UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE DEL FARMACO

DIPARTIMENTO DI SCIENZE BIOMEDICHE

CORSO DI LAUREA MAGISTRALE IN

PHARMACEUTICAL BIOTECHNOLOGIES

BIOTECNOLOGIE FARMACEUTICHE

Preliminary study of the Zinc-Finger RAN-Binding protein 3 (ZRANB3)

Supervisor: **Prof. Dorianna Sandonà**

Padova, Italy

Co-Supervisor: **Dr. Matteo De March**

UNG, Slovenja

Student: **Aymane El Fellah**

Matriculation N*: 2040341

ANNO ACCADEMICO 2023/2024

UNIVERSITÀ DEGLI STUDI DI PADOVA

Second Cycle Degrees in Pharmaceutical Biotechnologies

Preliminary study of the Zinc-Finger RAN-Binding protein 3 (ZRANB3)

Supervisor: **Prof. Dorianna Sandonà**

Padova, Italy

Co-Supervisor: **Dr. Matteo De March**

UNG, Slovenja

ANNO ACCADEMICO 2023/2024

Student:

Aymane El Fellah

Matriculation N*: 2040341

The genome that we decipher in this generation is but a snapshot of an ever-changing document. There is no definitive edition.

Matt ridley (1988), Genome Autobiography: A Personal View of Scientific Discovery, p. 61

Acknowledgements

I would like to express my deepest gratitude to all those who supported and guided me throughout the journey of completing my Master's thesis.

First and foremost, I extend my heartfelt thanks to the University of Padova for providing me with an exceptional academic environment and the resources necessary to pursue my studies. I am profoundly grateful to my supervisor, Professor Dorianna Sandonà, from the Department of Biomedical Sciences - University of Padova whose invaluable guidance, insights, and encouragement have been precious. I also wish to thank Professor Stefano Salmaso and Professor Patrizia Polverino, the heads of the program, for their support and leadership throughout my academic journey.

I am sincerely thankful to the University of Nova Gorica - Slovenia for hosting me as part of the Erasmus+ grant. Special thanks go to my co-supervisor, Dr. Matteo De March, whose rigorous assistance, immense help, and kindness both in and out of the lab have been crucial to my success. His dedication and mentorship have greatly enriched my academic and personal growth.

Additionally, I would like to express my gratitude to Professor Barbara Gatto for her participation in my thesis defense and for being an inspiring guide in various subjects during my coursework. Her support and encouragement have been greatly appreciated.

I am also deeply thankful to my family members, my parents and siblings, whose unwavering support from day one has been my greatest strength. Their constant encouragement and belief in me have been a source of immense motivation. I also extend my gratitude to my relatives and friends for their continuous support and understanding throughout this journey.

To all those mentioned and many others who have contributed in various ways, I offer my sincerest thanks. This thesis would not have been possible without your support and encouragement.

Thank you.

CONTENTS

List of Figures

List of Tables

Abbreviations and symbols

DDT - DNA Damage Tolerance

DDR - DNA Damage Response

RFR - Replication Fork Reversal

ZRANB3 - Zinc Finger RANBP2-Type Containing 3

HLTF - Helicase-Like Transcription Factor

SWI/SNF - Switch/ Sucrose Non-Fermentable Protein

ATP - Adenosine Triphosphate

DNA - Deoxyribonucleic Acid

RNA - Ribonucleic Acid

dsDNA - Double stranded DNA

ssDNA – Single stranded DNA

PCR - Polymerase Chain Reaction

SDS-PAGE - Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

EDTA - Ethylenediaminetetraacetic Acid

BSA - Bovine Serum Albumin

BLAST - Basic Local Alignment Search Tool

FASTA - Fast-All

INDEL - Insertion/Deletion

PDB - Protein Data Bank

NCBI - National Center for Biotechnology Information

EMBL - European Molecular Biology Laboratory

GO - Gene Ontology

APS - Ammonium Persulfate

Bis - N,N'-Methylenebisacrylamide

TAE - Tris-Acetate-EDTA

DTT - Dithiothreitol

IMAC - Immobilized Metal Affinity Chromatography

ABSTRACT

To efficiently duplicate their genomic content, cells must overcome DNA lesions that interfere with processive DNA replication. These lesions may be tolerated and, for instance, bypassed rather than directly removed or repaired, to allow the continuity of DNA replication on template DNA.

One of the main roles of ZRANB3 is its involvement in DNA damage response, particularly DNA Damage Tolerance (DDT). Upon a replication stress, such as a DNA lesion and consequently the fork stalling, ZRANB3 interacts with the Proliferating Cell Nuclear Antigen (PCNA) regulating the recruitment of DNA repair factors to the damage site (Weston, Peeters, et Ahel 2012). Indeed, it has been suggested that ZRANB3 functions as a scaffold protein, coordinating the assembly of protein complexes, such as HLTF, the E3-ubiquitin ligases and the DNA mismatch repair protein MSH6.

For this reason, ZRANB3 is mainly considered an oncosuppressor in cancer, for instance breast cancer (BC) and colorectal cancer (CRC). Decreased ZRANB3 expression has been associated with poor prognosis in cancer patients, suggesting also its potential as prognostic biomarker for this disease.

The current knowledge on the structural and functional relationship at the basis of ZRANB3 function is limited. Therefore, the need of an immediate in-deep characterization and a fulfilled understanding of the precise role of ZRANB3. in the context of genome stability is evident.

SUMMARY

The Zinc finger RAN-binding domain-containing protein 3 (ZRANB3) is a multifunctional protein encoded by the ZRANB3 gene in humans. It plays a crucial role in DNA repair and replication stress response, making it a significant target in diseases with impaired DNA repair mechanisms, such as cancer. ZRANB3 functions by processing DNA intermediates that arise during replication stress, and thus maintaining genome stability. Despite its importance, the structural basis of ZRANB3 and its action are yet unknown.

The aim of this study was to conduct a preliminary analysis on ZRANB3 through the use of bioinformatic tools and experimental techniques, with the primary goal to produce a suitable amount of protein for structural studies, in particular X-ray crystallography and Cryo-Electron Microscopy (Cryo-EM).

We performed sequence, sequence-structure and structure alignments using Clustal Omega, PROMALS3D and DALI light, respectively, and identified the putative ZRANB3 DNA binding residues. AlphaFold was utilized to predict the 3D structures of the full-length wild-type (FL WT) protein, which revealed key structural features. In particular, our attention focused on the C-terminal Substrate Recognition Domain (SRDct) and the ATPase motor domain (RecA-like 1/2), which are necessarily for the protein function. Indeed, the latter was used as substrate for docking a dsDNA molecule based on the information described above.

Experimentally, the FL WT ZRANB3 gene along with SRDct and RECA1A2 domains were cloned into the pET28+ vector and initially tested for expression in small-scale in BL21 cells. Further, attempts to optimize the protein expression in large scale resulted in the formation of inclusion bodies and/or insoluble aggregates. Finally, we cloned the FL WT ZRANB3 in yeast P. *pastoris* aimed at Cryo-EM structural investigation.

These preliminary experimental results serve as groundwork for optimization of the large-scale protein preparation for structural studies, which potentially can aid to the development of therapeutic strategies targeting cancer tissues.

1. INTRODUCTION

1.1. The cell cycle

Eukaryotic cell cycle is composed by highly regulated series of events that leads to cell division and duplication. It consists of four main phases (Figure 1): G1 (Gap 1), S (Synthesis), G2 (Gap 2), and M (Mitosis). During the G1 phase, the cell grows and synthesizes proteins necessary for DNA replication. The S phase involves the replication of DNA, doubling the cell's genetic material. In the G2 phase, the cell continues to grow ensuring all DNA is correctly replicated and undamaged, while preparing itself for phase M. This encompasses both Mitosis, where the cell's nucleus divides, and Cytokinesis, where the cytoplasm divides, resulting in two identical daughter cells. The cell cycle is controlled by checkpoints and various proteins, including cyclin-dependent kinases (CDKs) and cyclins, which ensure that each phase is completed accurately before the next phase begins, preventing errors that could lead to cell death or cancer.

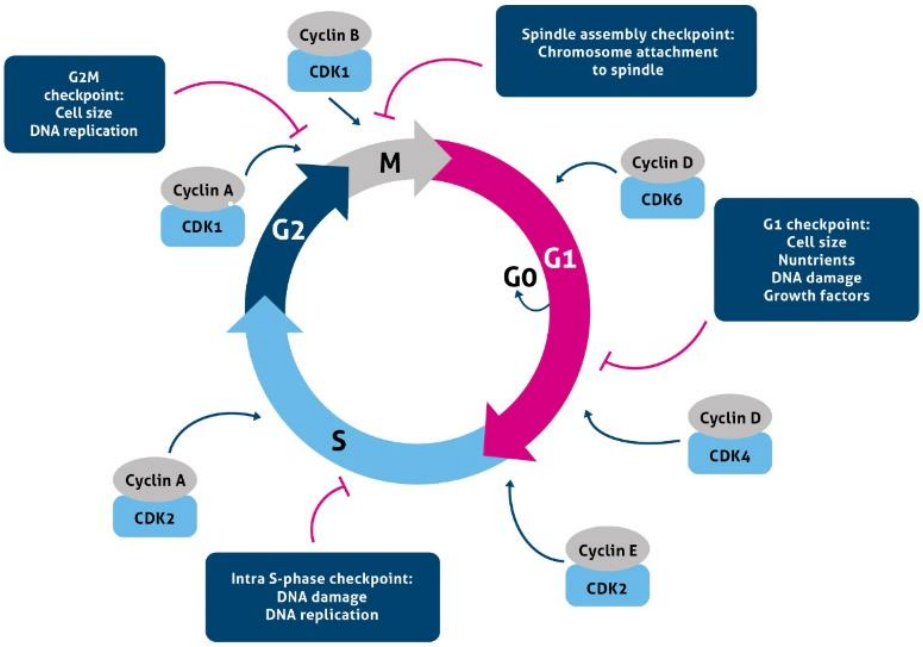


Figure 1. Cell cycle and checkpoints Control

1.2. Eukaryotic DNA Replication

DNA replication in Eukaryotes starts form multiple origins of replication (ORCs) and occurs in distinct steps involving several key players. Helicases unwind the double-stranded DNA

(dsDNA), Single-Strand Binding Proteins (SSBs) stabilize the unwound DNA, DNA polymerases add nucleotides to the growing DNA strand, the primase synthesizes short RNA primers to provide starting points for DNA synthesis and the ligase, which role is to seal the gaps between DNA fragments.

The entire process of DNA replication is composed by three steps.

Initiation. The initiation of DNA replication is facilitated by the helicase complex, which includes Cdc45, MCM, and GINS (CMG) that binds to the origin recognition complex (ORC) and unwinds the DNA at the replication fork. MCM consists of six subunits: MCM2, MCM3, MCM4, MCM5, MCM6, and MCM7, which form a ring-like structure that encircles the DNA and utilizes ATP hydrolysis to move along the DNA and separate the two strands, providing the core helicase activity. Cdc45 is a single protein that interacts with both the MCM complex and the GINS complex, playing a critical role in the recruitment and activation of the MCM helicase at the replication origin. It stabilizes the CMG complex and enhances the helicase activity. The GINS complex is a tetramer composed of Sld5, Psf1, Psf2, and Psf3, forming a stable scaffold that ensures proper assembly and function of the CMG helicase by maintaining its integrity and promoting efficient DNA unwinding. As the CMG complex unwinds the DNA, the SSBs bind to the exposed single strands to prevent them from re-annealing and to protect them from nucleases, ensuring that the replication machinery can synthesize new DNA strands efficiently. Together, these components orchestrate the intricate process of DNA unwinding and stabilization, allowing replication to proceed smoothly (Aparicio, T et al 2006).

Elongation. On the leading strand (3’ to 5’), the primase synthesizes a short RNA primer with about 10-12 nucleotides that serve as a starting point for DNA polymerase α, which extends the primer by adding a short stretch of DNA nucleotides. DNA polymerase δ then takes over to add DNA nucleotides continuously in the 5' to 3' direction. However, DNA polymerase δ cannot synthesize DNA in the 3' to 5' direction, implying that the lagging strand (5’ to 3’) being synthesized discontinuously in short Okazaki fragments. DNA polymerase δ adds nucleotides to the RNA primers on the lagging strand, creating these fragments, and DNA ligase joins them to form a continuous strand. The elongation depends also on the Proliferating Cell Nuclear Antigen (PCNA) protein**,** a homotrimeric ring-shaped protein that acts as a sliding clamp, enhancing the processivity of DNA polymerases by encircling the DNA and holding the polymerase in place. PCNA is leaded onto the DNA by the multi-subunit complex replication factor C (RFC), which

recognizes the primer-template junctions and uses ATP hydrolysis to open and close the sliding ring, thus ensuring its proper assembly on the DNA.

Termination. Once the entire DNA molecule has been replicated, RNA primers are removed by RNase H and replaced with DNA nucleotides. DNA2 nuclease and FEN1 (Flap Endonuclease 1) play roles in processing and removing RNA-DNA hybrids and flap structures during primer removal. DNA ligase then seals any remaining gaps between Okazaki fragments. The replication process concludes with proofreading by DNA polymerase δ and ε, which possess 3' to 5' exonuclease activity to ensure the accuracy of the newly synthesized DNA (Alberts, B. et al 2002).

1.3. Mechanisms to overcome DNA lesion

Despite its sophisticated regulation, the DNA replication process can undergo to errors, such as incorrect or missed bases, and/or mismatches. External factors like UV radiation or chemicals can also cause DNA lesions, disrupting the cell cycle. To maintain genomic integrity and prevent major diseases like cancer, cells may employ DNA Damage Tolerance (DDT) and/or DNA Damage Repair (DDR) mechanisms. The first allows cells to bypass DNA lesions during replication without repairing them, property that belongs to the second, ensuring replication can continue. DDT mechanisms (Figure 2) include Translesion Synthesis (TLS), Template Switching (TS), and Replication Fork Reversal (RFR).

TLS involves specialized low-fidelity DNA polymerases, such as polymerase η, ι, κ, and Rev1, which replicate across damaged sites with lower fidelity compared to high-fidelity replicative polymerases like DNA polymerase δ and ε (Leung, W et al 2019). These TLS polymerases are recruited to the site of the lesion through the mono-ubiquitination of the clamp protein PCNA, which plays a critical role in regulating DDT by ensuring that the appropriate polymerase is used for lesion bypass.

TS is an error-free process that uses the undamaged sister chromatid as a template for bypassing lesions. This mechanism involves proteins like Rad51 and Rad52, which facilitate strand invasion and synthesis using the undamaged template, ensuring accurate replication despite the presence of lesions.

RFR involves the formation of a chicken foot-like DNA structure, where the replication fork reverses, allowing the repair machinery to access the damaged DNA and restore a functional fork.

Both TS and RFR are initiated through the polyubiquitination of PCNA, which switches the template from the damaged strand to the newly synthesized complementary strand on the homologous sister chromatid.

Proteins such as SMARCAL1, ZRANB3, and HLTF play crucial roles in these three processes, with ZRANB3 particularly involved in fork reversal and remodeling, ensuring that replication can proceed despite the presence of DNA damage.

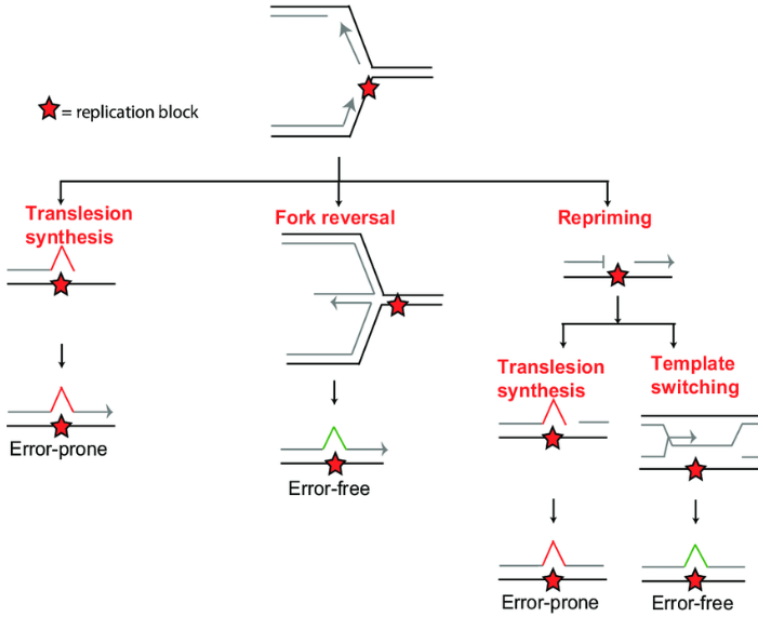


Figure 2. DDT pathways for tolerating a DNA damage or the replication fork stalling

DDR pathways directly repair DNA damage and include the Nucleotide Excision Repair (NER), specific for bulky lesions such as those caused by UV light, the Base Excision Repair (BER) that fixes small non-helix-distorting base lesions, the Mismatch Repair (MMR) that corrects small insertions or deletions during DNA replication by recognizing and fixing mismatched bases. Homologous Recombination (HR) repairs double-strand breaks (DSBs) using a homologous sequence as a template, making it highly accurate and it is mostly active during the S and G2 phases of the cell cycle. Non-Homologous End Joining (NHEJ) repairs DSBs by directly ligating the broken DNA ends without a homologous template, making it quicker but more error-prone. These pathways are essential for maintaining genomic stability and preventing mutations that could lead to diseases such as cancer.

1.4. ZRANB3 and its role

ZRANB3 belongs to the SNF2 family of ATP-dependent chromatin remodelers, which are more generally involved in gene transcription, DNA replication, and repair. These proteins manipulate protein-DNA interactions, altering the DNA structure and facilitating various cellular processes (Poole and Cortez, 2017). Similar to ZRANB3, SMARCAL1 and HLTF play crucial roles in these processes by promoting the stability and integrity of the replication fork and aiding in the repair of damaged DNA. However, unlike SMARCAL1 and HLTF, ZRANB3 only functions as endonuclease, disrupting D-loop formations and preventing erroneous homologous recombination events. The presence of multiple structural motifs, two ATPase motor domains (RecA-like 1 and RecA-like 2), the PCNA-interacting protein (PIP-box) and AlkB homolog 2 PCNA-interacting (APIM) domains, the substrate recognition domain (SRD) and the NPL4 zinc finger domain (NZF) allow it to bind to ubiquitinated PCNA and localize to damaged replication forks. ZRANB3 can then promote error-free lesion bypass (Lina Cipolla et al.2016) or fork regression activity, stabilizing the replication forks and reducing the accumulation of single-stranded DNA. More specifically, ZRANB3's ATP-dependent chromatin remodeling activity facilitates the recruitment of specialized polymerases, such as polymerase η, ι, κ, and Rev1, capable of bypassing DNA lesions and restarting replication. Its endonuclease activity induces a DNA break in the double-stranded region of the replication fork 2nt from the branching point. Cleavage by ZRANB3 exposes a free 39-OH group, which can be extended by DNA polymerase to remove the replication-blocking DNA lesion. This leads to the formation of the 59 overhanging DNA flap, which can be processed by the activity of FEN1. Following nick sealing and reversal of regressed forks, DNA replication resumes on the undamaged DNA template (Weston et al. 2012). The specific action pf ZRANB3 in DDT is reported below.

In TLS. This process begins when PCNA is monoubiquitinated by the E3 ubiquitin ligase RAD18 at the K164 residue through the activation of HLTF. This modification signals the recruitment of Pol η for bypassing the lesion. The steps are: 1) DNA damage is detected, 2) RAD18 monoubiquitinates PCNA at K164, 3) Monoubiquitinated PCNA recruits ZRANB3, 4) ZRANB3 uses its chromatin remodeling activity to facilitate the recruitment of TLS polymerases, 5) TLS polymerases bypass the DNA lesions, allowing replication to continue, 6) HLTF assists in the process by unwinding DNA and aiding in ubiquitination.

In TS and HR. This process relies on mediator proteins such as RAD6/18 and the UBC13, MMS2 ubiquitin ligases (Figure 3A), as well as on HLTF that in turn ubiquitinate PCNA (Vujanovic et al. 2017). ZRANB3 remodels chromatin to facilitate the use of the sister chromatid as a template for accurate DNA synthesis. Its endonuclease activity helps disrupt D-loop formations, preventing erroneous recombination events. The steps are: 1) DNA damage is detected during replication, 2) HLTF or SHPRH polyubiquitinates PCNA at K164, 3) Polyubiquitinated PCNA recruits ZRANB3 (Figure 3B), 4) ZRANB3 remodels the chromatin, allowing the sister chromatid to be used as a template, 5) ZRANB3's endonuclease activity disrupts D-loops to prevent erroneous recombination, 6) HLTF facilitates the process by maintaining the polyubiquitinated state of PCNA and aiding in DNA unwinding.

In RFR. ZRANB3 actively participates in remodeling the replication fork into a four-way holiday junction, a critical structure that allows the cell to manage and repair DNA damage encountered during replication. ZRANB3's ATPase activity provides the necessary energy to remodel the fork, while its endonuclease activity helps to resolve any secondary structures that may impede the process. This dual functionality ensures that the replication fork can be reversed and properly restored after the damage is repaired, thus maintaining the integrity and progression of the replication process. ZRANB3 interacts with polyubiquitinated PCNA through its PIP-box and APIM, while the NZF motif recognizes and binds to polyubiquitinated PCNA. This interaction stimulates the regression of replication forks, stabilizing them and minimizing the production of single-stranded DNA. By doing so, ZRANB3 helps prevent inappropriate homologous recombination, which could lead to genomic instability (Lina Cipolla et al.2016). The specific steps are: 1) DNA damage is encountered during replication, 2) ZRANB3 binds to polyubiquitinated PCNA via PIP-box and APIM motifs, 3) ZRANB3 uses its ATPase activity to remodel the replication fork into a Holliday junction, 4) The endonuclease activity of ZRANB3 resolves any secondary structures, 5) The replication fork is reversed and restored after repair, maintaining genomic stability.

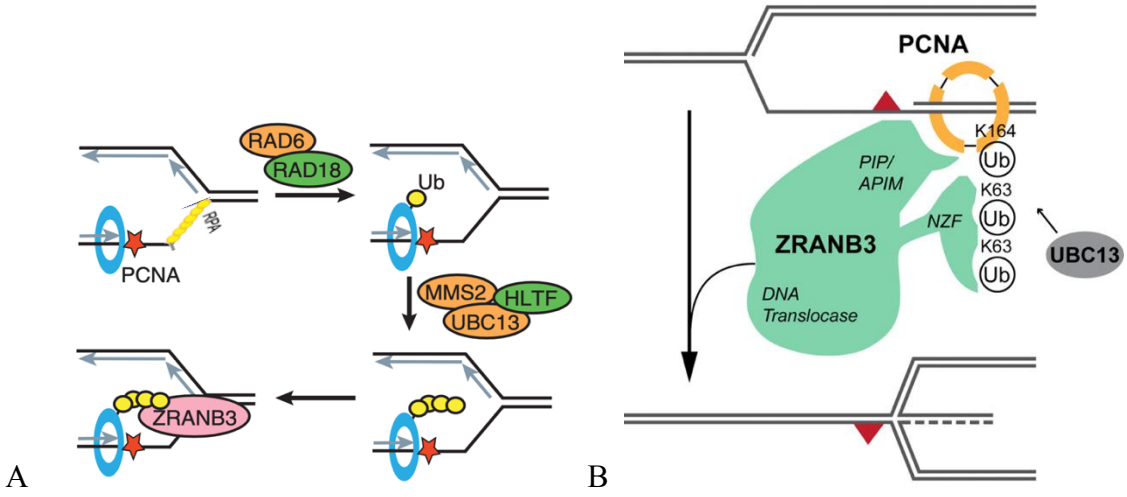


Figure 3. A. ZRANB3 recruitment to the lesion site and its involvement during TS and RFR. B. ZRANB3 interaction with polyubiquitinated PCNA at the replication Fork.

1.5. ZRANB3 in Cancer

For its crucial role in DNA response ZRANB3 is considered oncosuppressor. This mainly happens because of the partial or total loss of its activity due to the hypermethylation of its promoter region. In mouse models of Myc (the proto-oncogene basic helix-loop-helix transcription factor)-driven B-cell lymphomagenesis, ZRANB3 deficiency led to a 30% reduction in median survival time (Matthew V. Puccetti et al. 2019). Indeed, ZRANB3 function can be linked to that of MYC, well known for inducing high levels of DNA replication stress in cancer cells. The upregulation of MYC leads to an increase in ZRANB3 expression to cope with the heightened replication stress and maintain genomic stability. In the absence of ZRANB3, MYC-overexpressing cells exhibit increased DNA damage, replication fork collapse, and genomic instability, highlighting the critical role of ZRANB3 in responding to oncogenic stress (Matthew V. Puccetti et al. 2019). Despite this, ZRANB3 activity can be regulated i) the change of mRNA amount, ii) presence of structural variations on the chromosome region encoding for its gene and/or iii) presence of CNA (Copy Number Alterations), a term that refers to the change in the number of *ZRANB3* copies.

In breast cancer, approximately 60% of tumor samples showed significantly elevated ZRANB3 expression compared to normal tissues, correlating with more aggressive tumor phenotypes and poorer overall survival rates (Protein Atlas). Similarly, in ovarian cancer, about 55% of samples exhibited increased ZRANB3 levels, which are associated with reduced patient survival, suggesting its role as a prognostic marker. Figure 4 illustrates the frequency and types of ZRANB3 gene

alterations across various cancer types. The Y-axis represents the percentage of samples with alterations in ZRANB3, while the X-axis lists different cancer types, such as esophagogastric adenocarcinoma, bladder urothelial carcinoma, and uterine endometrial carcinoma, among others. The alterations are categorized into mutations (green), structural variants (red), amplifications (blue), deep deletions (purple), and multiple alterations (grey).

Esophagogastric adenocarcinoma exhibits the highest alteration frequency at approximately 12%, primarily due to amplifications and mutations. Bladder urothelial carcinoma and uterine endometrial carcinoma follow, with alteration frequencies between 6-10%, also dominated by amplifications and mutations. Colorectal adenocarcinoma shows similar trends, with notable occurrences of mutations and structural variants. Less frequent but still significant alterations are observed in cancers such as lung squamous cell carcinoma, head and neck squamous cell carcinoma, and breast invasive ductal carcinoma. These cancers show a mix of mutations, amplifications, and structural variants, albeit at lower frequencies compared to the top cancer types.

Overall, Figure 5 underscores the diverse and significant alterations in the ZRANB3 gene across various cancers, highlighting its potential role in cancer progression and its viability as a therapeutic target. Nevertheless, these data also underline the need of further research at this level to characterize specifically these *ZRANB3* alterations and their correlation with cancer. Targeting ZRANB3 could enhance the efficacy of existing treatments and improve patient outcomes.

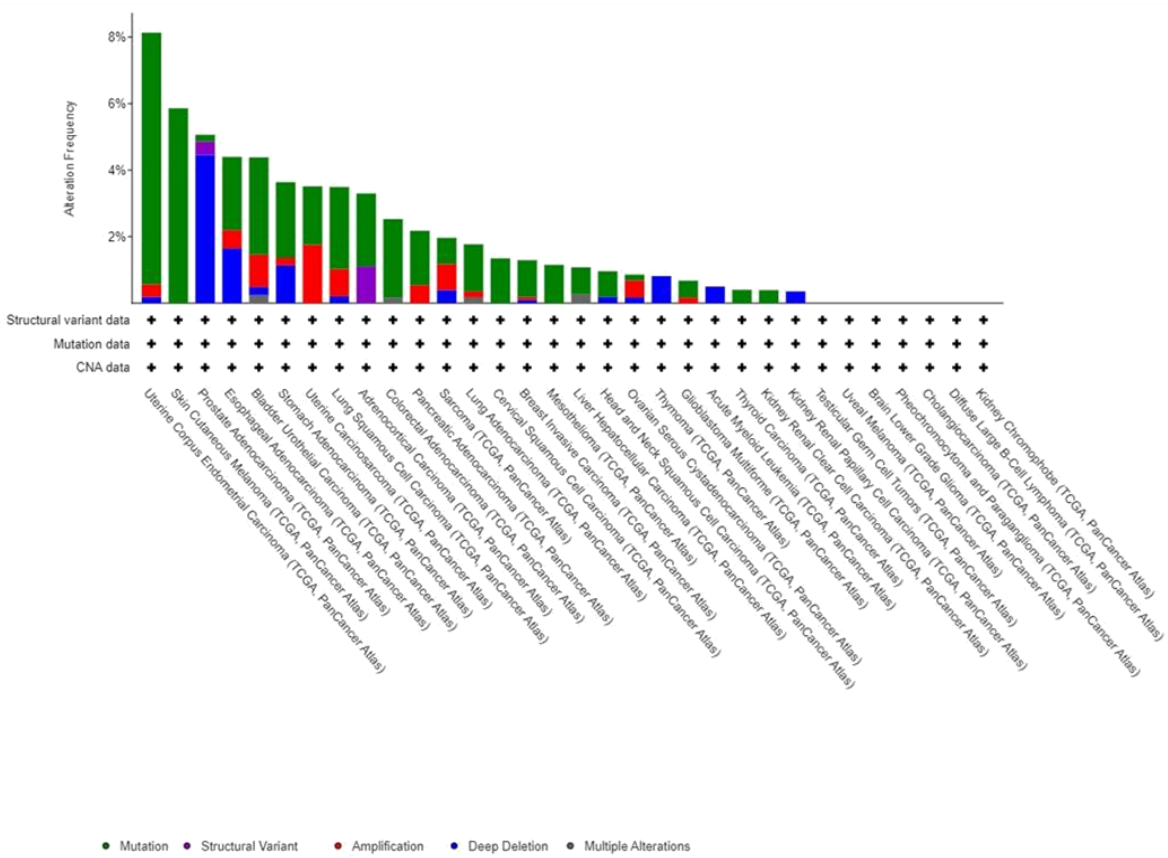


Figure 4. Mutation Frequency and Types in Various Cancer Studies

2. MATERIALS & METHODS

PART I: Bioinformatic analysis

Bioinformatics can provide the computational tools and techniques for analysis of complex biological data. This section outlines the bioinformatic methodologies employed to analyze the sequence and structure of critical ZRANB3 domains.

2.1. Sequence alignment

It allows the understanding of the evolutionary relationships and functional similarities among proteins. We used Clustal Omega (http://www.clustal.org/) and MAFFT (https://mafft.cbrc.jp/) software. The first provides a quick and efficient initial alignment, while the second enhances the alignment's accuracy through iterative refinement. The final alignments were visualized using both Jalview and ESPript (https://espript.ibcp.fr/) software for better interpretation. The following parameters were used:

- Scoring matrix for amino acid sequences: BLOSUM62
- Scoring matrix for nucleotide sequences: 1PAM / κ=2 used when aligning closely related DNA sequences.
- Gap opening penalty: 1.0 – 5.0
- Offset value: 0.0 – 1.0

2.2. Sequence-structure alignment

It allows the characterization of functional and evolutionary relationships among protein sequences based on their three-dimensional structures. One of the most powerful software tools for this purpose is PROMALS3D (http://prodata.swmed.edu/promals3d/), which specifically aligns multiple protein sequences by integrating both sequence information and structural data. The PROMALS3D advanced algorithms utilize probabilistic models and profile-profile comparisons. By incorporating the structural information of proteins, PROMALS3D can detect and align conserved regions that may not be obvious from sequence data alone, thereby improving the reliability of the alignments.

PROMALS3D adjusts alignment parameters to optimize performance, such as the structure-based gap penalties (Figure 5). This ensures that the alignments reflect both evolutionary relationships and functional similarities. Additionally, the software can handle diverse and complex protein

families by leveraging both local and global structural features, making it particularly useful for studying proteins with low sequence similarity but conserved structural motifs. The settings of PROMALS3D were adjusted to achieve a good balance between speed, accuracy, and the use of both sequence and structural information, ensuring high-quality multiple sequence alignments that are crucial for subsequent analyses (Figure 5).

```
Alignment parameters:

        Identity threshold above which fast alignment is applied: 0.6
        Weight for constraints derived from sequences: 1
        Weight for constraints derived from homologs with structures: 1.5
        Weight for constraints derived from input structures: 1.5

Parameters for profile-profile comparison:
        Weight for amino acid scores: 0.8
        Weight for predicted secondary structure scores: 0.2

Parameters for deriving sequence profiles from PSI-BLAST searches:
        PSI-BLAST iteration number: 3
        PSI-BLAST e-value inclusion threshold: 0.001
        Identity cutoff below which distant homologs are removed: 0.25
        Maximum number of homologs kept for PSI-BLAST alignment: 300

Parameters for detecting and using homologs with 3D structures (homolog3d):
        PSI-BLAST e-value cutoff against structural database: 0.001
        Identity cutoff below which 3D structures are not used: 0.2
        Align homologs with 3D structures by programs: dali - not used; fast - used; tmalign - used
        Realign target-homolog3d using profile-profile alignment: yes

Parameters for pairwise alignments between input 3D structures:
        Align input structures by programs: dali - not used; fast - used; tmalign - used

Parameters for aligning sequences within groups in the first alignment stage:
        Align sequences within groups in the first alignment stage by: mafft
```

Figure 5. PROMALS3D alignment parameters

2.3. Identification of structural homologues

The alpha-fold model of ZRANB3 was used as a query in the DALI server (http://ekhidna2.biocenter.helsinki.fi/dali/) to find for structural homologs in the Protein Data Bank (PDB) (https://www.rcsb.org/). DALI calculates the distance matrices that represent the spatial distribution of amino acids and compares these with those calculated from the similar PDB structures. The Z-score in DALI reflects the statistical significance of the similarity, thus higher Z-scores indicating stronger similarity. Indeed, the selection of the output homologues is based on the Z-scores, the lower RMSD value (indicating better alignment), and the higher identity match (%) on the aligned regions, alongside the length of the alignment (nres). The output structures were further overlapped to the ZRANB3 model in the PyMOL software (https://pymol.org/).

2.4. Homology and ab-initio modelling

Homology Modeling

It generates structural models based on homologous proteins with known structures. To model the structure of ZRANB3, we used first SWISS-MODEL (https://swissmodel.expasy.org/) and identified homologous templates from the Protein Data Bank (PDB); then we aligned the ZRANB3 protein to these templates. Finally, I-TASSER (https://zhanggroup.org/I-TASSER/) was employed as threading algorithm to identify templates and align sequences followed by iterative fragment assembly and ab initio modeling for regions without reliable templates. The final models were refined through iterative simulations, scored, and clustered to select the most reliable ones. Indeed, SWISS-MODEL provides a straightforward approach based on sequence-template alignment, while I-TASSER offers an iterative and more detailed modeling process, combining threading, ab initio methods, and structural assembly. Hereafter, a summary of the details and options used for both programs:

SWISS-MODEL. Input: The ZRANB3 sequence. Search: target structural homologues in the Protein Data Bank (PDB) that have a high sequence similarity. Selection: templates with the highest sequence identity (>30%) and coverage (>70%) are selected.

I-TASSER. Input**:** ZRANB3 sequence. Threading**:** I-TASSER uses multiple advanced threading algorithms, such as MUSTER, SP3, SPARKS-X, and PROSPECT2, to scan the Protein Data Bank (PDB) for potential template structures. These algorithms align the ZRANB3 sequence against the sequences of known protein structures in the PDB, searching for regions of significantly conserved structural motifs. The threading process evaluates the fitness of each alignment based on both sequence homology and structural compatibility, identifying templates that provide the most reliable structural framework for modeling ZRANB3.

Ab Initio Modeling with AlphaFold

AlphaFold (https://alphafold.ebi.ac.uk/) is the most powerful software for ab-initio modeling of protein structures. It is based on artificial intelligence and machine learning algorithms, which apply principles of physics and knowledge-based potentials to generate 3D folds. AlphaFold can be used as integrated in ChimeraX (https://www.rbvi.ucsf.edu/chimerax/) via ColabFold2 tramlined channel, thereby leveraging Google's GPU resources and, thus, allowing for efficient and rapid protein structure predictions timescale and accuracy. Here below the step-by-step process used in this study for ZRANB3:

1. Input: the ZRANB3 sequence in *fasta* format; pmax iterations 3; model type "monomer" for single protein chain rather than for complex or multimers.
2. Run: execute the cells in the Colab notebook to predict the structure of ZRANB3. The notebook uses Google's TPU/GPU resources for efficient computation.
3. Analysis of the results: The plot illustrates the predicted local distance difference test (pLDDT) scores, with the regions in blue indicating high confidence while in red low confidence in the predicted protein structure. This color gradient helps identify regions of the protein that are modeled with varying levels of accuracy, with blue areas being more reliable.

2.5. Putative DNA binding sites analysis

BioLiP2 (https://zhanggroup.org/BioLiP/), a comprehensive database for protein-ligand binding interactions, was used to predict possible ZRANB3-DNA binding sites using the sequence of the RecA-like 1/2 domain as input. Indeed, the software predicts potential binding sites by leveraging known protein-ligand interaction data to identify similar binding motifs in the input sequence. The predicted binding sites from BioLiP2 were then manually mapped onto the target structure identified by the DALI search and the use of PyMOL. The conservation of each site as well as any substitutions were then tabled.

2.6. Molecular Docking

The software BIOVA Discovery Studio (DCS) (https://discover.3ds.com/) and HDock (http://hdock.phys.hust.edu.cn/) were used as follow:

- Identification of the binding sites: this can be done using the "Define and Edit Binding Site" tool, which helps to locate and visualize the binding pockets.
- Generation of the spheres: the "Receptor-Ligand Interactions" tool generates SBD (Sphere-Based Docking) spheres around the identified binding sites. These spheres represent the potential interaction hotspots where the ligand is likely to bind.
- Loading of protein and ligand file in HDock: the prepared protein structure file from Discovery Studio is imported. The binding sites should be correctly imported and visible.
- Loading of the ligand or substrate molecule (PDF format)
- Configuring of the parameters: Set up search algorithm and scoring functions.
- Run

- Examination of the binding modes: look at the binding affinity scores and interaction details such as hydrogen bonds, hydrophobic interactions, and electrostatic interactions.
- Visualization of all interactions: visualization tools within HDock to view and analyze the docking poses, focusing on the interaction between the ligand and the binding site residues.
- Analysis of the results: the best docking model was selected based on the highest docking score (most negative), the highest confidence score, and the most favorable interaction details. These are defined as follow:

Docking Score represents the predicted binding affinity between the protein and the ligand. A lower value indicates stronger binding affinity. A model with a significantly negative docking score suggests a stable and favorable interaction between the protein and ligand.

Confidence Score reflects the reliability of the predicted binding mode. Higher confidence scores indicate that the predicted pose is more likely to be accurate. It is based on the consistency and reproducibility of the docking results.

Ligand RMSD (Root Mean Square Deviation) measures the average distance between the atoms of the docked ligand and a reference structure. Lower RMSD values indicate that the docked pose closely matches the expected binding mode, suggesting a more precise and reliable docking result.

Interaction Residues at the interface between the protein and the ligand are examined. The presence of critical interactions such as hydrogen bonds, hydrophobic contacts, and electrostatic interactions are crucial for binding. By analyzing these interactions, we can determine the functional relevance of the docking pose.

2.7. PPI network analysis

The analysis of the protein-protein interaction (PPI) network for ZRANB3 was conducted using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (https://string-db.org/), which provides a comprehensive interaction map based on various sources, including experimental data, databases, text mining, and computational predictions. This provided insights into the functional context and biological processes associated with ZRANB3 interactions. The search used as input the ZRANB3 query name and was run in a flowchart as follows:

- the confidence score threshold was set to 0.7 indicating that only interactions with strong supporting evidence are considered, focusing on high-confidence interactions to reduce noise.
- the network visualization includes 30 nodes (proteins) representing ZRANB3 and its

interacting partners.

- 112 edges (interactions) were shown, representing both direct and indirect interactions between these proteins.
- the functional enrichment analysis allowed to identify several significant partners and functional pathways, including DDT and DDR.

PART II: Protein production

The pET28+_ZRANB3, containing kanamycin resistance and in frame with the 6xHis tag at the N-terminus, was ordered from TWIST Bioscience (https://www.twistbioscience.com/), thus the FL WT ZRANB3 gene was already cloned between the EcoRI and XhoI restriction sites.

2.8. Gibson cloning

Gibson Assembly (Gibson, D.G et al 2009) is a method that allows for the seamless joining of multiple DNA fragments in a single, isothermal reaction. It relies on the use of overlapping sequences at the ends of each DNA fragment, which are then processed by an exonuclease to create single-stranded overhangs that can be annealed to complementary sequences. A DNA polymerase fills in the gaps, while the DNA ligase seals the nicks, resulting in a fully assembled DNA construct.

We used specific primers (Table 1) - designed with the software Snapgene (Figure 6) (https://www.snapgene.com/) - and temperatures for the amplification (PCR – *see section 2.8.1*) of SRD-ct (aa 721-1079), RecA1/A2 (aa 38-472) domains and the FL WT ZRANB3 for yeast plasmid (Table 2).

Constructs	Primers
SRD-ct	F1 vector pET28: 5’ GATCCGGCTGCTAACAAAGCCCG 3’ R1 vectort pET28: 5’ GCCGGGGCCCTGGAACAG 3’ F2 SRDct: 5’GGAGGTGCTGTTCCAGGGCCCCGGCGGACTTACTTCTCAACCTGGAAATGAG 3’ R2 SRDct: 5’ GGGCTTTGTTAGCAGCCGGATCTCATTACTTTTTTACCAGAAAGCGCGTAATGT 3’

Rec A1/A2	F1 vector pET28: 5’ GATCCGGCTGCTAACAAAGCCCG 3’ R1 vectort pET28: 5’ GCCGGGGCCCTGGAACAG 3’ Rec A1 Forward: 5’GGAGGTGCTGTTCCAGGGCCCCGGCGATCGCTTACGCGCTAAACTT 3’ Rec A2 Reverse: 5’ GGGCTTTGTTAGCAGCCGGATCTCATTAAGCGCCCATTCAGTGTGCTA 3’
FL WT ZRANB3 in Yeast.	F1.yeast vector: 5' GGACATCATCATCATCATTGAGTTTGTAGC 3' R1.yeast vector: 5' AGCTTCAGCCTCTCTTTTCTCGAGAGATA 3' ZRANB3 Forward : 5' TCTCGAGAAAAGAGAGGCTGAAGCTCCGCGTGTACACAACATTAAAAAGTCCTTG 3' ZRANB3 Reverse: 5'CGACATTACGCGCTTTCTGGTAAAAAAGGGACATCATCATCATCATCATTGAG 3'

Table 1. Primers designed for cloning.

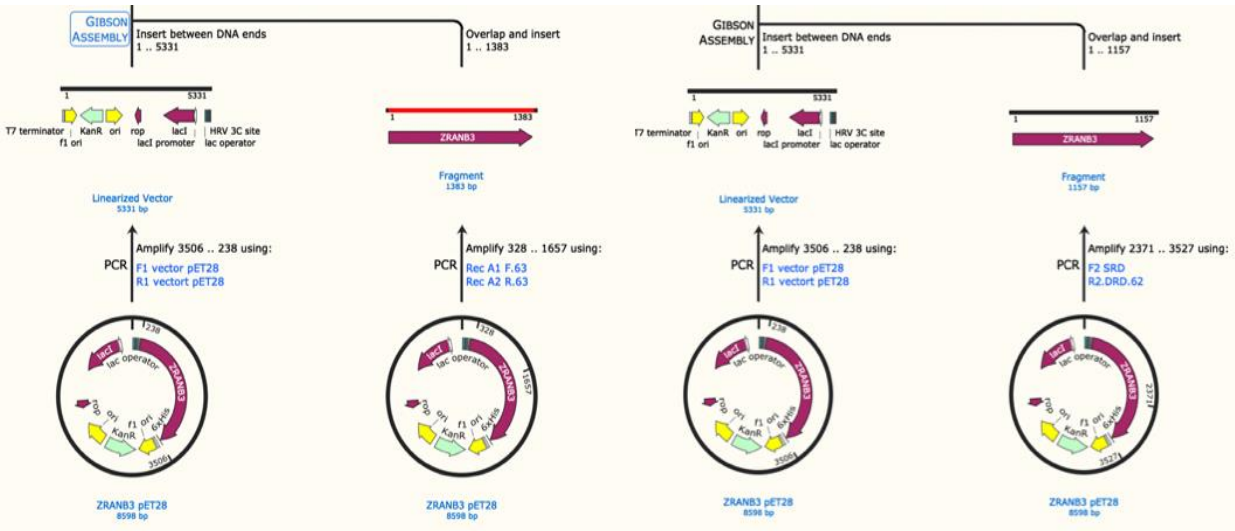


Figure 6. Snapgene session of the Gibson assembly for SRDct and RecA1/A2

PCR amplification

The Polymerase Chain Reaction (PCR) amplifies a specific region of a DNA strand (the DNA target). Most PCR methods amplify DNA fragments of between 0.1 and 10 kilo base pairs (kbp) in length, although some techniques allow for amplification of higher fragments. A basic PCR set-up requires several components and reagents, including: 1) a DNA template that contains the DNA target region to amplify, 2) the Taq DNA polymerase, that remains intact during the high-

temperature DNA denaturation process, 3) two DNA primers that are complementary to the 3'-ends of each of the sense and anti-sense strands of the DNA target, 4) specific primers that are complementary to the DNA target region are selected beforehand, 5) the dNTPs from which the DNA polymerase synthesizes a new DNA strand, 6) the buffer solution providing a suitable chemical environment for optimum activity and stability of the DNA polymerase, and 7) the bivalent cations, typically magnesium Mg^{2+} or Mn^{2+} concentration (Sambrook, J., & Russell, D. W. 2001).

The protocol specifically used was the following:

1. Prepare a reaction mix for each PCR reaction
   - DNA template: [1] µL (typically 10-100 ng of template DNA)
   - Forward primer: [0,75] µL (typically 0.2-1 µM final concentration)
   - Reverse primer: [0,75] µL (typically 0.2-1 µM final concentration)
   - RepliQa mix: [12,5] µL
   - Sterile distilled water: [10] µL (to achieve the final reaction volume, e.g., 25 µL)
2. Mix the components gently by pipetting up and down to ensure thorough mixing
3. Dispense the reaction mix into PCR tubes or plates, taking care to prevent contamination
4. Place the tubes or plates in a thermal cycler (Figure 7)
5. Program the thermal cycler with the appropriate temperature settings for denaturation, annealing, and extension steps (Table 2)
6. Analyze the PCR products on agarose gel electrophoresis to confirm the presence from size of the amplified DNA fragments
7. Cut the corresponding DNA band of interest and extract the DNA from the gel using the VWR life science KIT

Figure 7. Polymerase chain reaction machine used in the process

Step	Temperature	Time	cycle
Start	98°C	10 sec	1
denaturation	98°C	1 min	--------
Annealing	65°C on each cycle)	1 min	40
Extension	68C	20 sec	--------
Final extension	68°C	5 min	1

Table 2. Thermal cycler settings for specific DNA amplification.

Transformation of competent cells

Making Ca++ competent cells (Chang AY, et al. (2017)

- Grow a bacteria prec. in the appropriate medium, 37, overnight
- Grow a dilution 1:400 of the ON culture in 250 ml of the same medium
- Grow at 37 °C, until OD reaches 0.2-0.3
- Chill the entire culture on ice for 15min, and $CaCl_2$ 0.1M and 0.1M $CaCl_2$ + 15% glycerol
- Pellet the bacteria in pre-chilled tubes or bottles (3300 g, 10 min, +4 °C)
- Resuspend all the cell pellets in total amount of 40 ml of cold $CaCl_2$
- Keep the tube in ice for 30min
- Pre-chill sterile small tubes to aliquot the cell

- 0.5ul tubes for 200-400ul aliquots (=2-4 reactions)
- 1.5ml tubes for 100ul (=1 reaction)

- Pellet the bacteria (3300g, 10min, +4°C)
- Discard the supernatant and resuspend cells in 5ml of cold CaCl2 + 15% glycerol
- Aliquot the bacteria and store at -80°C

Competent cell transformation

- Thaw on ice for 15min the bacteria from the -80C stock (100ul/each transformation)
- Put to dry 1 petri dish for each transformation reaction (with appropriate antibiotic)
- Add 1µL to 100µL of competent cells.
- Incubate for 30 min on ice. Set the tube-bath at 42°C and pre-warm 1ml medium/reaction
- Heat shock for 45sec at 42°C then recover for 2 min on ice
- Add 1ml of medium to each tube and incubated for 45min at 37°C
- Plate 150 µL/10cm plate and grow at 37°C overnight.

Gibson assembly

In sterile conditions, we proceed as follows:

1. *Assembly Reaction:* the GeneArt™ Gibson Assembly® Kit (https://www.thermofisher.com/) provides a streamlined and efficient method for assembling DNA fragments with overlapping ends. This protocol involves the use of a specialized mix that includes DNA polymerase, DNA ligase, and exonuclease activity to facilitate the seamless assembly of multiple DNA fragments. In each 5ul MIX, we added 1:3 vector to insert ratio. The exonucleases create single-stranded overhangs on the DNA fragments, and the complementary overhangs allow the fragments to hybridize.
2. *Incubation:* the reaction mixture was incubated at 50C° for 20 min to allow the DNA fragments to anneal and the DNA polymerase to fill in the gaps, creating a continuous DNA molecule.
3. *Transformation:* the assembled DNA was transformed into DH5 alpha competent bacterial cells.
4. *Plating:* the transformed competent cells were striked on agar with the right antibiotic plates and incubated at 37C overnight for colonies to grow.

Colony PCR

It is used to verify the success of the Gibson through the amplification of the DNA insert directly from bacterial colonies transformed with the Gibson product. The protocol used was:

- *Colony Selection:* after transforming the assembled DNA into bacterial cells and spreading them on a selective agar plate, we left the colonies to grow overnight at the appropriate temperature
- *Pick Colonies:* by using a sterile toothpick or colony-picking tool we picked up a small portion of a bacterial colony from the agar plate.
- *Prepare PCR Master Mix:* in small PCR tubes we added:

10ul of master mix (Green tag)

8ul of steril distilled water

1ul of specific primers for the DNA insert (forward of colony-PCR and reverse of the insert) flanking regions to the PCR master mix allowing the amplification of the insert.

- *Cell growth:* in sterile condition, we used sterile toothpick the picked colony and swirl it in a tube containing sterile PCR-grade water. Tubes or wells containing the PCR master mix and then mixed gently. Then we used the same tips and rinsed it again with pipetting in a tube containing usual LB for cell growth purpose.
- *Performing PCR:* we placed the PCR tubes into a thermocycler and set up the appropriate PCR cycling conditions. The cycling parameters include denaturation, annealing, and extension temperatures, as well as the number of cycles (Table 3).
- *Analyze PCR Products:* we run the PCR products on a DNA agarose gel.
- *Gel electrophoresis analysis:* we loaded the PCR products onto an agarose gel along with appropriate components. In our case we used a 100ml of 1x TAE, 1,50gm of 0.8% agarose gel, stained with 7ul of DNA specific Ethidium bromide dye, 4ul Dna tracker and 7.5ul of DNA ladder.
- *UV light Results:* we compared the size of the PCR products obtained from the colonies to the expected size of the assembled DNA.

Step	Temperature	Time	cycle
Start	95°C	2min	1
Denaturation	95°C	1 min	
Annealing	65°C each cycle	1 min	35
Extension	72°C	1 min	

Final extension	72°C	7 min	1

Table 3. Thermal cycler settings for colony PCR

Sample sequencing

Samples were sent to GENEWIZ (https://www.genewiz.com/) for next-generation sequencing (Amplicon EZ service) using the Sanger method (https://clims4.genewiz.com/Customer-Home/Index). The results obtained, both as *fasta* file and chromatoferogram file, were check using Snapgene.

2.9. Protein Expression and purification

Small scale expression

This step is useful for the identification of the optimum expression parameters for a given target. There are several parameters that can be used to optimize conditions: the temperature, the time, the concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG), a specific agent that induces protein expression by binding to and inactivating the lac repressor, thereby allowing transcription of genes under the control of the lac operator, and the OD cell-growth value. Our protocol lasted 2 days:

Day 1 - bacterial induction

1. Pick a colony of trasformed bacteria and grow it in 10 ml of LB + antibiotics
2. Grow at 220 rpm, 37 °C until $OD_{600\ nm}$ is 0.4 (about 4 hours)
3. Keep 50 µL of cell suspension for SDS Page electrophoresis (uniunduced control)
4. Incubate 30 minutes (the cells will reach $OD_{600} = 0.6$)
5. Add IPTG (1mM final concentration) to induce the protein expression and grow overnight, at the appropriate temperature C, 220rpm.

Day 2 - bacterial lysis and purification

1. Pellet the cells at 10000-13000 rpm for 3 min
2. Discard the supernatant and did the freeze and thaw of the tube at -80 °C, 3 times
3. Resuspend the bacterial pellet in 350 µl of RB Resuspension Buffer (500 mM NaCl, 50 mM Tris pH 8.0, 5 mM $MgCl_2$)
4. Add Lysozime and DNAse to each sample and incubated 30 min at RT
    - DNAse final concentration: 10 ug/ml

- lysozyme final concentration: 1 mg/ml

5. Sonicate for 1 min, put on ice for 1 min (3 times)
6. Keep 50 µL of the total lysate for SDS Page electrophoresis.
7. Spin the lysate for 2 min at 13‘000 rpm.
8. Add all the supernatant to the nickel Histag-Beads (ThermoFisher) and incubate for 30 min under mixing.
9. Wash the beads twice with 500ul of WB (50 mM Tris pH 8.0, 500 mM NaCl, 15 mM imidazole)
10. Resuspend the beads in 50ul of PBS (20 mM phosphate pH 7.4, 150 mM NaCl without Ca^{2+} and Mg^{2+}).

Large scale expression

Scaling up expression of proteins involves transitioning from small-scale production to a larger production system with the exact same conditions in order to generate larger quantities of a target protein. This process is essential when there is a need of large protein amounts aimed at various applications, i.e., structural studies.

Protein purification

The protocol involved 1) cell lysis for 30min on ice (buffer: 50mM Tris/HCl pH 8.0, 500mM NaCl, 15mM Imidazole, 5mM ZnCl2, 5mM MgCl2, 1mg/mL Lysozime, 1U DNAaseI, 1U AEBSF protease inhibitors), 2) cell sonication (1 min on, 1 min off, power 9, intermittent 9, 3 times each step), 3) pellet at 13000 rpm for 30 min at 4C, and 4) chromatographic steps (see sections 2.9).

Purification of insoluble aggregates

In those cases where the protein resulted not soluble from step 3) we preceded as follow:

1. The pellet was solubilized in appropriate detergent buffer (0.2% n- lauroylsarcosil, 40mM Tris pH 8) (Figure 8A, B) for 40 hours at room temperature (40mL for 1g pellet)
2. The pellet was washed twice with dH_2O and centrifuged (13 000 rpm, 45 min, 4 ° C)
3. The supernatant was purified according to step 4)

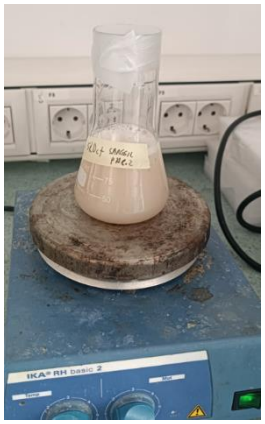
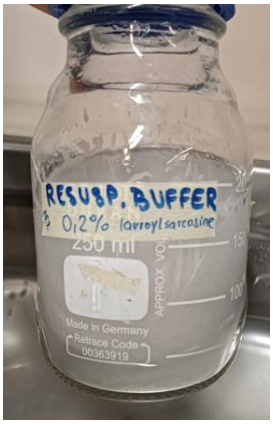
A  B 

Figure 3. ZRANB3 pellet with insoluble aggregates (A) resuspended in buffer with 0.2% n- lauroylsarcosil (B)

Metal affinity and Desalting

We used AKTA pure machine system (Figure 9) available at the Laboratory of Environmental and Life Sciences and followed the protocol described hereafter:

- the supernatant was loaded into a TALON column, with a matrix based on nickel beads at low flow rate to favor the binding of 6xHis tagged proteins; consequently, all impurities were removed using buffer 50mM Tris/HCl pH 8.0, 500mM NaCl, 15mM Imidazole (buffer A).
- the elution was done with buffer A + 500mM imidazole by applying a linear gradient of it from 0% to 100%
- the peak corresponding to the ZRANB3 protein was then loaded into a desalting column to remove the imidazole concentration (50mM Tris/HCl pH 8.0, 50mM NaCl, 2mM DTT, 5% glycerol); indeed, imidazole can denature the protein when kept in the buffer for longer time.

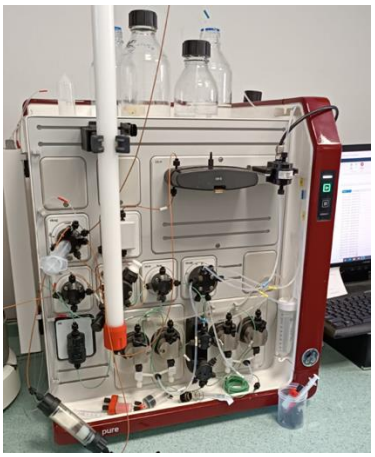
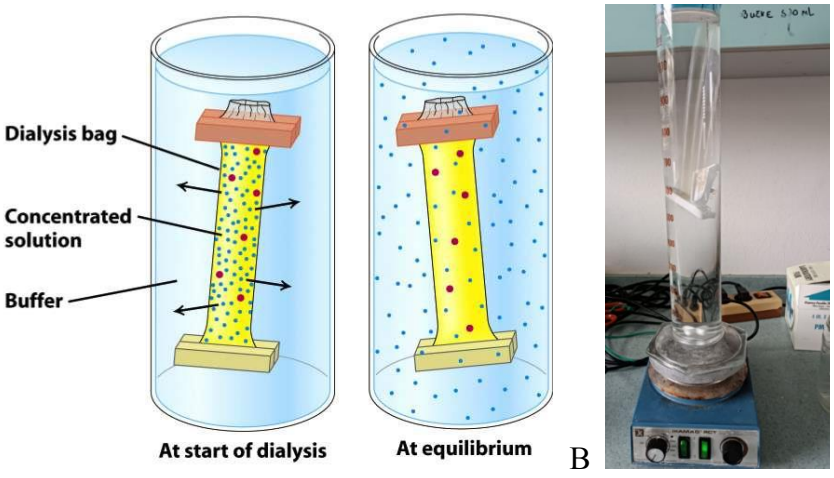


Figure 9. The AKTA pure machine system at Laboratory of Environmental and Life Sciences.

Ion exchange and dyalisis

This technique separates proteins based on their net charge. It involves the use of a solid stationary phase with charged groups and a mobile phase (buffer) that contains ions of opposite charge. The stationary phase can be either anion exchange resin (for negatively charged proteins) or cation exchange resin (for positively charged proteins). The process is highly selective, as proteins with different charges interact differently with the stationary phase, allowing for their separation.

The sample desalted in 50mM Tris/HCl pH 8.0, 50mM NaCl, 2mM DTT, 5% glycerol (buffer A) was loaded into a MonoQ column (Quiagen – anion exchanger). All our constructs had pI > 8.5, thus, they were negatively charged at pH 8.0. They were eluted using a linear gradient between 0.3 and 0.5M of NaCl (buffer B: buffer A + 1M NaCl. Where needed, we applied a dialysis step (Figure 10) to change the final buffer for downstream applications.



A B

Figure 10. Example of our dialysis experiment, theory (A) and practice (B).

2.10. Protein characterization by SDS-Page

SDS polyacrylamide gel electrophoresis (SDS-PAGE) involves the separation of proteins based on their size. By heating the sample under denaturing and reducing conditions, proteins become unfolded and coated with SDS detergent molecules, acquiring a high net negative charge that is proportional to the length of the polypeptide chain. When loaded onto a gel matrix and placed in an electric field, the negatively charged protein molecules migrate towards the positively charged electrode and are separated by a molecular sieving effect. After visualization by a protein-specific staining technique, the size of a protein can be estimated by comparison of its migration distance with that of a standard of known molecular weight. The concentration of acrylamide used for the gel depends on the size of the proteins to be analyzed. Low acrylamide concentrations are used to separate high molecular weight proteins, while high acrylamide concentrations are used to separate proteins of low molecular weight. Improved resolution of protein bands is achieved by the use of a discontinuous gel system having stacking and separating gel layers.

Table 4 highlights the reagents used for the stacking (the proteins enter the gel) and the running (proteins run properly) phases. Each sample was "labeled" with 7ul of loading buffer containing SDS and heated for 5min at 95C° ensuring proper denaturation (Figure 11).

Visualization of protein bands was carried out by incubating the gel with a staining solution. The two most commonly used methods are Coomassie and silver staining. Silver staining is a more sensitive staining method than Coomassie staining, and is able to detect 2–5 ng protein per band on a gel. Coomassie staining, though less sensitive, is quantitative and Coomassie-stained proteins can be used for downstream applications.

Separation/running gel		
Stock solution	Final volume	
Acrylamide 40%	21%	2.5 ml
Tris- hcl 1.5 M pH8.8	375mM	3ml
SDS 10%	0.01%	100µl
APS10%	0.01%	100µl
H_2O	Up to 15ml in the final volume	4.3ml

TEMED	20µl	10µl

Stacking gel		
Stock solution	Final volume	
Acrylamide 40%	6%	0.500ml
Tris- hcl 0.5 M pH 6.8	125mM	0.520ml
SDS 10%	0.01%	40µl
APS10%	0.01%	40µl
H_2O	Up to 10ml in the final volume	2.88ml
TEMED	10µl	4µl

Table 4. Stacking and running gel preparation protocol

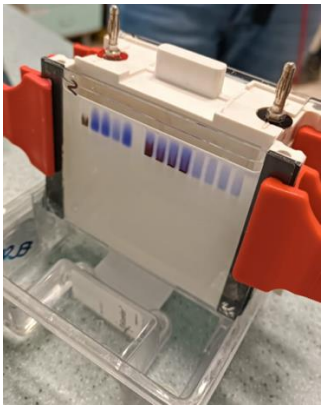


Figure 11. SDS page gel electrophoresis

Figure 12 summarize the entire protein preparation flowchart used for all constructs of this thesis, from cloning to protein expression, purification and characterization.

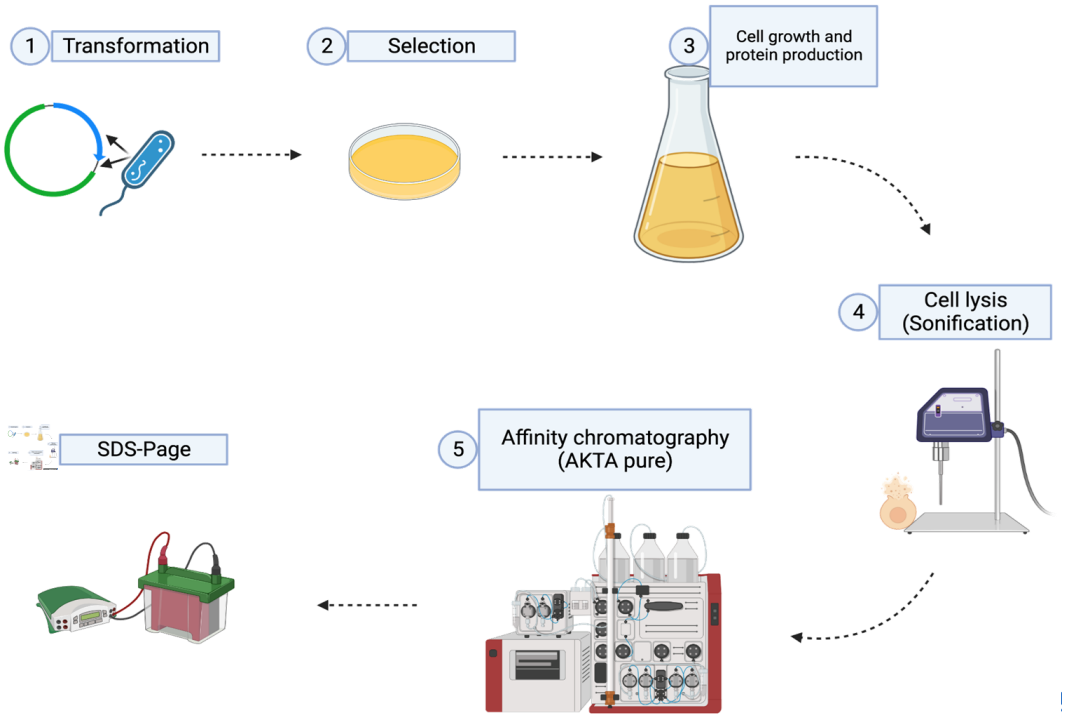


Figure 12. Full large-scale production and purification workflow conducted

3. RESULTS AND DISCUSSION

PART I: Bioinformatic Analysis

3.1. The full length wild-type ZRANB3

ZRANB3 is a multifunctional protein characterized by several distinct domains, each contributing to its role in DNA repair and replication stress response (Figure 13).

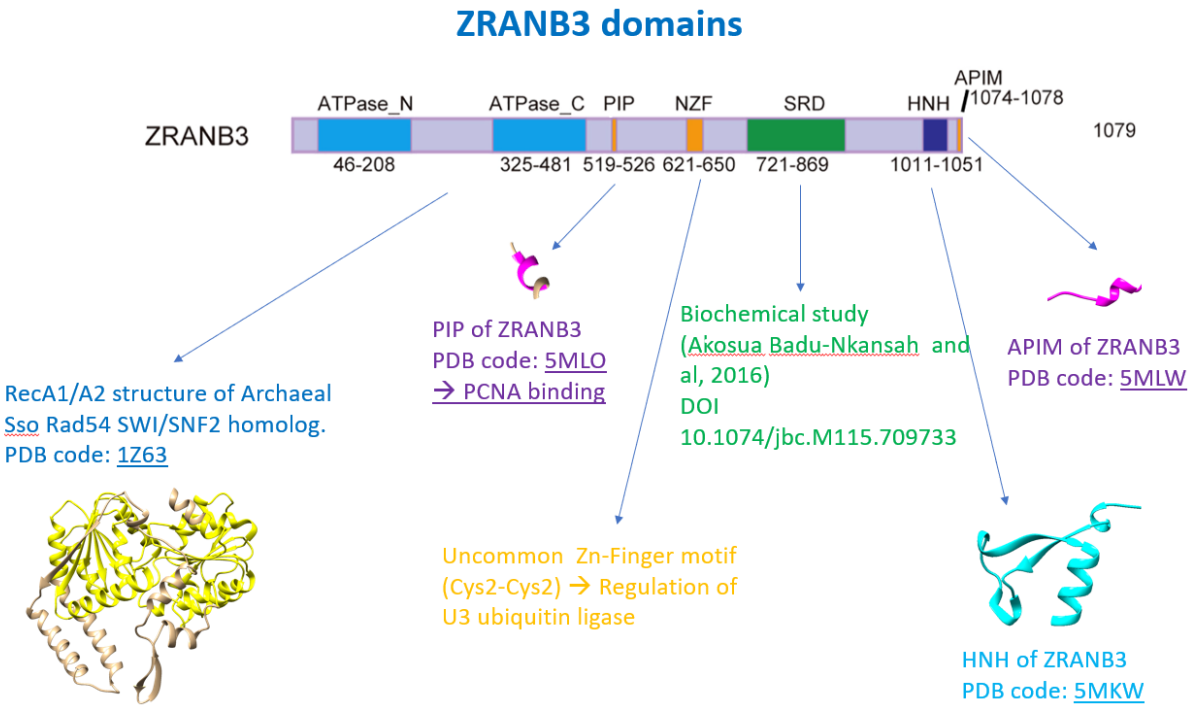


Figure 4. Zranb3 Domains and characteristics based on Motif-finders

ATPase_N and ATPase_C (46-208, 325-481): this is the helicase-like core, crucial for ATP-dependent DNA translocase activity. It has ancestor on the bacterial RecA protein; thus, it is also called RecA-like 1/2, and is homologue of analogous to the archaeal SWI/SNF2.

PIP (519-526): the PCNA-Interacting Protein (PIP) motif binds to PCNA, a processivity factor for DNA polymerase. This interaction is essential for targeting ZRANB3 to sites of replication stress.

NZF (621-650): this Zinc Finger domain, containing an uncommon Cys2-Cys2 motif, which regulates the U3 ubiquitin ligase activity and influences protein stability and interactions.

SRD (721-869): known as SMARCAL1-Related Domain is implicated in DNA repair processes.

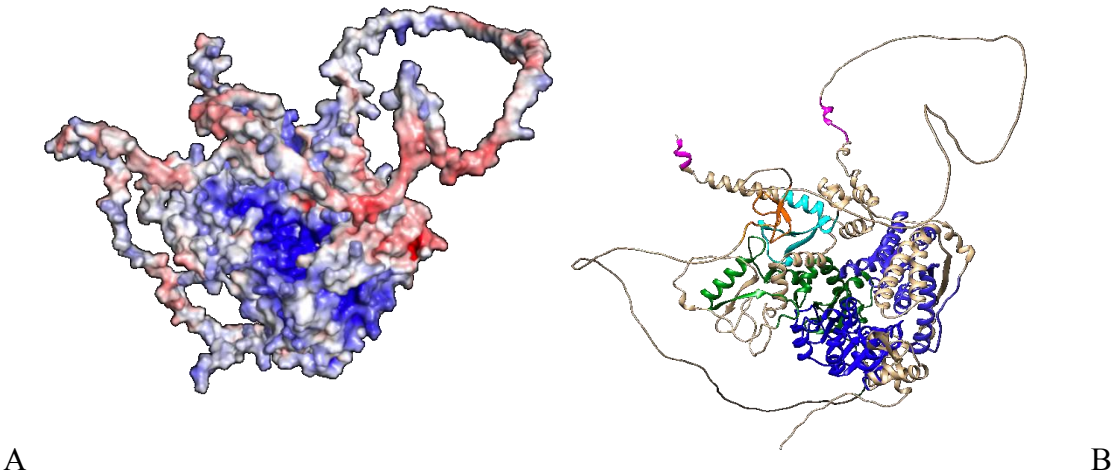
HNH (1011-1051): it is the nuclease domain, in particular endonuclease aiding in the resolution of DNA structures during repair.

APIM (1074-1078): the AlkB homolog 2 PCNA-Interacting Motif (APIM) supports interaction with PCNA, enhancing the recruitment to DNA damage sites.

The AlphaFold model

The use of AlphaFold to predict the structure of ZRANB3 provided significant insights into its molecular characteristics. Indeed, the AlphaFold model of ZRANB3 revealed detailed structural features such as – in particular - its electrostatic surface (Figure 14A) with positive and negative charges crucial for the protein's interaction with DNA and other partners. The overall positively charged cavity in between the two RecA lobes indicated the DNA-binding potential. This positive charge can attract the negatively charged phosphate backbone of DNA, facilitating the binding and stabilization of DNA within the catalytic site of ZRANB3. This binding is crucial for the protein's function in DNA repair, allowing it to effectively recognize DNA lesions. Additionally, the AlphaFold model provided a detailed view of the folding of each domain in ZRANB3 (Figure 14B). The ATPase_N domain (residues 46-208, yellow color) is predominantly composed of 15 alpha-helices and 6 beta-strands, forming a compact and globular structure that hosts the catalytic site for ATP binding and hydrolysis. The presence of beta-sheets within this domain provides stability and rigidity. Similarly, the ATPase_C domain (residues 325-481) is characterized by a mixture of alpha-helices and beta-strands. The interaction between the ATPase_N and ATPase_C domains forms a central cleft that binds DNA, as evidenced by the positively charged cavity between the two RecA lobes (Figure 14A). The PIP box domain (residues 519-526, purple color), consists mainly of short peptide motifs that interact with PCNA (Proliferating Cell Nuclear Antigen). This partially structured domain is typical for peptide motifs that need flexibility to fit into the PCNA binding pocket, facilitating the recruitment of ZRANB3 to sites of DNA replication and repair. The NZF domain (residues 621-650, yellow) contains an uncommon Zn-Finger motif characterized by Cys2-Cys2 coordination, primarily composed of short beta-strands and loops that coordinate the zinc ion and stabilize the domain's structure. The zinc finger motif's flexibility allows it to interact with various protein partners, suggesting its role in the regulation of the U3 ubiquitin ligase pathway essential for DNA damage response. The SRD (Serine-Rich Domain, residues 721-869, green) consists predominantly of disordered regions with few alpha-helices. This unstructured/flexible region may become ordered upon binding to DNA or other regulatory

proteins. Also, it may play a role in protein-protein interactions and post-translational modifications. The HNH domain (residues 1011-1051, cyan) features a well-defined structure with several alpha-helices and beta-strands forming a compact nuclease active site. The positively charged regions enable the cleavage of damaged DNA strands within the HNH domain likely involved in binding the DNA substrate to facilitate precise cutting and processing of DNA during repair. The APIM (AlkB homolog 2 PCNA Interacting Motif, residues 1074-1078, violet) has minimal secondary structure allowing efficient fit into the PCNA binding pocket. The detailed structural organization of ZRANB3, as revealed by AlphaFold, underscores the coordinated interaction between its domains to ensure proper DNA repair function. The flexibility observed in certain domains, such as the SRD and PIP box, highlights the protein's ability to adapt and interact with various partners, enhancing its functional versatility. Overall, the AlphaFold model provides a comprehensive view of ZRANB3's domain architecture.



A B

Figure 14. A. Electrostatic Surface of Zranb3 Protein. B AlphaFold model (domains).

PPI network analysis

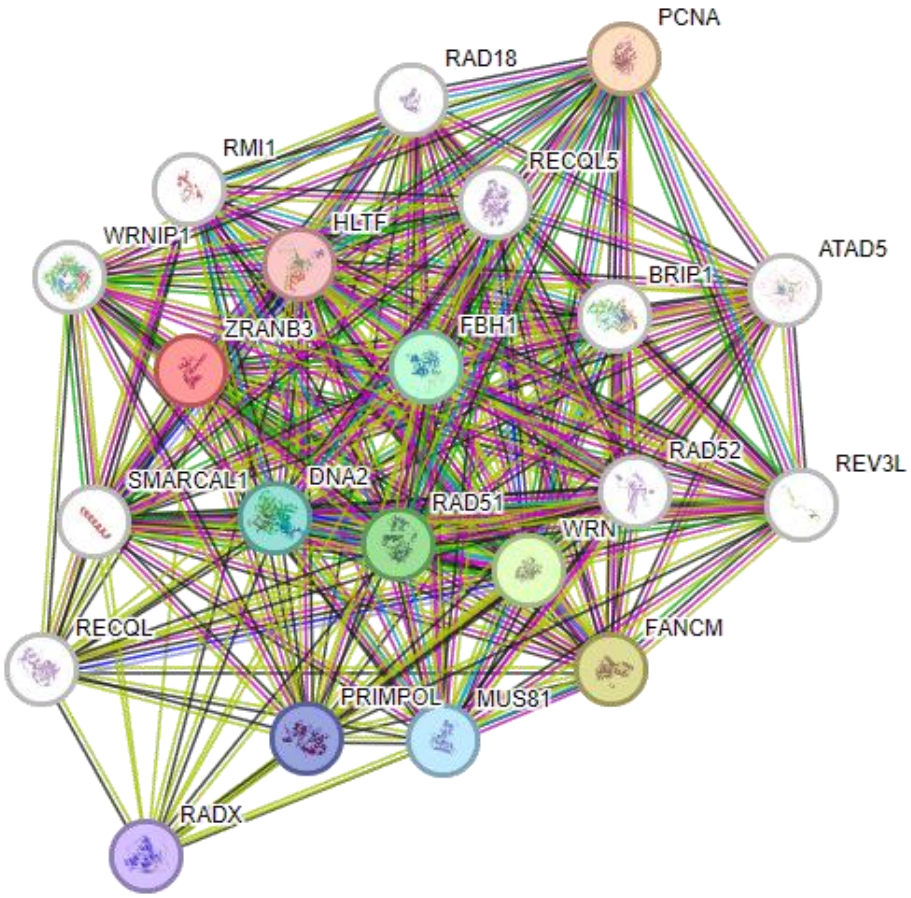


Figure 5. the PPI network for ZRANB3, highlighting its potential interaction partners. The interactions are scored based on the confidence level, with thicker lines indicating stronger interactions.

The analysis of the protein-protein interaction (PPI) network for ZRANB3 was conducted using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins). STRING provides a comprehensive interaction map based on various sources, including experimental data, databases, text mining, and computational predictions (Figure 15). Table 5 summarizes the key interacting partners of ZRANB3 identified through STRING, along with their functions and interaction scores.

Protein	Full Name	Function	Score
PCNA	Proliferating Cell Nuclear Antigen	Auxiliary protein of DNA polymerase delta, involved in the control of eukaryotic DNA replication and repair.	0.954
FANCM	Fanconi Anemia Group M Protein	DNA-dependent ATPase, part of the Fanconi anemia core complex, involved in DNA repair and stability.	0.835
WRN	Werner Syndrome ATP-Dependent Helicase	Multifunctional enzyme with helicase and exonuclease activity, involved in DNA repair and replication.	0.790
RAD51	DNA Repair Protein RAD51 Homolog 1	Plays a key role in homologous recombination and DNA repair through strand exchange.	0.780
FBH1	F-Box DNA Helicase 1	Involved in genome maintenance, acts as an anti-recombinogenic helicase, and part of the SCF(FBH1) E3 ubiquitin ligase complex.	0.774
DNA2	DNA Replication Helicase/Nuclease	Key enzyme in DNA replication and repair, processes Okazaki fragments and involved in double-strand break repair.	0.738
MUS81	Crossover Junction Endonuclease	DNA structure-specific endonuclease, involved in processing branched DNA structures during replication and repair.	0.733
PRIMPOL	DNA-Directed Primase/Polymerase	Facilitates DNA replication fork progression and tolerates replication-stalling lesions by bypassing them.	0.726
RADX	RPA-Related Protein RADX	Single-stranded DNA-binding protein, stabilizes replication forks, and regulates RAD51 accumulation.	0.725
HLTF	Helicase-Like Transcription Factor	Has helicase and E3 ubiquitin ligase activities, involved in error-free postreplication repair and maintenance of genomic stability.	0.717

Table 5. ZRANB3 Interaction Partners from STRING Software

Among those that are physical partners:

1. **PCNA** forms a sliding clamp around DNA, acting as a scaffold to recruit other proteins, to the sites of DNA synthesis and repair. The binding of ZRANB3 to PCNA must be tightly regulated to ensure that ZRANB3 is localized only to sites of replication stress. This regulation is crucial to prevent ZRANB3 from outcompeting other PCNA-interacting partners that are essential for normal DNA replication elongation. Indeed, ZRANB3 binds selectively to polyubiquitinated PCNA, a modification that typically occurs in response to replication stress. Interestingly, sequence alignment of ZRANB3 revealed an uncommon zinc-finger motif (Cys2-Cys2).

This motif primarily regulates the interaction with U3 ubiquitin ligase and promotes protein interactions (Figure 16).

2. **FANCM**, part of the Fanconi anemia pathway, is crucial for DNA repair and maintaining chromosomal stability. Its interaction with ZRANB3 suggests a collaborative role in resolving DNA damage and replication stress.

3. **RAD51**, essential for homologous recombination, facilitates the exchange of DNA strands during repair. Its interaction with ZRANB3 underscores a role in homologous recombination repair pathways.

4. **MUS81,** functions as an endonuclease that resolves DNA structures during replication stress; its interaction with ZRANB3 indicates a role in processing stalled or collapsed replication forks.

5. **HLTF,** with dsDNA translocase and E3 ubiquitin ligase activities, plays a role in post-replication repair and maintaining genomic stability. Its interaction with ZRANB3 suggests involvement in ubiquitin-mediated repair pathways.

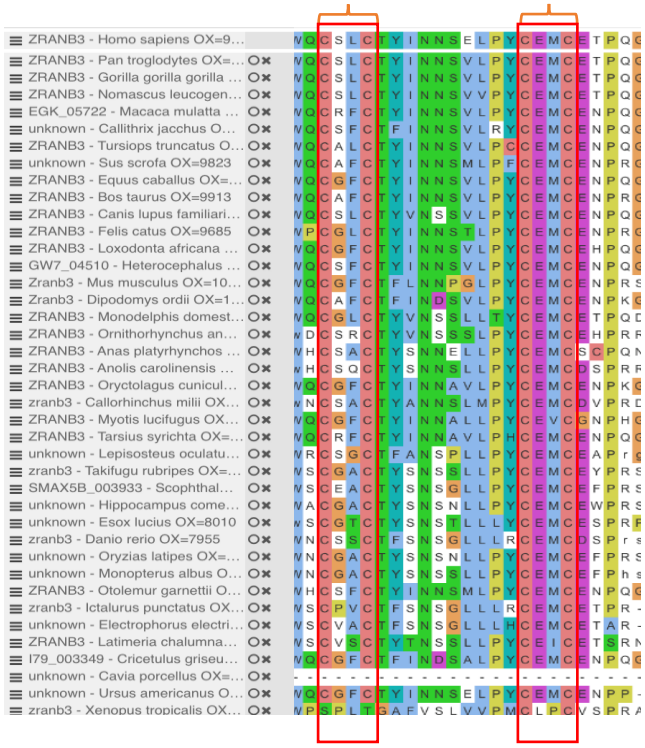


Figure 16. ZRANB33 sequence alignment of ZNF motif.

The cBIO portal for Cancer Genomics Data

Table 6 reports the ZRANb3-linked cancer genomics data (cBIO Portal) across various cancer types, specifically the cancer type, mutation type (e.g., missense mutation, frame shift deletion), frequency, change of amino acid, the gene location and the protein domain affected (Figure 17). This preliminary analysis highlights the varied nature of ZRANB3 mutations and their potential impact on protein function across different cancer types. Each mutation type—missense, frame shift, in frame insertion, nonsense, and splice site—presents unique challenges to the stability and activity of the ZRANB3 protein. Understanding these mutations' specific impacts can inform targeted therapeutic strategies and further research into the role of ZRANB3 in cancer biology.

Cancer Type	Mutation Type	Frequency	Amino Acid Change	Location	Domain
Lung Cancer	Missense_Mutation	20	p.E76K	76	ATPase
Breast Cancer	Frame_Shift_Del	15	p.Q312fs	312	HNH domain
Colorectal Cancer	In_Frame_Ins	10	p.A456_A457insV	456	HNH domain
Liver Cancer	Nonsense_Mutation	8	p.R123*	123	ATPase
Ovarian Cancer	Splice_Site	5	p.R240_splice	240	HNH domain
Lung Cancer	Missense_Mutation	15	p.G12D	12	N/A
Breast Cancer	Frame_Shift_Del	10	p.V600E	600	N/A
Colorectal Cancer	In_Frame_Ins	5	p.L858R	858	N/A
Liver Cancer	Nonsense_Mutation	2	p.T790M	790	N/A
Ovarian Cancer	Splice_Site	3	p.D538G	538	N/A

Table 6. Frequency and Types of ZRANB3 Mutations Across Different Cancer Types

Missense Mutations

- **p.E76K (Lung Cancer)**: this mutation involves the substitution of glutamic acid (E) with lysine (K) at position 76. This change occurs within the ATPase domain, which is critical

for the protein's function in energy transfer and molecular interactions. Alterations in this domain can significantly affect protein stability and function, potentially disrupting cellular processes that rely on ATPase activity.

- **p.G12D (Lung Cancer)**: the substitution of glycine (G) with aspartic acid (D) at position 12, although not within a specific domain, can still impact the protein's structure and function. Similar mutations in other genes (e.g., KRAS) have shown significant oncogenic potential, suggesting a possible pathogenic role for this mutation.

Frame Shift Deletions

- **p.Q312fs (Breast Cancer)**: this frame shift deletion at position 312 within the HNH domain leads to a shift in the reading frame, resulting in a completely altered protein sequence downstream of the mutation site. Such changes often lead to a loss of normal protein function, which can contribute to cancer progression by disabling normal DNA repair mechanisms or other critical cellular functions.

In Frame Insertions

- **p.A456_A457insV (Colorectal Cancer)**: the insertion of a valine (V) between positions 456 and 457 within the HNH domain can affect the structural integrity and functional capacity of the domain. The HNH domain is typically involved in DNA binding and cleavage, so alterations here may compromise the protein's ability to interact with DNA, potentially leading to genomic instability.

Nonsense Mutations

- **p.R123* (Liver Cancer)**: this nonsense mutation introduces a premature stop codon at position 123, truncating the protein. Located within the ATPase domain, this truncation likely results in a non-functional protein, eliminating the essential activities mediated by this domain and contributing to oncogenesis.

Splice Site Mutations

- **p.R240_splice (Ovarian Cancer)**: this mutation affects the splicing machinery, potentially leading to aberrant splicing of the ZRANB3 mRNA. Such disruptions can result in the production of malfunctioning protein variants or the complete loss of protein expression, affecting the cellular processes that depend on the correct function of ZRANB3.

Future studies should focus on functional assays and clinical correlations to validate these preliminary findings and elucidate the precise mechanisms by which ZRANB3 mutations contribute to cancer development and progression.

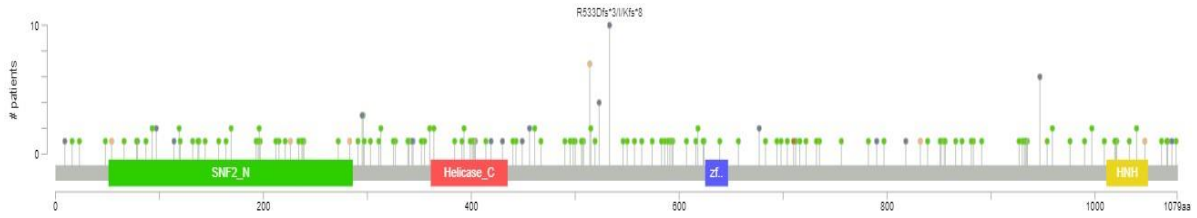


Figure 17. Mutation Distribution Across ZRANB3 Protein Domains

3.2. The Substrate Recognition Domain fused to the C-term (SRDct)

The SRD (Serine-Rich Domain) of ZRANB3 is implicated in the regulation of protein-protein interactions and post-translational modifications. The C-terminal part of the entire ZRANB3 is known to be essential for its nuclease activity, facilitating the cleavage and processing of damaged DNA strands. In the study reported by Akosua Badu-Nkansah et al. (2016), the authors demonstrated that mutations within the SRD domain led to a significant loss of DNA-binding, ATPase, and nuclease activity. This study provided a detailed analysis of the functional implications of these mutations, confirming the essential role of the SRD domain in maintaining ZRANB3's enzymatic activities necessary for DNA repair. Moreover, it has been reported that the HARP domain of SMARCAL1 exhibits annealing helicase activity, crucial for remodeling stalled replication forks, while the HIRAN domain of HLTF is involved in recognizing and binding to the 3' ends of DNA, facilitating its remodeling activity. These findings suggest a possible similar structural and functional adaptation for ZRANB3 SRD, implying that it may also play a critical role in DNA binding and remodeling.

The sequence alignment (Figure 18A) of ZRANB3 SRD within mammals shows several conserved regions. Any mutations, as mentioned in the graph (in orange), could lead to the loss of DNA-binding, ATPase, and nuclease activity (Badu-Nkansah et al., 2016). Interestingly, a fusion of the ZRANB3 SRD to the ATPase domain is sufficient to reconstitute a minimal ZRANB3 enzymatic unit that can catalyze fork reversal, which confirms this hypothesis. Taken together, this information stimulated our curiosity regarding whether a fused construct made by the SRD and C-terminal regions could have been folded in a way to ensure certain functions. Therefore, we provided a computational analysis of this fused domain and further tried to express and purify it.

The SRD AlphaFold model (Figure 18B) highlighted several key features: the presence of multiple alpha helices and beta sheets typical of DNA binding domains, and regions of high confidence

in the model corresponding to conserved motifs identified in the sequence alignment. These regions could suggest functional importance in DNA binding or interactions with other proteins. This structural information provides a foundation for further experimental validation and exploration of the SRD's role in ZRANB3's DNA repair activities.

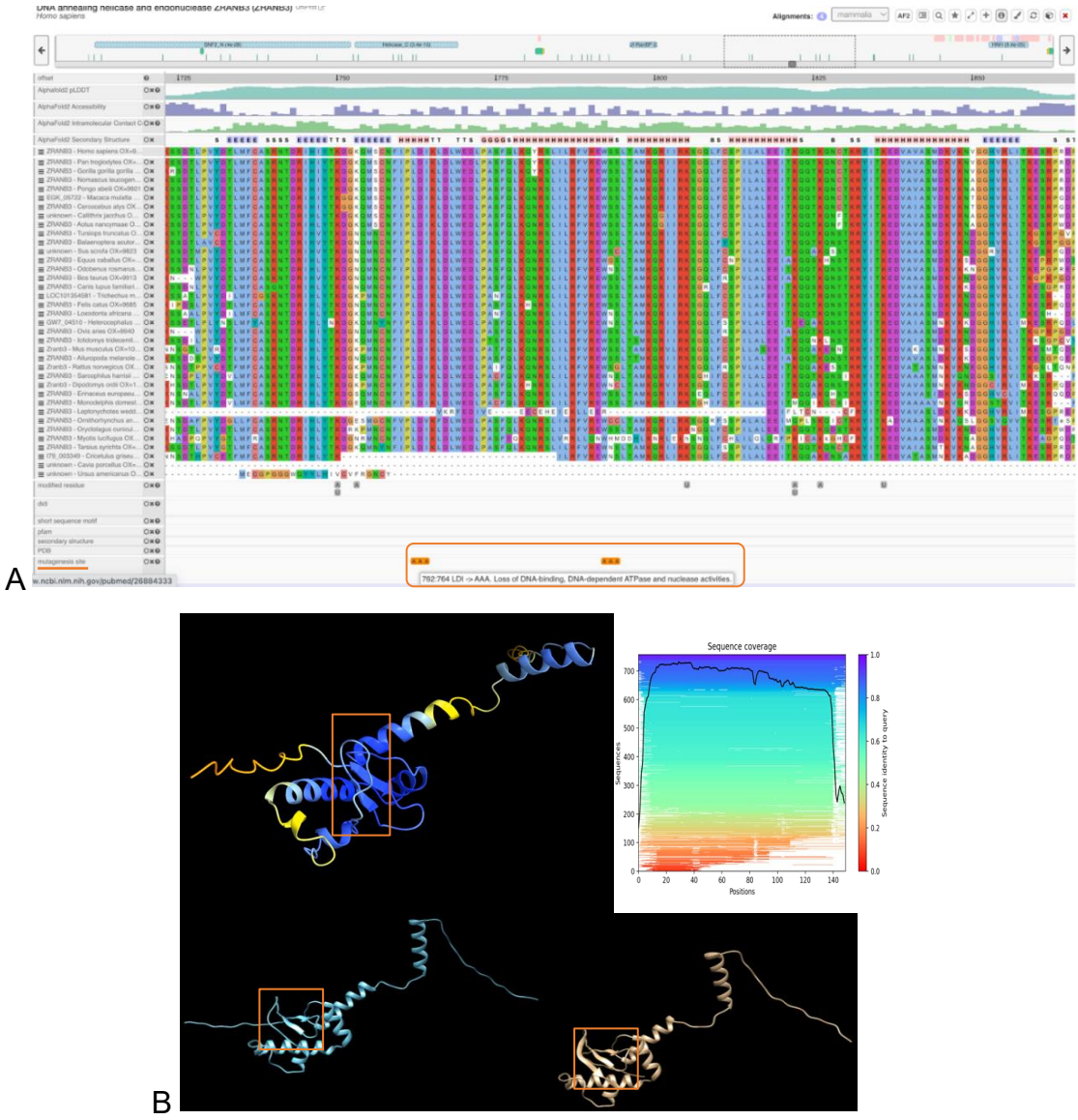


Figure 18. A. Sequence alignment of zranb3 protein with all across Mammalia species. B. AlphaFold models of SRD domain of ZRANB3

3.3. The ATP-dependent motor domain RecA-like 1/2

Sequence-sequence and sequence-structure alignment

First, I took the pdf file of the RECA1A2 domain as input for the structure homology analysis in both DALI and ZHANG LAB (I-tasser) web-based servers. The results obtained shows all the similar structures to our input in the pdb database (Table 7). The idea is to select the common highest score structures from the results. Indeed, the top 9 structures reported are based in score order of their ID which refers to the percentage of identical residues between the sequences being compared, while the Z-score reflects the significance of the alignment.

PDB ID	Description	Z	rmsd	lali	nres	%id
8atf-G	MOLECULE: INO80 ATPASE	35.6	2.3	376	442	26
8b3d-b	MOLECULE: DNA-DIRECTED RNA POLYMERASE SUBUNIT	34.5	2.4	400	520	22
7enn-K	MOLECULE: CHROMODOMAIN-HELICASE-DNA-BINDING PROTEIN 1-LIKE	29.1	2.8	385	539	28
7vdt-A	MOLECULE: ISOFORM 2 OF TRANSCRIPTION ACTIVATOR BRG1	28.7	2.8	388	617	27
7y8r-I	MOLECULE: The nucleosome-bound human PBAF complex	26.9	2.6	379	732	27
8av6-G	MOLECULE: RUVB-LIKE HELICASE	23.7	2.5	387	724	26
7egp-A	MOLECULE: SWI/SNF CHROMATIN-REMODELING COMPLEX SUBUNIT SWI1	23.6	2.8	383	750	25
1z63-A	MOLECULE: Sulfolobus solfataricus SWI2/SNF2 ATPase core in complex with dsDNA	22.7	4.4	220	468	23
6kw5-Q	MOLECULE: CHROMATIN STRUCTURE-REMODELING COMPLEX SUBUNIT RS	22.7	3.2	386	548	27

Table 7. Top 9 homologous structures selected based on the Z-score, %id

All SWI2/SNF2 enzymes share a conserved catalytic ATPase domain, indicating that the principal ATP-driven event on DNA is conserved among diverse family members, for this reason I further did a sequence-structure alignment of this region of ZRANB3 against the other SNF2 protein families using PROMALS3D (Figure 19). ZRANB3 contains two key domains, each with a "RecA" type α/β subdomain, referred to as 1A and 2A, respectively. These subdomains feature

a central β sheet flanked by α helices. The seven classical sequence motifs involved in ATP hydrolysis and DNA binding in helicases are located in the loop regions of these subdomains. Specifically, subdomain 1A contains motifs I, Ia, II, and III, while subdomain 2A contains motifs IV, V, and VI.

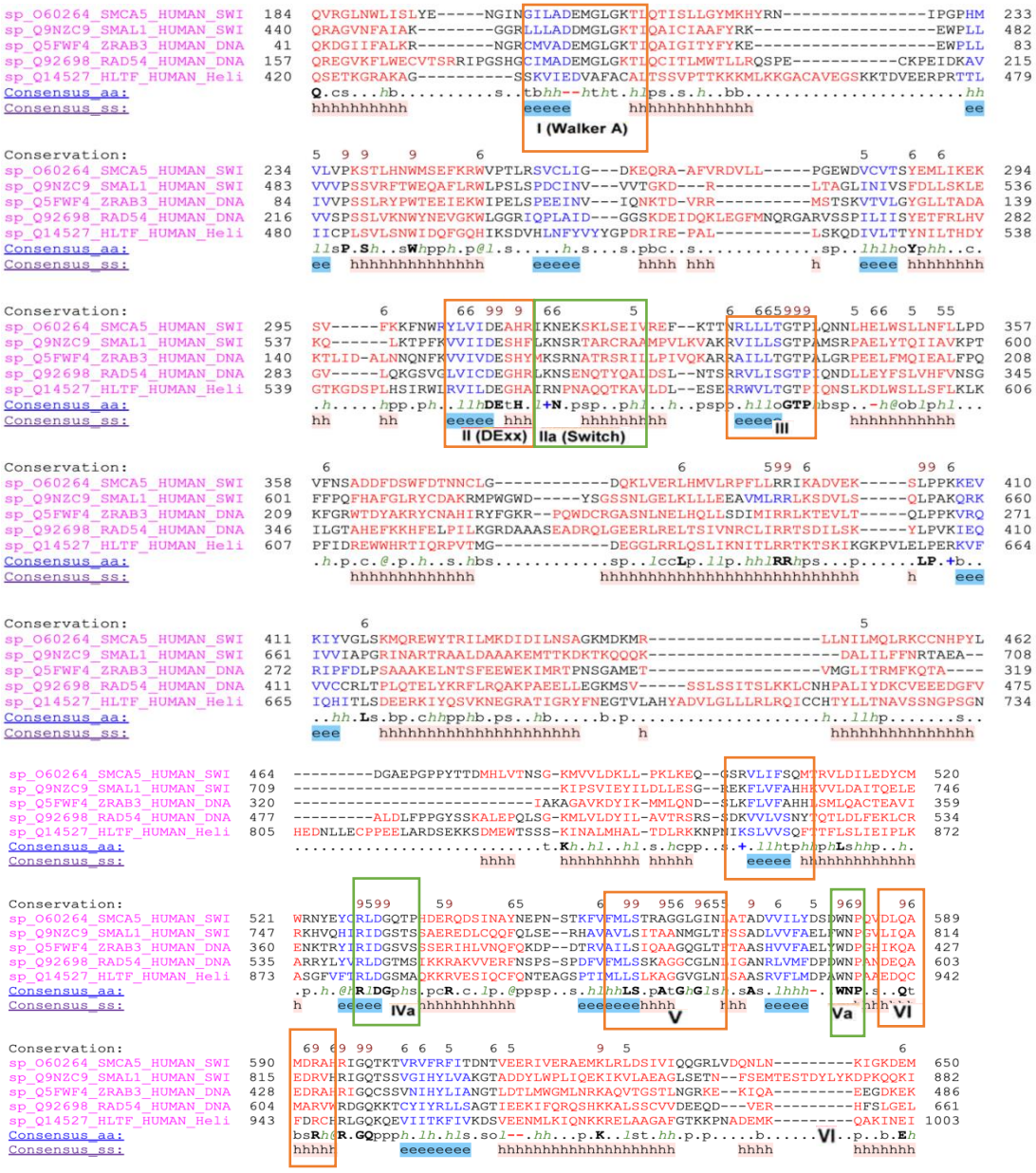


Figure 19. Structure-based sequence alignment of the catalytic domain of ZRANB3 with other SWI2/SNF2 proteins.

A significant structural feature of ZRANB3 is the deep cleft that separates domain I from domain II, as shown by the AlphaFold model (Figure 20). The primary ATP binding site is formed by

motifs I (Walker A), II (DExx), and III within subdomain 1A, located in the active site cleft. In contrast, motifs IV, V, and VI in subdomain 2A are uniquely positioned on the outside of ZRANB3 rather than in the active site cleft, due to a distinctive 180° flip of domain 2 compared to other helicases. In the SWI/SNF chromatin remodeler family, the RecA1 and RecA2 subunits function as ATPase subunits. These domains undergo conformational changes essential for their role in modifying chromatin structure and repositioning nucleosomes. In the closed conformation, the ATPase domain binds tightly to ATP or ADP, forming a compact structure. In the open conformation, the nucleotide-binding pocket is accessible, allowing for nucleotide exchange.

These conformational changes are crucial for advancing double-stranded DNA in the active site of SWI2/SNF2 enzymes in the same direction that single-stranded DNA is transported in helicases. ATP hydrolysis subsequently relaxes the structure, enabling domain 1 to rebind DNA, exchange ADP for ATP, and allow domain 2 to bind at a newly translocated upstream DNA site. This dynamic process facilitates the essential functions of DNA repair and chromatin remodeling (Dürr et al. 2005).

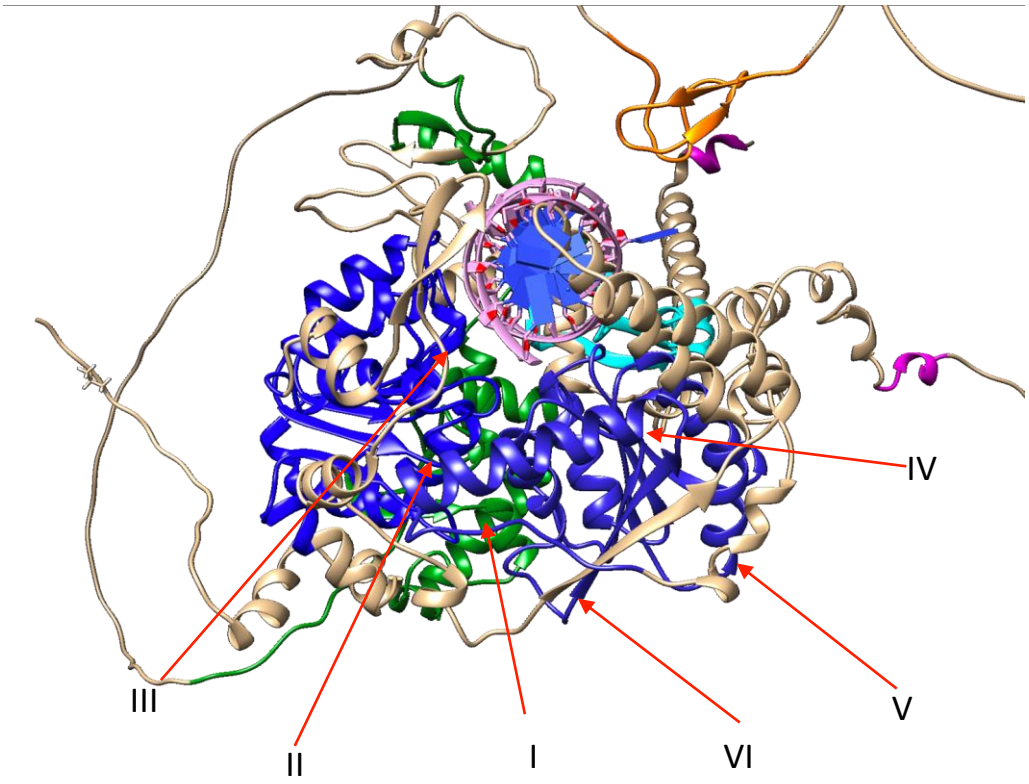


Figure 20. Homology model of ZRANB3 RecA1/A2 with dsDNA. 90° oriented

RecA-like 1/2 binding to substrate

The homologous structures in complex with a dsDNA found in the PDB were used to identify the putative in-silico DNA binding residues belonging to this domain. Specifically, I used the top 10 structures reported in Table 7 and overlapped them on the ZRANB3 RecA-like 1/2 using PyMol. BioLiP2 was used to identified the putative DNA binding residues (Table 8):

8B3D

Structure Description	Resolution (Å)	Binding Site	Original Residue Numbers (PDB)	Residue Numbers (Reindexed)
Structure of the Pol II-TCR-ELOF1 complex	2.6	BS1	R652 N653 A656 A657 N680 F796 W936 N937 R975 K979	R157 N158 A161 A162 N185 F301 W428 N429 R467 K471
Structure of the Pol II-TCR-ELOF1 complex	2.6	BS2	T575 H598 S624 S626 Y627 M793 F796 S797 I800 S864 R865 G887 R894 T913 V916	T80 H103 S129 S131 Y132 M298 F301 S302 I305 S356 R357 G379 R386 T405 V408

7VDT

Structure Description	Resolution (Å)	Binding Site	Original Residue Numbers (PDB)	Residue Numbers (Reindexed)
Structure of human chromatin-remodeling PBAF complex bound to a nucleosome	2.8	BS1	R885 K887 C891 K892 H900 W1178	R175 K177 C181 K182 H190 W460
Structure of human chromatin-remodeling PBAF complex bound to a nucleosome	2.8	BS2	L812 R841 E861 Y862 K865 M1036 N1037 M1105 T1106 G1128 R1135 S1155 R1157 A1158	L102 R131 E151 Y152 K155 M318 N319 M387 T388 G410 R417 S437 R439 A440

8AV6

Structure Description	Resolution (Å)	Binding Site	Original Residue Numbers (PDB)	Residue Numbers (Reindexed)
Structural mechanism of extranucleosomal DNA readout by the INO80 complex	4.68	BS1	K1113 S1117 S1118 N1141 R1629 W1650 N1651 K1693	K153 S157 S158 N181 R651 W672 N673 K715
Structural mechanism of extranucleosomal DNA readout by the INO80 complex	4.68	BS2	A1030 Q1087 L1088 M1258 T1579 R1580 D1600 G1601 S1627 R1629 A1630 G1631	A70 Q127 L128 M280 T601 R602 D622 G623 S649 R651 A652 G653

8ATF

Structure Description	Resolution (Å)	Binding Site	Original Residue Numbers (PDB)	Residue Numbers (Reindexed)
Structural mechanism of extranucleosomal DNA readout by the INO80 complex	3.45	BS1	K1113 S1118 R1119 W1650 N1651 R1685 K1693	K153 S158 R159 W390 N391 R425 K433
Structural mechanism of extranucleosomal DNA readout by the INO80 complex	3.45	BS2	R1059 K1060 R1063 Y1095 M1262 T1579 G1601 R1608 S1627 A1630	R99 K100 R103 Y135 M264 T319 G341 R348 S367 A370

7EGP

Structure Description	Resolution (Å)	Binding Site	Original Residue Numbers (PDB)	Residue Numbers (Reindexed)
Structure of the SWI/SNF complex bound to the nucleosome	6.9	BS1	L825 E874 Y875 F1048 N1049 M1112 G1135 H1136 K1138 R1164 A1165	L278 E327 Y328 F479 N480 M543 G566 H567 K569 R595 A596
Structure of the SWI/SNF complex bound to the nucleosome	6.9	BS2	R880 R898 N901 S904 K905 R1164	R333 R351 N354 S357 K358 R595

7ENN

Structure Description	Resolution (Å)	Binding Site	Original Residue Numbers (PDB)	Residue Numbers (Reindexed)
Structural basis of ALC1/CHD1L autoinhibition and the mechanism of activation by the nucleosome	2.8	BS1	R178 S184 L185 N444	R151 S157 L158 N404
Structural basis of ALC1/CHD1L autoinhibition and the mechanism of activation by the nucleosome	2.8	BS2	R133 E154 K310 Q312 K320 M372 T373 D394 R402 S420 R422 A423	R106 E127 K270 Q272 K280 M332 T333 D354 R362 S380 R382 A383

6KW5

Structure Description	Resolution (Å)	Binding Site	Original Residue Numbers (PDB)	Residue Numbers (Reindexed)
Structure of the RSC complex bound to the nucleosome	10.13	BS1	L528 P554 E577 Y578 K581 N754 Q815 T817 G839 S840 R846 S866 R868 A869	L101 P127 E150 Y151 K154 N305 Q366 T368 G390 S391 R397 S417 R419 A420
Structure of the RSC complex bound to the nucleosome	10.13	BS2	N604 S607 K608 L609 Q631 W889	N177 S180 K181 L182 Q204 W440

7Y8R

Structure Description	Resolution (Å)	Binding Site	Original Residue Numbers (PDB)	Residue Numbers (Reindexed)
Structure of nucleosome-bound human PBAF complex	4.4	BS1	L812 P838 R841 Y862 L1035 M1036 M1105 T1106 G1128 R1135 S1155 A1158	L279 P305 R308 Y329 L467 M468 M531 T532 G554 R561 S581 A584
Structure of nucleosome-bound human PBAF complex	4.4	BS2	R885 K887 N888 K892 T899 R1157 W1178 N1179	R352 K354 N355 K359 T366 R583 W604 N605

1Z63

Structure Description	Resolution (Å)	Binding Site	Original Residue Numbers (PDB)	Residue Numbers (Reindexed)
X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA	3.0	BS1A	H524 R527 T541 A543 R547 R728	H93 R96 T110 A112 R116 R297
X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA	3.0	BS2A	K573 I574 K681	K142 I143 K250
X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA	3.0	BS1B	H524 R527 T541 A543 R547 R728	H93 R96 T110 A112 R116 R297
X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA	3.0	BS2B	K573 I574 K681	K142 I143 K250

Table 8. All binding sites of the 9 homologous structures

Next, by combining this information with that coming from PROMALS3D, I have reported the final resulted conserved residues and their conservation or substitution, respectively (Table 9), and the possible relevant binding site sorted by occurrence (Table 10).

Importantly, a 100% conservation can be reported for S(88), Q(398-399), W(419) and R(377), and , interestingly the G(133-134). Gly is small and provides flexibility due to its lack of a side chain, which is useful in tight or highly flexible regions of the protein. The substitution with Glutamine, that has a larger side chain that can form hydrogen bonds, might reflect an evolutionary adaptation where enhanced binding interactions or stability in the protein structure is beneficial. This could relate to the environmental conditions or specific cellular or molecular requirements of the organism. Other important highlights are:

- KS (Lysine, Serine) to KN (Lysine, Asparagine) at positions 163-164

Serine to asparagine substitution could reflect a need for a stronger or more specific interaction through hydrogen bonding, possibly due to the nature of the interacting molecules or the complexity of the cellular environment in the organism.

- AT (Alanine, Threonine) to SS (Serine, Serine) at positions 167-168

Substituting alanine, a non-polar molecule, with serine, a polar molecule with a reactive side chain, could suggest an adaptation where increased interaction with the aqueous environment or other polar molecules is favorable. This might enhance the protein's functional capacity in catalysis or binding.

- QA (Glutamine, Alanine) to RA (Arginine, Alanine) at positions 398-399

Arginine is highly versatile in protein interactions due to its positive charge and capacity to form multiple hydrogen bonds and salt bridges. The substitution could be driven by a need for stronger or more specific interactions, possibly in response to different regulatory or structural roles that proteins assume in those organisms.

- WD (Tryptophan, Aspartate) to WN (Tryptophan, Asparagine) at positions 419-420

Replacing aspartate (negatively charged) with asparagine (neutral but polar) could indicate an adaptation where a less charged but still polar environment is beneficial. This might affect how the protein interacts with its ligands or other proteins, potentially altering binding affinity and specificity.

Structures	RECA1A2 Residue	Aligned Structure Residue
8b3d	S 88	T 575
	Q 113	H 598
	KS 163-164	RN 652-653
	AT 167-168	AA 656-657
	R 194	N 680
	HL 347-348	SR 863-864
	G 370	G 886
	R 377	R 893
	QA 398-399	QA 914-915
	WD 419-420	WN 935-936
	M 458	R 974
	K 462	K 979
7vdt	SS 88-89	LS 812-813
	E 860	H 598

Structures	RECA1A2 Residue	Aligned Structure Residue
	Y 161	R 884
	K 163	K 886
	AT 167-168	CK 890-891
	W 419	W 1176
8av6	S 89	S 1029
	GL 133-134	QL 1087-1088
	KS 163-164	KS 1113-1114
	AT 167-168	SS 1117-1118
	R 194	N 1141
	LS 348-349	TR 1579-1580
	DG 369-370	DG 1598-1599
	S 396	S 1625
	QAA 398-400	RAG 1627-1629
	W 419	W 1648
7egp	SS 88-89	LS 824-825
	Y 161	R 897
	KS 163-164	KN 899-900
	AT 167-168	SK 903-904
	H 347	M 1110
	DG 369-370	DG 1132-1133
	QA 398-399	RA 1157-1158
7enn	Y 161	R 178
	AT 167-168	SL 184-185
	K 315	K 320
	HL 347-348	MT 372-373
	DG 369-370	DG 394-395
	R 377	R 402
	S 396	S 420
	QA 398-399	RA 422-423

Structures	RECA1A2 Residue	Aligned Structure Residue
	D 420	N 444
6kw5	SS 88-89	LS 528-529
	GL 133-134	EY 577-578
	S 164	N 604
	L 192	Q 631
	DGS 369-371	DGS 838-840
	S 396	S 866
	QA 398-399	RA 867-868
	W 420	W 889
8atf	S 89	S 314
	L 134	L 371
	I 155	I 388
	KS 163-164	KS 396-397
	AT 167-168	SS 400-401
	TG 187-188	TG 418-419
	DG 369-370	DG 883-884
	R 377	R 891
	S 396	S 910
	QA 398-399	RA 912-913
	WD 419-420	WN 933-934
	K 462	K 972
7y8r	SS 88-89	LS 812-813
	L 134	Y 862
	Y 161	R 885
	KS 163-164	KN 887-888
	HL 348-349	MT 1105-1106
	DG 369-370	DG 1127-1128
	R 377	R 1135
	S 396	S 1155

Structures	RECA1A2 Residue	Aligned Structure Residue
	QA 398-399	RA 1157-1158
	WD 419-420	WN 1178-1179

Table 9. Aligned binding sites of RecA1A2 with structures

Residue	Aligned Structure(s)
S 88	5 occurrences
Q 113	1 occurrence
KS 163-164	7 occurrences
AT 167-168	5 occurrences
R 194	8 occurrences
HL 347-348	4 occurrences
G 370	2 occurrences
R 377	7 occurrences
QA 398-399	9 occurrences
WD 419-420	3 occurrences
M 458	1 occurrence
K 462	2 occurrences

Table 10. Relevant possible binding sites based on their occurrence

Molecular docking of RecA-like 1/2 to dsDNA

Based on the information reported above, I docked the ATPase motor domain to a dsDNA molecule using HADDOCK (Figure 21). More specifically, the residues selected as 'active' for docking based on their high frequency of occurrences were S88 and K163-S164. I also set these residues in the parameters as potential binding sites to refine the docking process, aiming to achieve a model that accurately represents the molecular interaction dynamics.

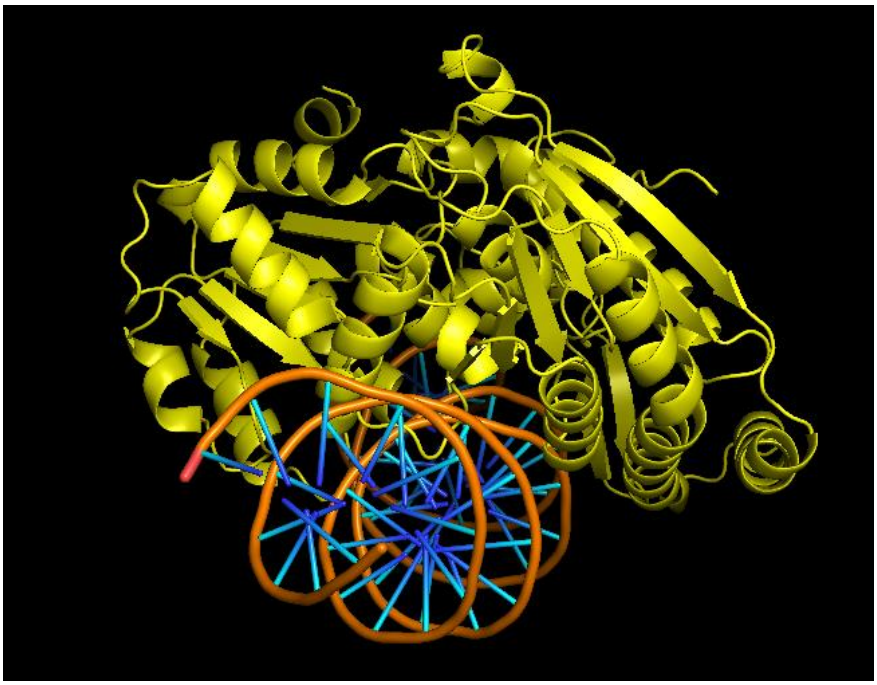


Figure 21. Best model by HADDOCK illustrates possible docking of RecA1A2 with DNA

The best model as result of the docking was selected considering:

Docking Score: Measures the predicted effectiveness of a ligand binding to a target. Lower scores (more negative) indicate better binding affinity.

Confidence Score: Reflects the reliability of the docking prediction, with values closer to 1 indicating higher confidence.

Ligand RMSD (Root Mean Square Deviation in angstron): Quantifies the deviation of the ligand's position from a reference, where lower values suggest a more accurate model.

Interface Residues: Identifies which model residues are involved in the ligand binding, differing across models.

Next, using chimera, I analyzed the specific H-bond network of the putative in-silico DNA binding residues. Indeed, the ideal hydrogen bonds are typically characterized by D→A distances of about 2.5 to 3.5 angstroms and D-H→A distances around 2.0 angstroms. From the data obtained on hydrogen bonding interactions, the following residues show the strongest hydrogen bonds, indicated by the shortest D→A distances and D-H→A distances, which are typically associated with more stable and stronger interactions. These, in the case of RecA-like 1/2 – dsDNA complex is:

```
H-bonds (donor, acceptor, hydrogen, D..A dist, D-H..A dist):
SER 88.A OG    DG 70.S O3'    SER 88.A HG     2.289  1.433
SER 88.A OG    DC 71.S OP2    SER 88.A HG     3.049  2.517
TYR 161.A OH   DC 69.S O3'    TYR 161.A HH    2.510  1.652
ARG 169.A NH1  DC 69.S O4'    ARG 169.A HH12  3.268  2.709
ARG 169.A NH2  DC 69.S O4'    ARG 169.A HH22  2.874  2.231
ARG 171.A NH2  DG 61.S OP2    ARG 171.A HH22  3.356  2.526
TYR 418.A OH   DA 68.S OP2    TYR 418.A HH    3.182  2.545
DG 79.S N1     GLU 376.A OE1  DG 79.S H1      2.847  2.073
```

Other H-bonds with slightly longer distances include:

- **ARG 169.A** (donor) with **DC 69.S** (acceptor) - Both NH1 and NH2 groups of ARG 169.A are involved, with D→A distances of 3.268 Å and 2.874 Å respectively.
- **ARG 171.A** (donor) with **DG 61.S** (acceptor) - 3.356 Å.
- **TYR 418.A** (donor) with **DA 68.S** (acceptor) - 3.182 Å

Instead, the list of the long-range electrostatic interactions and Van Der Waals forces, essential for the proper folding, stability, or function are:

```
atom1           atom2           overlap  distance
TYR 365.A HH    DA 78.S O3'     1.488  1.012
DA 68.S C2      ARG 169.A HH21  1.351  1.349
DA 78.S C3'     TYR 365.A OH    1.333  1.867
DA 78.S O3'     TYR 365.A OH    1.310  1.690
DA 78.S C4'     TYR 365.A OH    1.251  1.949
DA 78.S H4'     TYR 365.A OH    1.077  1.423
ARG 169.A NH2   DA 68.S N3      1.056  2.194
ARG 169.A NH2   DA 6&.S C2      1.034  2.291
DA 68.S N3      ARG 169.A HH21  0.989  1.636
DA 78.S C2'     TYR 365.A OH    0.935  2.265
TRP 234.A NE1   DG 58.S P       0.907  2.589
DA 68.S H2      ARG 169.A HH21  0.881  1.119
TYR 161.A HH    DC 69.S O3'     0.848  1.652
DC 69.S H2'     TYR 161.A HE1   0.832  1.168
DG 57.S OP2     ILE 225.A CD1   0.831  2.369
DG 58.S P       TRP 234.A HE1   0.815  2.056
DG 76.S N2      ARG 367.A NH2   0.775  2.475
```

The presence of specific residues like TYR 161, ARG 169, and GLU 376 in both the hydrogen bond data and contact data underscores their crucial role in stabilizing and facilitating the interaction

between the protein and DNA. This dual presence is not merely coincidental but indicates a functional necessity where these residues likely serve as key points of contact that mediate the biochemical interactions necessary for biological activity.

From a biochemical perspective, residues that frequently appear in both hydrogen bonding and contact analyses are prime candidates for further investigation for mutational analysis. In the next session, the experimental work on the cloning, protein expression and purification aimed at the characterization of the FL WT ZRANB3, the SRDct domain and the RecA-like 1/2 domain is presented.

PART II: experimental data

3.4. Large-scale expression tests of FL WT ZRANB3

First, I have transformed the BL21 cells with the pET28+ plasmid containing the ZRANB3 Fl WT gene obtained from TWIST Bioscience (https://www.twistbioscience.com/) and got positive resistant colonies on LB-Agar plates coated with Kanamycin (KANA) antibiotic (Figure 22).

Figure 6. Agarose-LB plate with BL21 colonies transformed with ZRANB3 WT FL gene cloned into the pET28+ plasmid received from TWIST bioscience.

To test the level of protein expression I set-up 3 different conditions, based on IPTG concentration and temperature/time of expression, and tested 3 colonies for each of them. For the sick of clarity, Figure 23 reports the results obtained for 1 colony with respect to the 4 different conditions. At 20°C O.N. and 0.1 mM IPTG, the band corresponding to a MW compatible with ZRANB3

(150KDa) is present in the supernatant (SN) and the flow-through (FT). Increasing IPTG concentration to 1 mM clearly decreases expression levels. The induction temperature change to 37°C (and incubation time to 4h) shows similar effects, except when using 1mM IPTG that seems to destabilize the process and, thus, push the protein into the pellet (P). SDS-PAGE gel analysis was conducted to compare expression under different conditions. Unfortunately, the protein did not bind efficiently the His-tagged beads used for protein purification (EL), a condition that often happen, especially with large proteins or protein-protein complexes that may be flexible or highly dynamics.

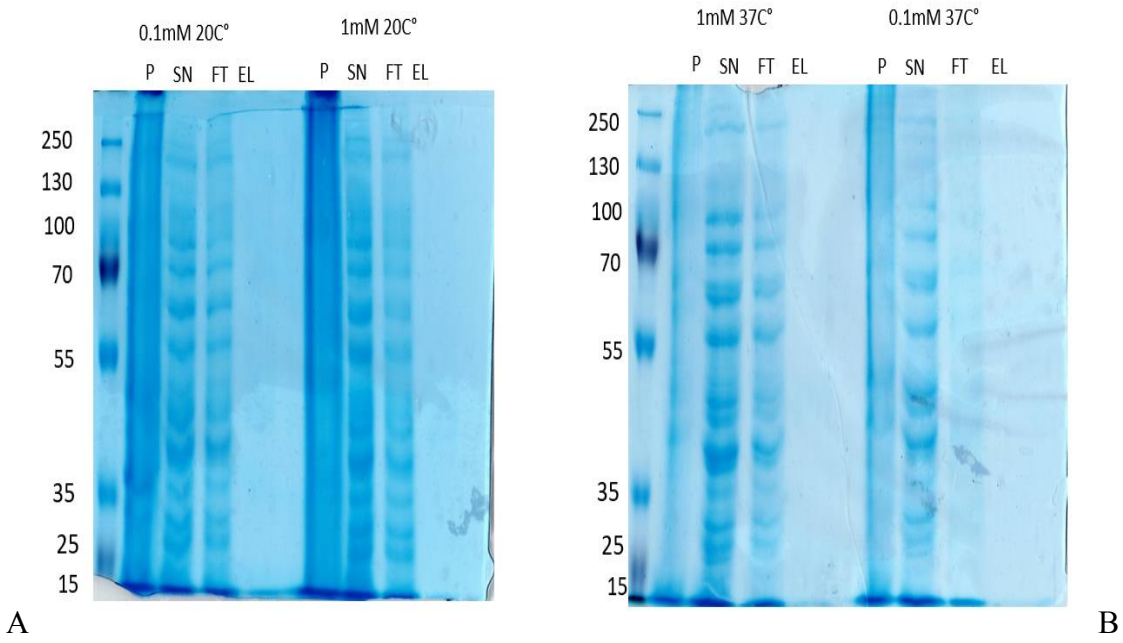


Figure 7. SDS-PAGE Gel 10% of ZRANB3 FL WT large scale expression in BL21 cells. Condition A: 20°C O.N. with 0.1 mM,1mM IPTG. Condition B: 37°C O.N. with 0.1 mM,1mM IPTG

3.5. Production of SRDct

The target regions for amplification of the SRD-ct domain was (721 to 1079bp). Figure 24A shows the agarose gel of the amplification products (lanes 2-3). The presence of clear, distinct bands confirmed successful PCR amplification of the target DNA., as well as the intensity of the bands suggested absence of nonspecific products. Results from colony PCR are highlighted in Figure 24B, which clearly show successful experiment, through the comparison with the DNA ladder on the left. Indeed, the presence of bands in the colony PCR lanes indicates that the colo-

nies contain the SRDct insert. On the contrary, variations in band intensity could reflect differences in the efficiency of DNA extraction from the colonies or the amount of template DNA. The samples were then sent to GENEWIZ and the sequence quality was assessed using fastQ files, showing high-quality reads with a typical minor decline towards the end. After confirming the integrity of the Gibson cloning, I transformed SRDct in BL21 cells for expression purpose (Figure 24C). Therefore, I prepared 10ml of LB with KANA for preculture. After 24h incubation, the cells were ready for growth and induction with IPTG. After the OD reached 0.8, I set-up 4 conditions for small-scale expression tests (see below).

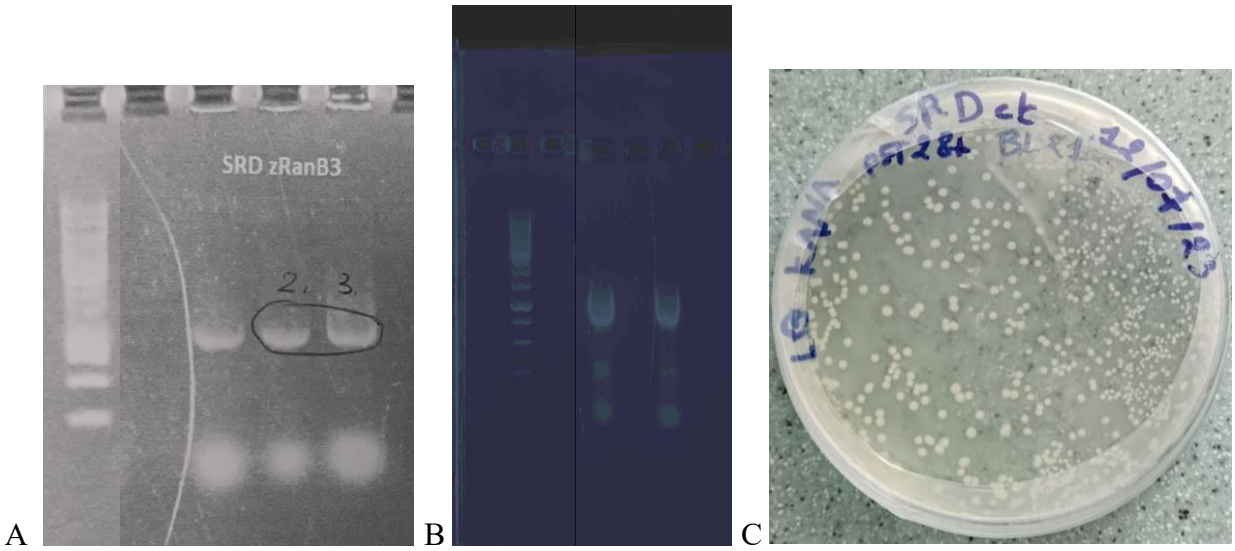


Figure 24. A. DNA gel Electrophoresis of and SRD domain. B. Colony PCR Gel of SRDct. C. BL21 colonies with Pet28+ SRDct

Small-scale expression

Figure 25 shows the SDS-page gel relative to the four different expression conditions, varying IPTG concentration and temperature/time of expression. At 20°C overnight with 0.5 mM IPTG, a strong band at approximately 45 kDa indicates robust expression of SRDct. Increasing the IPTG concentration to 0.5 mM results in a noticeable decrease in expression levels. When shifting to 37°C for 4 hours, protein expression is lower for both 0.1 mM and 0.5 mM IPTG, with the higher concentration showing a significant reduction, likely due to protein destabilization and aggregation. This last is highly observable in the Pellet and Total lysate. It is also possible to see that protein stacked within Nickel beads in condition of 0.1 IPTG in 20C ON. These results suggest that the

optimal conditions for SRDct expression are at 20°C overnight with 0.5 mM IPTG and 37°C overnight with 0.1 mM providing the best balance for efficient protein production without causing destabilization or aggregation.

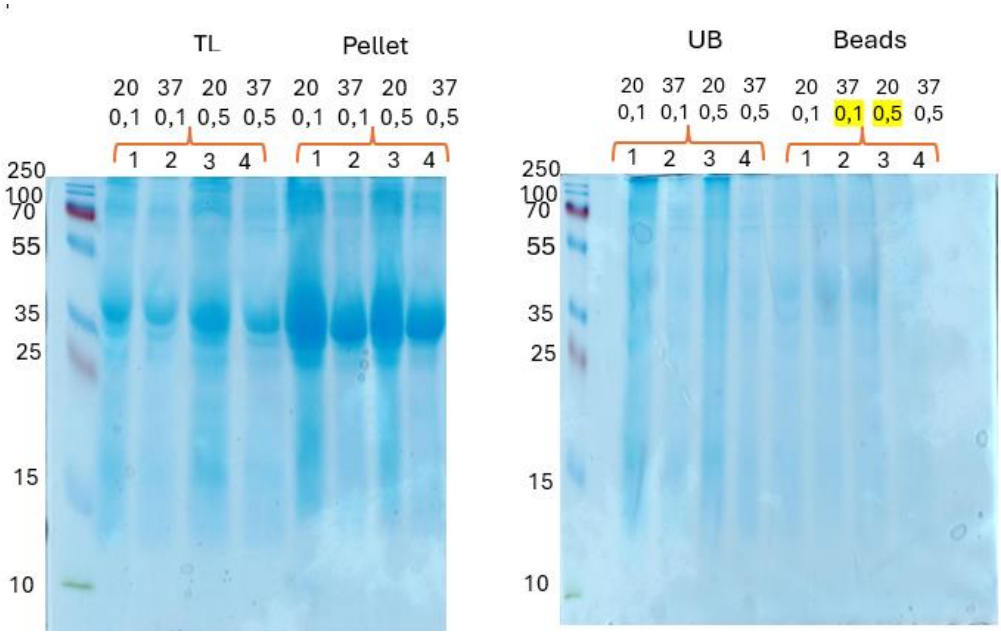


Figure 25. SDS-PAGE Gel of SRDct Small-scale expression.

Large-scale expression

I've used 1L of LB for better and larger protein expression. As observed in Figure 26A, the band relartive to the MW of SRDct are in the flow-through (FT) and supernatants (SN), indicating possible aggregation or disorder/flexibility. The majority of the protein remained unbound. Interestingly, treatment with n-lauroylsarcosine, a mild detergent, helped protein solubilization, resulting in the appearance of a small band (Figure 26B). This suggests that the protein might have been partially insoluble or aggregated, and the detergent helped to maintain it in a more soluble and stable form.

In conclusion, the large-scale expression and purification of SRDct faced challenges due to inefficient binding and potential solubility issues. Optimization of purification conditions and solubilization treatments, like the use of n-lauroylsarcosine, are essential for slightly improving yield and quality of the target protein. Despite this, this process needs further optimization in the case of structural analyses.

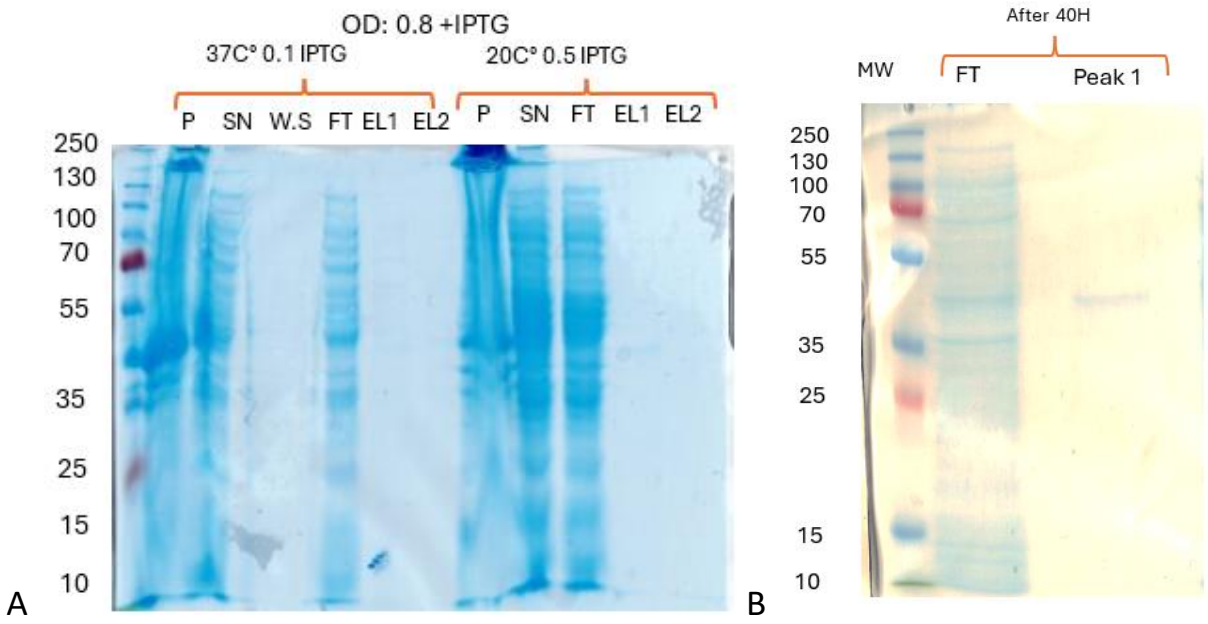


Figure 26. A. SDS-PAGE Gel of SRDct Large scale expression. B. SDS-PAGE Gel of SRDct large scale expression after 40h of treatment with n- lauroylsarcosine.

3.5. Production of RecA-like 1/2

The target region for amplification of the RECA1/A2 domain was from base pairs 38 to 472. Using SnapGene software, four primers were designed for each construct, with two primers targeting the vector and two targeting the insert. Agarose gel electrophoresis was performed on the amplification products to verify the successful PCR amplification of the RECA1/A2 insert. The presence of clear, distinct bands on the gel confirmed successful amplification, with the DNA ladder on the left serving as a size reference to determine the lengths of the PCR products. Variations in band intensity might reflect differences in DNA extraction efficiency or the quantity of template DNA used. The presence of these distinct bands allowed the identification and selection of positive colonies for further examination. The amplified samples were confirmed by GENEWIZ next-generation sequencing. Therefore, I transformed BL21 cells for expression purposes (10ml of LB with KANA antibiotics for preculture).

Small-scale expression

Figure 27 shows the SDS-PAGE gel relative to the four different expression conditions, varying IPTG concentration and temperature/time of expression. At 20°C overnight with 0.1 mM IPTG, a strong band at approximately 49 kDa indicates robust expression of RecA1/A2. Increasing the

IPTG concentration to 0.5 mM at 20°C resulted in a slight decrease in expression levels. When shifting to 37°C for 4 hours, protein expression was lower for both 0.1 mM and 0.5 mM IPTG, with the higher concentration showing a significant reduction, likely due to protein destabilization and aggregation. This is particularly evident in the Pellet and Total lysate samples. Additionally, the protein appears to be effectively bound within Nickel beads under the condition of 0.1 mM IPTG at 20°C overnight. These results suggest that the optimal conditions for RecA1/A2 expression are at 20°C overnight with 0.1 mM IPTG, providing efficient protein production without causing destabilization or aggregation.

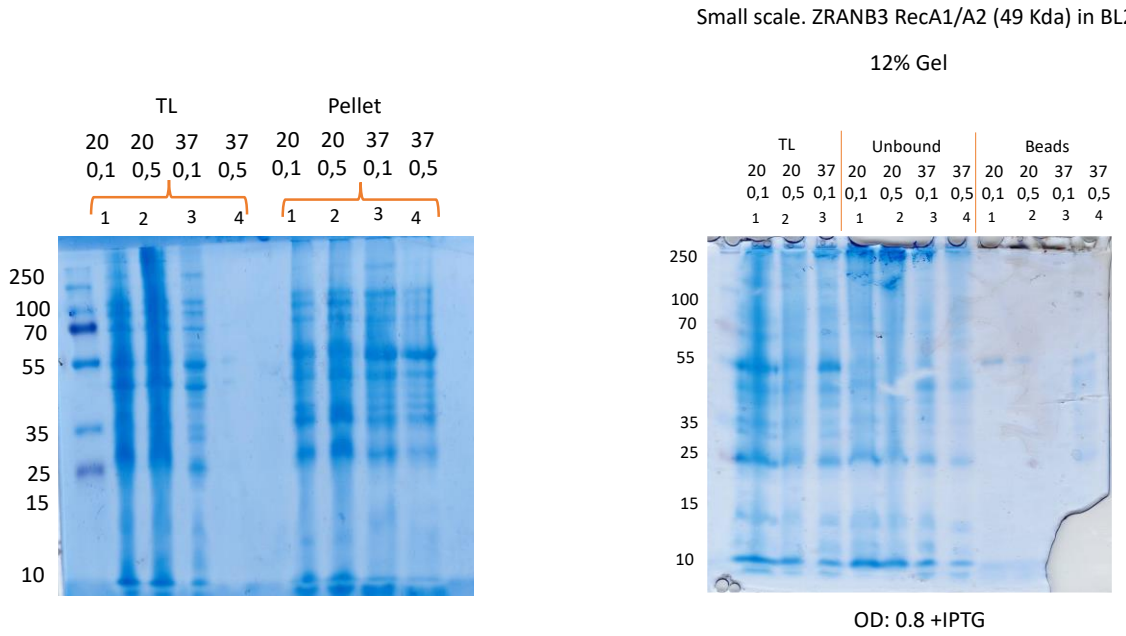


Figure 27. SDS-PAGE Gel of RecA1A2 Small scale expression

Large scale expression

For large scale production (1L of LB), condition of 0.1 mM IPTG at 20°C overnight was selected. In the first purification trial protein was aggregated and eluted in the Pellet. Therefore, I used 8M urea to solubilize inclusion bodies or aggregates (Figure 28). However, even with the urea treatment, the majority of the protein remained unbound, as evidenced by the strong bands in SN and FT. The example chromatogram from IMAC TALON affinity step is shown in Figure 29. The overall yield and purity of the protein were still not optimal. Repeated attempts, including the use of ion exchange chromatography, n-lauroylsarcosine and urea treatment did not fully resolve these problems. Further optimization of purification conditions and solubilization treatments is essential to improve the yield and quality of the target protein for structural analyses.

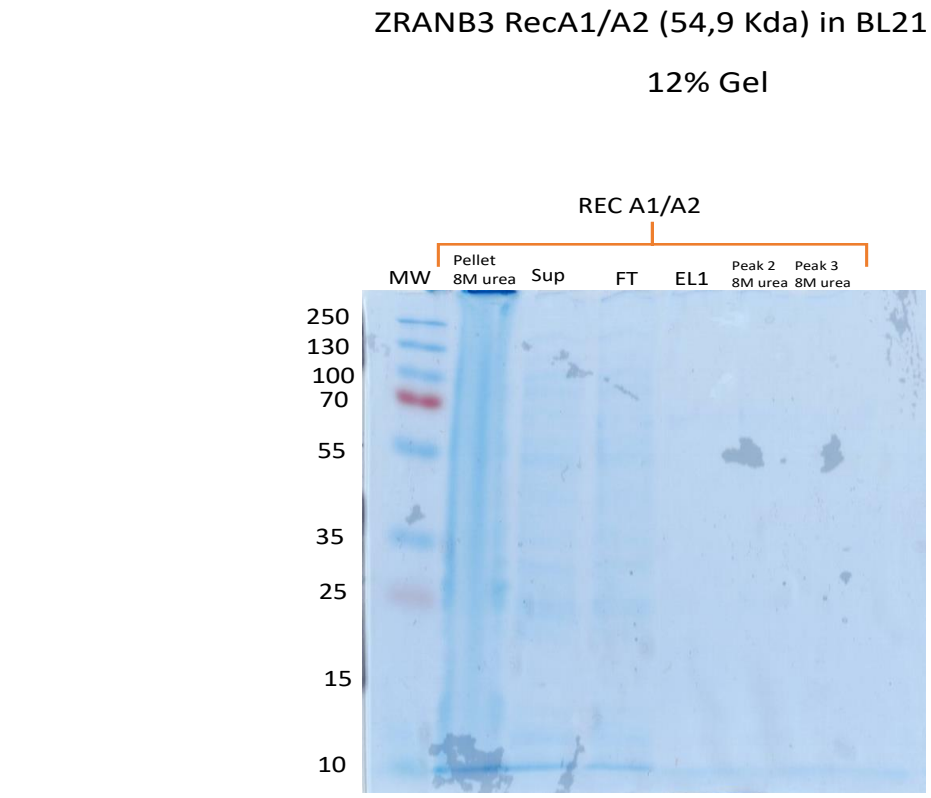


Figure 28. SDS-PAGE Gel of RecA1A2 Large scale expression

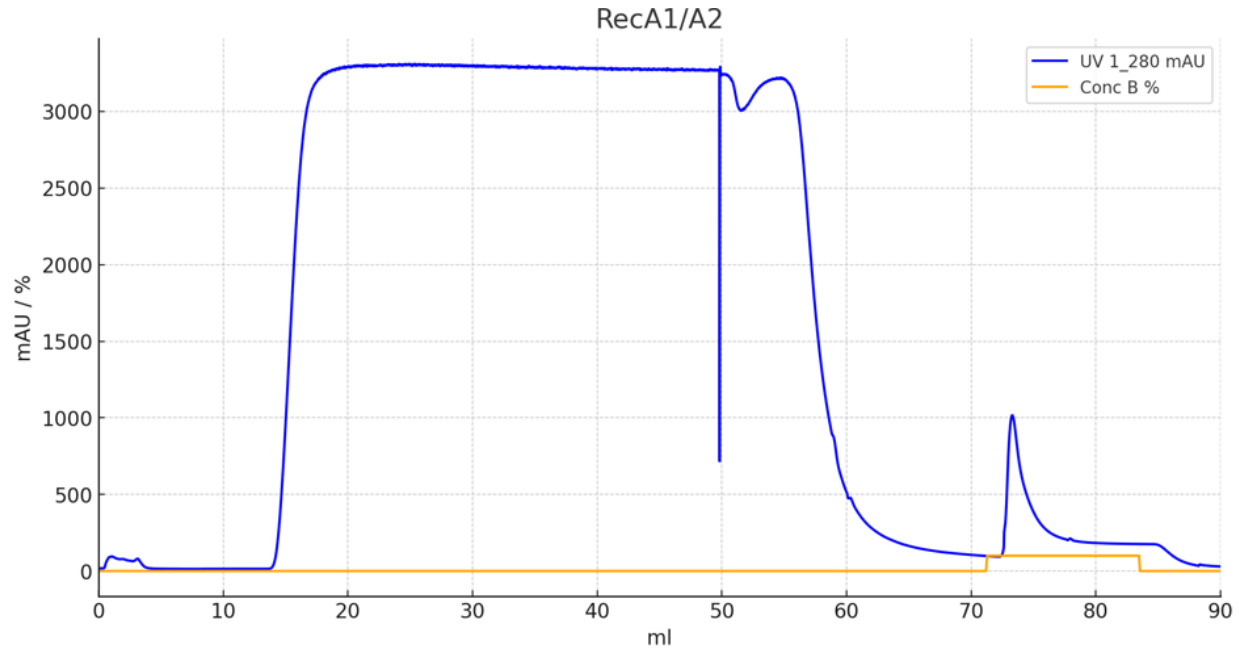


Figure 29. Example of IMAC AKTA pure Chromatograph from RecA1A2 purification

3.6. Cloning of ZRANB3 WT FL in yeast P. *pastoris*

The cloning of ZRANB3 WT FL into the yeast *Pichia pastoris* was achieved using the Gibson assembly method, with the gene inserted into the vector plasmid pPICZalphaB to enable protein expression in yeast. Amplification of the construct into the pPICZalphaB vector (Figure 30A) and colony PCR were successful (Figure 30B).

The agarose gel displayed distinct band at the expected size of the ZRANB3 WT FL insert, confirming the successful incorporation of the gene into the plasmid. Further confirmation of successful cloning was obtained through sequencing. Given that the Sanger sequencing technology used has a read length limitation (typically around 800-1000 bases), it was unable to read the entire length of the full-length of the construct, which exceeds 3000 bases. To overcome this, the sequencing was performed in overlapping segments, and the remaining sequence of the construct was manually checked. High-quality sequencing reads resulted 98% predominantly, confirming the accuracy of the cloned ZRANB3 WT FL sequence. Due to the relative short time of the Erasmus+ and to the absence of an electroporator at that time it was not possible to proceed further with tha P. *pastoris* transfection.

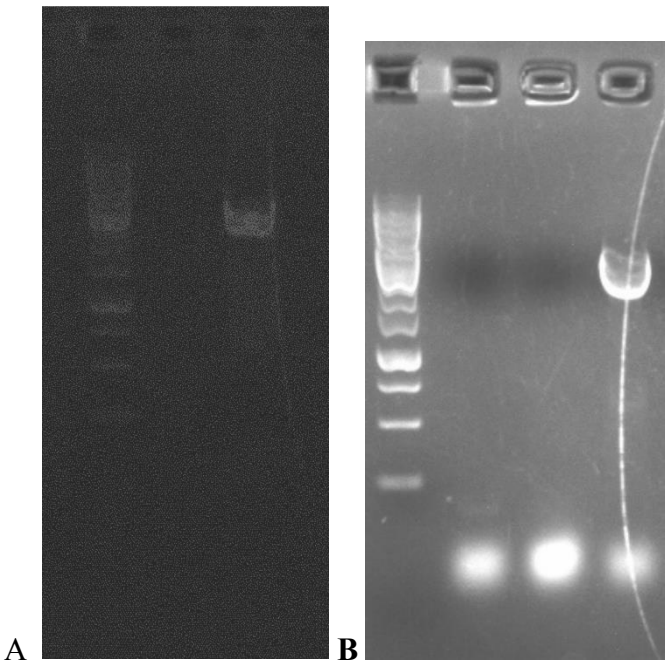


Figure 30. A. DNA amplification of FL WT Zranb3. B. Colony PCR Gel of the pPICZalphaB plasmid.

4. CONCLUSIONS

In this work I focused on the fork remodeler SNF2-family ATPase translocase ZRANB3. ZRANB3 is characterized by one ATPase motor domain composed by two RecA-like lobes, two PCNA binding motifs, the PIP-box and the APIM, the Substrate recognition domain (SRD) and the nuclease domain (NZF). The entire architecture allows the protein to bind to the dsDNA and move in the direction of the replication fork progression, while redirecting protein partners to the site of lesion and remodeling chromatin structure also through the cleavage of the 3'-end ssDNA.

Overall, this renders ZRANB3 a crucial epigenetic/epigenomic factor during DNA repair, particularly DNA Damage Tolerance (DDT). Indeed, ZRANB3 participates in translesion synthesis (TLS) through the activation of HLTF-dependent PCNA monoubiquitylation, template switching (TS) and homologous recombination (HR) through the D-loop formation, and replication fork reversal (RFR) through the formation of the 4-way chicken junction. A failure of its action, due to environmental causes (UV, drugs, etc..), epigenetic or genetic alterations, often results in cancer development, suggesting an oncosuppressor role for this protein. The lack of sufficient functional and, especially, structural information pushed us to start a preliminary work aimed at understanding the molecular basis of ZRANB3 action in the cell.

Within this context and during my internship, I have characterized the architecture of ZRANB3 using sequence-sequence, sequence-structure, structure alignment tools and motif scan predictors. Through the use of threading and ab-initio algorithms I've modelled the 3D-structure of the FL WT protein, as well as few functionally-critical constructs implicated in the dsDNA translocation (the RecA1/A2 domain) and in the substrate recognition (the SRDct). The analysis of homologous protein-DNA interfaces highlighted specific putative DNA binding residues, which are positively charged and some of them found mutated in cancer cells.

In parallel to this, I've conducted wet lab work cloning the FL WT and the two constructs in bacterial plasmid pET28+ and testing their expression in BL21 cells and in small-scale. Unfortunately, the transition to large-scale production was not successful and lead to the production of insoluble aggregates that were possible to disaggregate only in small amount. However, the cloning of the FL WT ZRANB3 into the yest vector pPICZalphaB-SapL3 for protein production and secretion in P. *pastoris* was successful. This will be use to produce high yield of protein and in the more physiological fold, ensuring the presence of all post-translational modifications (PTMs) to be further used for structural investigations.

BIBLIOGRAPHY

Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002.

Aparicio, T., Ibarra, A., & Méndez, J. (2006). Cdc45-MCM-GINS, a new power player for DNA replication. Cell Div, 1, 18.

Badu-Nkansah, Akosua et al. (2016). Identification of a Substrate Recognition Domain in the Replication Stress Response Protein Zinc Finger Ran-binding Domain-containing Protein 3 (ZRANB3). The Journal of Biological Chemistry, 291(15), 8251-7.

Chang, A.Y., Chau, V., Landas, J.A., & Pang, Y. (2017). Preparation of Calcium Competent Escherichia coli and Heat-Shock Transformation. Journal of Experimental Microbiology and Immunology, 1, 22-25.

Cipolla, Lina et al. (2016). The Regulation of DNA Damage Tolerance by Ubiquitin and Ubiquitin-Like Modifiers. Frontiers in Genetics, 7, 105.

Dürr H, Körner C, Müller M, Hickmann V, Hopfner KP. (2005). X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. Cell, 121(3), 363-73.

Follonier, C., Oehler, J., Herrador, R., & Lopes, M. (2013). Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions. Nat Struct Mol Biol, 20, 486-494.

Gibson, D., Young, L., Chuang, RY. et al. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods, 6, 343–345.

Kile, A. C. et al. (2015). HLTF's ancient HIRAN domain binds 3' DNA ends to drive replication fork reversal. Mol Cell, 58(6), 1090-1100.

Leung, Wendy et al. (2018). Mechanisms of DNA Damage Tolerance: Post-Translational Regulation of PCNA. Genes, 10(1), 10.

Lemacon, D. et al. (2017). MRE11 and EXO1 nucleases degrade reversed forks and elicit MUS81-dependent fork rescue in BRCA2-deficient cells. Nat Commun, 8, 860.

Neelsen, K. J. et al. (2013). Deregulated origin licensing leads to chromosomal breaks by rereplication of a gapped DNA template. Genes Dev, 27, 2537–2542.

Poole LA, Cortez D. (2017). Functions of SMARCAL1, ZRANB3, and HLTF in maintaining genome stability. Crit Rev Biochem Mol Biol, 52(6), 696-714.

Puccetti MV, Adams CM, Kushinsky S, Eischen CM. (2019). Smarcal1 and Zranb3 Protect Replication Forks from Myc-Induced DNA Replication Stress. Cancer Res, 79(7), 1612-1623.

Quinet, A., Lemacon, D., & Vindigni, A. (2017). Replication fork reversal: players and guardians. Mol Cell, 68, 830-833.

Sambrook, J., & Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual. 3rd Edition, Vol. 1, Cold Spring Harbor Laboratory Press, New York.

Schmid, J. A. et al. (2018). Histone ubiquitination by the DNA damage response is required for efficient DNA replication in unperturbed S phase. Mol Cell, 71(6), 897-910.e8.

Scully, R., Panday, A., Elango, R., & Willis, N. A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. Nat Rev Mol Cell Biol, 20, 698-714.

Vujanovic M, Krietsch J, Raso MC, et al. (2017). Replication Fork Slowing and Reversal upon DNA Damage Require PCNA Polyubiquitination and ZRANB3 DNA Translocase Activity. Mol Cell, 67(5), 882-890.e5.

Weston R, Peeters H, Ahel D. (2012). ZRANB3 is a structure-specific ATP-dependent endonuclease involved in replication stress response. Genes Dev, 26(14), 1558-72.