supramolecular assemblies, where their ATPase activity is not always required. Such versatility has long posed a mystery, since no specific function can be attributed to these proteins, and yet they appear to be critical for the regulation of an array of complexes, particularly related to the control of gene expression and DNA damage response. It is possible that one of the RuvBLs' functions within larger complexes could be to recognize transient binding partners, translating cellular needs to the rest of the complex, and thus regulating its activity. The fact that only heteromeric RuvBL1/2 complexes have been found to date adds another layer of complexity to their roles in transcription regulation. An interesting suggestion as to the function of RuvBLs has been put forward by Zhou and colleagues<sup>19</sup>, based on the fact that the unstable Ino80 insert could be stabilized by interactions with the yeast RuvBL1/RuvBL2 hexamer, whereby the RuvBL1/RuvBL2 complex could have proteins as the most relevant substrates, instead of DNA. Indeed, so far a DNA-binding activity has been observed *in vivo* for both RuvBL1, RuvBL2 and their complex<sup>4,7,20</sup>, but cellular functions so far identified mostly pertain to its function as a chaperone in larger multi-protein complexes<sup>10,21-23</sup>. The aggresome-related function of human RuvBLs lends further support to this hypothesis<sup>10</sup>, as both RuvBL1 and RuvBL2 were found to interact directly with synphilin-1. This hypothetical function is reminiscent of that of the AAA+ ATPase katanin p60. This protein forms a hexameric complex with a central pore rich in basic residues, which are known to interact with the unstructured, acidic tails of tubulins<sup>24</sup>.

When ATP binding is necessary for the dodecameric complex function, it is possible that the nucleotide drives the movement of domain II (of RuvBL2, at least) in a sequential way, causing a circular tilting of the barrel-shaped complex.

The consequences of this movement in the interacting partners of the RuvBL complex are still to be determined.

During the timeframe of the work described in this dissertation, the advances in the area obtained by other groups, were significant. Notably, the crystallographic structure of the truncated human RuvBL2 was obtained by the group of Petukhov<sup>6</sup>, and the groups of Karl-Peter Hopfner and Christoph Müller<sup>1,2</sup> determined simultaneously the crystallographic structure of the RuvBL1/RuvBL2 complex from the thermophilic fungus *Chaetomium thermophilum*. More recently, the cryo-EM structure of the yeast R2TP complex was determined at 8.37 Å<sup>25</sup>, and the cryo-EM structure of RuvBL1/RuvBL2 hexamer in complex with the Ino80 insert was determined to about 12 Å resolution<sup>19</sup>. Our group, through the work described in this dissertation (summarily represented in **fig. 4.7**), has contributed to the advancement of the area through the determination of the full length structure of human apo RuvBL2, thus filling the gap on structural information about the regulatory domain II. We have complemented the structural information with data on the oligomeric behaviour of *hsRuvBL2*, providing some insight into how the use of tags may affect this protein, with putatively significant consequences at the regulatory level, since different oligomeric forms of the protein may be differently regulated. We also aimed to answer the question of what structural changes occur upon DNA binding. Since *hsRuvBL2* is monomeric at very low concentrations, we resorted to electron microscopy to observe that the presence of ssDNA promotes the oligomerisation of initially monomeric *hsRuvBL2*. Whether that oligomerization occurs around the ssDNA strand is still to be determined. However, considering the mode of action of ring-shaped helicases, it can be speculated that that option is quite likely.



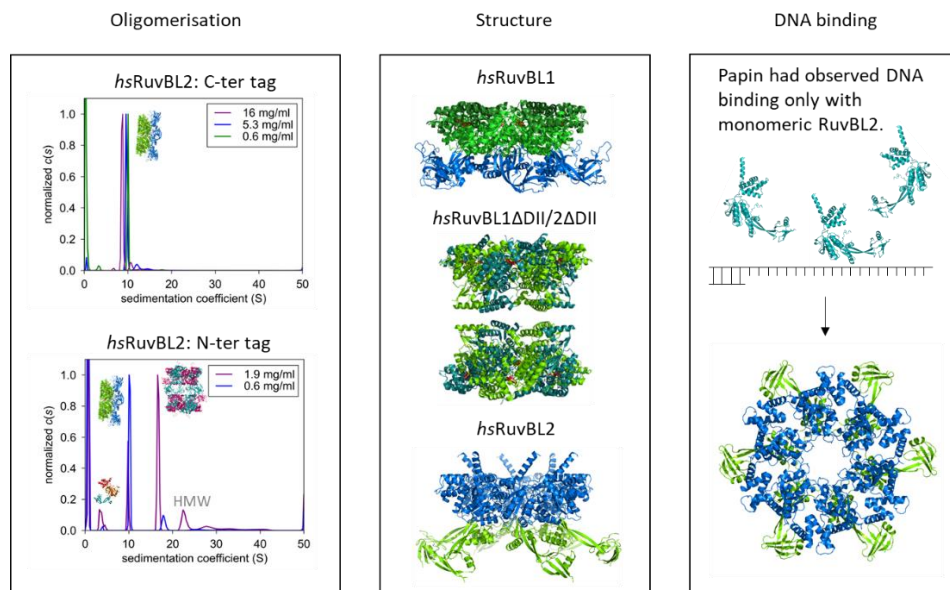


Figure 4.7 – Main knowledge gaps that were filled with the work described in the present dissertation.

## 4.2 Concluding remarks and future perspectives

With this work, we tackled the question of what makes RuvBL transcription factors similar and interchangeable, but also (and especially) what makes them different on a molecular level such that, when both are present simultaneously, they can be differently recognized by upstream regulators. To achieve this, we solved the missing structure of full-length human RuvBL2, and thus unveiled important structural features that differentiate RuvBL1 and RuvBL2. The differences in charge distribution reflect the surface exposure of distinct motifs, which may underlie their occasional non-interchangeability, even though they are able to perform the same biochemical activities.

Overall, this work is a significant contribution to the area of RuvBL studies, and of great interest for pharmaceutical applications. Since RuvBL2 is directly

involved in many pathways involved in cancer and other pathologies, the future goals will involve steps to be taken in the short, medium and long term:

The logical way to proceed with this project on a **short term** seems to be approaching it through several fronts. Importantly, it is necessary to analyse the structure and activity of specific *hsRuvBL2* mutants, chosen for their putative influence on ATPase activity, in order to corroborate the proposed mechanism. In this regard, on the short term, it seems reasonable to:

a) Produce mutants of the N-terminus Histidines H40 and H42, involved in the interaction with ATP, and of the conserved residues involved in electrostatic interactions between the N-terminus and the domain II, and study their biochemical activities and structure.

b) Produce an ATP-binding site mutant (of the Walker A and B motifs), in order to obtain a structure with trapped nucleotide, both of *hsRuvBL2* and of the *hsRuvBL1/RuvBL2* complex.

These studies will also be extended to *hsRuvBL1*, in order to further explore structural characteristics on the basis of differences in activity and regulation.

On the **medium to long term**, we will study the interaction of *hsRuvBL2* and its homolog *hsRuvBL1* with specific binding partners, respectively HIF-1 $\alpha$  and c-Myc, in order to understand how these interactions may impact cells, with a particular interest on cancer cells. To accomplish this, a reasonable course of action includes:

a) The determination of the structure of the HIF-1 $\alpha$ -*hsRuvBL2* and c-Myc-*hsRuvBL1/hsRuvBL2* complexes, with the purpose of understanding its effects at the cellular level, and study ways to disrupt or regulate their interaction, by analysing their interface.

b) Performing high-throughput compound screenings for these targets using Surface Plasmon Resonance, with the final goal of producing a compound that can be altered to have enhanced affinity to target protein-protein interaction (PPI). Upon identification of strongly interacting compounds, co-crystallization can be performed in order to observe how the compounds interact with the interface pocket, and improve their binding affinity in a directed process. These compounds can then be tested in *hsRuvBL1*- or *hsRuvBL2*-overexpressing cancer cell lines and hypoxic cells, in order to study the interference in cell viability and resistance.

c) Analysis of the stability of *hsRuvBL2* in the presence of a library of compounds, in order to find interacting compounds that may be improved to regulate the activity of this target. We may accomplish this goal through the use of HT Differential Scanning Fluorimetry Assays to identify the first hits (compounds that interact with the target protein) and follow with activity assays using Surface Plasmon Resonance. Upon identification of strongly interacting compounds, co-crystallization can be performed in order to observe how the compounds interact with the binding pocket. From those structure(s), the compound(s) can be improved to use the binding pocket to full capacity, increasing its binding affinity, in an iterative process.

In parallel, also in the medium to long term, we aim to study the crystallographic structure of larger complexes, such as the R2TP and INO80. For these studies, the most rational approach is probably the collaboration with groups that are able to complement the X-ray crystallography limitations, such as groups working with resort to electron microscopy. In this way, we can tackle the structure of subcomplexes, previously identified through crosslinking and activity assays, done either by us or in collaboration. To produce such multiprotein assemblies, we

*Chapter 4*

will attempt both co-expression and co-incubation of individually expressed proteins.

## References

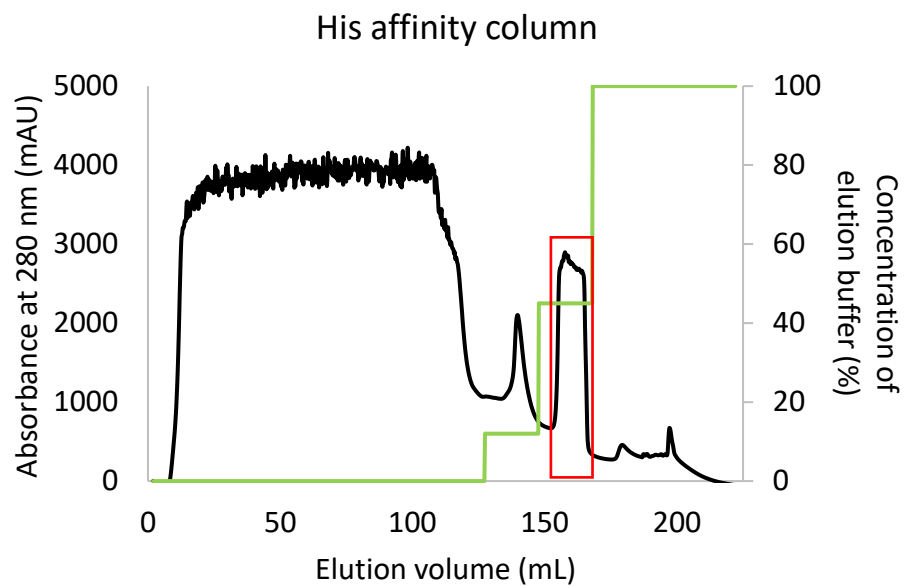
1. Silva-Martin, N. *et al.* The Combination of X-Ray Crystallography and Cryo-Electron Microscopy Provides Insight into the Overall Architecture of the Dodecameric Rvb1/Rvb2 Complex. *PLoS One* **11**, e0146457 (2016).
2. Lakomek, K., Stoehr, G., Tosi, A., Schmailzl, M. & Hopfner, K.-P. Structural Basis for Dodecameric Assembly States and Conformational Plasticity of the Full-Length AAA+ ATPases Rvb1/Rvb2. *Structure* **23**, 1–13 (2015).
3. Gorynia, S. *et al.* Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex. *J Struct Biol* **176**, 279–291 (2011).
4. Matias, P. M., Gorynia, S., Donner, P. & Carrondo, M. A. Crystal structure of the human AAA+ protein RuvBL1. *J Biol Chem* **281**, 38918–38929 (2006).
5. Rottbauer, W. *et al.* Reptin and Pontin Antagonistically Regulate Heart Growth in Zebrafish Embryos. *Cell* **111**, 661–672 (2002).
6. Petukhov, M. *et al.* Large-scale conformational flexibility determines the properties of AAA+ TIP49 ATPases. *Structure* **20**, 1321–1331 (2012).
7. Papin, C. *et al.* 3'- to 5' DNA unwinding by TIP49b proteins. *FEBS J.* **277**, 2705–2714 (2010).
8. Clarke, T. L. *et al.* PRMT5-Dependent Methylation of the TIP60 Coactivator RUVBL1 Is a Key Regulator of Homologous Recombination. *Mol. Cell* **65**, 900–916.e7 (2017).
9. Cheung, K. L. Y., Huen, J., Kakihara, Y., Houry, W. A. & Ortega, J. Alternative oligomeric states of the yeast Rvb1/Rvb2 complex induced by histidine tags. *J. Mol. Biol.* **404**, 478–92 (2010).
10. Zaarur, N. *et al.* RuvBL1 and RuvBL2 enhance aggresome formation and disaggregate amyloid fibrils. *EMBO J* **34**, 2363–2382 (2015).
11. Ren, J. *et al.* Overexpression of Reptin in renal cell carcinoma contributes to tumor malignancies and its inhibition triggers senescence of cancer cells. *Urol. Oncol. Semin. Orig. Investig.* **31**, 1358–1366 (2013).
12. Mallegol, J. *et al.* Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma. *Gastroenterology* **132**, 1866–1876 (2007).
13. Haurie, V. *et al.* Adenosine triphosphatase Pontin is overexpressed in hepatocellular carcinoma and coregulated with Reptin through a new posttranslational mechanism. *Hepatology* **50**, 1871–1883 (2009).
14. Rousseau, B. *et al.* Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma. *Hepatology* **46**, 1108–1118 (2007).
15. Ménard, L. *et al.* In vivo silencing of Reptin blocks the progression of human hepatocellular carcinoma in xenografts and is associated with replicative senescence. *J. Hepatol.* **52**, 681–689 (2010).

16. Lee, J. S. *et al.* Negative regulation of hypoxic responses via induced Reptin methylation. *Mol. Cell* **39**, 71–85 (2010).
17. Brown, J. M. Exploiting the hypoxic cancer cell: mechanisms and therapeutic strategies. *Mol. Med. Today* **6**, 157–162 (2000).
18. Brown, J. M. in *Methods in enzymology* **435**, 297–321 (2007).
19. Zhou, C. Y. *et al.* Regulation of Rvb1/Rvb2 by a Domain within the INO80 Chromatin Remodeling Complex Implicates the Yeast Rvbs as Protein Assembly Chaperones. *Cell Rep.* **19**, 2033–2044 (2017).
20. Gorynia, S. *et al.* Structural and functional insights into a dodecameric molecular machine - The RuvBL1/RuvBL2 complex. *J. Struct. Biol.* **176**, 279–291 (2011).
21. Kakihara, Y. & Saeki, M. The R2TP chaperone complex: its involvement in snoRNP assembly and tumorigenesis. *Biomol Concepts* **5**, 513–520 (2014).
22. Morrison, A. J. & Shen, X. Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nat Rev Mol Cell Biol* **10**, 373–384 (2009).
23. McKeegan, K. S., Debieux, C. M. & Watkins, N. J. Evidence that the AAA+ proteins TIP48 and TIP49 bridge interactions between 15.5K and the related NOP56 and NOP58 proteins during box C/D snoRNP biogenesis. *Mol. Cell. Biol.* **29**, 4971–4981 (2009).
24. Johjima, A. *et al.* Microtubule severing by katanin p60 AAA+ATPase requires the C-terminal acidic tails of both  $\alpha$ - and  $\beta$ -tubulins and basic amino acid residues in the AAA+ring pore. *J. Biol. Chem.* **290**, 11762–11770 (2015).
25. Rivera-Calzada, A. *et al.* The Structure of the R2TP Complex Defines a Platform for Recruiting Diverse Client Proteins to the HSP90 Molecular Chaperone System. *Structure* (2017). doi:10.1016/j.str.2017.05.016

*Supplementary  
information*

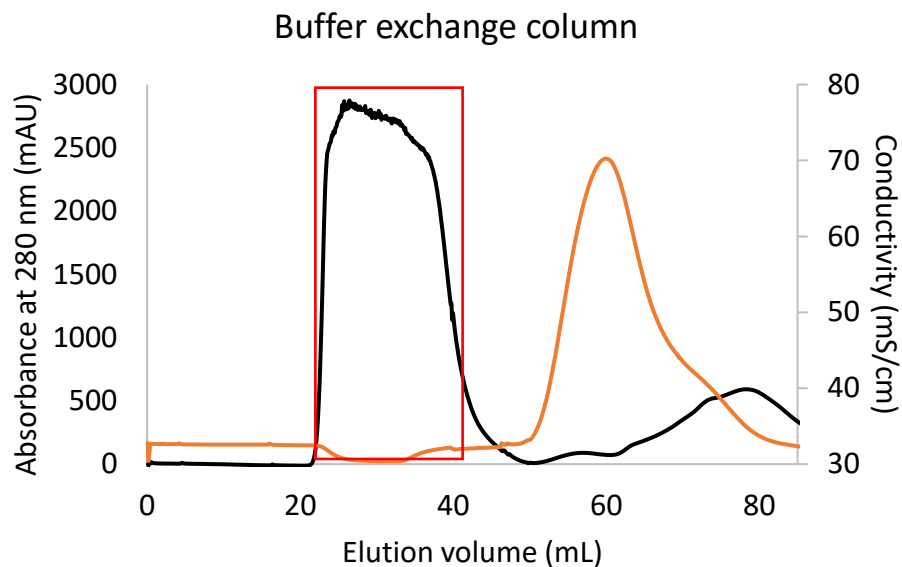
---

## SI.1 Chromatograms of RuvBL2 purification

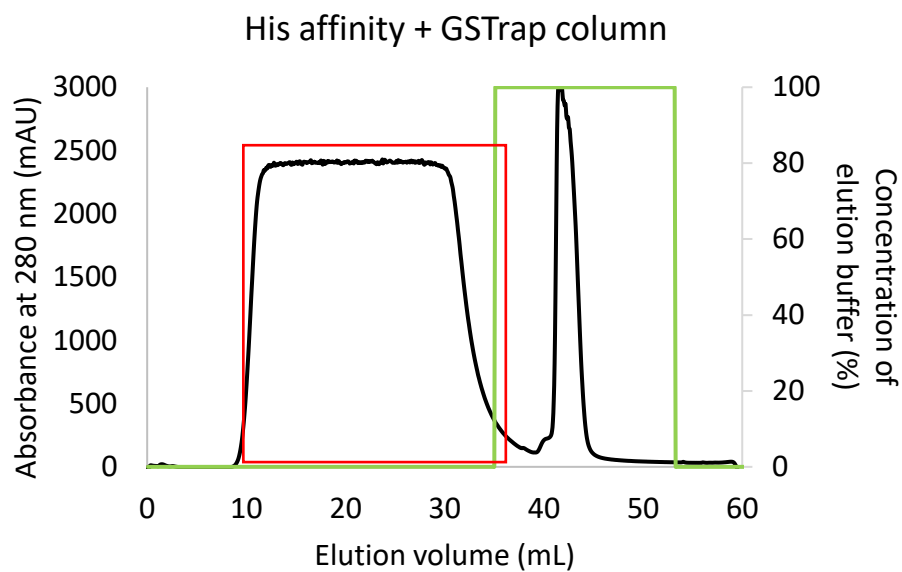


**SI Figure 1. First purification step: HisTrap HP (GE Healthcare).** Black line: absorbance at 280 nm (left axis); green line: concentration of buffer B (right axis). RuvBL2 is eluted almost pure on the second step of elution with buffer B (red box).

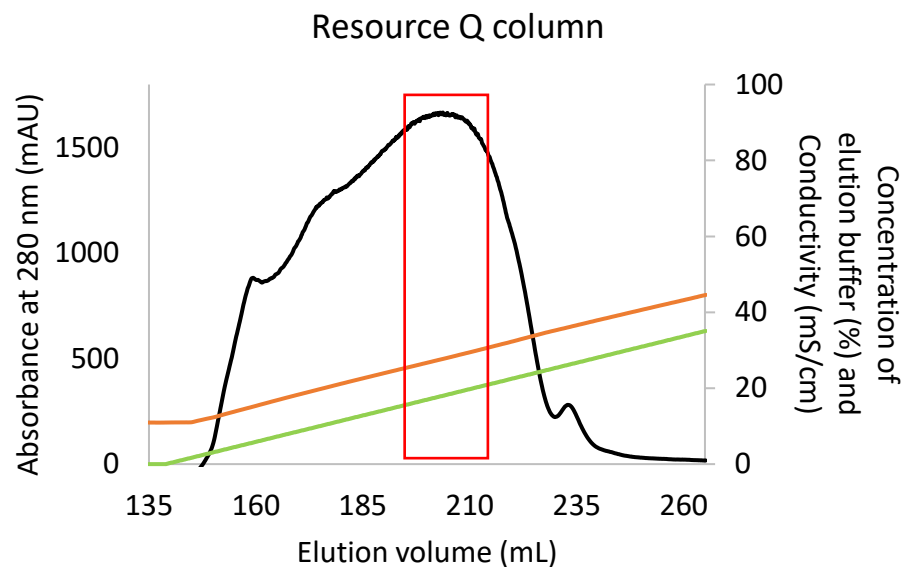




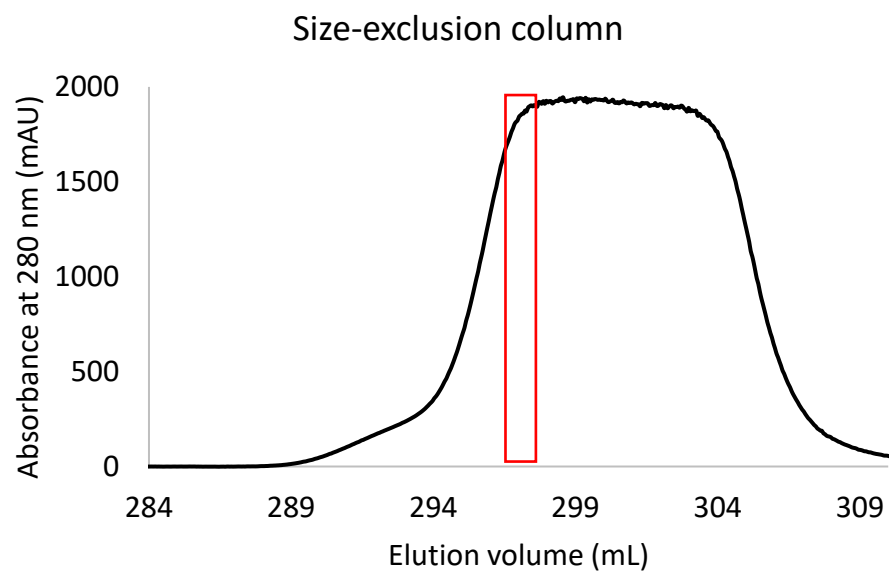
**SI Figure 2. Second purification step: HiPrep 26/10 Desalting (GE Healthcare).** Black line: absorbance at 280 nm (left axis); orange line: buffer conductivity (right axis). The red box indicates the collected peak.



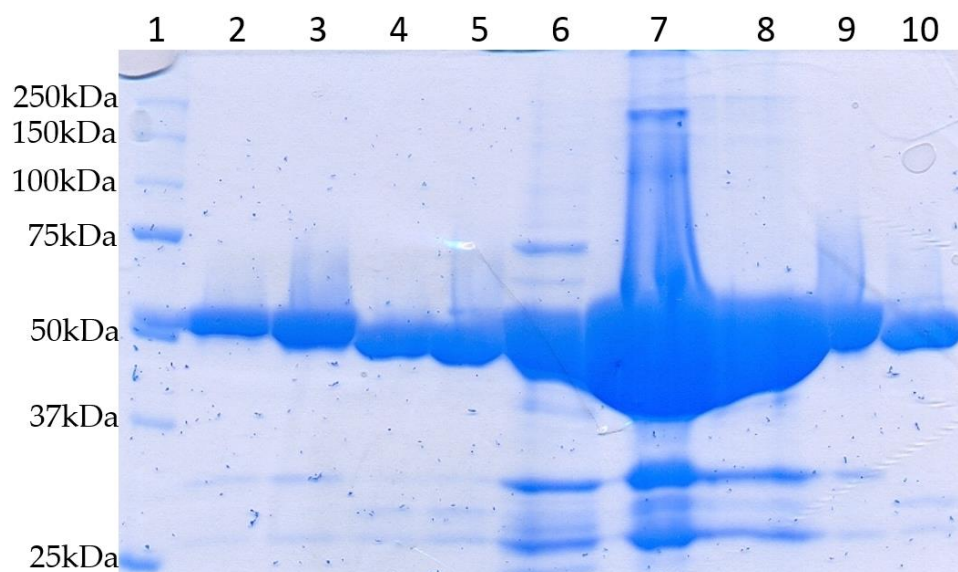
**SI Figure 3. Third purification step: HisTrap HP in tandem with a GSTrap 4B column (GE Healthcare).** Black line: absorbance at 280 nm (left axis); green line: concentration of elution buffer (right axis). The red box indicates the collected peak.



SI Figure 4. Fourth purification step: 6 mL Resource Q column (GE Healthcare). Black line: absorbance at 280 nm (left axis); green line: concentration of elution buffer (right axis). Orange line: sample conductivity (right axis). The red box indicates the collected peak.

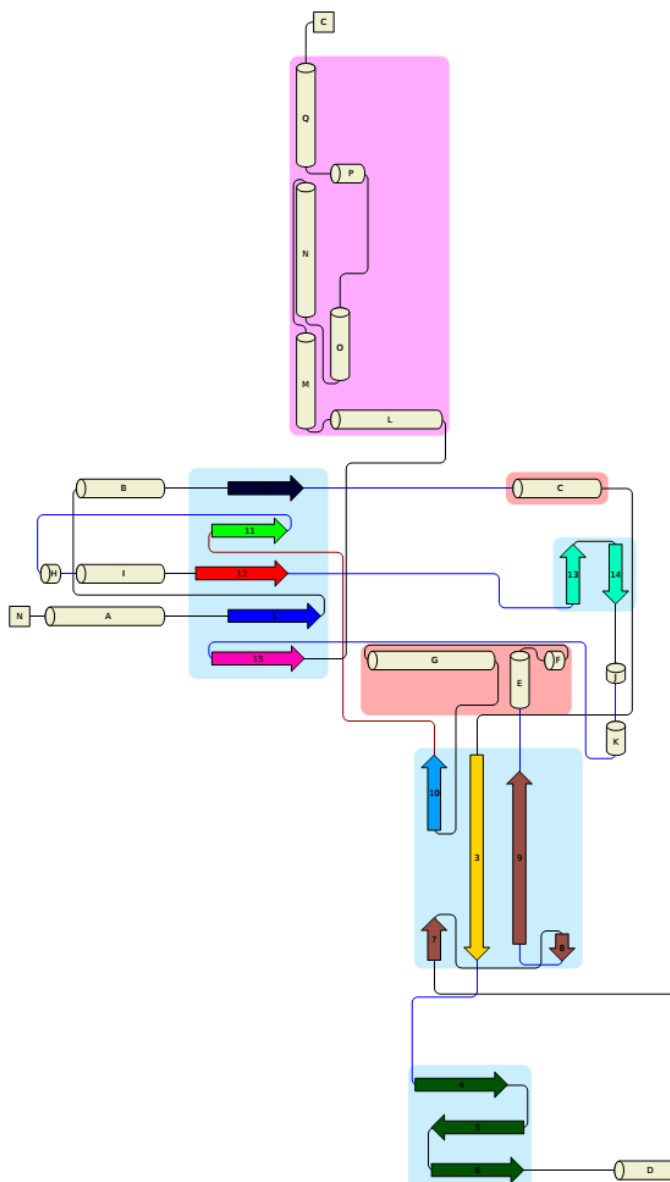


SI Figure 5. Inset of the fifth purification step, a HiLoad 16/60 Superdex 200 pg (GE Healthcare). Black line: absorbance at 280 nm. The red rectangle indicates the collected fraction, which produced crystals diffracting to 2.8 Å.

**SI.2 SDS-PAGE gel of RuvBL2 purification**

**SI Figure 6. Sodium dodecyl sulphate polyacrylamide gel electrophoresis analysis of samples from steps of RuvBL2 purification.** 1 - BioRad Un-stained Protein Markers, 161-0363 (8  $\mu$ L); 2, 3 - Size-exclusion pool (tagged RuvBL2) (0.2 and 0.4  $\mu$ L); 4,5 - Size-exclusion pool (untagged RuvBL2) (0.16 and 0.32  $\mu$ L); 6 - Pool eluted with 170 mM Imidazole (cleaning step); 7 - Pool collected from HisTrap (18  $\mu$ L); 8 - Pool eluted with 1 M Imidazole (cleaning step); 9 - Sample injected in SEC (tagged RuvBL2, 1  $\mu$ L); 10 - Sample injected in SEC (untagged RuvBL2, 1  $\mu$ L).

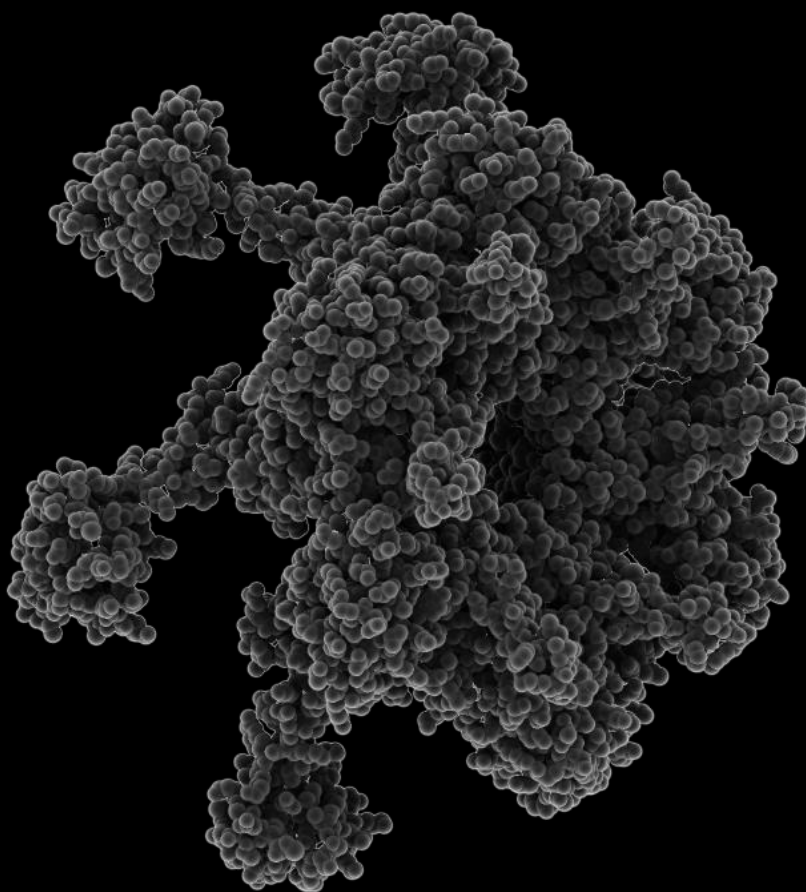
### SI.3 *hsRuvBL2* structure diagram



**SI Figure 7. Cartoon diagram of *hsRuvBL2*, obtained with Pro-origami.** Individual subdomains formed by  $\beta$ -sheets are highlighted in blue, and those formed by  $\alpha$ -helices are highlighted in pink. Sheets from a continuous string have the same colour. Clearly identified are the two clusters of  $\beta$ -sheets from the ATPase core (top), the linker to the OB-fold within the domain II (middle), and the OB-fold itself (bottom). All clusters of  $\alpha$ -helices present are part of the ATPase core.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal  
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

[www.itqb.unl.pt](http://www.itqb.unl.pt)



Oeiras, November, 2017

Structural insights into the human multifunctional protein RuvBL2

Sara Teresa Neves da Silva

