

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwigs-Maximilians-Universität München

**Structural and Biochemical
Studies on Holliday Junction
Recognition by the SLX4-MUS81
Complex**

Carina Christine Vraschek

aus

Offenbach am Main, Deutschland

2018

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Frau Prof. Elena Conti, PhD betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 13.08.2019

Carina Christine Vraschek

Dissertation eingereicht am 13.08.2018

- | | |
|----------------------|----------------------------|
| 1. Gutachterin | Prof. Elena Conti, PhD |
| 2. Gutachter | Prof. Dr. Andreas Ladurner |
| Mündliche Prüfung am | 10.10.2018 |

Contents

Summary	1
1 Introduction	3
1.1 DNA Damage and Repair	3
1.2 DNA Double-Strand Break Repair	5
1.2.1 The Significance of Double-Strand Breaks	5
1.2.2 The Sensing of DSBs	6
1.2.3 Non-Homologous End Joining	7
1.2.4 Homologous Recombination and Its Subpathways	9
1.3 The SLX4-SLX1-MUS81-EME1-Complex	14
1.3.1 Domain Architecture of SLX4	15
1.3.2 The Interaction Partners of SLX4	17
1.3.3 The MUS81-EME1 Nuclease	22
1.3.4 Cleavage Properties of the SM-Complex	25
1.3.5 Regulation of the SM-Complex	27
1.3.6 Comparison Between the Human and the <i>C. elegans</i> SM-Complex	28
1.3.7 Involvement of the SM-Complex in Genome Integrity and Disease	29
1.4 Aim of the Thesis	31
2 Results	33
2.1 Expression and Purification of the Human SM-Complex	33
2.1.1 Purification of hsMUS81-EME1	33
2.1.2 Expression and Purification of hsSLX4 and hsSLX1	38
2.2 Phosphorylation Site Mapping of hsSLX4 ¹³⁰⁰⁻¹⁸³⁴	42
2.2.1 Phosphorylation sites of hsSLX4 ¹³⁰⁰⁻¹⁸³⁴	44
2.2.2 Phosphorylation Site Mapping under the Influence of DNA Damage	49
2.3 Purification of the <i>C. elegans</i> SM-Complex and Crystallization Attempts	50
2.3.1 Purification of the Full-Length <i>C. elegans</i> Proteins	50
2.3.2 Limited Proteolysis of ceSLX4 and ceMUS81-EME1 and Crystallization Trials	52
2.3.3 Crystallization Trials of ceSLX4 ^{SAP}	56
2.3.4 Crystallization Trials of ceSLX4 ^{SAP} in Complex with Interaction Partners	58
2.4 Structural Analysis of ceSLX4 ^{SAP} and ceMUS81 ^N by NMR	59
2.4.1 HSQC analysis of ceSLX4 ^{SAP} and ceMUS81 ^N	59
2.4.2 NMR-Model of ceSLX4 ^{SAP}	62

2.5	Biochemical Characterization of MUS81-EME1 and SLX4 ^{SAP}	68
2.5.1	DNA Cleavage Activity of MUS81-EME1	68
2.5.2	Interaction between ceSLX4 ^{SAP} and ceMUS81 ^N	70
2.5.3	DNA Binding Properties of ceSLX4 ^{SAP} and ceMUS81 ^N	75
2.5.4	Mapping of the ceSLX4 ^{SAP} DNA Binding Region	85
3	Discussion	89
3.1	Reconstitution of the SM-Complex <i>in vitro</i>	89
3.2	Post-Translational Modification of hsSLX4	90
3.3	The ERCC4 Typ Nuclease MUS81-EME1	92
3.4	Interaction between ceMUS81 ^N and ceSLX4 ^{SAP}	94
3.5	Biological Implications for the DNA binding of ceSLX4 and ceMUS81	96
3.6	A Modified HJ Resolution Model	100
3.7	Open Questions and Outlook	101
4	Materials and Methods	103
4.1	Materials	103
4.1.1	Consumables and Chemicals	103
4.1.2	Media	103
4.1.3	Bacterial Strains and Cell Line	104
4.1.4	Antibiotic Solutions	104
4.1.5	Vectors and Plasmids	104
4.1.6	Oligonucleotides	107
4.1.7	Buffer for Large Scale Plasmid DNA Preparation	109
4.1.8	Buffers for Protein Purification	110
4.1.9	Equipment	112
4.1.10	Bioinformatic tools and software	112
4.2	Methods	114
4.2.1	Cloning Procedures	114
4.2.2	Protein Expression	119
4.2.3	Protein Purification	122
4.3	Biochemical Assays	127
4.3.1	Determination of Protein Concentration and Purity	127
4.3.2	Polyacrylamide Gel Electrophoresis	127
4.3.3	Native Gel Electrophoresis	127
4.3.4	Limited Proteolysis	127
4.3.5	Mass Spectrometry Analysis	128
4.3.6	Treatment with DNA Damaging Agents	128
4.3.7	Phosphosite Identification by MS	129
4.3.8	<i>In vitro</i> Kinase assay	130
4.3.9	Cleavage Assay	130
4.3.10	Electrophoretic Shift Assays	131
4.3.11	SEC-based Protein-Protein Interaction	132
4.3.12	Static Light Scattering	132

4.3.13 Analytical Ultracentrifugation	133
4.3.14 Thermal Shift Assay	133
4.3.15 Fluorescence Anisotropy	133
4.3.16 Nano Differential Scanning Fluorimetry	134
4.3.17 Crystallization Procedures	135
4.3.18 NMR Spectroscopy	136
4.3.19 Structural Model Calculation	137
Acknowledgements	139
Abbreviations	142
Appendix	147
List of Figures	160
List of Tables	163
Bibliography	165

Summary

The genomic integrity of a cell faces many challenges through various agents, that damage DNA in many ways. Double-strand breaks (DSBs) are particularly serious as they can lead to gross genome rearrangements, loss of heterozygosity and apoptosis. Depending on cell cycle stage and availability of repair factors different ways exist in order to repair the breaks. One pathway is homologous recombination, which uses sister chromatids as intact template for repair. In this repair process the two homologous DNA strands are connected in a structure called Holliday Junction (HJ), which has to be resolved before the cell divides. The resolution can be carried out by structure-specific endonucleases, like SLX4-SLX1 and MUS81-EME1. These four proteins form a complex (SM-complex) and resolve the HJ in a nick and counter-nick mechanism, where SLX1 introduces the first cut and MUS81-EME1 introduces the second cut. SLX4 is a large protein comprising several domains and serves as a platform protein interacting with several other proteins involved in DNA repair or cell cycle control. So far it has not been fully understood, how SLX4 coordinates the nucleases SLX1 and MUS81-EME1 on the HJ substrate. This study presents a biochemical and structural characterization of the SM complex with a focus on the SLX4^{SAP} and MUS81 N-terminal domain. I could express and purify *C. elegans* SLX4^{SAP} and MUS81^N from bacterial cells and by testing different DNA substrates, resembling intermediates in double-strand break repair, I showed that both domains are promiscuous DNA binding domains. DNA binding of ceSLX4^{SAP} turned out to be in a cooperative mode, as has been reported also for other SAP domains. Earlier studies showed that MUS81^N and SLX4^{SAP} interact with each other. Here, the binding affinity between them could be quantified. The interaction with MUS81^N was abolished in a mutant of SLX4^{SAP}, which also revealed a 3.5-fold reduced DNA binding. The mutational analysis demonstrated that the interaction between the DNA-binding and the MUS81-interaction area is only partially overlapping. The DNA binding site was further delimited by shorter constructs. The systematical truncations revealed, that contrarily to expectations the most conserved part of the SAP domain among several SLX4 orthologs was alone not sufficient for DNA binding. In band shift assays with the SAP domain alone, residues from 467 to 560 were needed for DNA binding. Interestingly in complex with MUS81^N the SLX4⁴⁶⁷⁻⁵³⁰ construct showed stronger binding than the individual constructs. An NMR model of SLX4^{SAP} could be derived and was structurally compared to SAP domains from other proteins and displayed a high flexibility at the C-terminus. Structural and biochemical comparison showed that SLX4^{SAP} is a typical SAP domain with hydrophobic residues between the helix-loop-helix fold and positively charged residues with the invariant glycine in the loop region as well as cooperative DNA binding. Based on the comparison with SAP domains from other proteins, the SLX4^{SAP} is an important component of the SM-complex for the stabilization of the interaction between SM-complex and

DNA. Including the results in this thesis, the current model for HJ cleavage by the SM-complex can be extended. While the first cut on the HJ is carried out by SLX1, SLX4^{SAP} stabilizes the HJ and interacts with the N-terminal domain of MUS81. This interaction and stabilization leads to the second cleavage reaction carried out by MUS81-EME1. The involvement of two nuclease active sites and several DNA-binding domains lead to the disentanglement of the HJ and contributes to the genomic maintenance.

1. Introduction

1.1 DNA Damage and Repair

In every living organism, DNA with its four different nucleotides carries the genetic information for the formation of proteins, cellular structures and whole organisms. It encloses the genetic code for the development, functioning and survival of all known living organisms and many viruses. Every time the cell divides, the DNA has to be duplicated in a process called replication and the genetic information has to be preserved and passed on to the daughter cells. Defects in the genome impair DNA metabolism, e.g. replication, transcription and recombination. Missing or defective repair of the damage can lead to chromosomal rearrangements, fusion, deletion, mutations or chromosomal loss. Various internal and external DNA damaging factors put the integrity of the genome under pressure. This damage can sum up from 10^4 to 10^6 DNA damaging events per cell per day [1]. Taken together, the stability of the cell's genome faces many threats from endogenous and exogenous sources and has to make sure each damage is repaired before the next cell division.

Endogenous sources for DNA damage can be reactive oxygen species or intermediate products from metabolic processes (e.g. aldehydes), but also errors resulting from replication. Exogenous sources are for example UV and X-ray radiation as well as chemotherapeutic drugs. The genomic integrity is harmed on different levels by damage to the nucleobases of DNA (e.g. abasic sites, mismatches, insertions / deletions, oxidative modifications, methylations), damage to the DNA backbone (chemical adducts, single-strand breaks, double-strand breaks) and interstrand crosslinks (ICLs) [1]. ICLs are characterized by covalent links between the bases on opposite strands, which can lead to a blockage of replication. Additionally, protein adducts are another toxic DNA lesion, which arise from the covalent linkage of proteins (e.g. topoisomerases) with DNA [3]. For each type of DNA damage particular routes of repair have evolved, e.g. base excision repair, mismatch repair, nucleotide excision repair, double-strand break repair (Figure 1.1 A) [1, 4, 5]. Multiple pathways, dealing with a wide range of DNA lesions, have been characterized in the last decades of DNA damage research and are reviewed in [1, 2, 4, 6].

The DNA repair pathways, forming the DNA damage response (DDR), are highly conserved from bacteria to humans [7–9]. Together they prevent cell cycle arrest, apoptosis of cells and uncontrollable cell growth (Figure 1.1 B). The cell cycle is divided in several stages, G₁, S, G₂ and M. The passage from one stage to the next stage is highly regulated and depends on cell cycle checkpoints, which make sure that progression through the cycle is only occurring, when certain conditions are fulfilled. Currently, three checkpoints are known, G₁ checkpoint, G₂/M checkpoint and metaphase checkpoint, which all sense defects during the process, induce an arrest

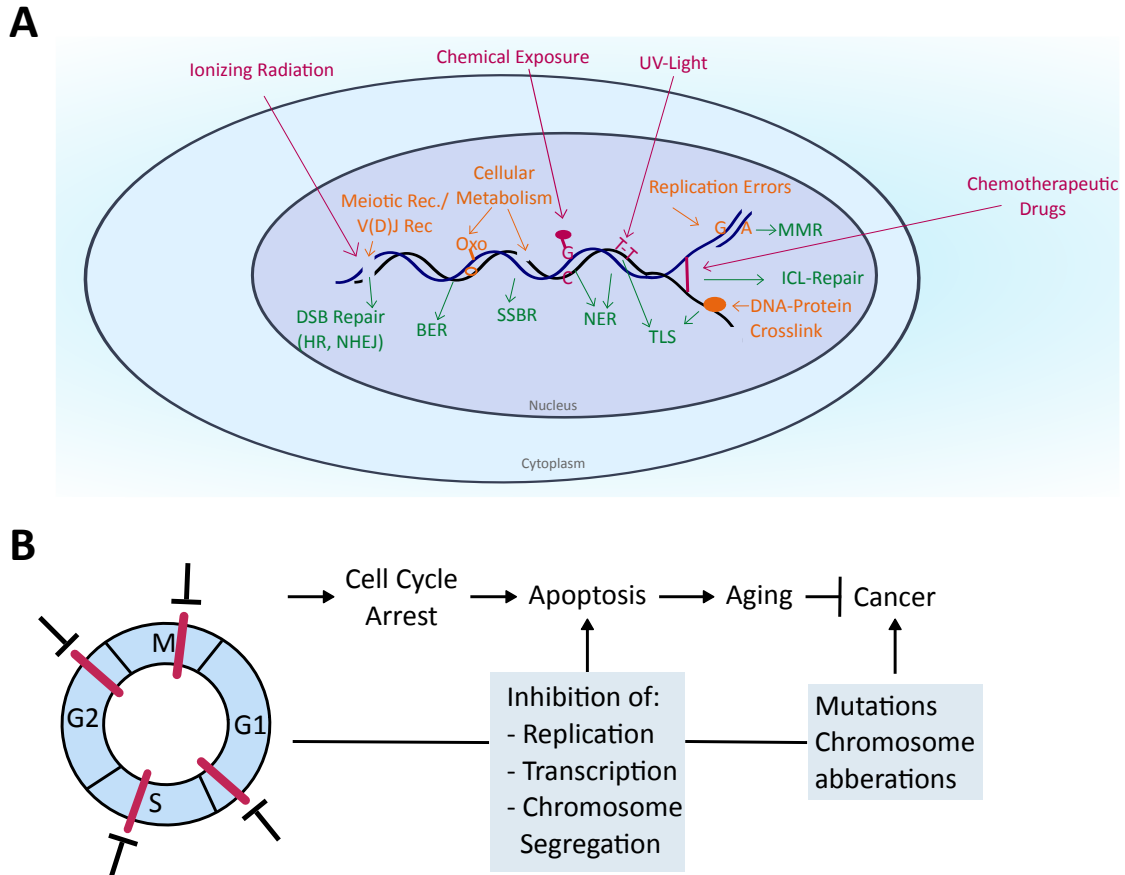


Figure 1.1: The complexity of DNA damage repair. (A) DNA damage is induced by various endogenous and exogenous sources. Common exogenous DNA damaging agents are UV radiation, chemicals and chemotherapeutic drugs. Endogenous damaging agents originate from the chemical instability of the nucleotides, cellular metabolites and stalled replication forks. For the different lesions several pathways have evolved to counteract the DNA lesions. For example base modifications, like 8-oxo-guanine arising from oxidation of guanine or uracil emerging after cytosine deamination, are repaired by base excision repair. Stalled or collapsed replication forks and ionizing radiation can generate strand breaks. (B) Adopted from [2]. Defects in the DNA can lead to a halt at different cell cycle stages (G_1 , S, G_2 , M phases). Furthermore, DNA lesions defer DNA metabolism and can lead to cell death, if replication and transcription can not be carried out in a timely manner due to unresolved DNA lesion. Too many double-strand breaks can lead to apoptosis initiation. Increased apoptosis can contribute to premature aging. On the other hand the development of cancerous cells is driven, if cells do not undergo apoptosis and accumulate DNA errors. HR Homologous recombination; NHEJ Non-homologous end joining; VDJ Variable diverse joining recombination during antibody diversification; BER Base excision repair; SSBR Single-strand break repair; NER Nucleotide excision repair; TLS Translesion synthesis; MMR Mismatch repair; ICL Interstrand crosslink

and regulate the activity of cell cycle kinases, the cyclin-dependent kinases. Cyclin dependent kinases ensure that all phases of the cell cycle are executed in the correct order. An erroneous DNA repair can lead to a halt in cell cycle progression, which in turn means that the repair of DNA lesions has to be carried out before the cell enters mitosis [10–12]. Cells have evolved complex signaling pathways to arrest the progression of the cell cycle in presence of DNA damage, thereby providing increased time for repair and tolerance mechanisms to operate [4, 7]. This has also an effect on DNA metabolism and long-term DNA injuries lead to permanent changes in the DNA sequence [2] (Figure 1.1 B).

To summarize, the DNA damage response is essential to prevent premature aging and the development of cancer. Failures of the repair proteins can give rise to inherited disease, e.g. Fanconi anemia (FA) [13, 14], Xeroderma pigmentosum [15] and Werner syndrome [16, 17]. In the case of too many DNA lesions, cells in multi-cellular organisms initiate programmed cell death and eliminate themselves from the population [4, 18].

1.2 DNA Double-Strand Break Repair

1.2.1 The Significance of Double-Strand Breaks

Double-strand breaks (DSBs) are one of the most deleterious type of DNA damage as they can lead to disruption of genetic information and chromosomal aberrations, causing cancer. It was demonstrated in yeast cells that even a single DSB left unrepaired could lead to cell death [19–21]. DSBs can arise in various ways, including as a result of spontaneous acts like the collapse of stalled replication forks during S phase, exposure to DNA damaging agents and repair of incomplete telomeres after failure of telomerases [5]. DSB repair is important for the repair of replication-associated damage [22] and as a subsequent step in the repair of interstrand crosslinks [23]. In any way, the repair of DSBs is highly regulated as the cell has to ensure that repair starts at the right time, place and in the right manner. Furthermore, double-strand breaks do not only arise from erroneous acts, but can also be programmed by the cell [24]. Programmed DSBs are carried out in order to generate genetic diversity, which plays a role in meiosis and in the immune system. During prophase of meiosis I programmed DSBs are produced by the protein SPO-11 [25–27]. In meiosis I chiasmata are formed, which result from the crossing-over between homologous chromosomes [28]. These physical attachments between homologous chromosomes are essential for the accurate separation in meiosis I [28]. Without the formation of the chiasmata the chromosomes can not be segregated properly leading to infertility and miscarriages in humans as well as birth defects in newborns [29]. Hence, homologous recombination fulfills two functions in meiosis, it generates genetic diversity among the progeny derived from common parents and ensures proper chromosome segregation [5]. The vertebrate immune system employs V(D)J recombination to rearrange variable (V), diverse (D) and joining (J) gene segments leading to diversification of antigen-binding regions in the immunoglobulins and T-cell receptors [30]. In this process, the lymphoid specific RAG (recombination activating gene) proteins initiate the recombination and introduce a double-strand break at specific recognition sequences [30–33]. Furthermore, DSBs are created during immunoglobulin heavy chain class switching [34]. In this process the constant region of the antibody is changed in order to generate antibodies with different effector function

[34]. Thereby, the exons encoding the initially expressed IgH constant region is replaced by a set of downstream exons, which changes the antibody class from e.g. IgM to IgG1 [34].

Eukaryotic cells possess two major repair pathways for DSBs, non-homologous end joining (NHEJ) and homologous recombination (HR) (Figure 1.3 A, B) (NHEJ reviewed in [35–37] and HR is reviewed in [8, 12, 26, 38, 39]). Under certain conditions, DSBs can also be repaired by microhomology mediated end joining (MMEJ) and single-strand annealing (SSA) [40]. The pathways are complementary and operate optimally under different circumstances [41]. The single steps involved in the DSB repair were described firstly by Szostak et al. [42]. In the following subsection the DNA double-strand break repair will be outlined.

1.2.2 The Sensing of DSBs

In order to protect the cell from genome instability, DNA damage has to be sensed first. The recruitment of the different DDR factors to the DNA lesion is initiated by sensor proteins, which recognize the lesion and activate the DDR [7, 16]. Several factors influence repair options, e.g. cell cycle (availability of homologs), extent of damage and availability of factors. Thus, the choice of repair options is provided by the order and the timing of the recruitment [16]. In general, the assembly of the DDR response depends on a broad spectrum of posttranslational modifications, e.g. phosphorylation, ubiquitination, poly(ADP-ribosylation) and others [16].

Double-strand breaks are detected by at least four independent sensors: Ku70/80, PARP (poly(ADP-ribose) polymerase), RPA (replication protein A) and MRN (MRE11-RAD50-NBS1) [16], which decide about the exact DSB repair pathway being taken and the outcome of the repair process.

DSBs are rapidly bound by the Ku70/80 heterodimer, which loads onto the DNA ends [16]. Yeast Ku is proposed to counteract and prevent extensive DNA-end resection, which favors NHEJ [36, 43]. There is evidence, that the DNA damage response protein p53-binding protein 1 (53BP1) can work together with the replication timing regulator factor 1 (RIF1) and both block end resection [36, 44]. Thus, in G1 phase, NHEJ is favored by more than 50-fold for the repair of DSBs owing to the level of Ku and the suppression of extensive end resection by CtIP (CtBP-interacting protein; CtBP C-terminal binding protein) and MRN [36]. Even during S- and G₂ phase the resection machinery has to overcome the presence of Ku at DNA ends either by outcompeting Ku or by processing of DNA ends to a point at which Ku binding is less favored [36, 45].

Members of the PARP family are activated by the DSB and catalyze the addition of poly-ADP-ribose chains on proteins to recruit DDR factors to chromatin at breaks [16]. PARP promotes the microhomology mediated end joining pathway, which is a backup to the classical NHEJ pathway [16]. PARP is thought to mediate the initial accumulation of the MRN complex [16]. Limited DSB resection then leads to association of CtIP in the MMEJ pathway [16].

The MRN-complex is recruited to the DSB at the beginning of the repair process and consists of three proteins, MRE11 (meiotic recombination 11), RAD50 (radiation sensitive) and NBS1 (Nijmegen breakage syndrome 1). [46, 47]. MRN localizes at the DSB ends via RAD50 [48] and maintains the tethering of the DSB ends to each other via the coiled-coil tails of RAD50 [49]. In association with CtIP and BRCA1 (breast cancer associated gene 1), MRN mediates initial stages of DSB resection to promote HR in S and G₂ phase [16]. Extensive resection and formation of 3'

ssDNA ends leads to RPA accumulation, which stabilizes ssDNA regions [16]. The formation of RPA-ssDNA complexes plays an important role in the activation of the kinase pathway.

Following the recognition of the DNA lesion by sensor molecules in HR, the master response is then the activation of the ATM (ataxia telangiectasia mutated) and ATR (ATM and Rad3-related) proteins [46], whereupon ATM is primarily activated by DSBs and ATR responds to a broad spectrum of DNA damage [46]. ATR and its partner protein ATRIP (ATR interacting protein) is activated following recruitment to RPA-coated ssDNA regions at stalled replication forks and DSBs [16]. The activation of ATM is regulated by MRN, where the NBS1 subunit associates with ATM [16, 46]. ATM is proposed to phosphorylate mediator proteins, like CHK2 (Checkpoint kinase 2) and p53 as well as several other DDR factors [7, 16]. In the chromatin, flanking the DSB, ATM phosphorylates histone variant H2AX [46]. The phosphorylated H2AX is required for the accumulation of more DNA repair proteins and chromatin-remodelling complexes around the DSB [46]. After the recruitment of the first DNA damage sensors, the decision between various possibilities has to be made, e.g. apoptosis, immune surveillance or DNA repair [16].

1.2.3 Non-Homologous End Joining

In the decision for a repair of the DSB, there are different options for the repair pathway possible. One pathway is the non-homologous end joining pathway (NHEJ), which mediates the direct ligation of broken DNA ends and usually involves minimal DNA end processing [41] (Figure 1.2 A). In a first step the DNA ends are bound by the Ku70/Ku80 heterodimer, which recruits and activates the (DNA-PKcs) [51] forming the DNA-PK complex. DNA-PKs stabilize the DSB ends and prevent end resection [52, 53]. Through sequential phosphorylation events DNA-PK is stabilized at the DNA ends and gives access to end processing enzymes, like ARTEMIS [16]. ARTEMIS will cut DNA substrates at the boundaries between single-stranded and double-stranded DNA [36, 54, 55]. Another nuclease, which might contribute to the repair of DSBs, is APLF (aprataxin and PNKP (Polynucleotide kinase/phosphatase)-like factor) [36]. DNA Polymerases λ and μ of Pol X-family polymerases interact with Ku and incorporate dNTPs [36]. The terminal deoxynucleotidyltransferase (TdT) is another Pol X family member, but it is only expressed in early B-lymphocytes and T lymphocytes during V(D)J recombination [36]. Ligation is then carried out by the XLF (XRCC4-like factor)-XRCC4 (X-ray complementing chinese hamster gene 4)-Ligase IV complex [56] (Figure 1.2 B). The alignment and the base pairing of the free DNA overhangs are critical steps to initiate the ligation [35]. Additionally, PAXX (paralogue of XRCC4 and XLF) interacts with Ku and it might be involved in the ligation of the NHEJ subpathway for blunt ends [36]. NHEJ is flexible towards the modification of free DNA ends, which makes it susceptible for small insertions and deletion [35, 36] (Figure 1.2 B). NHEJ does not need a template for repair and can operate throughout the cell cycle [36]. However, the direct ligation of the DNA ends harbors the risk of ligating the wrong ends and / or deleting DNA sequences [57]. Furthermore, NHEJ during DNA replication can be deleterious because of the intrinsic increased incidence of breaks, whose false joining could lead to chromosomal rearrangements [57–59]. When the HR machinery is defective, the NHEJ-mediated mutagenic repair can turn into a major contributor to genomic instabilities and tumorigenesis [57, 60, 61]. An alternate end-joining pathway comes into place, when classical NHEJ is impeded, and relies on micro-homologies for the joining reaction, involving

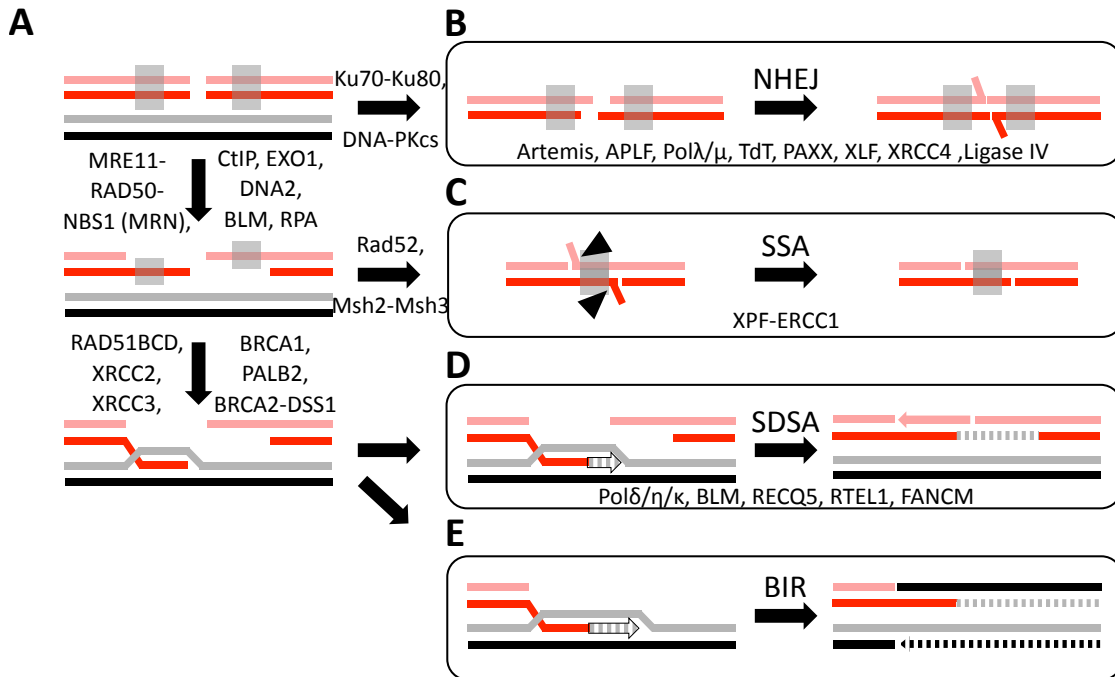


Figure 1.2: Pathways in double-strand break repair. Homologous recombination pathway steps from resection to D-loop formation (A), Non-homologous end joining (B) and different pathways choices of Homologous recombination (C - E). (A) In the Homologous recombination (HR) pathway the 5' ends of the broken ends are resected in order to expose single-stranded DNA. ssDNA is covered by replication protein A (RPA), which is replaced by RAD51 in the next step. Then RAD51-dependent homology search and strand invasion forms a displacement loop (D-loop). Thereby DNA synthesis is primed from the 3'-OH end of the broken chromosome on an intact template. The following process of second-end capture and HJ formation is described in Figure 1.3. (B) Non-homologous end joining (NHEJ) DNA protein kinase (DNA-PK) consists of the heterodimer Ku70 and Ku80 and the DNA-PK catalytic subunit (DNA-PKcs). This pathway is alternative to homologous recombination and does not require 5' end resection. (C) The single-strand annealing pathway (SSA) uses flanking homology regions to bridge the DNA lesion [50]. This causes a deletion between the repeats and leads to mutagenesis. The ssDNA left over after annealing of the strands are being removed by the XPF/ERCC1 nuclease. (D) Synthesis-dependent strand annealing (SDSA) repairs the break by extending the D-loop and subsequently disrupts and reanneals it to the second end. SDSA leads to a non-crossover outcome. This process involves several DNA polymerases and helicases. (E) Break-induced replication (BIR) is employed after a replication fork encountered a nick. Then a single-ended double-strand break is formed. A different way of resolving the single-ended double-strand break is described in Figure 1.3. Overall, BIR can lead to a loss of heterozygosity. Adopted from [39].

MRN complex, PARP-1, DNA Polymerase δ [36], XRCC1 and DNA Ligase I or II [41, 62]. The usage of micro-homologies at the lesion in the alternate NHEJ pathway increases errors [35].

1.2.4 Homologous Recombination and Its Subpathways

HR allows the accurate repair of DSBs by using a sister chromatid as a template. 5' end resection and formation of a D-loop are the intermediate steps leading to the formation of Holliday Junctions, which have to be processed [42, 63]. As HR requires the presence of a homologous template and is less error-prone compared to NHEJ, it is the preferred repair pathway in S and G2 phase [64]. A prerequisite for the homologous recombination pathway is the resection of the broken DNA ends [16, 61]. Extensive resection is regulated by the cell cycle through cyclin-dependent kinases (CDKs) [36, 65]. These factors promoting end resection are more active during S and G2 phases, favoring HR when a sister chromatid is present [16, 36, 65]. In G1 phase end resection is suppressed, in order to give way for a DSB repair via NHEJ [50].

DSBs are sensed by the MRN complex, which promotes the activation of ATM and the preparation of DNA for HR or one of its subpathways [16] (Figure 1.2 A). MRN complex and CtIP generate in an initial resection reaction 3' ssDNA tails [36, 38]. For this, CtIP is phosphorylated by cyclin dependent kinases [44, 50, 66–68] and associates with BRCA1 [16]. Further end resection to generate longer ssDNA stretches is carried out by the Bloom syndrome RecQ-like helicase (BLM-helicase) together with Exonuclease 1 (EXO1) [16, 69]. Additionally, the DNA replication helicase/nuclease 2 (DNA2) has been shown to extend the 3' ssDNA tails [65, 69, 70]. The single-stranded DNA tails are covered by multiple copies of RPA complex [36, 71]. The RPA-nucleofilament allows a protection of the ssDNA and prevents secondary structure formation [36]. In the later course of HR, RPA will be replaced by the RecA homologue RAD51 to allow for homology search and strand invasion [36, 72].

Single-Strand Annealing (SSA)

The single-strand annealing pathway (SSA) is a RAD51-independent pathway (Figure 1.2 C). ssDNA sequences generated during DSB processing contain regions of homology at both sides of the DSB, which can be annealed and ligated [36, 73, 74]. First, extensive bidirectional resection reveals complementary ssDNA [73]. The complementary ssDNA tails are then annealed by mediation of the strand annealing protein RAD52 [36, 73, 75]. For removal of unannealed non-homologous portions of the 3' ssDNA the nucleotide excision repair complex XPF-ERCC1 (xeroderma pigmentosum group F - excision repair cross complementation group 1) and the mismatch repair complex MSH2-MSH3 (MutS Homolog 2-MutS homolog 3) are employed [36, 65, 73, 76, 77]. In yeast, it has been shown that Slx4 (synthetic lethal of unknown function 4) interacts with Rad1/Rad10, the yeast ortholog of XPF-ERCC1 [78, 79]. By nuclease action the intervening region and one of the repeats is deleted [73]. This pathway is error-prone and mutagenic as deletions and translocations can be easily generated [36, 73].

Strand Exchange and D-loop Formation

After the first step in HR with the 5' to 3' resection of the DNA ends and the generation of 3' single-stranded DNA [61, 65], the ssDNA is bound by the ssDNA-binding protein RPA [61]

(Figure 1.2 A, second arrow). BRCA1 promotes the recruitment of BRCA2 to the DSBs via the bridging protein PALB2 (partner and localizer of BRCA2) [61]. Next, BRCA1 and BRCA2 load the RAD51 recombinase onto the ssDNA to form a nucleoprotein filament [61]. BRCA2 thereby specifically promotes RAD51 filament assembly on ssDNA [5, 28, 61]. The interaction between RAD51 and BRCA2 is limited to S and G2 phase of the cell cycle by CDK-dependent phosphorylation of BRCA2 [16]. RAD51 belongs to a conserved class of recombinase enzymes, which mediate DNA pairing and strand exchange [72, 80]. Different RAD51 complexes are formed, which consists of various combinations of RAD51 paralogs [72, 80]. These pairs mediate RAD51 filament assembly [72, 80]. The displacement of RPA by RAD51 stimulates strand exchange [61, 74]. RAD51-nucleofilament promotes strand invasion into a homologous duplex to initiate repair synthesis [61, 74]. The homologous duplex is preferentially the identical sister chromatid [61, 74]. In the intact strand a Displacement-loop (D-loop) is formed [5, 42].

Arising from the formation of the D-loop, different subpathways are possible: synthesis-dependent strand annealing (SDSA), Break-induced repair (BIR) (Figure 1.2 D, E) or the formation of double Holliday Junctions (HJs) (Figure 1.3 A) [12].

Synthesis-Dependent Strand Annealing (SDSA)

During SDSA the newly synthesized DNA dissociates to anneal to the other 3' DNA end of the broken chromosome and double-strand break repair is completed by ligation, resulting in a non-crossing over product [73, 81, 82] (Figure 1.2 D). The Rad51 formed D-loop is extended by the replicative DNA Polymerase δ with loaded PCNA (proliferative cell nuclear antigen) [83]. This process is stimulated by the translesion polymerases Pol η and Pol κ [83]. SDSA is favored by the displacement of the invading strand from the D-loop [73]. In higher eukaryotes several helicases are involved in this process [73]. The BLM helicase along with TopIII α topoisomerase promotes non-crossover formation [84]. The RecQ5 helicase was shown to promote SDSA by disrupting RAD51-ssDNA filaments and thus counteracting the inhibitory effect of RAD51 on strand annealing activity of RAD52 [73, 85]. Yeast Mph1, the ortholog of FANCM (Fanconi anemia complementation group M) is proposed to dissociate the D-loop and promote the formation of the non-crossover product independent of the yeast BLM ortholog Sgs1 (Slow growth suppressor 1) [73, 86]. However, it has been suggested that another helicase, RTEL1 (regulator of telomere length 1) disrupts D-loops and promotes SDSA *in vivo* after displacement of the RAD51 filament and D loop dissociation [87]. The helicases involved, mediate the dissociation of the invading strand from the D-loop in order to avoid crossover formation [82]. Following D-loop unwinding, the nascent strand has also the opportunity to anneal with sequences other than those found at the other end of the broken chromosome, which can contribute to the mutagenesis associated with SDSA [88].

Break-Induced Repair (BIR)

The D-loop structure can assemble into a replication fork and copy the chromosome arm in a process called break-induced replication (BIR) [73, 89–91] (Figure 1.2 E). BIR is carried out by a migrating D-loop (bubble), where branch migration of an unresolved HJ promotes the displacement of a newly synthesized strand and therefore leads to non-crossover repair [92–94]. The BIR repair

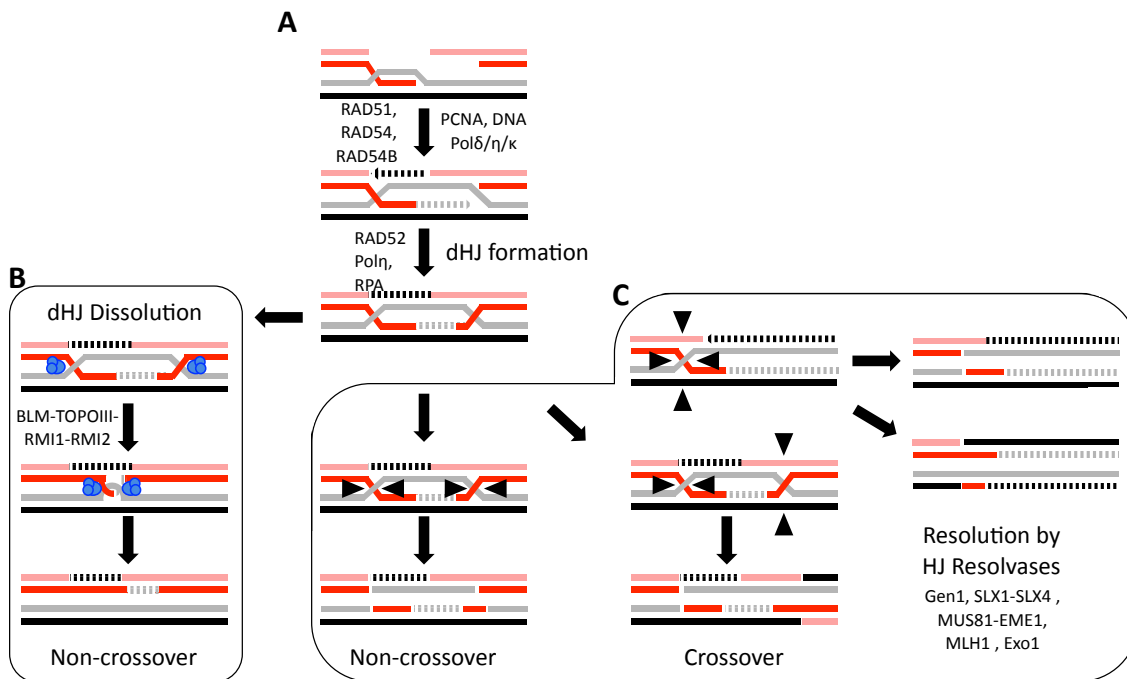


Figure 1.3: Processing of Holliday Junctions Homologous recombination after D-loop formation (A) ssDNA is covered by RAD51 and a RAD51 dependent homology search is performed. Upon strand invasion a D-loop is formed and DNA synthesis is primed on an intact template. After second-end capture the displaced strand of the D-loop anneals to the other resected 3' strand forming two nicked HJs and after ligation a double HJ. (B) The double HJ can be dissolved by the BLM/TopIII α /RMI1/2 complex leading to a non-crossover product. (C) Another way to process dHJs is the resolution by structure-specific endonucleases. This cleavage is precisely coordinated and leads to crossover or non-crossover products depending on the incision. During BIR the intact strand can be displaced, which has to be resolved, leading to a crossover or a non-crossover product. Adopted from [39].

pathway is especially important in cases of one-ended chromosomal breaks caused by replication fork collapse or shortening of telomeres [88, 91]. In telomerase-deficient cells BIR is an alternative pathway to maintain the chromosome ends [91]. Because the BIR repair pathway is a kinetically slow process it can be outcompeted by SDSA [73, 95]. However, as soon as BIR is initiated, the synthesis rate becomes comparable with normal replication [73, 95]. In contrast to SDSA the amount of DNA synthesis is much more extensive [88] and can proceed via a migrating bubble for hundreds of kilobases [92]. To date, most mechanistic insight into BIR is derived from studies on *Saccharomyces cerevisiae*. BIR has an additional need for replicative helicases (e.g. GINs complex, CDC45 and the MCM complex) [88] and DNA Polymerase α , Pol32 and the 3' to 5' DNA helicase Pif1 [93, 94, 96]. Pol32 is a non-essential subunit shared by Pol δ and the translesion polymerase ζ [96]. So far, the exact functions of Pol32 and Pif1 are not well understood, but in the BIR model of DSB repair Pol32 could act as a processivity factor of Pol δ and Pif1 might be important for the progression of the mobile D-loop [93, 96]. BIR fully restores the integrity of the chromosome, but it can lead to a loss of heterozygosity distal to the DSB [12, 97].

Formation and Processing of Holliday Junctions

In order to accommodate the invading and the donor ssDNA strand in the D-loop RAD51-dsDNA filaments are formed [28] (Figure 1.3 A). DNA synthesis is primed from the intact strand using the invading 3' end as primer [98]. Thereby, RAD51 dissociates from dsDNA to expose 3'-OH required for DNA synthesis [28]. DNA synthesis during D-loop extension is carried out by translesion polymerase η [99] and κ as well as the replicative polymerase δ [83, 88]. Human polymerase δ together with PCNA, a trimeric complex that encircles the DNA and prevents DNA polymerases from dissociation, can extend the D-loop up to 2 kb [83]. In this process, DNA strands have to be opened, where RAD54 is involved as a DNA translocase and as supercoiling motor [72]. RAD54 is an ATPase that is stimulated by RAD51 and generates positive supercoils ahead of protein movement and negative supercoils trailing it [72]. Supercoiling facilitates the homologous pairing reaction of the RAD51 nucleofilament and further DNA branch migration [72]. RAD54 has the ability to remove RAD51 from dsDNA [5, 100] and to remodel chromatin [72]. A RAD54 paralog has similar biological activities as RAD54 and plays a stronger role during meiosis [5].

After D-loop extension by Pol δ , Pol η and Pol κ [83], the current model proposes that the D-loop is stabilized by capturing the second end of the DSB, leading to the formation of Holliday Junctions (HJ) [42, 63] (Figure 1.3 A, lower part). The formation of HJ is important during meiosis, when crossovers lead to divergence of the genetic pool [12]. However, in somatic cells dHJs are an intermediate in recombinational DNA repair [101] and formation of dHJs can be blocked by RAD51 suggesting an inherent bias towards SDSA [12]. SDSA is preferred in mitosis leading to non-crossover products [28]. After D-loop formation, it is proposed that the displaced DNA strand is extended towards the sequences that are homologous to the second processed DNA end [102]. The displaced strand would then anneal to the ssDNA tail created by the resection at the other end of the break [42, 63, 102]. Subsequent DNA synthesis from this captured second end, branch migration and ligation generate two four-way structures composed of DNA strands base paired with old and new partners simultaneously [102]. The occurrence of two four-way structures is called double Holliday Junction (Figure 1.3 A).

So far, the whole process of the HJ formation, the involved proteins and their function is not well understood. It is known, that Pol η , RAD52 and RPA collaborate to promote D-loop extension, second-end capture and *de novo* DNA synthesis from the 3' terminus of the captured second DSB [103]. On the second resected end RPA is still bound to ssDNA and RAD52/RPA mediated annealing leads to the capture of the second DNA end, which in turn serves as a template for the initiation of DNA synthesis by Pol η [103]. In total, RAD52 is involved in the process of second-end capture [103–105]. The second-end capture forms one of the key intermediates of homologous recombination: the four-way DNA (Holliday) Junction.

The formation of Holliday Junctions is the second-last step in DNA double-strand break repair [42]. HJs were first proposed by Robin Holliday as four way branched DNA substrates [63]. They can be formed as a result of regression of stalled replication forks [106] and can be disentangled by two mechanisms. HJs can be either dissolved in a non-nucleolytic way by the BLM-TopIII α -RMI1/2 helicase-topoisomerase complex (Bloom - DNA Topoisomerase III α -RecQ mediated genome instability protein 1/2; yeast: Sgs1-Top3-Rmi1/2) [84, 107] or resolved by structure-selective endonucleases (MUS81-EME1, SLX1-SLX4, GEN1) [39, 108–115] (Figure 1.3

B, C).

The dissolution is initiated by the migration of the HJs towards each other [116] (Figure 1.3 B). The collapse / fusion of the two HJs forms a hemicatenated structure, whose decatenation leads to the formation of a non-crossover product [116]. In somatic cells, the dissolution pathway is favored for HJ removal as it prevents sister chromatid exchanges [84]. These sister chromatid exchanges are to be prevented as they could lead to loss of heterozygosity, which in turn incites cancer predisposition [108, 117–120]. Mutations in human BLM helicase give rise to Bloom’s syndrome, a genetic disorder associated with genomic instability and early-onset cancer [17, 84, 121]. The BLM helicase belongs to the family of RecQ helicases and forms an evolutionary conserved complex with human topoisomerase III α [84, 122]. TopoIII α , a type 1A topoisomerase [116], breaks and rejoins DNA to alter its topology [84]. In the decatenation, TopoIII α requires the ATPase activity of the BLM helicase to execute the reaction [84, 107, 116]. Branch migration activity is exhibited by the BLM helicase causing torsional stress [116]. This torsional stress is being alleviated by the action of the topoisomerase [116]. Besides BLM-TOPO3, the cofactors RMI1 / RMI2 are needed for efficient dissolution [116]. RMI1 interacts with BLM and TopoIII α and strongly stimulates the dissolution reaction, where it primarily functions in the decatenation step [107, 116]. RMI2 interacts with RMI1 and is important for the functionality of complex [116]. Interestingly, the BLM helicase carries out different functions during DSB repair. BLM can act on various substrates and dissolve various recombination intermediates. The ssDNA production after DNA damage is carried out by BLM [70, 123]. Furthermore, BLM displaces RAD51 from ssDNA [105] and dissolves in complex with a topoisomerase and RMI1/2 the HJ substrate as the last step of homologous recombination [84, 107].

The resolution of HJs is carried out by an ubiquitous but diverse group of metal ion-dependent endonucleases called Holliday Junction resolvases or structure-selective nucleases (see reviews [124–126]) (Figure 1.3 C). In general the resolvases are characterized by their ability to 1) bind Holliday Junctions in a structure-specific manner; 2) introduce a pair of symmetric nicks across the junction point; 3) generate nicks with 5’-phosphate and 3’ hydroxyl termini that are readily ligatable [127]. In most cases junction binding is associated with a distortion of the stacked-X conformation of the HJ to a more open structure [127–130]. The HJ resolution mechanism is very well studied for bacterial and bacteriophage resolvases [124, 131]. For some HJ resolvases, like bacteriophage T4 endonuclease VII and T7 endonuclease, simply the binding to the stacked X-conformation of the HJ is sufficient to trigger cleavage [125]. Some resolvases, e.g. *E. coli* RuvC (resistance to UV light), are not only structure selective, but the cleavage is specific and prefers a cognate sequence at the junction point [132, 133]. The RuvC resolvase might depend on the relocation of the Holliday Junction by branch migration so that the required sequence can be localized at the junction point [127]. The ATP dependent branch migration of the HJ is catalyzed by RuvAB, which associates with RuvC [127, 134].

In eukaryotes the resolution is more complex and involves different resolvases, which are structurally not related to each other. The human GEN1 (gap endonuclease 1) protein and the *S. cerevisiae* ortholog Yen1 were identified by mass spectrometry following fractionation of HeLa cell extracts and screening a yeast gene fusion library for HJ resolution activity, respectively [111]. GEN1 is a member of the Rad2/XPG (Xeroderma pigmentosum complementation group G) family of structure specific nucleases, characterized by an N-terminal and an internal XPG nuclease

domain and a helix-hairpin-helix domain [111]. HJs are cleaved by GEN1 in a symmetrical way leading to a ligatable product [111]. This cleavage mechanism is contrast to the asymmetric HJ cleavage seen for MUS81-EME1 (methyl methanesulfonate UV sensitive, clone 81 - essential meiotic endonuclease) [135], but it is functionally conserved with the bacterial HJ resolvase RuvC [111]. GEN1 activity is regulated by nuclear exclusion [117]. Thus, the cell makes sure, that GEN1 has access to recombination intermediates and cleaves them, when the nuclear membrane breaks down at mitosis [117, 136]. Mutations of GEN1 have been found in breast cancer patients, classifying *GEN1* as a cancer gene [111, 137]. GEN1 and SLX4 (synthetic lethal of unknown function 4) are synthetically lethal in human cells as a consequence of dysfunctional mitosis proceeding in the presence of unprocessed HJs [112]. In yeast, flies and humans GEN1 and MUS81-EME1 could be shown to work in redundant pathways [109, 138–140]. A knockout of them in a cellular background of BLM syndrome leads to severe chromosomal abnormalities [140]. In *C. elegans* this genetic interaction could not be detected [25].

The nuclease complex SLX4-SLX1-MUS81-EME1 (SM-complex) has activity on HJs [113–115, 141, 142]. The nucleases work in a nick and counter-nick mechanism with SLX1 (synthetic lethal of unknown function 1) introducing the first cut and MUS81-EME1 the second cut [141]. Recently, Wyatt et al. proposed that the complex forms a multi-nuclease complex consisting of SLX4-SLX1-MUS81-EME1-XPF-ERCC1 (SMX-complex) [143]. The cleavage mechanism of the SM-complex is unique among the Holliday Junction resolvases as the cleavage is carried out by two different nucleases. Usually, HJ resolution proteins bind four-way junctions in dimeric form and introduce symmetrical nicks [144]. For example, bacterial RuvC (Resistance to UV light) binds as a dimer on opposite faces of the junction [145, 146]. T4 endonuclease VII also forms a dimer in order to cleave the HJ [147–149]. The SM complex supplements the BLM-dependent double HJ dissolvase activity [112]. But even with an active BLM complex HJ resolution is essential and depends on SLX4 and GEN1 function [112]. The resolvases are responsible for the generation of crossovers as resolvase deficient cells exhibit reduced frequency of sister-chromatid exchange [150]. The following section will explain the properties of the SM complex and the domain architecture in more detail.

1.3 The SLX4-SLX1-MUS81-EME1-Complex

Originally, Slx4 was first discovered together with Slx1 in yeast synthetic lethal screens in absence of the dissolution complex Sgs1/Top3 (yeast ortholog to human BLM/TopoIII) [151]. The heteromeric complex was then functionally characterized as a structure specific endonuclease redundant to Sgs-Top3 [152]. Slx1-Slx4 form a complex and are involved in the replication of rDNA in yeast [152, 153]. The eukaryotic orthologs of yeast Slx4 were identified through several approaches [113–115, 154]. Two groups scanned for human SLX4 in a BLAST search using either a potential DNA binding domain or a conserved C-terminal domain as query [113, 115]. Another group's attention was drawn towards SLX4 as it was identified as a target of the ATM/ATR protein kinase [114]. The *Drosophila* SLX4 ortholog, MUS312, was also found by sequence comparison with other proteins containing a C-terminal conserved domain [154]. A rough domain assignment was carried out by sequence alignments with SLX4 orthologs from other species [113, 115], identifying five domains at the beginning (Figure 1.4 A).

Interaction partners were identified and their interaction site was roughly mapped by Yeast-two-hybrid screen [113–115, 154], immunoprecipitation [113–115] and proteomic analysis of SLX4 complexes [114]. These analysis revealed that SLX4 is a scaffold protein and through the assembly and coordination of several nucleases it is involved in diverse pathways of genome maintenance, including ICL repair, DNA replication, nucleolytic processing of HR intermediates, replication stress management and telomere maintenance [155] (Figure 1.4 B). Furthermore, a crystal structure of a yeast Slx1-Slx4 complex suggested how Slx1 is activated and it revealed how Slx1 binds on its interaction site in Slx4 [156] (Figure 1.5 A). The following subsections will give an overview of the domain architecture of SLX4 and its interaction partners as well as structural and biochemical information available about the heterodimers, SLX1-SLX4 and MUS81-EME1, which are the four constituents of the SM-complex.

1.3.1 Domain Architecture of SLX4

The 200 kDa human SLX4 protein, also called BTBD12, has a multi-domain structure, which allows binding of a wide range of interacting proteins and thereby controlling the activity at the target site. The initial domain organization was found by sequence alignments with other SLX4 orthologs by various groups [113–115] and was refined by interaction studies later on [157, 158].

At the N-terminus are two ubiquitin binding domains localized, UBZ-1 and UBZ-2 (Figure 1.4 A). These domains are important for recruitment of SLX4 to sites of replication-dependent interstrand crosslinks [14, 159–161]. UBZ-1 is required for the SLX4-dependent repair of ICLs and binds poly-ubiquitin chains with a preference for K63- over K48-linkages [14, 158–160]. Both UBZ domains are required for the resolution of Holliday Junctions under BLM depletion [160].

The MLR region was detected as a conserved motif required for the interaction between the *Drosophila* SLX4 ortholog MUS312 and the *Drosophila* XPF ortholog MEI9 [113]. It was thus named as MLR (MUS312-MEI9 interaction-like region) [113]. MEI9 (meiotic protein 9) is essential for meiotic recombination in *Drosophila melanogaster* [162]. The sequence in MUS312, which is known to associate with MEI9, is conserved among SLX4 orthologs and may serve as a general binding site for XPF [78, 113].

The BTB domain (Bric-a-brac, Tramtrack, Broad complex), sometimes also referred as POZ domain (pox virus and zinc finger) was originally identified in *Drosophila melanogaster* transcription regulators and pox virus zinc finger proteins [163]. BTB domains are in general responsible for protein dimerization or interaction with non-BTB-proteins [163]. SLX4^{BTB} might play a role in ICL repair, but its exact role remains elusive [158]. The dimer formation of SLX4^{BTB} is proposed to be important for the localization of SLX4 at genomic sites such as telomeres, the assembly and regulation of the nucleases and the foci formation of SLX4 at common fragile sites [157]. Another study showed by immunoprecipitation that SLX4^{BTB} mutants are deficient for SLX4 dimerization [164].

SIMs (SUMO-interacting motifs) in SLX4 are short stretches of hydrophobic residues flanked by acidic residues [164–166]. In general, they mediate noncovalent interaction with the small ubiquitin-like modifier (SUMO) and presence of multiple SIMs in a protein enables the protein to bind to poly-SUMO2/3 chains [166], which has been shown also for SLX4 [164, 165]. The SIMs promote the sumoylation of XPF and SLX4 itself and are important for managing replication

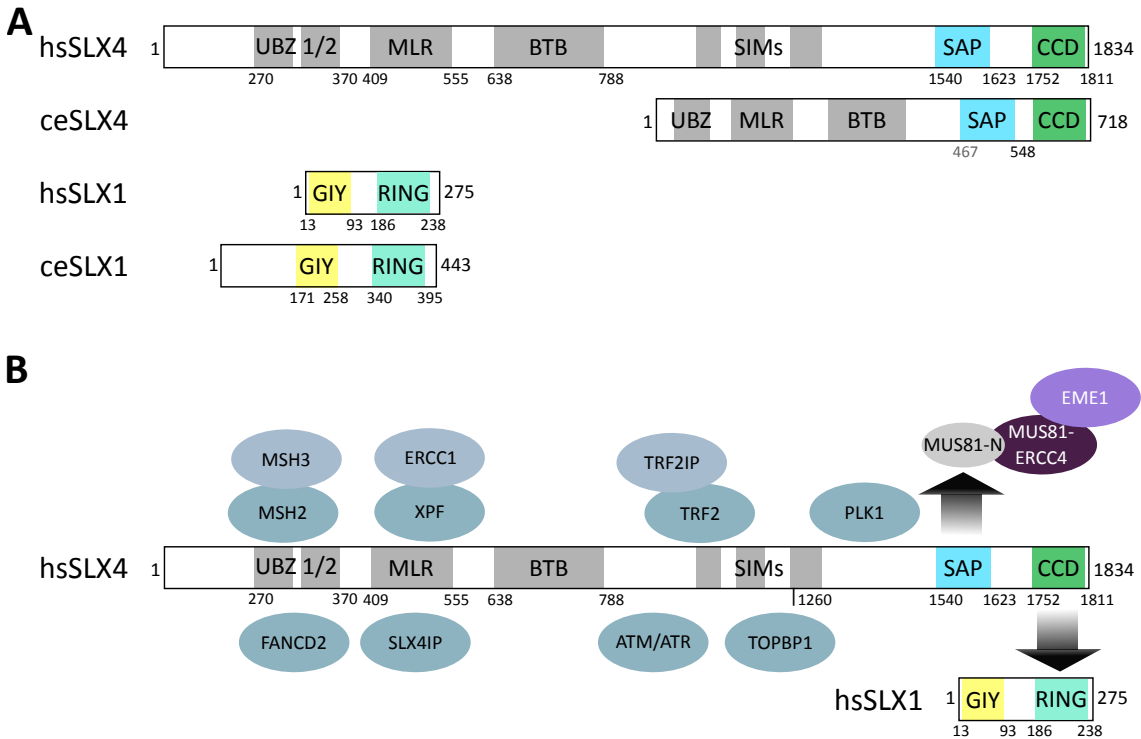


Figure 1.4: The SLX4 scaffold protein and its interaction partners. (A) Diagram showing the domain architecture of the human and *Caenorhabditis elegans* SLX4 and SLX1 proteins. UBZ: ubiquitin-binding zinc finger domain; MLR: MUS312-MEI9 interaction-like region; BTB: Bric a brac, Tramtrack, Broad-Complex domain; SIM: SUMO-interacting motif; SAP: SAF-A/B, Acinus and PIAS domain; CCD: conserved C-terminal domain; GIY: nuclease domain; RING: really interesting new gene (B) Interaction partners of SLX4 and their relative interaction site. MSH2/3 (MutS homologs 2/3) are involved in DNA mismatch repair. FANCD2 (Fanconi anemia complementation group D2) is involved in the repair of interstrand crosslinks. The so far uncharacterized SLX4 interacting protein (SLX4IP or C20orf94) interacts with the N-terminus of SLX4. The nuclease XPF-ERCC1 (xeroderma pigmentosum group F - excision repair cross complementation group 1) interacts via the MLR domain. The interacting kinase ATM / ATR (ataxia telangiectasia mutated / ataxia telangiectasia mutated and Rad3-related) is involved in sensing of DNA damage. SLX4 is involved in telomere maintenance through TRF2/TRF2IP (telomeric repeat-binding factor 2 / TRF2-interacting protein). The interaction with TOPBP1 is conferred via a single phosphorylation site at T1260. PLK1 (Polo-like kinase 1) is involved in cell cycle regulation. Through the C-terminal domains, SLX4 coordinates the activity of the Holliday Junction resolvases MUS81-EME1 (methyl methanesulfonate UV sensitive, clone 81 - essential meiotic endonuclease 1) and SLX1 (synthetic lethal of unknown function 1).

stress at genomic common fragile sites [164, 165]. However, the SUMOylation function of SLX4 does not solely depend on the SIMs, but also on the BTB domain [164]. Thus, SLX4 is not only a scaffold protein but it is also proposed to harbor a SUMO-E3-ligase activity [164]. So far, the functional impact of XPF sumoylation is unknown, but it has been described that sumoylation of the yeast XPF-ortholog Rad1 reduces its DNA-binding ability [126, 167]. Furthermore, the SIMs contribute to the localization of SLX4 at telomeres [165].

SLX4 contains an area at the C-terminal region that resembles a SAP motif (named after SAF-A/B, Acinus and PIAS) [168] (Figure 1.4 A). The SAP motif is characterized by two amphiphatic helices containing hydrophobic, polar and bulky residues separated by a conserved glycine in the loop region [168]. SAP proteins are found in a number of chromatin-associated proteins, such as the scaffold attachment factors SAF-A and SAF-B, PIAS family members and Acinus [168]. Furthermore the SAP motif is also found in other DNA repair proteins, including the non-homologous end joining protein Ku70, plant poly-ADP-ribose polymerase, and the post-replicative repair protein Rad18 [168]. Direct evidence that the SAP motif binds DNA has come from studies of SAF-A and Ku70 [169, 170]. The SAP domain can either be used as the main DNA binding component of a protein (e.g. SAF-A) [169] or as an additional stabilizing element (e.g. Ku70 and SpCCE1) [127, 170, 171].

The extreme C-terminal domain is the CCD domain (conserved carboxy-terminal domain) or referred to as SBD (SLX1 binding domain) [158]. This domain confers the interaction with the structure specific endonuclease SLX1 [113, 158]. The CCD and the SAP domain contain the highest degree of sequence similarity across SLX4 orthologs from fungi to mammals [78, 113–115, 154].

Taken together, the domains comprise a large area, where the different interaction partners of SLX4 are associated. This allows SLX4 to take part in different DNA repair pathways and other cellular pathways.

1.3.2 The Interaction Partners of SLX4

SLX4 has been described as a scaffold protein [113–115]. In general, scaffold proteins serve as a binding platform for proteins, thereby tethering them into complexes and helping to localize pathway components to the target cell compartments. The various domains in SLX4 serve as interaction sites with structure-selective endonucleases, but also with proteins involved in several cellular pathways [113–115, 154] (Figure 1.4 B). SLX4 helps in the successful recruitment of the endonucleases SLX1, MUS81-EME1 and XPF-ERCC1 to DNA and coordinates their action [126]. The interaction partners have been uncovered by several methods, like yeast-two hybrid or proteomic analysis and will be described in this subsection [113–115, 154].

Proteomic analysis identified MSH2/3 as interaction partners of SLX4 [114, 115]. The MSH2/3 proteins are involved in DNA mismatch repair and recognize insertions, deletions and mismatches [172]. So, far it is only known that MSH2/3 interact at the N-terminus of SLX4 between the residues 1 to 669 [114]. Neither the exact interaction site nor the function of this interaction has been characterized.

The Fanconi anemia complementation group D 2 (FANCD2) has been shown to interact with SLX4^{UBZ} in co-localization and immunoprecipitation studies [159]. Ubiquitylated FANCD2 recruits SLX4 to ICL sites [159]. FANCD2 is part of the Fanconi anemia (FA) pathway for the

repair of ICL and is ubiquitylated by the FA core complex, an E3 ubiquitin ligase. FANCD2 and its interaction partner FANCI constitute a central axis to connect the FA core complex with downstream functional units of the FA pathway [13].

SLX4IP (previously referred to as C20orf94) was identified as an N-terminal interaction partner of SLX4 [114]. The protein was shown to be deleted in childhood acute lymphoblastic leukemia [173]. To date, there is neither functional nor structural information about SLX4IP or its interplay with SLX4 available. Interestingly, SLX4IP appeared in the evolution of vertebrates and it is absent in invertebrates, but lacks any known protein interaction domains [114].

XPF-ERCC1 (xeroderma pigmentosum group F - excision repair cross complementation group 1) is one of the three interacting nucleases of SLX4 and belongs to the XPF/MUS81 nuclease family member [113–115, 174]. XPF-ERCC1 (Rad1-Rad10 in yeast) plays important roles in ICL repair, single-strand annealing, telomere maintenance and nucleotide excision repair [174–179]. The interaction with SLX4 is conferred via the MLR domain and the BTB domain [113, 158, 175]. Only a portion of cellular XPF interacts with SLX4 [115], the non-SLX4 bound XPF participates in other DNA repair pathways, such as nucleotide excision repair [158, 180]. Through the interaction with XPF-ERCC1, SLX4 is involved in ICL repair, specifically in the unhooking step [161, 175, 179]. The recruitment of the first UBZ domain in SLX4 to ICL sites depends on ubiquitylated FANCD2 and subsequently SLX4 recruits XPF-ERCC1 via the MLR domain [14, 160, 161, 179]. Leucine 230 in the hsXPF helicase domain is essential for the interaction with SLX4 and the recruitment to the damage site [175]. This interaction has high binding affinity [175]. A second interaction site between XPF and SLX4 is conferred transiently via SLX4^{BTB} and activates the ICL incision [175]. Finally, the nuclease domain in XPF recognizes and processes the ICL substrate [175]. Another study could show, that a truncated form of SLX4, comprising SLX4 UBZ, MLR and BTB domain, stimulates XPF-ERCC1 nuclease activity towards the processing of forked or flap containing DNA and is thus accelerating XPF-ERCC1 unhooking of crosslinked DNA [179]. Human cells with low levels of XPF or ERCC1 are sensitive towards UV irradiation and DNA cross-linking agents [158, 181, 182]. XPF-ERCC1 mutations are found in various genetic disorders, e.g. the autosomal recessive skin disorder Xeroderma pigmentosum, which is characterized by UV sensitivity and high predisposition to skin cancer [15, 183].

ATM / ATR kinases (ataxia telangiectasia mutated / ataxia telangiectasia mutated and Rad3-related) interact with SLX4 [79, 114, 184, 185]. Mammalian SLX4 is a key substrate of ATM in the germ line, playing a role in the spermatogonial stages of gametogenesis and in the entry and progression through prophase I [186]. The localization of SLX4 onto meiotic chromosomes is dependent on ATM [186]. In yeast, Slx4 is phosphorylated by the yeast ATM / ATR orthologs, Mec1 and Tel1, and is required for repair of DNA alkylation damage and single-strand annealing [79, 184]. ATM and ATR are both members of the phosphatidylinositol-3'-OH-kinase-like protein kinase family and are important for the sensing of DNA damage [46, 187, 188]. ATM is responding to DNA DSBs arising throughout the cell cycle and is regulated by the MRN complex, which prepares the DNA for HR [16, 45, 185, 188]. ATR responds to perturbations of the cellular DNA replication exposing ssDNA and is regulated by the ATR-interacting protein [45, 185, 188]. ATM and ATR initially phosphorylate mediator proteins, which amplify the DNA damage response by recruiting ATM / ATR substrates [16]. ATM / ATR either phosphorylate the substrates directly or by other kinases, like the CHK2 kinase, leading to a phosphorylation cascade and in the end

to cell cycle arrest [16, 187, 188]. Further, they are known to cross talk with each other and their function can also be redundant [46]. However, ATM / ATR are not only involved in DNA damage response, but coordinate a variety of cellular activities, from DNA replication and repair to transcription, metabolic signaling and RNA splicing [16, 185]. Mutations in the *atm* gene were first connected with the autosomal recessive disorder Ataxia-telangiectasia, characterized by cerebellar ataxia, telangiectases and lymphoid tumors [189]. ATR is mutated in Seckel syndrome, characterized by 'bird-headed' facial appearance and hematological abnormalities [188].

Through the interaction of SLX4 with TRF2/TRF2IP (telomeric repeat-binding factor 2 - TRF2 interacting protein), SLX4 is involved in telomere maintenance [114, 177, 178]. Telomeres are regions of repetitive DNA sequences at chromosome ends, which protect the ends from degradation. Because of the alternate and secondary DNA structures in telomeres, they pose a challenge in DNA metabolism [190]. TRF2 is a negative regulator of telomere length and key component in the control of nucleases at telomeres [126, 155]. It is required to shield telomeres from uncontrolled activity of structure specific endonucleases and it is required for recruitment of SLX4 and its associated nucleases to telomeres [126, 155, 178]. Thereby TRF2 interacts via a TRF2-binding-like motif (aa 1014 - 1028) in SLX4 [177, 178]. The interaction of TRF2 with SLX4 is reinforced by the SIMs [165]. It is proposed that SLX4 and its three associated nucleases localize at the telomeres [177, 178]. Thereby, SLX4 and TRF2 form a scaffold and assemble a telomere maintenance kit containing XPF-ERCC1, MUS81-EME1 and SLX1, whereas SLX4-SLX1 regulates the telomere length [177, 178, 191].

Interaction between SLX4 and TOPBP1 (DNA topoisomerase 2 binding protein 1) is conferred via the phosphorylation of Thr 1260 in SLX4 (in yeast Slx4 Ser 486) by CDK, which is conserved between yeast and humans [192–194]. Additionally, the interaction between SLX4 and TOPBP1 (yeast: Dpb11) requires the kinase ATR (yeast ortholog Mec1) [57, 195]. The complex Slx4-Dpb11 integrates two cellular signals in yeast. First, the cell cycle state is implied through Slx4 phosphorylation by Cdk1 and second, the presence of DNA damage is signaled through checkpoint kinase phosphorylation [194]. TOPBP1 is an essential scaffolding protein involved in DNA replication, DNA repair, transcriptional regulation and DNA damage checkpoint signaling [57, 192, 196, 197]. In yeast, the Slx4-Dpb11 complex acts as an regulator of joint molecule removal pathway by interacting with the structure-selective endonuclease Mus81-Mms4, in which Dpb11 acts as a bridge between Slx4 and Mus81-Mms4 (methyl methane sulfonate sensitive 4; yeast ortholog of MUS81-EME1) [194].

Proteomic analysis found the polo-like kinase 1 (PLK1) as an interaction partner of SLX4 [114, 115] (Figure 1.4 B). PLK1 is a conserved eukaryotic kinase, controlling multiple essential steps of mitosis and meiosis [198]. It is implicated for example in the activation of Cdk1-Cyclin B at mitotic entry and activation of the anaphase promoting complex [198]. PLK1 associates with SLX4 in part through a candidate polo-box binding motif centered on S1453 in SLX4 [114]. Furthermore, it has been demonstrated that PLK1 also interacts with SLX4IP and ERCC4 [114]. Depletion of SLX4 reduced the interaction between PLK1 and ERCC4, suggesting that SLX4 bridges the interaction between PLK1 and ERCC4 [114]. PLK1 regulates MUS81-EME1 activity in yeast and human [117, 199]. Furthermore, PLK1 phosphorylates SLX4 and its activity is required for the association between SLX4 and MUS81 [114, 200].

The human MUS81-EME1 complex interacts via the MUS81-N-terminus with SLX4^{SAP} [158,

201, 202]. In human cells SLX4 binds directly to MUS81-EME1, whereas in budding yeast the interaction needs the TOPBP1 ortholog Dpb11 as a bridge between the two partners [194]. The MUS81-SLX4 interaction plays a role in the resolution of HJs and in the processing of stalled replication forks [141, 158]. Mus81 was first discovered by Interthal and Heyer in a yeast two-hybrid screen with the recombinational repair protein Rad54B as bait [203]. Eme1 was identified as an interaction partner of Mus81 in a tandem affinity purification coupled to mass spectrometry approach in fission yeast [135]. The following subsection will cover MUS81-EME1.

SLX1 (synthetic lethal of unknown function (X)-1) interacts with the CCD domain in SLX4 [113–115, 158] (Figure 1.4 B). *slx1* was first discovered in a synthetic lethal screen for genes in absence of the budding yeast RecQ helicase homolog Sgs1 [151, 152, 204]. The BLM homolog Sgs1 unwinds recombination intermediates [69]. In higher eukaryotes SLX1 was identified by homology searches [113–115, 154].

SLX1 contains an N-terminal GIY-YIG nuclease domain (also called Uri nuclease), named after the conserved active site residues (Figure 1.4 A, 1.5) [153]. The GIY-YIG nuclease domain is found in all kingdoms of life, e.g. as UvrC in prokaryotic nucleotide excision repair or as homing endonuclease I-*TevI* in bacteriophage T4 [205]. The typical GIY-YIG nuclease conserved core contains a conserved Y in the first motif, an almost universal Y and a strongly conserved G in the second motif [205, 206]. Those residues are also shown in SLX1 proteins from different organisms (see Appendix for GIY-YIG domain 4.2). A mechanism of substrate binding and cleavage has been proposed for Slx1 based on the restriction enzyme Hpy188I [206] (Figure 1.5 B). A single metal ion is bound in the GIY-YIG active site and the metal ion is proposed to destabilize the substrate and stabilize the transition state [156]. The metal ion is located on the opposite side of the scissile phosphate, which is the target of a nucleophilic attack by a water molecule [156]. This water molecule on the other hand is likely being coordinated by the active site residues Tyr 14 and Arg 36 in Slx1 [156]. Further residues are involved in the stabilization of the transition state [156]. The C-terminal domain of SLX1 was designated as RING (really interesting new gene)-finger like domain after structural comparisons [153, 156, 207] (Figure 1.4). The RING finger Zn-binding domain is postulated to mediate protein-protein or protein-DNA interactions [205, 208].

Until now, there is only sparse structural information available about SLX1 and its interaction site with SLX4 [156, 207]. Crystal structures of Slx1 from two yeast homologs (*Candida glabrata* and *Schizosaccharomyces pombe*) have been solved [156, 207]. The Slx1 monomer consisted of the two domains, the N-terminal GIY-YIG nuclease domain and the C-terminal zinc finger domain [156] (Figure 1.5 A). These two domains interact with each other and are stabilized by a long α helix, which spans the entire molecule and provides a scaffold [156]. In the center of the nuclease is a β -sheet containing five strands, which are flanked by six α -helices [156]. The GIY-YIG nuclease structure is highly conserved with other GIY-YIG nucleases [156]. Mutation of either of the residues R34 and E74 in fission yeast, corresponding to the residues E82 and R41 in human SLX1 lead to inactivity of the nuclease [113, 114, 141, 153]. The C-terminal part of Slx1 showed the zinc finger domain, which is related to RING finger domains, like those found in E3 ubiquitin ligases or FANCL [156, 207]. Slx1 is proposed to form a dimer [156]. The dimerization of Slx1 is suggested to block its activity as the active site of each subunit would be partially blocked by the RING domain of the other subunit [156]. The crystal structure of Slx1 and Slx4^{CCD} showed a change in the RING domain compared to Slx1 alone and the CCD domain was positioned in the cleft between

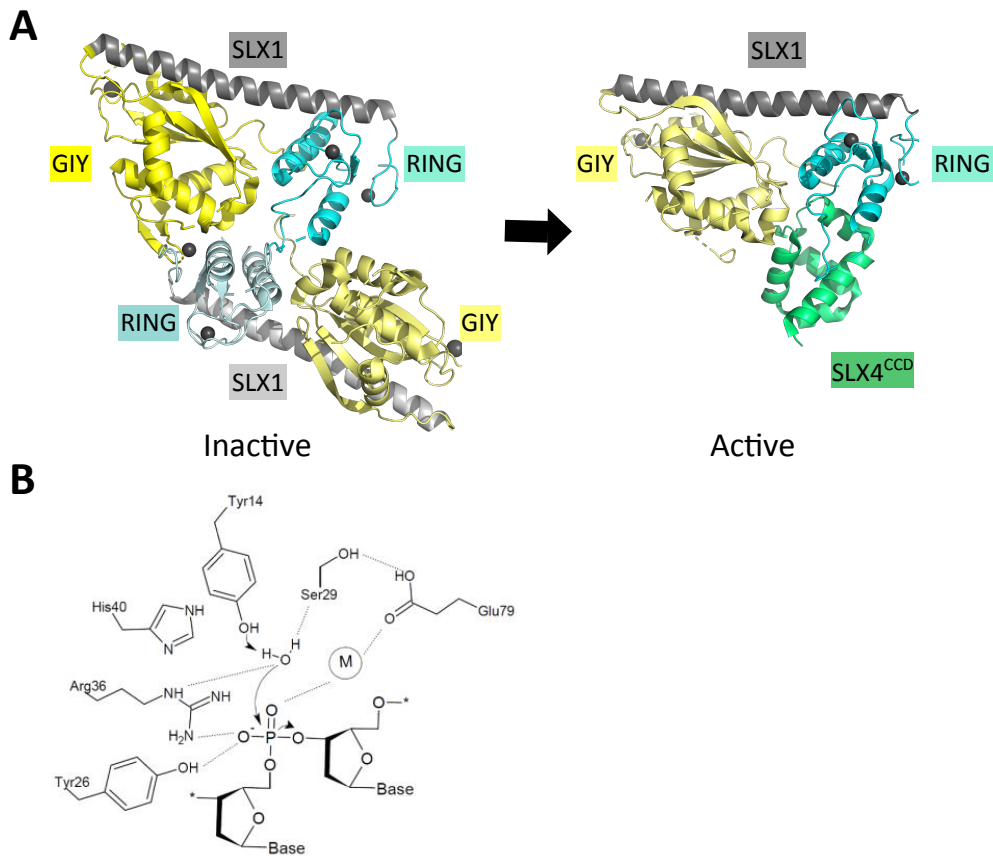


Figure 1.5: The structure of SLX4-SLX1 (A) Crystal structure of the yeast Slx1 dimer and the yeast Slx1-Slx4^{CCD} complex (PDB codes: 4XM5, 4XLG). Slx1 is proposed to form a homodimer, which blocks the active site. Slx1 is suggested to be activated by conversion of the homodimer into the SLX1-SLX4^{CCD}. (B) Schematic illustration of the proposed catalytic mechanism based on the restriction nuclease Hpy188I. Adopted from [156]

the nuclease and RING domain [156] (Figure 1.5 A). Slx4^{CCD} is folded as a helix-turn-helix domain with a four-helix bundle [207]. The complex formation of SLX1-SLX4 is conferred by hydrophobic interactions and hydrogen bonding [207] and ablated when residue C1805 in hsSLX4 is mutated or C1536R in murine SLX4 [178, 202]. The structure of *Candida glabrata* Slx4^{CCD} shows that a mutation to arginine at the same position leads to steric clashes and thus abrogating the interaction with Slx1 [156]. Several residues are giving a possible explanation why Slx1 homodimerization and interaction with Slx4 are mutually exclusive [156]. SLX1 in complex with SLX4 is involved in the processing of Holliday Junctions during homologous recombination [113–115, 141] as well as in telomere maintenance by preventing telomere lengthening [177, 178, 191]. Furthermore SLX1 could also be involved in the repair of ICLs as a 5' endonuclease [126]. SLX1 as part of the SMX-complex is involved in the cleavage of replication forks [143].

1.3.3 The MUS81-EME1 Nuclease

The first discovered eukaryotic Holliday Junction resolvases were the fission yeast and human MUS81-EME1 endonucleases [135, 203, 209]. They belong to the XPF-ERCC1 endonuclease family, which is only present in archaea and eukaryotes and conserved among them [142, 174]. This family can be structurally and functionally divided into catalytic and non-catalytic family members [174]. Catalytic family members share a common characteristic core of an ERCC4 endonuclease domain and a tandem helix-hairpin-helix (HhH)₂ domain [174] (Figure 1.6 A, B). The ERCC4 domain is distantly related to prokaryotic type II restriction enzymes [210, 211]. The nucleases containing an ERCC4 domain like XPF, MUS81 and FANCM form heterodimers with a non-functional catalytic domain subunit (ERCC4, EME1, EME2, FAAP24 (FA-associated protein 24 kDa)) [212]. In the non-functional subunits the key residues have been diverged to leave the subunit inactive [174, 210]. In humans, four heterodimeric complexes containing XPF family members have been identified, including XPF-ERCC1, MUS81-EME1, MUS81-EME2 and FANCM-FAAP24 [174]. The signature motif GDX_nERKX₃D is being proposed for the active site of XPF family members (see also Figure 1.6 A orange dots) [174, 212]. MUS81 and EME1 dimerize via their C-terminal (HhH)₂ domains [135, 213]. XPF-ERCC1 complex, a classical member of the described endonuclease family, processes splayed-arms, bubbles and stem-loop structures [174]. In contrast to XPF-ERCC1, MUS81-EME1 possesses the ability to process higher branched DNA structures, like HJs [126, 135, 209].

The human MUS81 protein has two more domains at the N-terminus in comparison to EME1, an additional HhH₂ domain and a winged helix(WH) domain (Figure 1.6 B). The first 86 amino acids in MUS81 contain the N-terminal HhH motif, which is responsible for the interaction with the SLX4^{SAP} domain and can in addition bind to DNA [112, 113, 143, 158, 201, 202]. In general, HhH motifs bind DNA in a non-sequence specific manner [218]. The binding takes place via the formation of hydrogen bonds between protein backbone nitrogens and DNA phosphate groups [218]. Here, as the same domain in MUS81 interacts with DNA and the interaction partner SLX4, it is so far not known how the interactions are regulated. A possible mechanism was proposed by Wyatt et al., that upon formation of the SMX complex the HhH domain is prevented to bind DNA substrate as the interaction between SLX4 and MUS81^{HhH} induces a conformational change, relaxes substrate specificity and activates its nuclease functions [143]. Recently, a nuclear

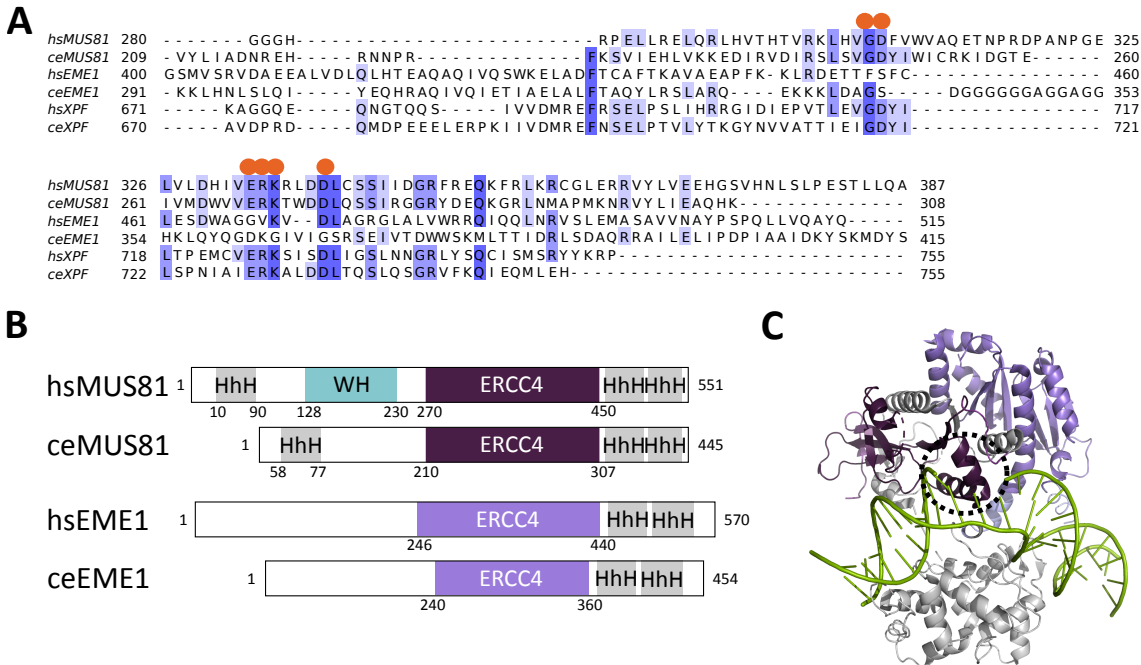


Figure 1.6: Sequence similarities, domain architecture and crystal structure of MUS81-EME1. (A) Sequence similarities between the ERCC4 domains of MUS81 and EME1 orthologs from different organisms. Dark blue indicates strictly conserved residues, white indicates variable residues. The orange dots indicate the key residues of the ERCC4 motif [214]. The orthologs were aligned with MUSCLE [215] and displayed using Jalview [216]. The alignment shows the sequences of *Homo sapiens* (hs), and *Caenorhabditis elegans* (ce). Uniprot accession numbers [217] from top to down: Q96NY9, P91153, Q96AY2, Q9GZG5, Q92889 and G5EBZ1. (B) Schematic representation of MUS81-EME1 domain organization for human and *Caenorhabditis elegans* proteins. Light grey boxes indicate the Helix-hairpin-helix (HhH) domains at the N- and C-terminus. The cyan box at MUS81-N-terminus depicts the winged helix (WH) domain. The dark purple boxes in the MUS81 proteins indicate the ERCC4 (excision repair cross-complementation group 4) domains with nucleic acid cleavage activity. The light purple boxes in the EME1 represent the ERCC4 domains, where important active site residues have been rendered to leave the protein inactive. Amino acid boundaries are shown as numbers. (C) Crystal structure of hsMUS81²⁴⁵⁻⁵⁵¹-hsEME1¹⁷⁷⁻⁵⁷⁰ in complex with 32-bp dsDNA with 3-nt 3' flap (green) (PDB code: 4P0R). The black dotted circle marks the cleavage site. The colors correspond to the domain organization laid out in (B).

magnetic resonance (NMR) structure of MUS81^{HhH} has been published, showing a stable folded domain comprised of a pair of anti-parallel α -helices that pack against in each other in an angle of 110° [143]. The second helix connects via a loop to the third and fourth helix, which then forms the HhH motif [143].

The WH domain in human MUS81 was structurally characterized by NMR as well [219]. The residues 128-230 fold into a helical domain that adopts a WH tertiary fold [219]. In general, the WH domains are a subclass of helix-turn helix DNA binding domains that are found in DNA-processing enzymes and many transcription factors [126]. The human MUS81 WH domain contains two functionally important elements of WH proteins, the recognition helix $\alpha 3$ and the beta-hairpin 'wing' motif [220]. The WH domain is highly basic with 12 Arg/Lys side chains, whereas two arginines are within the recognition helix, suggesting a possible charge complementary appropriate for binding DNA [219]. The winged helix domain has been shown to alter the DNA binding specificity of MUS81-EME1 complexes [219]. Furthermore, it has been suggested that both the amino-terminus and the recognition helix of the WH domain have a role in DNA binding [219]. Deletion of the WH domain reduces cleavage activity of MUS81-EME1, mutations in the WH domain reduce binding of the isolated domain to DNA *in vitro* and show a sensitivity towards DNA damaging agents in *S. pombe* [219]. The WH domain is essential for both meiosis and DNA damage tolerance in *S. pombe* [219].

Apart from the NMR structures of the N-terminal domains of MUS81, there are two crystal structures published containing the nuclease-(HhH)₂ core of the heterodimer MUS81-EME1 [213, 214]. The first structure of a DNA-free human-zebrafish hybrid MUS81-EME1 sheds light on the nuclease domain of MUS81, the nuclease-like domain of EME1 as well as on the two C-terminal Helix-hairpin-Helix ((HhH)₂) domains of each protein [213]. The nuclease core is shared with the PD-(D/E)XK superfamily and consists of a central β -sheet that is flanked by α -helices [213, 221]. The active site of the MUS81 contains carboxylate residues that coordinate divalent metal ions [213, 214, 222]. The metal-ions are essential for DNA hydrolysis in a two-metal ion mechanism [223]. Another structural study by the Cho lab showed at low resolution, how the nuclease recognized and cleaved several DNA flap-substrate [214]. Figure 1.6 C shows the overall structure of MUS81 (dark purple)-EME1 (light purple) in complex with 3' flap DNA adapted from [214]. The flap DNA substrate is being kinked by the N-terminal (HhH)₂ domains of MUS81 and EME1 together with the nuclease domain and the 3' end of the incision strand is guided to the active site [214]. DNA bending is achieved by conformational changes in MUS81-EME1 from a closed to an open form (Figure 1.6 C) [214]. Comparison of the structures with and without DNA substrate showed substantial differences as the relative orientation varies between the different crystal structures [213, 214]. DNA binding induces significant conformational change in the linkers, which transforms the MUS81-EME1 structure from a compact to an open state [214]. In pre-nick state, the DNA is packed between the EME1-HhH-C-terminal domain and the MUS81 nuclease domain, whereas after cleavage the DNA is bound to the EME1-ERCC4 like domain and the MUS81-(HhH)₂-C-terminal domain [214, 221].

So far, only low resolution structural information about XPF family members is available and further structural data is needed to understand how the N-terminal domain of MUS81 interacts with DNA substrates and positions in respect to the catalytic domain. The structural characterization of MUS81-EME1 has revealed a distinctive orientation of its nuclease and (HhH)₂ domains

toward DNA junction binding [213, 214]. However, both studies lacked the amino-terminal extension of MUS81 and EME1. The NMR studies of the N-terminal domains show the overall fold, but it still remains elusive how the N-terminal domains will interact with the DNA substrate. For other XPF family members a structure of the archaeal XPF homodimer in complex with a dsDNA fragment has been published, but the DNA is not interacting with the active site [224]. Structural information of the eukaryotic XPF-ERCC1 is available for the ERCC1 nuclease-like domain [225] and the HhH heterodimer [225, 226]. Furthermore, some structural information of FANCM-FAAP24 is available [227]. However, until now no structural information is available about the catalytic complex of XPF-ERCC1.

MUS81-EME1 can cleave diverse DNA substrates due to the active site with the DNA binding pocket and the differences in linker and location of the (HhH)₂ domain compared to XPF family members [214]. Therefore, MUS81-EME1 is involved in different DNA repair pathways. It plays a critical role in the genome maintenance, e.g. in the repair of internal cross-links and replication fork collapse in mitotic cells and in meiotic cross-overs [135, 209, 228]. Its role in the recovery of stalled or collapsed replication forks is evolutionary conserved [221]. Problematic replication forks can arise in regions of the genome that are inherently difficult to replicate, in response to replisome-blocking agents (ICLs) or when dNTP pools are diminished [221]. Thereby, MUS81 collaborates with BLM and ATR to convert the replication forks into transient DSBs that are rapidly repaired [126, 229]. Additionally, MUS81 can associate with and stimulate FEN1 (flap endonuclease 1) activity [230]. This interaction is dependent on the N-terminal winged helix of MUS81 [219]. MUS81-EME1 is involved in ICL repair and carries out the unhooking process of the DNA cross-link [228], thus creating a one-ended double-strand break. In murine cells is the role of MUS81 in ICL repair independent of its interaction with SLX4 [201]. Interestingly, in human cells, MUS81-EME1 mutants incapable of interaction with SLX4 are also defective in ICL repair [201]. Furthermore, MUS81-EME1 is upregulated at mitosis and it has been found in yeast, that Mus81 limits the error-prone process of break-induced replication [126].

1.3.4 Cleavage Properties of the SM-Complex

Through DNA replication and the different types of DNA damage occurring at molecular level [1], a huge variety of DNA substrates with distinct secondary structures arises. Therefore, the various nucleases have evolved to resolve the different DNA structures, each nuclease having its own cleavage range and specificity. The cleavage activity of the nucleases can be tested experimentally *in vitro* by labeled substrates. These substrates correspond to the different intermediates during DNA repair. Therefore, a short overview will be given about the cleavage properties of the nucleases associated with SLX4 in Figure 1.7.

First, SLX4 was described to form a complex with the GIY-YIG type endonuclease SLX1 in fission yeast for ribosomal DNA maintenance [153]. Biochemical data suggested then that the complex plays a role in the processing of stalled replication forks [231] and later Holliday Junction resolution activity was reported [113–115, 141]. The *S. cerevisiae* Slx4-Slx1 complex cleaves splayed arms, 5'flaps, replication forks (RF) and HJs [152]. The major cleavage site for the 5' flap lies in the 5' single-stranded arm at the junction between the single- and double-stranded DNA [152]. The fission yeast Slx1-Slx4 cleaves stem loops and HJs [153]. The human SLX4-SLX1







							other substrates	Citation
hsXPF-ERCC1		+					splayed arm, bubble, stem-loop	Ciccia et al. 2008 [175]
scRad1-Rad10		+					splayed arm, bubble, stem-loop	Ciccia et al. 2008 [175]
hsMUS81-EME1		+		+				Wyatt et al. 2013 [142], Ciccia et al. 2003 [143], Gwon et al. 2014 [217]
scMus81-Mms4		+		+		+	splayed arm	Boddy et al. 2001 [136], Ciccia et al. 2003 [143], Ciccia et al. 2008 [175]
hsSLX4-SLX1		+	+	+	+			Wyatt et al. 2013 [142]
ceSLX4-SLX1			+	+	+			Saito et al. 2012 [25]
spSlx4-Slx1					+		stem loop	Coulon et al. 2004 [154]
scSlx4-Slx1		+		+	+		splayed arms	Fricke et al. 2003 [153]

Figure 1.7: DNA substrates and cleavage properties of nucleases associated with SLX4 Cleavage activities of human XPF-ERCC1, its *Saccharomyces cerevisiae* ortholog Rad1-Rad10, the human MUS81-EME1 complex and its *Saccharomyces cerevisiae* ortholog Mus81-Mms4 as well as the SLX4-SLX1 complexes from human, *Caenorhabditis elegans*, *Shizosaccharomyces pombe* and *Saccharomyces cerevisiae* nuclease complexes. The XPF-ERCC1 / Rad1-Rad10 and the MUS81-EME1 / Mus81-Mms4 complexes belong to the XPF endonuclease family. SLX1 is a GIY-YIG nuclease.

shows activity towards Y forks, 3'- and 5'-flaps, RFs, stem loops and HJs [113–115, 141]. Both, yeast Slx1 and human SLX1 alone have only very weak nuclease activity, but in complex with Slx4 and SLX4 the cleavage rate is increased tremendously [114, 141, 153]. For the *C. elegans* SLX1-SLX4 complex, endonucleolytic activity could be detected for RFs, 3' flaps, 5'flaps and HJs [25]. However, the ceSLX1-SLX4 complex is not a bona fide HJ resolvase as it does not cleave HJ substrates with perfect symmetry as the human complex does [25]. The ceSLX1-SLX4 endonuclease is more reminiscent to the baker's yeast and the fission yeast complexes [25, 152, 153]. The difference in the cleavage activity is based on the different function *C. elegans* SLX1-SLX4 has, as it is involved in the resolution of meiotic crossovers [25]. Comparing the nuclease activities from different organisms, it shows that subtle differences between the orthologs of one nuclease complex exist.

In association with EME1 the MUS81 endonuclease is able to cleave different three- and four-way junctions containing a duplex downstream from a nick *in vitro* [232]. Human and yeast MUS81-EME1 / Mus81-Mms4 cleave nicked HJs, replication forks and 3'flaps [141, 174] (Figure 1.7). In addition to these substrates yeast Mus81-Mms4 cleaves also D-loops [213, 233]. These substrates for MUS81-EME1 form during mitosis and fission yeast meiosis, but also during processing of damaged replication forks.

When cleavage activity of SLX4 in complex with SLX1 and MUS81-EME1 was tested *in vitro*, a unique cleavage mechanism was detected: SLX4 brings the SLX1 and MUS81-EME1 nucleases during HJ resolution together giving rise to a coordinated nick and counter-nick mechanism [141, 143]. In the setting of the two active sites SLX1 introduces the first cut on Holliday junction and MUS81-EME1 introduces the second cut [141]. Both incisions occur within the lifetime of the SM-HJ DNA complex and the cleavage is faster than compared to the activity of the proteins

alone [141, 143]. Thus, SLX4 scaffold tethers MUS81-EME1 and SLX1 and thereby coordinates the activity of the nucleases on HJs [141, 202]. Recently, it has been shown that the N-terminal HhH domain in MUS81 regulates substrate specificity in MUS81-EME1 [143]. Interaction of SLX4 with MUS81 HhH domain, which can also bind DNA, relaxes nuclease specificity, thus activating nuclease functions [143]. Deletion of this N-terminal domain led to preferential cleavage of 3'flaps, but also showing weak activity on intact HJs [143]. In contrast to the wild-type enzyme MUS81-EME1 lacking the HhH domain showed decreased activity on RF and nHJs, but increased activity towards the unusual substrates 5'flap and splayed arms [143]. Thus, the N-terminus of MUS81 plays a critical role in the regulation of cleavage preference of the nuclease domain, when in complex with SLX4.

Furthermore, when the third SLX4-interacting nuclease, XPF-ERCC1, was tested in the context of the SMX-complex, it was shown that it does not take part in the cleavage reaction directly [143, 234]. The XPF-ERCC1 nuclease alone processes splayed-arms, bubbles and stem-loop structures [174]. But in the context of the trinuclease SMX complex XPF-ERCC1 increases HJ resolution activity compared to SM complex, thus playing a role in stabilizing the complex [143, 234].

HJ cleavage activity has been reported for the SM complex, but the general RuvC paradigm for HJ resolution cannot be applied [136, 146]. The *Escherichia coli* RuvC protein forms a homodimer and coordinates the two active sites to introduce a pair of symmetrical nicks into strands that are diametrically opposed across the junction [136, 146]. The reaction products are then a pair of nicked duplexes that can readily be ligated [136, 146]. The resolvase GEN1 shows comparable cleavage behavior as the prototypic HJ resolvase RuvC and cleaves 5'flaps, replication forks and HJs [111, 235]. In contrast, the SM-complex has been shown so far to work with a unique cleavage mechanism as it contains a pair of heterodimeric flap endonucleases, each contributing one active site to the cleavage [136]. The majority of the cleavage products can not be ligated, contains flaps and requires further processing [136]. Furthermore, human SLX4 provides a scaffold for the tri-nuclease complex [143] and has been shown to be the only example so far for a tri-nuclease complex [236]. Taken together, SLX4 is a context-dependent nuclease scaffold, that requires the activation of different modules depending on the lesion [158].

1.3.5 Regulation of the SM-Complex

The structure-selective endonucleases are an inherent threat to genome integrity because broken DNA ends can facilitate chromosome rearrangements and genome alterations, which are potentially tumorigenic [156]. Thus, the endonucleases have to be tightly regulated to make sure that the correct DNA substrate is cleaved at the right time [156]. The regulatory mechanisms are suggested to work at several levels [156]: protein expression, post-translational modification (Mus81-Mms4/Eme1, Yen1) [117, 118, 139], nuclear localization (Gen1) [237], conformational changes (FEN1), protein-protein interactions (XPG, XPF-ERCC1, SLX-MUS) [141] and dimerisation (SLX1) [156, 221]. Several studies carried out in yeast have been shown how the nucleases are regulated throughout the cell cycle [117–119, 136, 139]. Thereby, cell cycle stage specific phosphorylation of the nucleases and their interaction partners ensure temporal and spatial control as well as sequential activation [136]. During S phase, the HJ resolution activity of Mus81-Mms4 is low, but at mitosis onset it becomes elevated due to phosphorylation of Mms4 by cell cycles

kinases Cdc28/Cdk and Cdc5 [117–119, 136, 194, 238]. This temporal regulation of Mus81 channels a large proportion of joint molecules into the dissolution pathway [117, 238]. The dissolution leads to non-crossover outcome and is beneficial for mitotic cells [84, 194]. Additionally, the restriction of Mus81 activity may be necessary to achieve temporal separation of lesion bypass reaction and joint molecule resolution [120]. Mus81 could impede the bypass reaction, given its relatively broad substrate specificity to a range of DNA substrates (RF, D-loops, HJs) [194]. The regulation of the nucleases in human cells is in detail different from the regulation in yeast cells [136]. The EME1 subunit in the MUS81-EME1 complex becomes phosphorylated by cell cycle kinase CDK1 and PLK1 at the onset of mitosis [117, 136, 199]. SLX4 becomes phosphorylated by CDK1 and PLK1 [200]. These phosphorylations promote the association of MUS81-EME1 with SLX4-SLX1 to form the SM-complex [136, 141]. This direct interaction between SLX4 and MUS81 occur at prometaphase in response to the phosphorylations of MUS81-EME1 and SLX4 [114, 141, 200, 202, 234]. Hence, the two nucleases are combined to form a highly active HJ resolvase [136, 141].

1.3.6 Comparison Between the Human and the *C. elegans* SM-Complex

In this thesis the human and *Caenorhabditis elegans* proteins of the SM-complex were investigated. Therefore, this subsection will give a short comparison between the proteins of the SM complex from the different organisms. The nematode *Caenorhabditis elegans* is a well-established model organism, which has been used in a wide array of research questions, e.g. apoptosis, development and meiosis. During meiosis double-strand breaks occur and are easy to investigate as *C. elegans* follows a strict lineage of each larval cell [239]. Furthermore, *C. elegans* proteins are in general to 80% conserved to vertebrates [240]. The SLX4 ortholog in *C. elegans* is named HIM-18 (high incidence of males) which is characterized by an increased X chromosome loss [24]. For simplicity reasons HIM-18 is referred to herein as ceSLX4. ceSLX4 was discovered through a functional genomics approach in *C. elegans*. It has been shown to be essential for normal meiotic progression [24]. ceSLX4 facilitates processing of HR intermediates, which result from stalled replication forks and programmed meiotic DSBs [24].

Direct comparison of the domain organization between the human and worm proteins shows that the *C. elegans* proteins are in general shorter than the human proteins (Figures 1.4 A and 1.6 B), except for the SLX1 protein. ceSLX4 has only 40% size of the human ortholog and is much more compact as it lacks the long uncharacterized and disordered stretches between the domains. It shares 16% identity and 56% similarity with the human SLX4 [24]. Especially, the C-terminal SAP and CCD domains are highly conserved [113–115]. Interestingly, the *C. elegans* SLX1 protein contains a 168 amino acid N-terminal extension before the GIY-YIG nuclease domain (Figure 1.4 A), which can not be found in other SLX1 orthologs. Moreover, a bioinformatics analysis could not find a related sequence in any other protein, except for other nematode species or ascaris species. The *C. elegans* MUS81-EME1 proteins are shorter by approximately 100 amino acids (Figure 1.6 A). ceMUS81 does not contain a WH domain as human MUS81. So far, the WH domain has not been found in any nematode species [219]. An alignment between MUS81 orthologs from different organisms does not show any hints towards the domain in *C. elegans* (see Appendix 4.3).

The interaction between ceSLX4 and ceSLX1 has been shown by yeast two-hybrid assay [24,

241] and co-immunoprecipitation [25]. The interaction is based on both domains of ceSLX1, the RING finger domain and the nuclease domain, and the CCD domain of ceSLX4 [25]. In the nematode *C. elegans*, previous work found that ceSLX1 and the ceSLX4 are required for the resolution of some meiotic crossovers [25]. Studies showed that in *C. elegans* are two parallel, partially redundant pathways responsible for crossover formation - one is dependent on MUS81 and the other on XPF1 [176, 241, 242].

C. elegans SLX4 is required for HR-mediated repair at stalled replication forks and has a role in converting HR intermediates into crossover products [24]. ceSLX4 facilitates processing of HR intermediates resulting from replication fork collapse and programmed meiotic double-strand breaks in the *C. elegans* germline [24]. ceSLX1 is required for the repair of stalled/collapsed replication forks and meiotic DSBs [25]. The involvement of ceSLX1 in the repair of meiotic DSBs is not based on the resolution of the HJ itself, but it is involved in the regulation of the meiotic crossover distribution [25]. SLX1 is required for the suppression of crossovers at the center region of chromosomes [25]. For this, SLX1 could either act as a non-crossover specific HJ resolvase at the center region of chromosome V or the RING finger could act as an epigenetic reader [25, 241]. In *C. elegans*, SLX1 is involved in several genome maintenance pathways: replication support, re-establishment of a normal fork, ICL repair and nucleotide excision repair [25]. ceMUS81-EME1 is regulating the crossovers during meiosis together with SLX1 [241]. ceMUS81 and ceSLX1 have overlapping roles for DNA repair with HIM-6, the *C. elegans* ortholog of BLM [241]. In *C. elegans* meiosis XPF1 acts redundantly with MUS81 and SLX1 to resolve HJs into crossover products on chromosome arms, which are epigenetically marked by histone H3K9me [176, 241, 242]. Through the interplay between the structure-specific endonucleases they are able to have either positive or negative control on meiotic crossovers depending on the chromosome region [241].

1.3.7 Involvement of the SM-Complex in Genome Integrity and Disease

For mammalian cells Holliday Junction processing is of great importance. Cells depleted for BLM and SLX4, thus blocking the dissolution pathway and one resolution pathway, are synthetic lethal [112, 136, 141]. The same holds true for the depletion of SLX4 and GEN1, when blocking both resolution pathways [112, 136, 141]. Chromosomal abnormalities and mitotic defects lead to this mortality [112, 140, 141, 243]. Unresolved recombination intermediates fail to elicit checkpoint response, the resolvase-deficient cells enter mitosis with their sister chromatid bridges intact [234]. At anaphase the cells display ultrafine bridges, coated with RPA, which indicates ssDNA presence [150]. At mitotic division these bridges are broken, result in DNA breaks at G₁ and repair by NHEJ, but they lead ultimately to cell cycle arrest and cell death [150]. These kind of resolvase deficient system is somewhat artificial, but provides a model for the study of recombination intermediates [234]. The cellular defects caused by resolvase deficiency show that the resolvases are not simply a backup for the dissolution pathway, but are essential factors in chromosome segregation [234].

In budding and fission yeast, the deletion of *slx1* does not exhibit lethality or DNA damage sensitivity. The cells exhibit only defects in completion of rDNA replication [153, 204]. In human cells, siRNA based depletion of SLX1 increases DNA damage after treatment with ionizing radiation, camptothecin and DNA interstrand cross-linking agents [113–115, 154]. Knockout mice for

SLX1 did not exhibit any morphological, developmental or hematological defects and were fertile [202].

Fission yeast lacking either *mus81* or *eme1* produce spores that are inviable because of the defects in chromosome segregation during meiosis I [78, 135]. In contrast, deletion of *mus81* in budding yeast leads to a modest decrease in meiotic survival as budding yeast Mus81 is only responsible for a small subset of crossovers during meiosis [78]. So far, MUS81 has not been reported to be mutated in any human disorder [158]. However, knockout mice and cells are sensitive to cross-linking agents [158, 244, 245]. Two MUS81 mouse models have been generated [244, 245], but only one of them showed cancer predisposition [245]. Depletion of MUS81 from human cells leads to CPT sensitivity [114, 158]. In prostate cancer cells MUS81 has been detected as an alerter for the immune system to the presence of transformed cells [246]. MUS81 cleaves genomic DNA, which accumulates in the cytosol triggering the activity of the immune system [246].

Although, EME1 *-/-* mouse embryonic stem cells do not show changes in viability or growth rate, they exhibit aneuploidy, increased sister chromatid exchange upon DNA damage and increased levels of spontaneous chromosomal abnormalities [247].

S. cerevisiae cells lacking Slx4 are highly sensitive to the DNA alkylating agent MMS, leading to a blockage of replication and induction of the damage checkpoint signaling pathway [152]. SLX4 has been found to play a role in the Fanconi anemia pathway as FA complementation group P (FANCP) [14, 248, 249]. The FA pathway repairs ICLs and SLX4 with its associated nucleases has been found to play not only a role in the repair of double-strand breaks but also in the repair of DNA interstrand crosslinks (ICLs) [14, 161, 179, 249]. The rare recessive genetic disease FA is characterized by progressive pancytopenia, congenital and skeletal malformations, predisposition to hematologic and solid cancer [250]. Furthermore, sensitivity towards DNA crosslinking agents is a hallmark for diagnosis of FA [250]. Individuals carrying mutations in SLX4 manifest symptoms of FA by showing hypersensitivity towards the DNA damaging agents camptothecin and mitomycin C [14, 249]. So far 22 FA genes have been identified and FA signaling is crucial in the DNA damage response [13]. In FA patients relies the *in vivo* activity of XPF-ERCC1, MUS81-EME1 and SLX1 during DNA repair strictly on their association with SLX4 and the nucleases are important for DNA repair of distinct DNA lesion [112, 158]. Mice lacking SLX4 exhibit chromosomal instability with symptoms similar to Fanconi anemia [248]. So far, FA patients with different mutations in SLX4 have been described [14, 158, 249]. For example, siblings with deletions in the UBZ domain of SLX4 showed symptoms of FA and mutated SLX4 derived from this patients was not recruited to DNA damage [160]. Patients lacking SLX4 show a more severe developmental phenotype, e.g. bone marrow failure and death in the early twenties [14, 158, 249].

Mutations or knockouts in the genes encoding the nucleases lead to viable cells and organisms. This is in contrast with the rather strong phenotypes of SLX4 mutation or deletion. The nucleases are assumed to be redundant in the processing of DNA damage [159]. For the nucleases involved in HJ resolution, disease phenotypes have not been described as HJ resolution is a highly backed up process. If there is a mutation in e.g. SLX1, then HJ resolution can still be carried out by GEN1 [243]. However, as SLX4 is involved in several pathways and forms a platform for various proteins involved in different cellular processes, a deletion or a knockout has severe consequences for the integrity of the genome.

1.4 Aim of the Thesis

SLX4 is a scaffold protein and fulfills several roles in DNA damage repair and other cellular processes through its interactions with partner proteins. It has the remarkable ability to bind several DNA repair nucleases and is not a passive scaffold, but can act as a nuclease stimulating factor [179]. This versatility, its broad role in DNA damage response and the importance as a tumor suppressor makes SLX4 a very interesting object of study. So far, not much is known about how these nucleases interact with each other and in the context of a Holliday Junction. This thesis is focused on the C-terminal domains of SLX4 and the Holliday Junction resolvases, MUS81-EME1 and SLX1. In this context, it was of interest, how the individual DNA-binding modules are cooperating in order to recognize and lead to a cleavage of the substrate. The initial goal was the structural elucidation of the SM complex binding Holliday Junction by crystallography as so far only a low resolution structure of MUS81-EME1 in complex with flap DNA and a structure of SLX1 and SLX4^{CCD} without DNA are available [156, 214]. Therefore, expression of human and *C. elegans* orthologs in *E. coli* and mammalian suspension culture was established and purification protocols were installed. Expression and purification of a SLX4 C-terminal construct in mammalian cells resulted in a highly phosphorylated protein and the phosphorylation pattern of SLX4 in presence and absence of SLX1 was analyzed. Also the influence of DNA damaging agents was evaluated towards changes in phosphorylation pattern. The interaction between SLX4^{SAP} and the N-terminus of MUS81 was measured and quantified by several methods, e.g. analytical ultracentrifugation and thermofluor, as earlier studies demonstrated the interaction by co-immunoprecipitation and yeast-two-hybrid assay [201, 202]. Furthermore, it was known that the N-terminus of MUS81 can bind DNA [143, 219], thus DNA binding of *C. elegans* MUS81 N-terminus was further characterized. The SAP domain of SLX4 is referred to as a putative DNA binding domain because of its comparison to other SAP domains. However, this ability of SLX4 has not been demonstrated *in vitro* yet. Here, band-shift assays and fluorescence anisotropy show that SLX4^{SAP} indeed binds DNA with a slight preference for intact Holliday Junction. Furthermore, DNA binding of a complex of SLX4^{SAP} and MUS81^N showed higher DNA binding affinity than of the proteins alone with a clear preference of HJ and nicked HJ over dsDNA. Moreover, the area in SLX4^{SAP} was determined, which is responsible for DNA binding. In collaboration the NMR structure of SLX4^{SAP} was determined and compared with SAP domains from other proteins. A model how the SAP domain could bind DNA was discussed. Structural and biochemical information will help to answer the following questions: how is the complex of SLX4 and MUS81 differentiating the cleavage substrate? Which area in SLX4^{SAP} is involved in DNA binding? Finally, the results and conclusions from the experiments were discussed in the context of Holliday Junction resolution and the coordinated cleavage reaction by the four proteins.

2. Results

2.1 Expression and Purification of the Human SM-Complex

To obtain structural and biochemical information on the Holliday-Junction cleavage complex consisting of SLX4-SLX1-MUS81-EME1 (SM-complex), I tested expression of *Homo sapiens* proteins in *E. coli* and in mammalian cell expression system. The human proteins are particularly interesting as human SLX4 is found to be mutated in the rare genetic disease Fanconi anemia [14, 248, 249]. The first part of this chapter will describe the purification and crystallization process for the human proteins and in a later section the expression and purification of the *C. elegans* orthologous proteins will be described. As a general screening outline, it was always first tested, whether full length proteins could be expressed and purified and then shorter constructs were designed based on limited proteolysis and secondary structure prediction. Especially, for the human SLX4 protein the regions in the middle of the protein were intrinsically disordered.

2.1.1 Purification of hsMUS81-EME1

The structure-selective endonuclease MUS81-EME1 complex belongs to the XPF-ERCC4 family of nucleases and MUS81 contains the functional catalytically active ERCC4 domain whereas EME1 harbors the catalytically inactive domain [174]. MUS81-EME1 interact with each other via their C-terminal HhH domains [135, 213]. Previous publications about MUS81-EME1 showed that recombinant MUS81-EME1 are poorly expressed individually, but co-expression gave good yields [142, 251]. In order to overproduce full-length hsMUS81-EME1 (construct 1, Figure 2.3) I used a codon optimized synthetic gene, in which the codons are optimized for the codon bias and for an even distribution of GC content. The two proteins were cloned as a bicistronic construct with N-terminally His8-SUMO (small-ubiquitin related modifier)-tagged hsEME1 and untagged hsMUS81. The proteins were solubly expressed in *E. coli* BL21 gold cells and purified using IMAC (immobilized metal affinity chromatography), heparin affinity chromatography and SEC (size-exclusion chromatography) (Figure 2.1). SEC or gel filtration is a chromatographic method, which separates proteins according their size and so, it can be used for purification as well as analytical purposes. Larger proteins elute at shorter retention times and smaller proteins at longer retention times. Proteins forming a complex, are expected to elute together in one peak. Size exclusion chromatography of hsMUS81-EME1 showed one main peak and few smaller peaks. SDS-PAGE analysis revealed that the peaks, that elute at later time points, contained either degradation products or contaminants and the last peak was composed of the His8-SUMO tag. The main peak was analyzed by total mass spectrometry and showed to contain the FL protein of

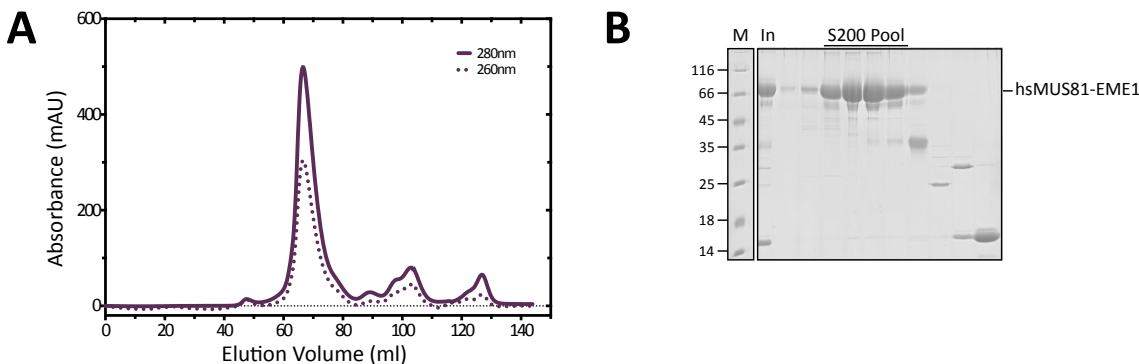


Figure 2.1: SEC analysis of hsEME1-MUS81 using a Superdex 200 16/600 column. (A) Size exclusion chromatogram of purified hsEME1-MUS81 showing an elution profile with a major peak and few smaller peaks visualized by UV absorbance at 280 nm and 260 nm. The major peak eluted at 66 ml containing the full length EME1 and slightly degraded MUS81. The other peaks eluted at 88.5 ml, 97.1 ml and 103.0 ml consisting of either various degradation products or contaminants. The last peak eluted at 127.0 ml and contained His8-SUMO tag. (B) SDS-PAGE shows the fractions of the main peak and the other peaks. The pooled fractions are marked by a line. (M) Protein marker in kDa. (In) Input protein sample injected on gel-filtration column.

EME1. Full length MUS81 was not detected, however, there were two slightly degraded fragments each lacking 11 kDa either at N- or C-terminus. They correspond to a MUS81 fragment lacking the N-terminal HhH domain and another MUS81 fragment lacking the C-terminal HhH domain. Despite the degradation issues, a limited crystallization screening was performed with various DNA substrates but did not yield any crystals.

In order to obtain constructs with a higher possibility to crystallize, I engineered several truncated constructs and tested them for expression and purification. For this especially the MUS81 construct was split up into two parts. Based on the already published structure of human MUS81-EME1 (PDB: 4P0P and 4P0Q) and secondary structure prediction more constructs containing the ERCC4 domain with the active site and the C-terminal HhH domains were designed and tested for purification (Figure 2.3). The published NMR structures of the human MUS81 N-terminal HhH and WH domain [143, 219] lead to the design of the hsMUS81²⁻²²⁵ construct combining the two domains (see below). To tackle the degradation problem of MUS81, the proteins were cloned into a bicistronic vector with N-terminal His8-SUMO tag (Figure 2.2 A) for hsEME1 and C-terminal CPD (cysteine protease domain)-TwinStrep tag for hsMUS81 (Figure 2.2 C). In general, a C-terminal tag on MUS81 should prevent C-terminal degradation and increase the yield of full length protein. The CPD tag is an inducible autocleaving protease tag [252, 253] (Figure 2.2 C) adapted from a *Vibrio cholerae* MARTX toxin. The addition of inositol-6-phosphate leads to an activation of the CPD protease and CPD cleaves at the target protein-CPD junction. The protease domain recognizes the amino acid sequence XXXL-AGGK and cleaves between the leucine and alanine residue leaving only a leucine residue left at the C-terminus of the target protein. After binding the target protein to an affinity column, tag cleavage allows a very convenient separation between target protein and the remaining tag. Furthermore, it has been shown before, that the CPD tag increases expression, solubility and stability of target proteins [253].

The constructs hsEME1²³³⁻⁵⁷⁰-MUS81²⁵⁶⁻⁵⁵¹, hsEME1²¹⁰⁻⁵⁷⁰-MUS81¹³³⁻⁵⁵¹ and hsEME1¹⁷⁸⁻⁵⁷⁰-MUS81²⁴¹⁻⁵⁵¹ were purified using IMAC, followed by affinity chromatography with Strep-Tag. The

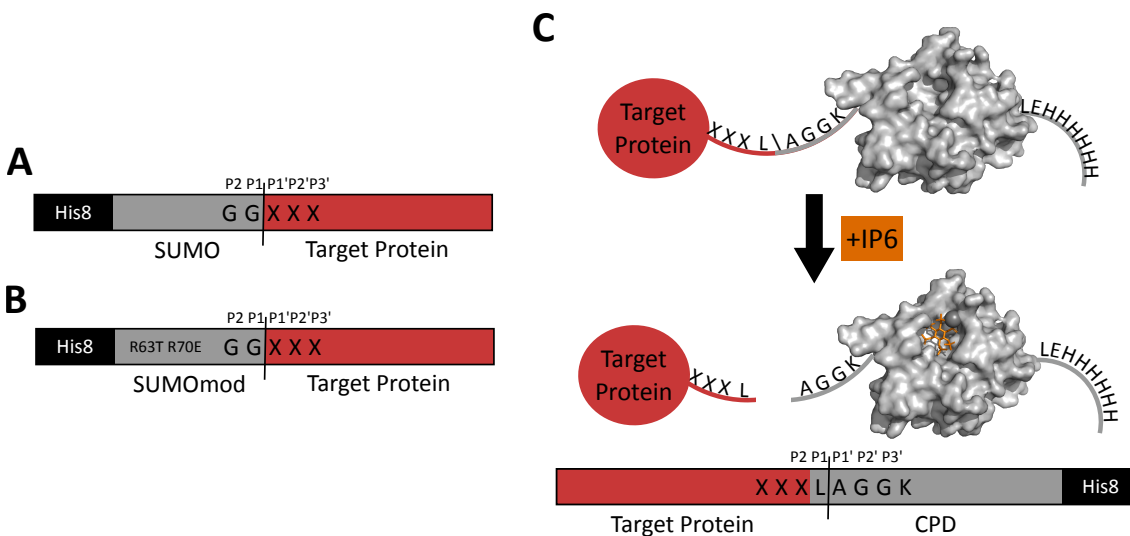


Figure 2.2: Fusion protein purification systems. (A) The small-ubiquitin related modifier (SUMO)-tag serves as a solubility enhancer for recombinant protein and is a protease recognition site for the protease Semp2 [254]. (B) For expression in mammalian expression system the SUMO tag was modified to prevent cleavage by endogenous Semp2 protease [255]. (C) Cysteine protease domain tag (PDB 3eeb [256]) is covalently linked to the target protein. Upon addition of inositol hexakisphosphate (IP6) the CPD protein undergoes cleavage at the P1 Leu site. The residues termed X are the C-terminal residues of the target protein. At P2 position the most common residues are A, G, S or W [252]. The auto cleavage site is marked by the black line between P1 and P1'.

proteins were eluted from the Strep column by activation of the CPD self-cleavage property with IP6 (Inositol-hexakisphosphate). However, in all four tested constructs, the cleavage activity was not sufficient as large amounts of CPD-Strep-tagged protein remained on the column and were eluted when the Strep-tag was displaced from the column material by desthiobiotin. Additionally, the removal of the SUMO-tag lead to decreased stability and solubility of the proteins. Taken together, all constructs had either precipitation issues after tag cleavage under the tested conditions or the CPD cleavage was not efficient and separation between tag and protein after cleavage was not possible.

Next, expression and purification of truncated constructs of hsMUS81-EME1 were tested only with N-terminal His8-SUMO tag on hsEME1 in the bicistronic vector and untagged hsMUS81. The published construct hsEME1¹⁷⁸⁻⁵⁷⁰-MUS81²⁴¹⁻⁵⁵¹ (Construct 2) was tested as a starting point [214]. The protein was purified by a combination of IMAC, weak cation exchanger and SEC and resulted in pure protein as judged by SDS-PAGE (Figure 2.4). The sample was subjected to crystallization with the published DNA substrates and our own designed DNA substrates. The goal of this reproduction was to evaluate whether the sample quality was high enough in order to produce crystals. hsEME1¹⁷⁸⁻⁵⁷⁰-MUS81²⁴¹⁻⁵⁵¹ in complex with one of the published substrates (17bp 5'flap) gave rise to crystals in 0.1 M MES pH 6.5, 5% Ethanol, 5 mM DTT. The crystal diffracted to 3.1 Å. The structure of this construct could be solved by molecular replacement with the published coordinates of 4POP using PHASER. But our structure did not show additional details than the already published structure. Other nHJ and HJ substrates, which were used for crystallization as well, did not give any crystal hits. Taken together these experiments show that

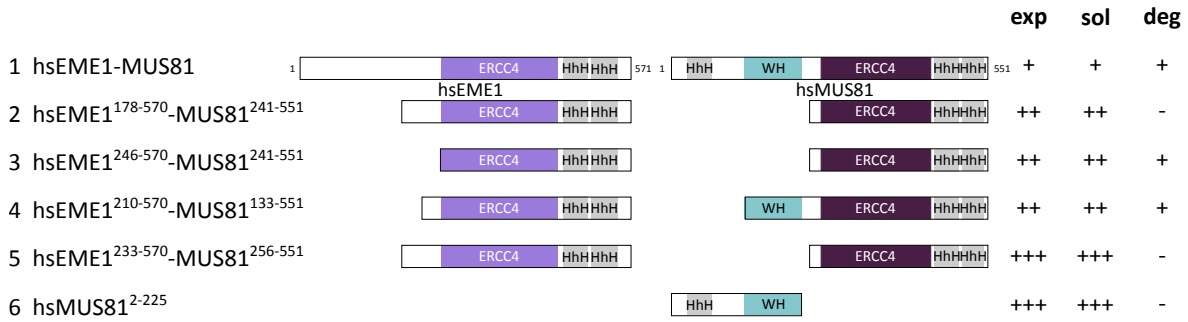


Figure 2.3: Schematic representation of hsEME1-MUS81 constructs the constructs were tested for expression (exp), solubility (sol) and degradation (deg) in *E. coli*. Success of the experiment is indicated by plus (positive) and minus (negative). Most of the constructs could be soluble expressed and did not show much degradation. All bicistronic constructs except for hsEME1²⁴⁶⁻⁵⁷⁰-MUS81²⁴¹⁻⁵⁵¹ were tested in two different vectors with different tags, N-terminal-His8-SUMO (pCB-bax04) and N-terminal-His8-SUMO-C-terminal-CPD-Twinstrep (pCB-bax25). The MUS81²⁻²²⁵ construct was expressed with N-terminal-His8-SUMO (pCB-bax04) and C-terminal-CPD-His8 (pCB-bax10) tag. However, the tested constructs did not yield any crystals with the DNA substrates. HhH helix-hairpin-helix (light grey); ERCC4 excision repair cross complementation group 4 (dark purple: catalytic active domain; light purple: catalytic inactive domain); WH winged helix (cyan).

the quality of the purified protein was suitable for crystallization but DNA substrates other than the published ones did not give rise to crystals.

As mentioned before, a construct from aa 2 to aa 225 was designed based on the already published NMR structures of hsMUS81 WH and HhH domain [143, 219] and secondary structural prediction (Figure 2.3, Construct 6). These two N-terminal domains of MUS81 were not included in the structural studies by Gwon et al. [214] as their cloned construct contained hsMUS81²⁴⁶⁻⁵⁵¹ and hsEME1¹⁷⁸⁻⁵⁷⁰. The WH domain has been found to interact with DNA and has an influence on positioning as well as on cleavage activity of MUS81-EME1 and MUS81-EME2 complexes [219]. Moreover, the HhH domain in human MUS81 is involved in DNA interaction, which has been shown by fluorescence anisotropy and NMR chemical shift titration [143]. However, both NMR structures do not contain a DNA substrate and there is so far no information of how these N-terminal domains interact with DNA substrates and how they assist in orienting DNA with respect to the active site domain of MUS81. Therefore, the next approach was to purify the N-terminal hsMUS81²⁻²²⁵ construct and crystallize it alone and in complex with DNA. The hsMUS81²⁻²²⁵ construct was very well expressed in *E. coli* and turned out to be highly soluble in *E. coli* using a His8-SUMO tag. The protein construct eluted as single peak with some tailing in gel filtration (Figure 2.4 B). Analysis of the protein identity by mass spectrometry analysis applying ESI-TOF mass spectrometry (electrospray ionisation - time-of-flight) showed the correct size of the protein construct. hsMUS81²⁻²²⁵ alone and in complex with various DNA substrates was subjected to crystallization trials.

The N-terminal DNA binding domains in the construct hsMUS81²⁻²²⁵ and the bicistronically expressed C-terminal constructs of MUS81-EME1 containing the ERCC4 domain were combined. Through the combination, a nearly full-length MUS81 was assembled on the DNA substrate with reduced degradation issues compared to the full-length expressed protein. By mixing the proteins and the DNA substrates in 1.0:1.0:1.1 molar ratio, the DNA binding domains and the active site would together bind the DNA substrate and were subjected to crystallization. However, none of

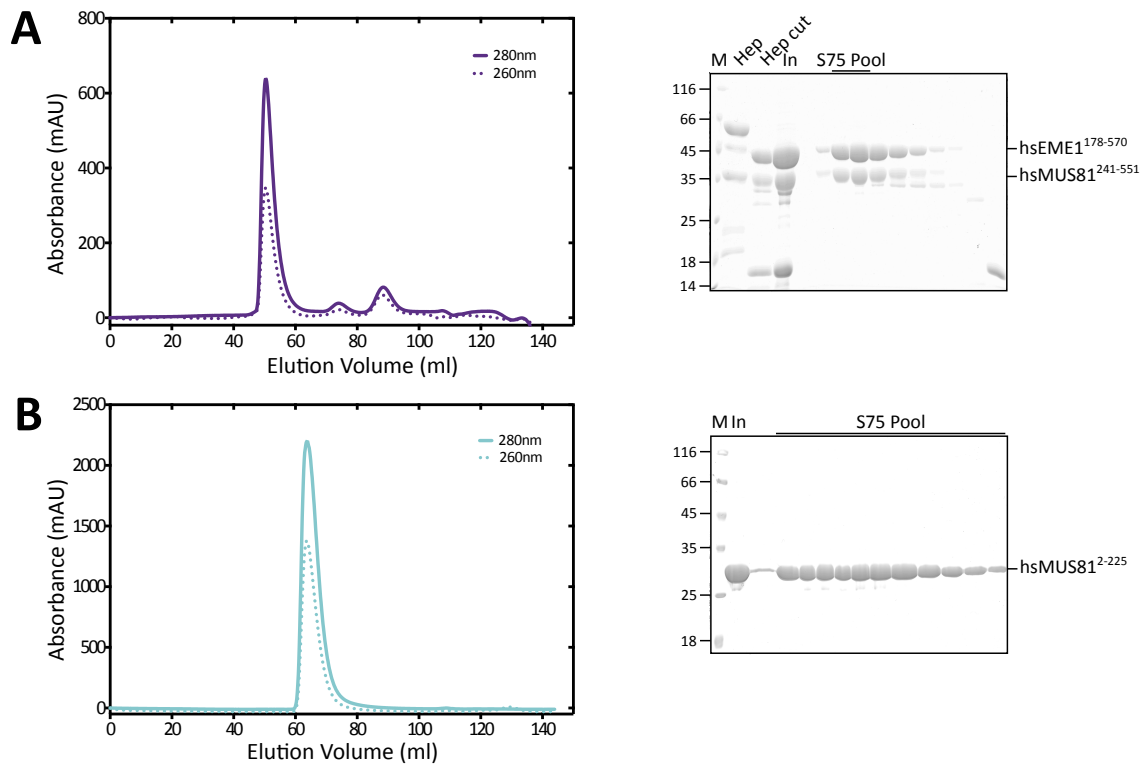


Figure 2.4: SEC analysis of hsMUS81 constructs based on limited proteolysis using a Superdex 75 16/600 column. (A) Size exclusion chromatogram of purified hsEME1¹⁷⁸⁻⁵⁷⁰-MUS81²⁴¹⁻⁵⁵¹ showing an elution profile with a major peak and few smaller peaks visualized by UV absorbance at 280 nm and 260 nm. The major peak elutes at 50.5 ml, the other peaks elute at 73.6 ml and 88.3 ml. The SDS-PAGE next to it shows the fractions of the main peak and the other peaks. (B) Size exclusion chromatogram of purified hsMUS81²⁻²²⁵ showing elution profile with a major peak visualized by UV absorbance at 280 nm and 260 nm. The major peak elutes at 63.8 ml. Right side, the SDS-PAGE shows the fractions of the main peak. The pooled fractions are marked by a line. (M) Protein marker in kDa. (Hep) Protein sample after Heparin column. (Hep cut) Protein sample after Snp2 cleavage. (In) Input protein sample injected on gel-filtration column.

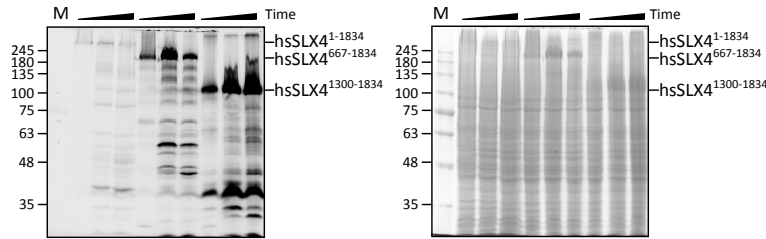


Figure 2.5: Expression test of different hsSLX4 constructs in HEK293 suspension culture. N-terminal His8-eGFP-SUMO tagged hsSLX4¹⁻¹⁸³⁴, hsSLX4⁶⁶⁷⁻¹⁸³⁴ and hsSLX4¹³⁰⁰⁻¹⁸³⁴ were expressed at 3:1 PEI:DNA ratio in Freestyle™ expression media supplemented with 1% FCS. Expression duration was one, two and three days. Samples of 0.2×10^6 cells/ml were applied to 10% SDS-PAGE and analyzed by in gel fluorescence (left) and Coomassie stain (right).

the trials containing hsMUS81²⁻²²⁵ lead to crystallization hits.

2.1.2 Expression and Purification of hsSLX4 and hsSLX1

SLX4 interacts as a platform protein with other proteins involved in DNA repair, especially structure-specific endonucleases like MUS81-EME1 and SLX1 [113–115]. The human 200 kDa SLX4 protein consists of six domains (Figure 1.4 A) [113–115, 158]. The C-terminus harbors the SAP domain, which interacts with the MUS81-EME1 nuclease complex [201, 202], and the CCD domain, which confers interaction with the GIY-YIG nuclease SLX1. Earlier biochemical studies described a nick and counter-nick mechanism for the four-protein complex on Holliday Junction substrate [141]. The expression of the full-length SLX4 protein was first attempted in *E. coli*, but did not yield any protein in pull down experiments, suggesting that the protein is not produced or insoluble. Secondary structure and disorder prediction indicated large unstructured regions. Therefore, expression of the full-length protein was tested in a mammalian expression system with HEK293 suspension culture. Expression of the full-length SLX4 and truncated SLX4 constructs, derived from secondary structure prediction, were tested with an N-terminal His8-eGFP-SUMO tag in small scale (Figure 2.5). For this, a small volume of cell culture was transiently transfected with the respective construct at different transfectant:DNA ratios and whole cell samples were taken after one, two and three days of expression. The samples were subjected to SDS gel electrophoresis and expression levels were detected by in gel fluorescence of the GFP tag. GFP was detected due to its stability towards SDS concentrations of up to 0.5% and omitting the boiling step for SDS-PAGE sample preparation [257]. For the full-length hsSLX4 only a weak signal was detectable at the respective molecular weight. Furthermore, after two or three days of expression, degradation of the protein was observed by SDS-PAGE. The N-terminally truncated constructs showed higher expression levels than the full-length protein, but with increasing duration of expression, a stronger breakdown of the protein occurred. The strongest expression level of the tested constructs was reached with hsSLX4¹³⁰⁰⁻¹⁸³⁴, which was used for further large scale expression in HEK cells and for purification.

Recombinant expression of hsSLX1 was attempted in *E. coli* but lead to the accumulation of insoluble protein in inclusion bodies. A common strategy to enhance protein solubility is the co-expression with an interaction partner. Therefore, co-expression with SLX4 as interaction partner

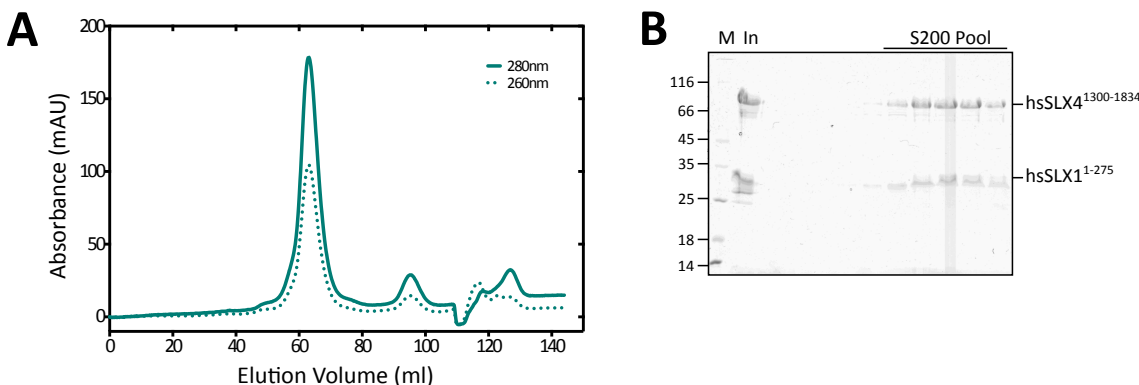


Figure 2.6: SEC analysis of hsSLX4¹³⁰⁰⁻¹⁸³⁴ co-expressed with hsSLX1¹¹⁻²⁷⁵ using a Superdex 200 16/600 column. (A) Size exclusion chromatogram of purified hsSLX4¹³⁰⁰⁻¹⁸³⁴ + hsSLX1¹¹⁻²⁷⁵ showing an elution profile with a major peak visualized by UV absorbance at 280nm and 260nm. The major peak elutes at 63 ml. The second peak elutes at 95.2 ml and contains the proteases. (B) SDS-PAGE shows the fractions of the main peak. The pooled fractions are marked by a line. (M) Protein marker in kDa. (In) Input protein sample injected on gel-filtration column.

was tested in HEK cells. Co-expression of hsSLX1 with hsSLX4¹³⁰⁰⁻¹⁸³⁴ construct resulted in soluble protein in small scale expression tests and in large scale expression. However, the yield of the expressed proteins was limited and a lot of expressed protein was still found in the insoluble fraction after lysis. In order to increase the amount of soluble protein, different methods of cell lysis (detergents, beads, sonication) were tested and assessed by fluorescence measurements. In the end, best lysis conditions were found with high salt lysis buffer containing the detergent NP-40 and sonication. hsSLX4¹³⁰⁰⁻¹⁸³⁴ and hsSLX1¹¹⁻²⁷⁵ were applied to a GFP-binder column (single domain anti-GFP fragment derived from heavy chain antibodies V_HH of Canelidae [258]) and after tag cleavage the eluted proteins were applied to gel filtration (Figure 2.6). SEC analysis showed one main peak at 63.0 ml and two smaller peaks at 95.2 and 126.9 ml. The smaller peak at 95.2 ml contained the proteases used for tag cleavage and the last peak could not be resolved by SDS-PAGE. The SDS gel showed slight degradation bands for both proteins.

The activity of the purified proteins was investigated in a cleavage assay testing different DNA substrates (Figure 2.7 A). The protein complex showed very weak cleavage on ssDNA substrate, medium activity on flap substrates and intact Holliday Junction. The strongest activity under the tested conditions was observed for a nicked HJ substrate. As the activity tests indicated that the purified proteins are active, I purified hsSLX4¹³⁰⁰⁻¹⁸³⁴ and hsSLX1¹¹⁻²⁷⁵ complex for crystallization trials. Crystallization with different DNA substrates was attempted with the active hsSLX1 protein and a construct with mutated active site residue Y17F, which was introduced based on the proposed cleavage mechanism and sequence alignments of SLX1 proteins from different organisms (Figures 1.5 B and 4.2). However, despite testing different protein concentrations and DNA substrates no condition yielded crystals.

As the hsSLX4¹³⁰⁰⁻¹⁸³⁴ construct was prone to degradation as seen in the SDS-PAGE (Figure 2.6), limited proteolysis was carried out for the complex with hsSLX1¹¹⁻²⁷⁵ in search for a more stable and compact construct. Limited proteolysis applies different proteases, which digest unfolded protein regions in a non-specific manner. Chain segments in regular secondary structures will be

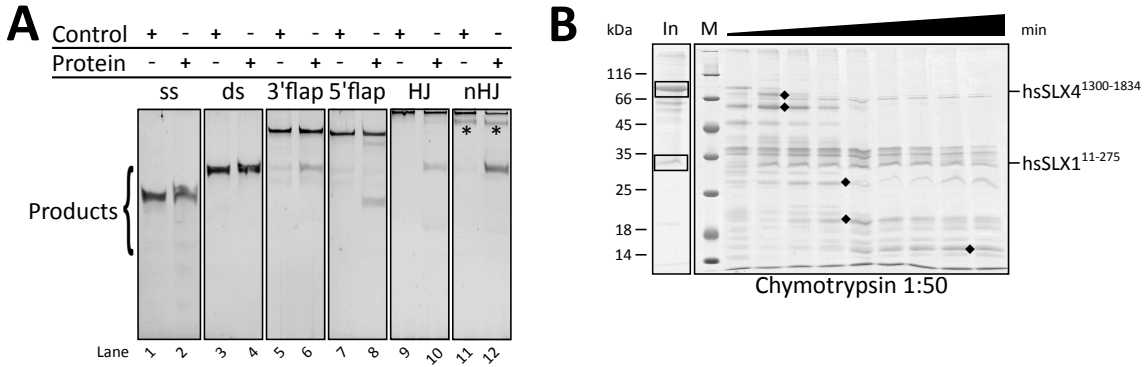


Figure 2.7: Activity test and limited proteolysis of $hsSLX4^{1300-1834}$ - $hsSLX1^{11-275}$. (A) Nuclease activity of $hsSLX4^{1300-1834}$ - $hsSLX1^{11-275}$ with different 5' FAM-labeled DNA substrates. 640 nM protein complex was incubated for 30 min at 37°C with 40 nM DNA substrate. The cleavage products were separated by native TB-PAGE and analyzed using a PhosphorImager. Bands marked with an asterisk indicate artifacts. (B) Time course of limited proteolysis for $hsSLX4^{1300-1834}$ - $hsSLX1^{11-275}$ with Chymotrypsin at 0.02 mg/ml concentration. Samples were taken after 1, 5, 10, 15, 30, 45, 60, 90 and 120 min. (In) Input protein sample used for proteolytic analysis. (M) Protein marker in kDa. Bands marked with a diamond on the right indicate stable $hsSLX4^{1300-1834}$ - $hsSLX1^{11-275}$ fragments and were sent for peptide fingerprinting.

less accessible to the active site of the protease as flexible loop regions. After testing five different proteases at three different concentrations, a time course with Chymotrypsin at 0.05 mg/ml was performed (Figure 2.7 B). The bands, which were analyzed by peptide mass fingerprinting for identification, are marked with a diamond at the side.

All bands sent to identification solely contained peptides of $hsSLX4$. $hsSLX1$ did not get further truncated in limited proteolysis experiments. Based on the results from limited proteolysis, new constructs were designed for $hsSLX4$ and tested for expression in *E. coli* with either N-terminal SUMO-tag or C-terminal CPD-tag (cysteine-protease domain) and in HEK293 culture with C-terminal mVenus-tag (Figure 2.8). The CPD-tag is an inducible auto-cleaving protease tag, which also enhances protein solubility (see section 2.1.1 and Figure 2.2). Different constructs were designed containing only the SAP, only the CCD domain or both of them together at different length (Figure 2.8).

Several $hsSLX4$ constructs showed medium expression levels with the CPD tag in bacterial expression (Constructs 8 - 13), especially the large $hsSLX4^{1300-1834}$ construct could be expressed (Construct 8). Thus, $hsSLX4$ constructs 8 - 13 in vectors with C-terminal CPD tag were purified by Ni affinity column with subsequent IP6 cleavage. The cleavage mix was loaded on ion exchange column, eluted by increasing salt concentration and subjected to size exclusion chromatography. However, in CPD-tagged $hsSLX4^{1300-1834}$ (Construct 8) and smaller constructs ($hsSLX4^{1706-1824}$ (Construct 12), $hsSLX4^{1748-1824}$ (Construct 13)) the protease recognized not only its target site but also other sites containing nonpolar residues. This resulted in a strong degradation pattern after induction of CPD-cleavage (see Appendix Figure 4.5). To check for this additional CPD recognition sites, Leu residues in $hsSLX4^{1300-1834}$ were counted to 39 and then checked whether the sequence around the Leu resembled similarity to the CPD recognition site [252] (see Appendix Figure 4.6). In total, nine sites in the $hsSLX4^{1300-1834}$ construct were found, which contained either Ala, Gly, Ser or Trp in P2 site and might therefore be recognized by the activated cysteine

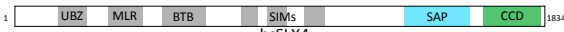
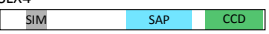









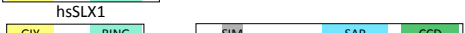

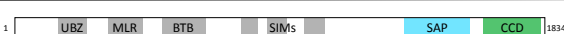





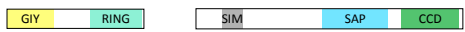




Bacterial Expression			exp	sol	deg
7	hsSLX4 ¹⁻¹⁸³⁴		-		
8	hsSLX4 ¹³⁰⁹⁻¹⁸²⁴		-		
9	hsSLX4 ¹⁴⁹²⁻¹⁸²⁴		-		
10	hsSLX4 ¹⁵¹⁶⁻¹⁸²⁴		-		
11	hsSLX4 ¹⁶⁴⁷⁻¹⁸²⁴		+	+	+
12	hsSLX4 ¹⁷⁰⁶⁻¹⁸²⁴		++	++	++
13	hsSLX4 ¹⁷⁴⁸⁻¹⁸²⁴		++	++	+
14	hsSLX4 ¹⁴⁹²⁻¹⁶⁴⁸		++	++	++
15	hsSLX4 ¹⁵¹⁶⁻¹⁶⁴⁸		++	++	++
16	hsSLX4 ¹⁵⁶⁵⁻¹⁶⁴⁸		+++	++	++
17	hsSLX1 ¹⁻²⁷⁵		++	-	
18	hsSLX1 ¹¹⁻²⁵⁸ +hsSLX4 ¹³⁰⁰⁻¹⁸²⁴		+	+	+++
19	hsSLX1 ¹¹⁻²⁵⁸ +hsSLX4 ¹⁷⁰⁶⁻¹⁸²⁴		+	+	++
HEK cell Expression			exp	sol	deg
20	hsSLX4 ¹⁻¹⁸³⁴		+	-	
21	hsSLX4 ⁶⁶⁷⁻¹⁸³⁴			+	++
22	hsSLX4 ¹³⁰⁰⁻¹⁸³⁴		+	+	-
23	hsSLX1 ¹⁻²⁷⁵		++	-	
24	hsSLX1 ¹¹⁻²⁷⁵ +hsSLX4 ¹³⁰⁰⁻¹⁸³⁴		+	+	+
25	hsSLX1 ¹¹⁻²⁵⁸ -hsSLX4 ¹³⁰⁰⁻¹⁸³⁴		-		
26	hsSLX1 ¹¹⁻²⁷⁵ +hsSLX4 ¹⁴⁸⁹⁻¹⁸¹⁸		+	+	++
27	hsSLX1 ¹¹⁻²⁷⁵ +hsSLX4 ¹⁶⁴⁷⁻¹⁸¹⁸		+	-	
28	hsSLX1 ¹¹⁻²⁵⁸ +hsSLX4 ¹⁶⁴⁷⁻¹⁸¹⁸		+	+	-
29	hsSLX1 ¹¹⁻²⁷⁵ +hsSLX4 ¹⁷⁰⁶⁻¹⁸¹⁸		+	+	++
30	hsSLX4 ¹⁷⁰⁶⁻¹⁸³⁴ -P2A-hsSLX1 ¹¹⁻²⁵⁸		+	+	++

Figure 2.8: Schematic representation of hsSLX4 and hsSLX1 constructs. The constructs were tested for expression (exp), solubility (sol) and degradation (deg) in *E. coli* and HEK293 cell culture. Bacterial expression was tested with either N-terminal His8-SUMO-tag or C-terminal CPD tag (upper part) and expression in mammalian suspension cells was carried out with C-terminal 3C-mVENUS-1d4 tag and / or with N-terminal His8-eGFP-SUMO (lower part). The success is indicated by + (positive) and - (negative). Most of the constructs were soluble, but especially the constructs in bacterial expression system were prone for degradation. The shorter constructs showed higher expression levels, but were also degradation prone.

protease. This creation of unwanted degradation products lead to the conclusion, that the CPD-tag was not applicable for hsSLX4 constructs.

Expression of constructs 14 - 16 comprising the SAP domain of hsSLX4 was carried out with a N-terminal His8-SUMO tag and in addition with C-terminal CDP-tag. Purification of the SUMO-tagged proteins using Ni column, anion exchange column and gel filtration lead up to 2 mg of protein with few degradation bands per l of expression culture. However, the tested protein constructs degraded over time, which can be potentially hindering for crystallization. Purification of the CPD-tagged SAP constructs (Construct 14 and 15) was carried out but lead to few degradation bands. In comparison the SUMO-tagged constructs had still less degradation bands. Crystallization trials were set up for the SAP domain constructs alone and in complex with DNA substrates, but they showed mainly precipitation and protein crystals could not be achieved. Hence, the strategy regarding the human proteins was shifted towards the expression of hsSLX1 and hsSLX4^{CCD} constructs in HEK cells.

hsSLX4 constructs containing the CCD domain were co-expressed in various combinations of hsSLX1 protein in mammalian cell culture (Figure 2.8 lower part, constructs 24 - 30) and reached low amounts of 0.2 mg per liter of expression culture after purification. Typically, purified protein amounts of 1 mg/ml can be reached in HEK suspension culture [259, 260]. Despite trying different tags the purified constructs showed strong degradation. Furthermore, the purified proteins were subjected to ESI-TOF mass spectrometry to assess the purity and it was found that the MS spectrum showed many peaks indicating heterogeneity. The peak differences in the spectrogram showed shifts of 80 Da post-translational modification, most likely phosphorylation. Further analysis of the phosphorylation of hsSLX4¹³⁰⁰⁻¹⁸³⁴ was performed and will be described in the following section 2.2. In summary, due to low expression level, degradation and heterogeneity of the different tested constructs, the hsSLX4 constructs did not gave hits in crystallization.

2.2 Phosphorylation Site Mapping of hsSLX4¹³⁰⁰⁻¹⁸³⁴

A construct comprising the C-terminal domains, hsSLX4¹³⁰⁰⁻¹⁸³⁴, could be expressed alone and together with hsSLX1¹¹⁻²⁷⁵ in HEK cells. Interestingly, the purified protein showed a peak series with differences of 80Da in ESI-TOF MS, indicating phosphorylation of the protein sample. As SLX4 is a platform protein and due to its interaction partners involved in different cellular pathways, its action is most likely regulated via post-translational modification. From literature it is known that phosphorylation of SLX4 has an influence on its association with its interaction partners [114, 141, 194, 195, 200, 261]. Wyatt et al. proposed a model in which CDK1-phosphorylation of SLX4 controls MUS81 association [141]. This could be underlined in a later study under the inhibition of Wee1 kinase interacting with CDK1 [200]. SLX4 is also a target of ATM / ATR kinases [79, 185]. However, the relevance of this phosphorylation remains elusive in mammalian cells. In yeast, Slx4 phosphorylation depends on Mec1 and Tel1 (ATR and ATM orthologs) in cells exposed to genotoxins [184, 195]. This phosphorylation was observed in cells arrested either in G₁ or G₂. [184]. Furthermore, PLK1 associates with SLX4 through a possible polo-box binding motif on Ser1453 [114]. Another example, showing the importance of phosphorylation, is the interaction between SLX4 and TOPBP1, which is conferred via phosphorylation at Thr 1260 by CDK [194]. This interaction has been also investigated with the yeast ortholog Dpb11 of TOPBP1 [193–195].

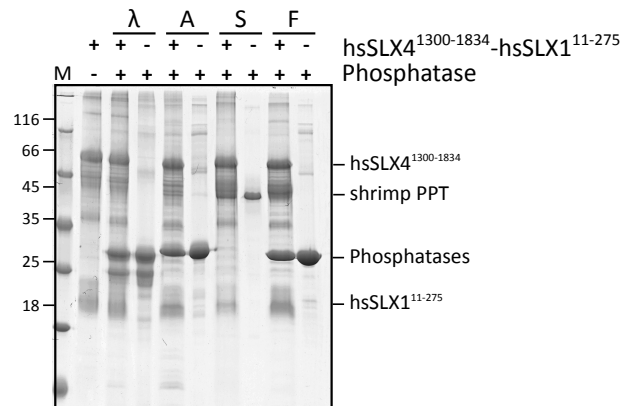


Figure 2.9: Electrophoretic shift analysis of hsSLX4¹³⁰⁰⁻¹⁸³⁴-hsSLX1¹¹⁻²⁷⁵ after phosphatase treatment. The dephosphorylation of purified hsSLX4¹³⁰⁰⁻¹⁸³⁴-hsSLX1¹¹⁻²⁷⁵ by various phosphatases causes limited electrophoretic shift in SDS-PAGE. hsSLX4¹³⁰⁰⁻¹⁸³⁴-hsSLX1¹¹⁻²⁷⁵ was incubated with λ phosphatase (λ), Antarctic phosphatase (A), shrimp phosphatase (S) and fast alkaline phosphatase (F). λ phosphatase (λ), Antarctic phosphatase (A) and fast alkaline phosphatase (F) ran at approximately 30 kDa and their labeling is summarized by "Phosphatases" on the right site of the gel.

This interaction is mediated by multiple phosphorylation sites in yeast Slx4 [193, 195] and cell cycle regulated as well as dependent of the CDK site Ser 486 in Slx4 [193, 194]. In yeast cells arrested in G₁ or G₂, Slx4 and the scaffold protein Rtt107 are heavily phosphorylated by Mec1, a DNA damage signaling kinase (ATR in humans) [184]. This phosphorylation by Mec1 is only carried out when each of the interaction partners is present [195]. The phosphorylation-regulated complex Rtt107-Slx4-Dpb11 plays an important role in the response to replication stress [261]. Furthermore, yeast Slx4 is not only involved in the resolution of HJs during mitosis, but plays a role in the resection of the 5' DNA ends by counteracting Rad9 and subsequent resection [261]. Thus Slx4 controls signaling events in order to facilitate HR [261]. As the yeast ortholog of Slx4 shows important phosphorylation sites, it was reasoned that the human SLX4 might also show prominent phosphorylation sites.

As a first and rapid experiment, the electrophoretic mobility shift of purified hsSLX4¹³⁰⁰⁻¹⁸³⁴ and hsSLX1¹¹⁻²⁷⁵ before and after phosphatase treatment was analyzed (Figure 2.9). The analysis of the electrophoretic mobility shift is a simple method to determine the ratio between phosphorylated and unphosphorylated protein. Same amounts of protein were treated with different phosphatases and their height on a SDS-PAGE was compared to an untreated protein sample. Four different phosphatases were applied, λ phosphatase, Antarctic phosphatase, shrimp alkaline phosphatase and fast alkaline phosphatase. The enzymes differ in their reactivities towards serine/threonine and tyrosine phosphorylation. λ phosphatase has dual specificity on serine/threonine and tyrosine phosphates [262]. Antarctic phosphatase shows similar activity on both types of phosphorylation sites [263]. Fast alkaline phosphatase has a preference for phosphotyrosine residues [263]. Shrimp alkaline phosphatase can be applied for dephosphorylation of DNA ends and proteins [263]. For each phosphatase the respective protocol was applied in order to have maximum enzyme activity. After phosphatase treatment only a limited shift between the samples could be observed (Figure 2.9). Treatment with λ phosphatase (λ) and shrimp phosphatase (S) did not show a shift in electrophoretic mobility. Treatment with Antarctic phosphatase (A) and fast alka-

line phosphatase (F) showed a marginal increase in electrophoretic shift and the protein bands for the dephosphorylated hsSLX4 construct ran few millimeters lower than the control band for the hsSLX4 construct. In summary, analysis of electrophoretic shift with a SDS-PAGE did not give an unambiguous result for the phosphatase of choice and the extent of dephosphorylation. These results hit the limitation of the method. Literature reported that phosphorylation-induced shifts in electrophoretic mobility are most likely due to local context-dependent effects on the flexibility of the protein chain rather than to changes in molecular weight [264]. Furthermore, the decrease in charge and the increase in SDS-binding after de-phosphorylation might not have been strong enough to induce the shift in electrophoretic mobility. Thus, the approach for the examination of phosphorylation sites was changed and will be described in the next subsection.

2.2.1 Phosphorylation sites of hsSLX4¹³⁰⁰⁻¹⁸³⁴

First the bioinformatic analysis tool GPS 3.0 [265] was applied for the prediction of phosphorylation sites in hsSLX4¹³⁰⁰⁻¹⁸³⁴ taking the kinases CDK1, PLK1 and ATM/ATR into account and showed in total 47 predicted phosphorylation sites for CDK1, PLK1 and ATM/ATR (Figure 2.10 A). The GPS 3.0 online prediction tool searched for minimal CDK1 consensus sites (S / TP) and found 25 phosphorylation sites. For PLK1 and its consensus sequence (D / E - X - S / T - hydrophobic aa - X - D / E) the software found 11 phosphorylation sites and for ATM / ATR (T / SQ) 11 sites. The search for kinase consensus sequences shows that there is the theoretical possibility that SLX4 can be heavily phosphorylated. Furthermore, literature suggests that SLX4 is phosphorylated and we also found it heavily phosphorylated. Therefore, it is likely that heterogeneity of SLX4 originates from phosphorylation. A reason for the inability to crystallize the constructs is caused by the presence of these post-translational modifications. In order to examine the amount of phosphorylation of SLX4 and to assess the influence of phosphatases regarding de-phosphorylation a proteomic approach was applied. Different phosphatases were tested in order to find a condition with low or missing phosphorylation. The phosphorylation sites of hsSLX4¹³⁰⁰⁻¹⁸³⁴ in absence and presence of hsSLX1 were examined subsequently. Furthermore the phosphorylation sites of hsSLX4¹³⁰⁰⁻¹⁸³⁴ in presence of hsSLX1 treated with different DNA damaging agents were investigated.

hsSLX4¹³⁰⁰⁻¹⁸³⁴ was expressed alone and co-expressed with hsSLX1¹¹⁻²⁷⁵ in transiently transfected HEK293 suspension culture, purified by GFP-binder affinity chromatography and peptide finger printing as well as mapping of phosphorylation sites was carried out by Dr. Nagarjuna Nagaraj in the Biochemistry Core Facility at the Max-Planck Institute of Biochemistry (Martinsried, Germany).

For the solely expressed SLX4 construct six phosphorylation sites could be identified (4x Ser, 2x Thr), whereof four were located close to the SIM 3 motif and two were between the SAP and the CCD domain (Figure 2.10 B upper panel). Comparison with the predicted phosphorylation sites showed that all six phosphorylations are located in CDK1 consensus sequences and thus likely been introduced by CDK1. In co-expression with hsSLX1¹¹⁻²⁷⁵ SLX4 was much stronger phosphorylated with 18 detected sites, 5x Thr and 13x Ser (Figure 2.10 C upper panel). Most of phosphorylation sites were located between the domains, except for one site (S1567), which was in the SAP domain. Phosphorylation of S1453 for PLK1 interaction [114] was not detected, however

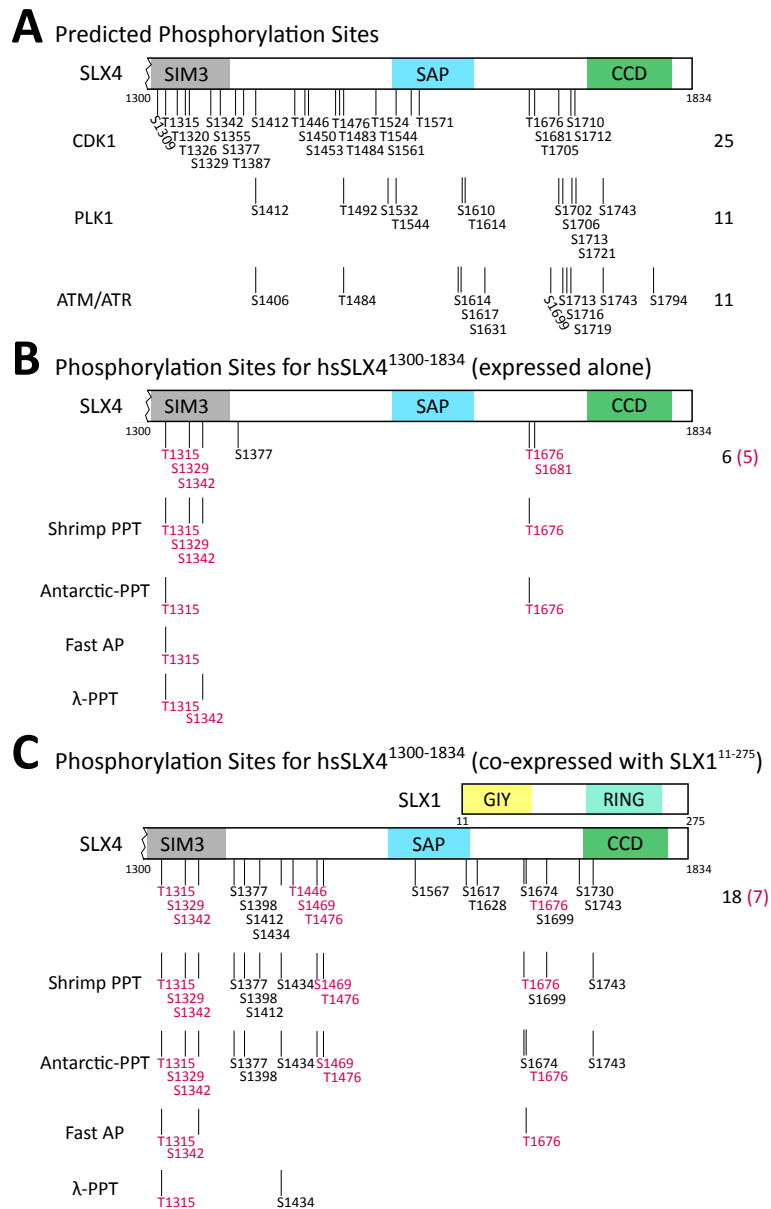


Figure 2.10: Phosphorylation site mapping of hsSLX4¹³⁰⁰⁻¹⁸³⁴ alone and co-expressed with hsSLX1¹¹⁻²⁷⁵. (A) Phosphorylation sites based on phosphorylation site prediction by GPS 3.0 (<http://gps.biocuckoo.org>) specific for CDK1, PLK1 and ATM / ATR phosphorylation. (B) hsSLX4¹³⁰⁰⁻¹⁸³⁴ was expressed alone in asynchronous HEK cell culture, purified and phosphorylation sites were analyzed by MS fingerprinting. Different phosphatases (see below) were used in order to remove the phosphorylation sites. (C) hsSLX4¹³⁰⁰⁻¹⁸³⁴ was co-expressed with hsSLX1¹¹⁻²⁷⁵ in asynchronous HEK cell culture, purified and phosphorylation sites were analyzed by peptide fingerprinting. For removal of phosphorylation sites different phosphatase were used for treatment: shrimp phosphatase (Shrimp PPT), antarctic phosphatase (Antarctic-PPT), fast alkaline phosphatase (Fast AP) and λ phosphatase (λ PPT). For peptide fingerprinting a protein amount of 40 μg for control and phosphatase treated samples was digested in solution with GluC and Trypsin to increase peptide coverage. For analysis only peptide intensities above 1×10^7 and phosphosites above 75% probability showing up in all triplicates were taken into account. Magenta residues indicate phosphorylation sites found in the database Phosphosite Plus [266]. The analysis of phosphorylation sites is based on triplicate measurements.

the adjacent residues T1446 and S1469 were phosphorylated in presence of SLX1. Peptides for hsSLX1¹¹⁻²⁷⁵ were identified in the peptide mass analysis but phosphorylations were not detected. Furthermore, phosphorylation of the CCD domain, the interaction site between SLX4 and SLX1, was not observed. In the course of this study, another investigation was published on the effect of the mutation of predicted CDK consensus sites around the SAP domain (S1469, T1476, T1524, T1544, T1561, T1571) [200]. Duda et al. used stably expressed SLX4 full length protein, mutated the sites to alanine and the interaction with MUS81 was abolished [200]. The SLX4-MUS81 association is controlled by CDK1-mediated phosphorylation of SLX4 and this phosphorylation is coupled to mitotic entry [200]. The mentioned phosphorylation sites except for S1469 and T1476 were not detected in the presented phosphorylation screen.

Comparing the experimentally detected phosphorylation sites with the predicted sites showed that one third of them were not shown in the prediction. The other sites were mostly predicted to be phosphorylated by CDK1, with the exemption of S1412, which could be both a CDK1 and a PLK1 site, two ATM / ATR sites (S1617 and S1699) and S1743, which could be either a PLK1 or an ATM / ATR site.

Next, the found phosphorylation sites were compared with the database Phosphosite Plus [266]. This database is a resource for experimentally observed post-translational modifications and comprises data from journal articles and cell signaling technologyTM scientists. This database served as a resource for the recovery rate of our analysis but also as comparison of the conditions in which the sites were identified. The here presented analysis identified 19 unique phosphorylation sites found in hsSLX4¹³⁰⁰⁻¹⁸³⁴ from expression conditions alone and in co-expression with SLX1¹¹⁻²⁷⁵. Only eight of these sites were also retrieved from the Phosphosite Plus database (Figure 2.10 B, C magenta residues and Table 2.1). When hsSLX4¹³⁰⁰⁻¹⁸³⁴ was expressed alone five of the six detected phosphorylation sites were also listed in the database. Interestingly, when the phosphorylation sites of the co-expressed hsSLX4¹³⁰⁰⁻¹⁸³⁴ with its 18 detected sites was searched, only seven of them could be found. The phosphorylation site mapping revealed under the conditions of the co-expression with hsSLX1¹¹⁻²⁷⁵ not only a different phosphorylation pattern compared to before listed phosphorylation sites, but also new sites were identified under the influence of hsSLX1¹¹⁻²⁷⁵ co-expression. When the phosphorylation sites from Phosphosite Plus database were retrieved, most of the sites were detected in tissues derived from leukemia or breast cancer. This is an interesting finding as leukemia as well as breast cancer are associated with defects in HR. Especially phosphorylations at T1315 and S1329 were detected in many proteomic studies and are the most commonly found phosphorylation sites of hsSLX4 in *in vivo* studies (Table 2.1). But so far, there are no further investigations reported, what the role of those specific phosphorylation sites is.

For the yeast ortholog of Slx4 eight phosphorylation sites have been identified [195, 261]. Most of the phosphorylation sites are located at the C-terminal part of the yeast protein and are introduced upon DNA damage by the Mec1 / Tel1 kinases (see also Appendix 4.7). However, upon comparison with the used hsSLX4¹³⁰⁰⁻¹⁸³⁴ construct, four of the eight sites were located in the human construct. Three of the yeast sites were not found phosphorylated in the human ortholog. Only S627 in scSLX4 could be identified as phosphorylation site S1567 in hsSLX4, which is located in both proteins in the SAP domain. In yeast Slx4 this site is together with six other phosphorylation sites important for the interaction with the TOPBP1 ortholog Dpb11 [195]. Mutation of this site together with six more residues abolishes phosphorylation by Mec1

Table 2.1: Comparison of experimentally found phosphorylation sites in hsSLX4¹³⁰⁰⁻¹⁸³⁴ with the Database Phosphosite Plus [266] (site group ID). The table gives only a selection of the references given for each phosphorylation site. Only the tissues and cell lines with the most frequent references are depicted. MS mass spectrometry, Jurkat T lymphocyte cells, HEK293 human embryonic kidney cells, HeLa cervix carcinoma cells

Phosphorylation Site	Site Group ID	Method of characterization	Disease Tissue	Cell Line	Number of references
T1315	472085	MS	breast cancer, leukemia, acute myelogenous leukemia	bone marrow, Jurkat, HeLa	50
S1329	472088	MS	breast cancer, leukemia, acute myelogenous leukemia	HEK293, bone marrow, HeLa, Jurkat	43
S1342	8110962	MS	n.a.	Jurkat	1
T1446	34352045	MS	n.a.	Jurkat	1
S1469	3225800	MS	various breast cancer types, lung cancer	HeLa, breast cell lines, pulmonary cell lines	19
T1476	4738110	MS	n.a.	n.a.	1
T1676	31082723	MS	breast cancer	HeLa, HEK293	5
S1681	31081478	MS	n.a.	HeLa, erythroid cell line	1

and interaction with Dpb11 [195]. In a recent study with human SLX4, mutation of all predicted CDK phosphorylation sites in the SAP domain to alanine abolished interaction with MUS81 [200]. Mutation of the residues to aspartic acid or glutamic acid has not been tested so far. The SAP domain as a phosphorylation site might be involved in important protein-protein interactions.

The detection of the phosphorylation sites showed the reason for the heterogeneity of the protein sample. In order to obtain a uniform sample with reduced post-translational modification for crystallization, I co-expressed and purified hsSLX4¹³⁰⁰⁻¹⁸³⁴-hsSLX1¹¹⁻²⁷⁵ and solely expressed hsSLX4¹³⁰⁰⁻¹⁸³⁴, which were treated with four different phosphatases after purification. The same phosphatases as for electrophoretic mobility shift assay were applied to use their different reactivities towards serine/threonine and tyrosine phosphorylation. For each phosphatase the respective protocol was applied in order to have maximum enzyme activity. For treatment of hsSLX4¹³⁰⁰⁻¹⁸³⁴ expressed alone, a strong reduction in phosphorylation was achieved. The best results were observed when applying either λ -phosphatase or fast alkaline phosphatase resulting in only one phosphorylated residue (T1315) (Figure 2.10 B lower panel). In co-expression with hsSLX1, the phosphatases removing most phosphate groups were λ -phosphatase and Fast-AP (Figure 2.10 C lower panel). Antarctic phosphatase removed a few phosphates, but shrimp phosphatase removed nearly no phosphate group, which made it not suitable for protein dephosphorylation. Interestingly, the phosphorylation of T1315 could never be removed, which suggests a location within a non-accessible region. Furthermore, this site has the most references found for hsSLX4 in the database Phosphosite Plus [266].

In subsequent purifications, hsSLX4 protein constructs expressed in HEK cells were de-phosphorylated with either λ -PPT or FastAP and were submitted to crystallization: However, despite the reduced phosphorylation they did not give rise to crystals.



Figure 2.11: Heat Map cluster of hsSLX4¹³⁰⁰⁻¹⁸³⁴ phosphopeptides in co-expression with hsSLX1¹¹⁻²⁷⁵ under influence of different DNA damaging agents. Each column of phosphosite intensities represents the average of triplicate measurements. Numbers on the site indicate the position of the detected phosphorylation site in hsSLX4¹³⁰⁰⁻¹⁸³⁴. The bar on the left site refers to domain organization of the construct. The heat map depicts log₂ to of the localization probability of the phosphorylation of a certain amino acid. Dark pink rectangles indicate high intensity and thus localization probability of the respective phosphorylation site. Dark green rectangles indicate low intensity of appearance. Grey spots refer to phosphorylation sites which could not be detected (n.d.). Control: co-expression of hsSLX4¹³⁰⁰⁻¹⁸³⁴ and hsSLX1¹¹⁻²⁷⁵ without addition of DNA damaging agents; Eto: incubation for 1 h with 5 μ M Etoposide; HU: incubation with 2 mM hydroxyurea for 3 h; MMC: incubation for 1 h with 0.1 μ g/ml with mitomycin C; MMS: incubation for 1 h with 2 mM methylmethanesulfonate.

2.2.2 Phosphorylation Site Mapping under the Influence of DNA Damage

In a next step, the expression of hsSLX4¹³⁰⁰⁻¹⁸³⁴ under the influence of different DNA damaging agents was tested in order to find a condition with a distinct distribution of phosphorylation sites. The strategy behind this screening was to detect whether the addition of different DNA damaging agents could lead to a more homogenous phosphorylation pattern and in addition to higher protein yield in expression. The focus of this approach goes beyond the mapping, it was also anticipated to find a damaging agent, which would induce the phosphorylation of a site, which is not detected in untreated cells. In this approach, based on the mapping further experiments would be carried out. The hypothesis is that upon DNA damage SLX4 will be recruited to the damage site and change its conformation to bind its interaction partners. Earlier publications noted that scSlx4 becomes phosphorylated in response to DNA damage at all cell-cycles stages [184]. In electrophoretic mobility shift experiments of scSlx4 from cells treated with different DNA damaging agents the most pronounced shift was caused by MMS (methyl methanesulfonate) [184]. After exposure of cells to MMS, yeast Slx4 is phosphorylated *in vivo* at Thr72, Ser289 and Ser329 [79]. Replication stress induced by MMS leads in yeast cells to phosphorylation of Slx4 by Mec1 (human ATR) and upon phosphorylation to interaction with Dpb11 (human TOPBP1) [57].

The tested DNA damaging agents induce different kind of DNA damage, which are processed differently in the cells and involve SLX4 in the repair process. Hydroxyurea (HU) is depleting the dNTP pool in the cell, leading to stalled replication forks and subsequent double-strand breaks [267]. The crosslinking agent mitomycin C (MMC) introduces covalent links within one strand or between two DNA strands and leads ultimately to the formation of sister chromatid exchanges [16, 112]. The application of the alkylating agent methyl methanesulfonate (MMS) blocks the progression of the replication forks [113]. The topoisomerase II inhibitor etoposide causes double-strand breaks due to blockage of the re-ligation in DNA unwinding [268, 269]. Any of the DNA damage induced by the different agents might be enough to trigger a certain phosphorylation pattern.

For the phosphosite analysis under DNA damage influence, hsSLX4¹³⁰⁰⁻¹⁸³⁴ was coexpressed with hsSLX1¹¹⁻²⁷⁵. After over night expression, the DNA damaging agents were added and incubated for the respective time indicated in Materials & Methods Table 4.25. Afterwards, the proteins were purified in a one-step purification with GFP-binder beads and subjected to phosphorylation site analysis, carried out by Dr. Nagarjuna Nagaraj (MPIB).

Overall, strong differences between the phosphorylated sites in SLX4 from untreated cells and cells treated with different DNA damaging agents could not be detected in the sense of switching phosphorylation on and off (Figure 2.11). In comparison to the untreated cells, most sites were still phosphorylated under the influence of the agents but with slight changes in intensity. Small differences in phosphorylation intensity was seen between untreated and treated cells. T1315 is always phosphorylated at high intensity under any condition. The same is true for S1434 (low intensity), S1450 (low intensity), S1453 (high intensity), S1617 (high intensity) and S1743 (high intensity). This indicates that the phosphorylation of these sites is not regulated by any tested DNA damage. Under the influence of MMS T1326 is phosphorylated, while phosphorylation of this residue is absent in the other conditions. Thr on position 1326 is followed by Pro, which

indicates a CDK1 phosphorylation site. Another striking difference is the absence of phosphorylation on S1716 in MMC and MMS, a ATM/ATR predicted phosphorylation site. Treatment with hydroxyurea shows the lowest differences compared to control cells. This is quite surprising as one would have expected that the treatment with the ICL-inducing agent hydroxyurea leads to a high phosphorylation of SLX4. ICLs are being repaired in the FA pathway and SLX4 is involved in subsequent steps of the ICL repair pathway. Published results showed furthermore that Slx4 from yeast cells treated with HU has a decreased electrophoretic mobility compared to untreated cells and thus a higher phosphorylation level [184].

In general many phosphorylation sites were still found and there was no condition tested, which showed extreme reduction or increase in phosphorylation (Figure 2.11). The heat map compares the relative appearance of each phosphorylation site towards a baseline (see Figure 2.11). The dark pink rectangles refer to an increased intensity of this phosphorylation site, whereas the dark green rectangles indicate a weak intensity. In general no big differences between the control sample and the treated samples could be detected.

With the detection of the phosphorylation sites the different peaks in ESI-TOF MS after purification could be explained. Some of the phosphorylation sites could be removed by phosphatase treatment, but especially when co-expressed with SLX1 many phosphates could not be removed. Due to the heterogeneity of the proteins and the degradation problems, crystallization with the human proteins was not further attempted. As another strategy the *Caenorhabditis elegans* proteins were tested for crystallization.

2.3 Purification of the *C. elegans* SM-Complex and Crystallization Attempts

Additionally to the human proteins, *Caenorhabditis elegans* homologs of SLX1, SLX4 and MUS81-EME1 were investigated. The *C. elegans* homologs are in general shorter than the human proteins, especially the SLX4 protein (see Figures 1.4 and 1.6 B). However, the SAP and the CCD of SLX4, the GIY-YIG nuclease and RING domain in SLX1 as well as the ERCC4 domains in MUS81-EME1 are conserved between the orthologs [113, 115]. The *C. elegans* SLX4 ortholog is involved in meiotic chromosome segregation and is named HIM-18, high incidence of males, characterized by an increased X chromosome loss [24]. Herein HIM-18 is referred to as ceSLX4. *C. elegans* is a model organism to study the DNA damage response as many orthologs of human DNA repair factors are found in nematodes with an reduced complexity of the system. Furthermore, *C. elegans* is a well known target to study germ cell development and meiosis [239] and lots of information about the four proteins is available [24, 241, 242].

2.3.1 Purification of the Full-Length *C. elegans* Proteins

The respective genes were cloned by PCR from a *C. elegans* cDNA library (a kind gift of Karsten Klage). Initial small scale expression tests of the *C. elegans* proteins were carried out and all full-length constructs of the four proteins showed expression (Figure 2.12). The large scale purification was carried out in a three-step process using IMAC, ion exchange chromatography and SEC. Figure 2.13 shows the gel filtration elution profiles and the respective SDS gels.

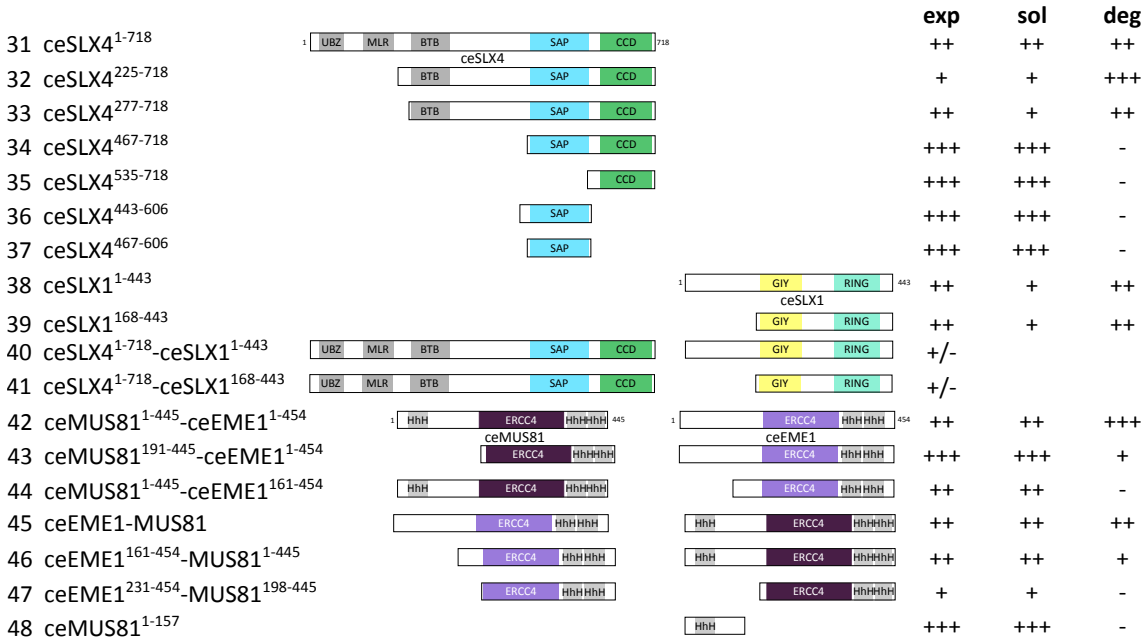


Figure 2.12: Schematic representation of ceSLX4, ceSLX1 and ceEME1-MUS81 constructs. The different constructs were designed based on limited proteolysis experiments and tested for expression (exp), solubility (sol) and degradation (deg) in *E. coli* with N-terminal His8-SUMO-tag. The success is indicated by + (positive) and - (negative). Most of the designed constructs were expressed very well and soluble. Few of them showed strong degradation.

The 81.1k Da full-length *C. elegans* SLX4 or HIM-18 (hereafter referred to as ceSLX4) could be solubly expressed in *E. coli* cells and was purified with Ni affinity, cation exchange chromatography with subsequent gel filtration (Figure 2.13 A). In comparison to the human SLX4 protein, the ceSLX4 protein could be purified to yields of 2 mg protein per liter of bacterial culture and showed only few degradation bands.

ceSLX1 contains a 168 amino acid N-terminal extension in front of the GIY-YIG nuclease domain, which is not present in most other SLX1 orthologs and was predicted to be disordered by the globular domain prediction tool GlobPlot [270]. In a bioinformatical search the 168 aa N-terminal extension was only found in SLX1 orthologs from other nematodes (e.g. *Caenorhabditis brenneri* Uniprot ID EGT50705) and some *Ascaris* species (e.g. *Ascaris suum* F1KVZ3). However, the *Ascaris* SLX1 orthologs were not present after a Blast search, but searching SLX1 for *Ascaris* species in the Uniprot database revealed the similarity to the nematode SLX1. The N-terminal extension is a special hallmark of SLX1 orthologs from worms and initially two constructs were designed comprising the full-length ceSLX1 (Figure 2.12 Construct 39) and a truncated version lacking the N-terminal extension (Figure 2.12 Construct 40). Both constructs could be solubly expressed and were purified by Ni affinity step and Heparin column. The shorter ceSLX1 construct showed aggregation after Heparin column, which could be seen from visual inspection by white precipitation, and showed degradation analyzed on SDS-PAGE. In the last step of purification using SEC (Superdex 75 10/300 GL, GE Healthcare) ceSLX1 full-length construct eluted at 8.3 ml corresponding to the void volume of the column (Figure 2.13 B). As the theoretical molecular weight of full-length ceSLX1 is 50.5 kDa, the volume of 8.3 ml indicates a state of soluble aggregates. As

the SLX1 protein forms a dimer in solution [156], it was already expected that SLX1 will elute in a lower elution volume than the monomer size. Next, cleavage activity was tested by mixing ceSLX1 and ceSLX4 in a 1:1 ratio, but activity could not be detected. To increase the stability of the ceSLX1 constructs, expression of ceSLX1 constructs was attempted in a bicistronic vector containing the N-terminally tagged interaction partner ceSLX4. Two different ceSLX1 constructs were cloned and tested for solubility (Figure 2.12 Constructs 41 and 42). After Ni Pulldown only ceSLX4 could be identified in the soluble fraction but not the ceSLX1 constructs (indicated by +/- in Figure 2.12). As another strategy to increase stability of the proteins, ceSLX1¹⁶⁸⁻⁴⁴³, ceSLX4 and ceMUS81-EME1 were co-purified. In the course of the purification, the four proteins showed even stronger degradation compared to the purification alone. Additionally, the four proteins did not show complex formation in size exclusion chromatography. Because of lacking enzymatic activity and long term protein stability issues, no further experiments with ceSLX1 were carried out. Additional reasons will be discussed in section 3.1.

Individual expression and purification of ceMUS81 and ceEME1 was only possible for ceMUS81, but not for ceEME1. The purified ceMUS81 alone did not show any cleavage activity towards tested substrates (Appendix Figure 4.8). Earlier it has been shown that for fission and baker's yeast the formation of a complex of MUS81 and EME1 is essential for cleavage of DNA substrate [135, 271]. The cleavage mechanism applied by the heterodimer seems to be conserved. Furthermore, the stability of the MUS81-EME1 complex is as suggested by their intertwined structure [214]. The inability of ceMUS81 to cleave DNA substrate is consistent with the before published results that MUS81 needs EME1 as co-factor for activity. Furthermore, in regard to the expression yield it has been shown earlier that recombinant MUS81-EME1 are poorly expressed individually, but co-expression gave good yields [142, 251]. Hence, they were expressed from a bicistronic vector in *E. coli* cells. The full-length proteins were purified and showed only few degradation bands and a mixture of SUMO-tagged EME1 as well as cleaved EME1 (Figure 2.13 C). Crystallization was attempted in presence and absence of DNA substrates, but did not yield any hits. Additionally, crystallization of the full-length constructs of ceSLX4 alone and ceMUS81-ceEME1 alone and bound to different DNA substrates was attempted but did not give rise to any hits.

2.3.2 Limited Proteolysis of ceSLX4 and ceMUS81-EME1 and Crystallization Trials

As the full-length proteins of ceSLX4 and ceEME1-MUS81 showed degradation, limited proteolysis was used to search for a more stable construct (Figure 2.14). First, the proteases Trypsin, Chymotrypsin, Elastase, Subtilisin and GluC with broad substrate specificity were applied to the ceSLX4 construct at three different concentrations (0.1, 0.01 and 0.001 mg/ml). The first cleavage test is performed in order to find the most useful protease and the optimal substrate:protease ratio. Based on the preliminary result a time course with Elastase at 0.02 mg/ml and Chymotrypsin at 0.01 mg/ml was carried out (Figure 2.14 A). The limited proteolysis experiment with SLX4 was performed in the absence of DNA, as the protein is suggested to work as a protein-protein interaction hub but does not bind DNA on its own [141, 143, 156]. Based on the results of limited proteolysis considering structural domain and disorder prediction several N-terminally truncated constructs were designed (Figure 2.13). A detailed description of the constructs and the rationale

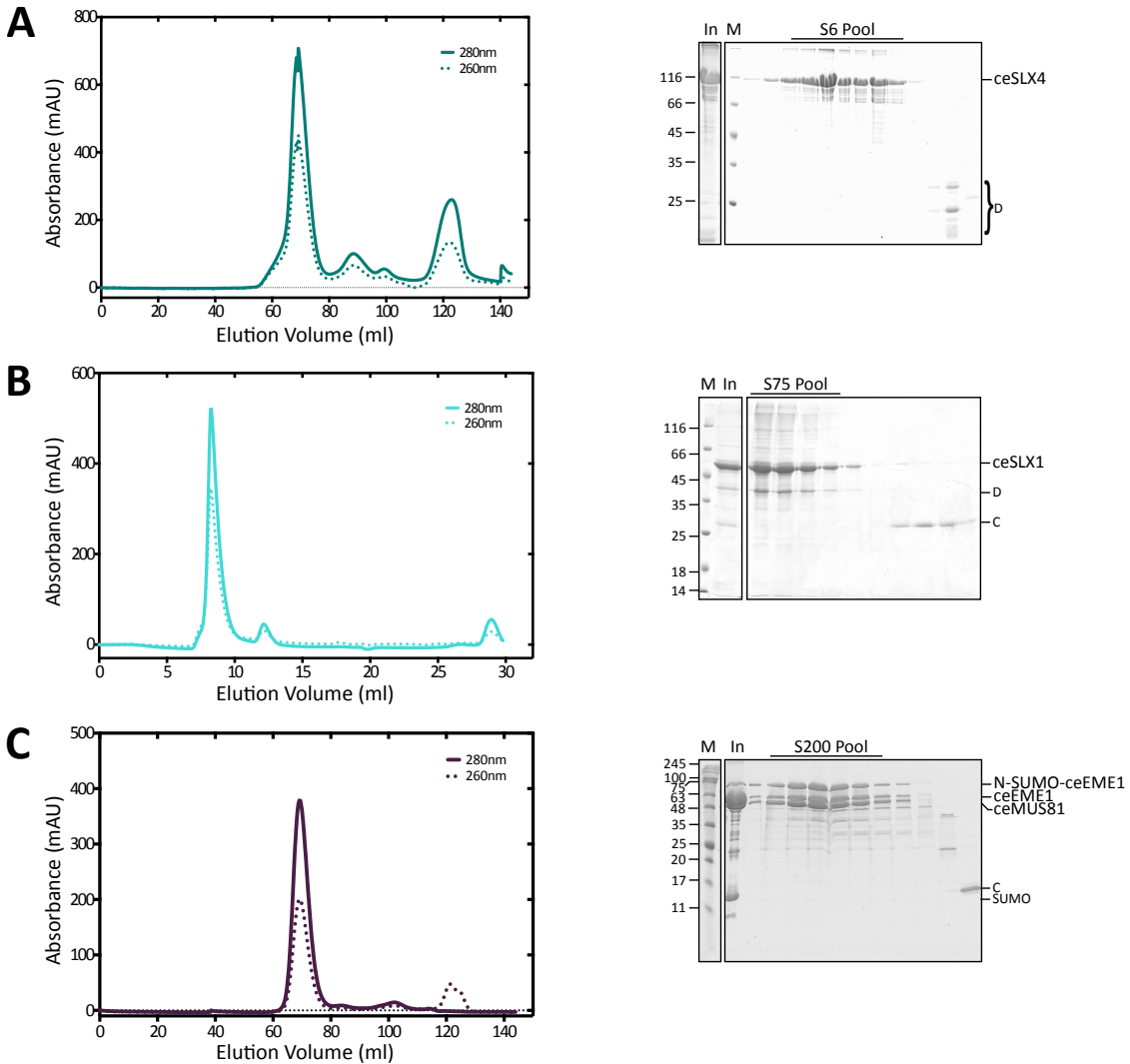


Figure 2.13: SEC analysis of full-length ceSLX4, ceSLX1 and ceEME1-MUS81 proteins. (A) Size exclusion chromatogram of full-length ceSLX4 using a Superose 6 16/600 column showing an elution profile with one major peak at 39.3 ml, containing the full-length ceSLX4 protein and smaller peak at 88.6 ml & 122.6 ml, corresponding to degradation products. The SDS-PAGE shows the fractions of the main peak and the other peaks. (B) SEC analysis of ceSLX1 using a Superdex75 10/300 column showing the elution profile with one major peak at 8.3 ml and two smaller peaks at 12.2 ml and 28.9 ml. The SDS-PAGE shows the fractions of the first two peaks. (C) Gel filtration profile of bicistronically expressed ceEME1-MUS81 using a Superdex 200 16/600 size exclusion column. The SDS-PAGE shows the fractions of the main peak (69.4 ml) and the smaller peaks (101.7 ml and 114.1 ml). All chromatograms are visualized by absorption at 280 nm and 260 nm. Pooled fractions are marked by a line. (M) Protein marker in kDa. (In) Input protein sample injected on gel-filtration column. (D) Degradation products (C) Contaminant band (SUMO) cleaved off SUMO tag.

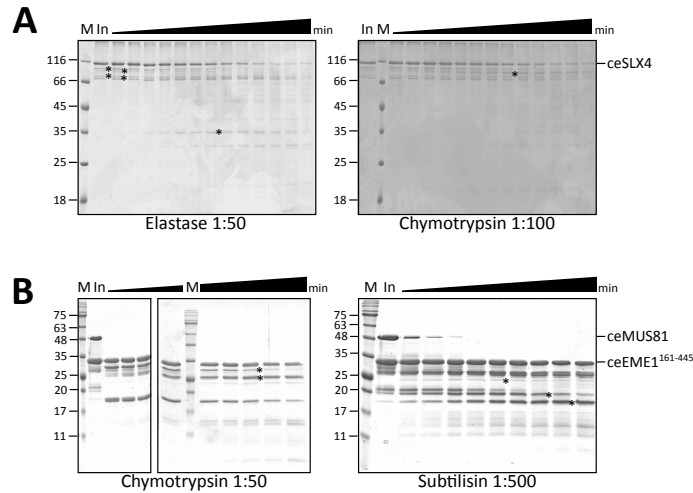


Figure 2.14: Limited proteolysis experiments for ceSLX4 and ceEME1¹⁶¹⁻⁴⁴⁵-ceMUS81. (A) Time course of limited proteolysis for ceSLX4 with Elastase at 0.02 mg/ml and Chymotrypsin at 0.01mg/ml concentration. (B) Time course of limited proteolysis for ceEME1¹⁶¹⁻⁴⁴⁵-ceMUS81 in complex with nicked Holliday Junction substrate 4w4x16n3 with Chymotrypsin at 0.02 mg/ml and in complex with nicked HJ substrate 4w4x15n3 with Subtilisin at 0.002 mg/ml concentration. Samples were taken after 5, 10, 15, 30, 45, 60, 90, 120 and 180 min. (In) Input protein sample used for proteolytic analysis. (M) Protein marker in kDa. Bands marked with an asterisk on the right indicate stable fragments and were sent for peptide fingerprinting.

will be given in the next paragraphs.

Similarly, the construct ceEME1¹⁶¹⁻⁴⁵⁴-ceMUS81¹⁻⁴⁴⁵, which was designed according to alignment with truncated hsEME1 from Gwon et al [214], was subjected to limited proteolysis. The proteolytic experiment was performed in presence and absence of nicked Holliday Junction DNA substrate to gain structural information about the mode how MUS81-EME1 complex binds the DNA substrate. In the first proteolysis experiments with the five different proteases at three concentrations, slight differences in the bands appearing could be seen. The truncated EME1 construct stayed stable in every condition, except for subtilisin at the highest concentration. The most differences between substrate-bound and free proteins was seen for the band pattern at low molecular weight (lower than 17 kDa). However, as the aim was laid on larger constructs, analysis of limited proteolysis experiments was more focused on the large discrete bands of the DNA-bound proteins. Next, a proteolytic time course was carried out in presence of nicked Holliday Junction with Chymotrypsin at 0.02 mg/ml and with Subtilisin at 0.002 mg/ml (Figure 2.14). Again, the EME1 construct stayed stable during the course of the experiment. The marked bands in figure 2.14 were sent to identification by protein fingerprinting. Analysis of the peptide fragments from each band revealed that only MUS81 was cleaved by the tested proteases. According to the results of the limited proteolysis experiments and *in silico* structural prediction, different constructs of MUS81-EME1 were designed. In addition, the position of the two proteins was also switched in the expression vector (Figure 2.12).

The new constructs containing different versions of truncated and full-length MUS81 and EME1 in the bicistronic vector. Mainly, they were truncated from the N-terminus, where a large unstructured region is predicted between the HhH domain and the ERCC4 domain. The constructs were expressed very well and were purified by Ni-affinity chromatography, weak cation exchanger (Hep-

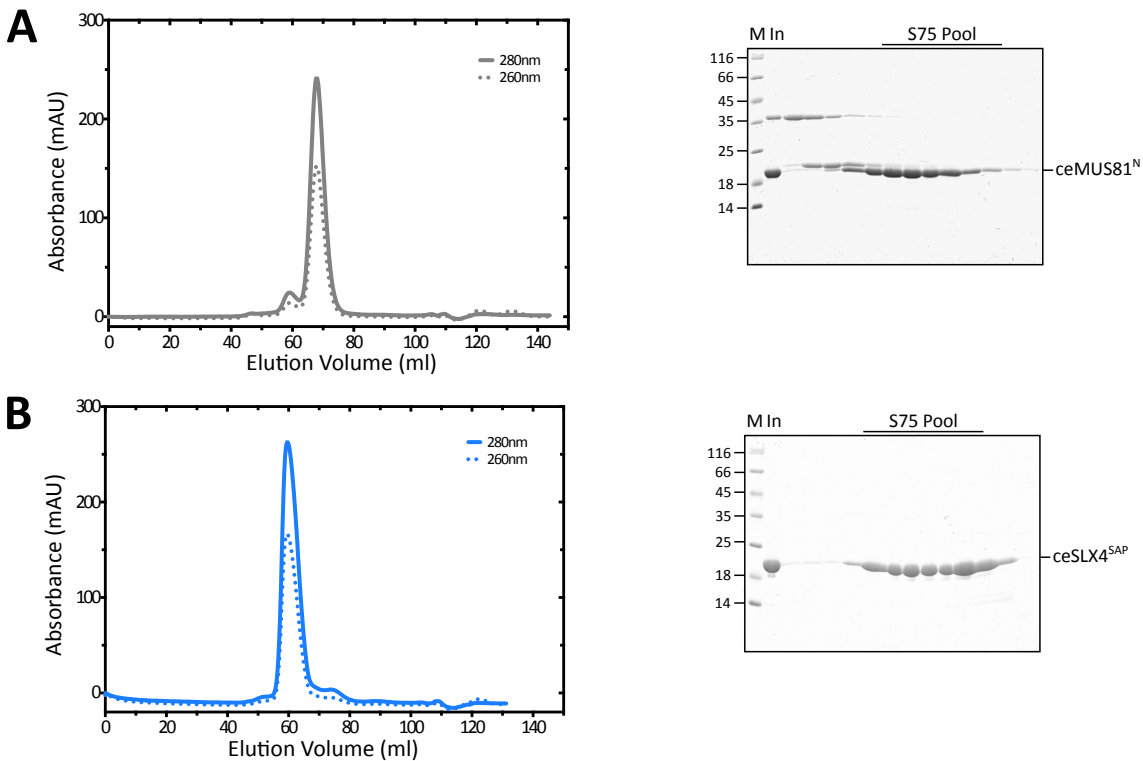


Figure 2.15: SEC analysis of ceMUS81¹⁻¹⁵⁷ and ceSLX4⁴⁴³⁻⁶⁰⁶. (A) Size exclusion chromatogram of the N-terminal construct of ceMUS81 (1 - 157) using Superdex 75 16/600 column. The elution profile shows a main peak and a small pre-peak. The main peak elutes at 67.8 ml and contains ceMUS81¹⁻¹⁵⁷. The small pre-peak elutes at 59 ml and contains a contaminant protein. The SDS-PAGE shows the fractions of the pre-peak and the main peak. (B) SEC analysis of ceSLX4⁴⁴³⁻⁶⁰⁶ using a Superdex 75 16/600 column. The protein elutes as symmetric main peak at 59.6 ml with a smaller peak afterwards at 74.7 ml. The SDS-PAGE shows the fractions of the main peak containing ceSLX4⁴⁴³⁻⁶⁰⁶ and of the peak afterwards containing degradation products. Chromatograms are visualized by absorption at 280nm (straight line) and 260nm (dashed line). Pooled fractions are marked with a line. (M) Protein marker in kDa, (In) Protein sample injected on gel filtration column.

arin column) and size exclusion chromatography. For crystallization, active site mutations of MUS81 were used to form complexes with DNA substrates. They were submitted to crystallization trials in presence and absence of DNA substrates (HJ, nHJ), but did not give rise to any hits.

In literature it was reported that MUS81 N-terminus can bind DNA [143, 219]. Limited proteolysis gave also rise to a construct containing only the N-terminus of ceMUS81 (Figure 2.12 Construct 49). Purification was carried out with NiNTA chromatography, Heparin column and size exclusion chromatography (Figure 2.15 A). The N-terminal construct of ceMUS81 could be purified to 20 mg of protein per liter of bacterial expression and did not show degradation after SDS-PAGE analysis. ceMUS81¹⁻¹⁵⁷, later on referred to as ceMUS81^N, was subjected to crystallization with and without DNA substrate, but did not give rise to any crystals.

After the new design of the ceSLX4 constructs based on limited proteolysis results and domain prediction, the new constructs were tested for expression, purification and degradation. Two constructs of ceSLX4, ceSLX4²²⁵⁻⁷¹⁸ and ceSLX4²⁷⁷⁻⁷¹⁸ (Figure 2.12 Constructs 33, 34), lacked

the first two domains and showed some degradation after purification as the unstructured part between the BTB domain and the SAP domain was still a target for proteases during purification. Then constructs 35 to 38, containing both or one of the C-terminal domains, were tested for expression. Construct 35, ceSLX4⁴⁶⁷⁻⁷¹⁸, expressed very well, did not show degradation bands in SDS-PAGE and was submitted for crystallization. It gave rise to initial microcrystals and phase separation at 20 mg/ml in 50 mM Tris pH 8.0, 20 mM Sodium oxalate and 2.2 M Sodium malonate (Figure 2.17 A). The micro crystals were harvested, crushed and used for matrix seeding. After two days a large crystal appeared in 50 mM Tris pH 8.0, 0.2 M Calcium acetate and 26% PEG 8000 (Figure 2.17 B). The crystal was measured with synchrotron radiation and did not show any diffraction. An analysis of the crystal by SDS-PAGE failed, as it completely dissolved during preparation of the sample. As the protein was highly soluble, it could be highly concentrated, which lead to high viscosity of the sample. This made setting up crystallization trials challenging as the protein solution could not be pipetted precisely.

Therefore, the strategy to induce crystal growth was changed to search for a different formulation buffer, which increases protein stability. In order to do so, nanoDSF (nano differential scanning fluorimetry) was used to screen different buffers and find the buffer exhibiting the highest melting temperature. nanoDSF allows a sensitive and practical way to monitor the thermally induced unfolding of protein samples [272, 273]. Here, internal Trp fluorescence was used to monitor the increase in fluorescence upon unfolding. The transition occurs at the melting temperature (T_m) and depends on the thermostability of the protein under the tested conditions. The T_m can shift under different buffer conditions, like in the case of hsSLX4⁴⁴³⁻⁷¹⁸ (Figure 2.16 A). The buffer containing Bis-Tris-Propane at pH 7.5 leads to a higher melting temperature than the Hepes buffer at pH 8.0. The melting curve thus shows a condition, which is beneficial for protein stability. Here, different salt concentrations and pH values have been tested (Figure 2.16 B). Low salt concentration (50 - 100 mM) decreased the thermal stability a lot and a pH of 8.0 showed lowest melting temperature and therefore lowest protein stability. In addition to sodium chloride, potassium chloride and potassium acetate were tested for stability studies, but did give rise to higher melting temperatures (data not shown). The highest thermal stability with highest melting temperature was seen at pH 7.5 with Bis-Tris-Propane buffer and 200 mM NaCl (Figure 2.16). This buffer condition was used in subsequent crystallization experiments for the formulation of the protein construct. Although the protein construct was tested extensively, the shown crystal (Figure 2.17 B) could neither be reproduced nor could other crystal forms be obtained.

The construct ceSLX4⁵³⁵⁻⁷¹⁸ containing only the CCD domain, which interacts with SLX1, expressed to large amounts (13 mg protein per liter of bacterial cell culture). It was tested in crystallization trials, but did not give any hits or nucleation starting points which could have been used for further optimization.

2.3.3 Crystallization Trials of ceSLX4^{SAP}

The SAP domain of ceSLX4 is the interaction site for MUS81 [201, 202] and other proteins, containing SAP domains, were shown to bind DNA [127, 274, 275]. DNA binding of the ceSLX4 SAP domain will be discussed in section 2.5.3. Two different constructs of the ceSLX4 SAP domain were initially designed and could be purified in large amounts (> 30 mg protein per liter of bacterial

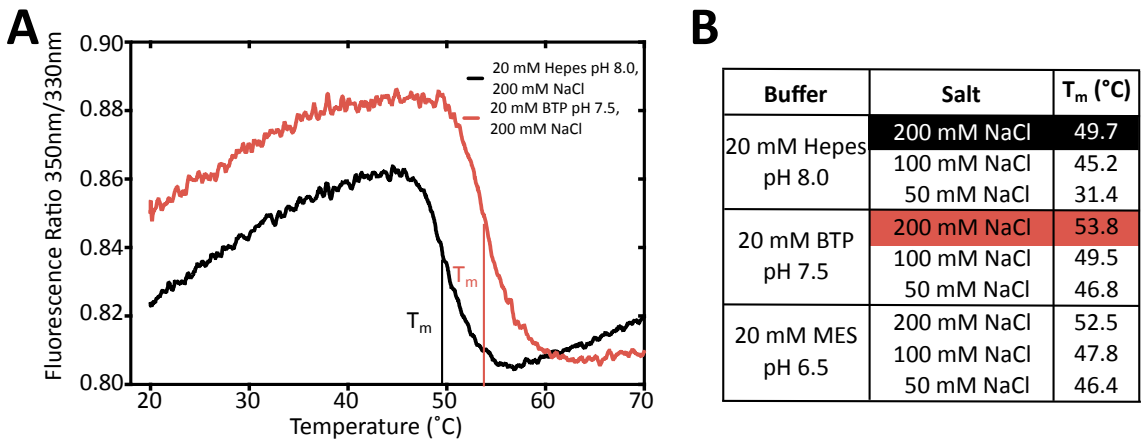


Figure 2.16: Stability test of ceSLX4⁴⁴³⁻⁷¹⁸. (A) The melting curves for ceSLX4⁴⁴³⁻⁷¹⁸ were measured by nano differential scanning fluorimetry with a temperature range from 20° to 90°. For clarity the x-axis was cut at 70°. (B) The table summarizes different conditions tested for optimal stability of the protein. Shown are three buffer systems at different pH values tried each with three different NaCl concentrations.

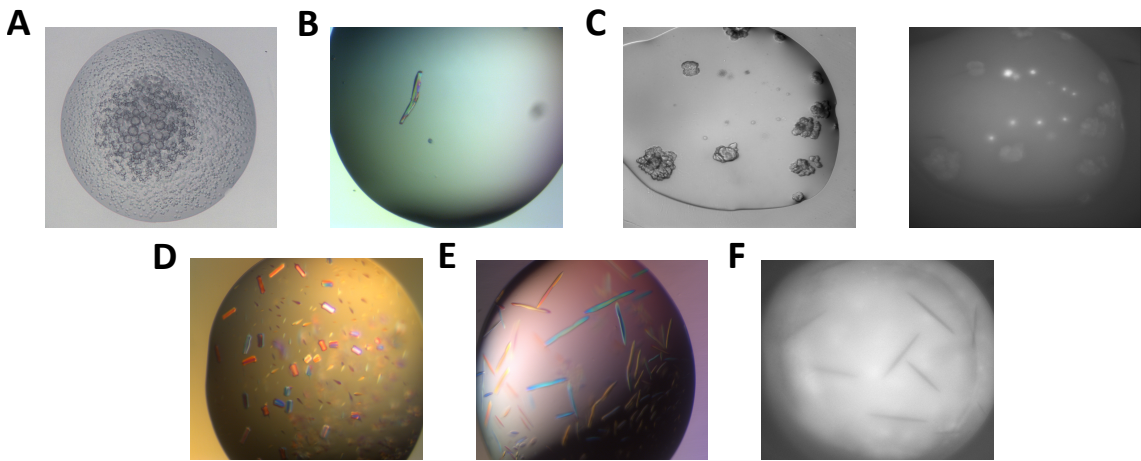


Figure 2.17: Crystallization of different ceSLX4 constructs. (A + B) Crystallization of ceSLX4⁴⁶⁷⁻⁷¹⁸. Initial microcrystals appeared after 2 days in 50 mM Tris pH 8.0, 20 mM Sodium oxalate and 2.2 M Sodium malonate (A). One crystal (B) appeared after 2 days in 50 mM Tris pH 8.0, 0.2 M Calcium acetate, 26% (w/v) PEG 8000 after matrix seeding with microcrystals shown in (A). (C) Crystallization of ceSLX4⁴⁴³⁻⁶⁰⁶ at 34.7 mg/ml in 0.1 M Sodium citrate pH 5.5, 20% (w/v) PEG 3000. The left picture shows amorphous crystals, which appeared after 2 days, and spherulites, which appeared after 13 days. The UV fluorescence picture of the same drop (right) shows strong fluorescence for the spherulites and weak fluorescence for the amorphous crystals. (D) Crystallization of birefringent ceSLX4⁴⁶⁷⁻⁶⁰⁶ in complex with ceMUS81^{K270M}-ceEME1¹⁶¹⁻⁴⁵⁴ at in 20% (w/v) PEG 3350, 0.2 M Magnesium chloride. (E) + (F) Crystallization of ceSLX4⁴⁶⁷⁻⁶⁰⁶-ceMUS81^{K270M}-ceEME1¹⁶¹⁻⁴⁵⁴-nHJ substrate (4w4x16n1) in 18% (w/v) PEG 3350 and 0.2 M Calcium acetate (E) Crystals viewed under polarized white-light illumination showed birefringence. (F) Crystals viewed under UV light showed absorption of the UV light.

culture) and without any degradation (Figure 2.15 B). These constructs posed to be also very stable and highly soluble. They could be concentrated easily up to 50 mg/ml. Crystallization trials of the protein alone (Figure 2.17 C) and in complex with DNA were attempted. The crystallization of the protein alone lead to the formation of amorphous crystals and spherulites (Figure 2.17 C left). They were tested for protein content with a UV microscope for crystal inspection (Figure 2.17 C right). The spherulites appeared bright white under UV light like protein crystals would appear. The amorphous crystals showed weak UV fluorescence. In the first crystallization trials amorphous crystals formed latest after 2 days. The fast growth of amorphous crystals indicates that nucleation is induced very fast and then proteins enter the phase of supersaturation. Frequently, drops containing either one of the SAP constructs showed skin formation. In order to slow down the nucleation and to prevent formation of amorphous crystals, a vapor diffusion rate control experiment with a mixture of different oils was attempted. In the sitting drop vapor diffusion setup the reservoir was covered with a paraffin oil: silicon oil ratio of 80:20, which allowed only limited drop evaporation as high paraffin oil percentages allow a slower rate of drop evaporation. The diffusion between drop and reservoir is slowed down, thus decreasing crystal growth. A crystallization plate was set up, where the reservoir was covered with 80% paraffin oil and 20% silicone oil, which resulted in a strong decrease of nucleation and only precipitation was visible but no microcrystals. In another trial to slow down growth of amorphous crystals, 5% glycerol was added to the protein solution, which increased the solubility of the protein and lead to crystallization drops with precipitation and phase separation. The amorphous crystals and the spherulites depicted in Figure 2.17 C were used for microseed matrix seeding. However, only precipitation and amorphous crystals were detected in subsequent crystallization trials with the applied seeding technique. Additionally, nanoDSF measurements were carried out for the construct ceSLX4⁴⁴³⁻⁶⁰⁶ as well. NanoDSF measurements rely on the intrinsic fluorescence mainly of tryptophane and tyrosine residues at 330 and 350 nm. The SAP construct contains only one tyrosine residue and fluorescence was therefore too weak for recording of thermal unfolding curves. Then buffer conditions with the low salt concentrations and high pH were adopted from the screening with the ceSLX4⁴⁶⁷⁻⁷¹⁸ construct. Crystallization of the ceSLX4^{SAP} constructs in complex with DNA substrates lead to very thin feather-like, long and fragile crystals. They were attached strongly to the bottom of the drop and detaching lead to fragmentation of the crystal. Mostly, when analyzing the crystals under the UV-microscope, they absorbed the UV fluorescence, which indicates DNA crystals (similar to Figure 2.17 F). Seeding, addition of various DNA substrates and different protein:DNA ratios did not gave any hits. In summary crystallization of ceSLX4^{SAP} alone as well as with DNA did not lead to diffraction quality crystals.

2.3.4 Crystallization Trials of ceSLX4^{SAP} in Complex with Interaction Partners

As the SAP domain and the MUS81-EME1 constructs alone or with DNA did not lead to any crystallization hits, another crystallization strategy was applied, which combines ceSLX4^{SAP} with ceMUS81^N and DNA substrates. The rationale behind was to stabilize unstructured regions in the proteins by their substrate and their interaction partner. This setup did not yield any crystals,

even though various DNA substrates were tested.

In another approach the before mentioned strategy was extended by crystallization trials of ceSLX4^{SAP}, full-length MUS81 (wild type and active site mutants) and different EME1 truncations with various DNA substrates. Using full-length MUS81 was important here, as the N-terminus of MUS81 interacts with SLX4^{SAP} and the C-terminal HhH domains are the interaction site with EME1 (Figures 1.4 B, 1.6 B)[135, 201]. In this way, a complex of MUS81 interacting with EME1, the DNA substrate in the active site of MUS81^{ERCC4} and SLX4^{SAP} stabilizing the interaction would be achieved. Three dimensional crystals were obtained (Figure 2.17 D) and conditions around 20% PEG 3350 and 0.2 M Magnesium chloride were tested with different cations in addition. In several conditions containing PEG 3350 and calcium acetate long feather-like crystals appeared (like in Figure 2.17 E). Testing them for protein content under the UV microscope, showed the absorbance of the light by black streaks. This indicated a crystal content of DNA only. Crystals containing glycerol as cryo-protectant were analyzed at the synchrotron and they did not diffract. To reduce skin formation and nucleation, crystallization drops with various DNA substrates and precipitant conditions were setup in a microbatch under oil experiment with 50% paraffin oil and 50% silicone oil. However, under the tested conditions microcrystals for further optimization or larger crystals could not be achieved. Taken together, different approaches have been tried in order to get diffraction quality crystals, but under the tested conditions, no protein-DNA substrate crystal could be achieved. Therefore, a different approach was chosen to obtain structural information.

2.4 Structural Analysis of ceSLX4^{SAP} and ceMUS81^N by NMR

After extensive crystallization trials of both the human and *C. elegans* proteins with and without different DNA substrates, so far none of the tested conditions and crystallization methods gave rise to crystals with diffraction quality. The used constructs for ceSLX4^{SAP} and MUS81^N were quite stable and are highly soluble. Furthermore, with a size of 18.0 and 18.5 kDa, respectively, the constructs are in a size range that is applicable for nuclear magnetic resonance (NMR). NMR measurements and assignments were carried out by our collaborator Dr. André Mourão at the Helmholtz Center Munich (Protein Expression and Purification Facility).

2.4.1 HSQC analysis of ceSLX4^{SAP} and ceMUS81^N

In order to analyze the proteins for their suitability for NMR spectroscopy, a two-dimensional ¹H¹⁵N-HSQC (heteronuclear single quantum coherence) experiment was carried out. In this experimental setup the ¹H and ¹⁵N shifts are measured as a correlation between the nitrogen atom of a HN group and the directly bound proton. Each signal in the spectrum represents a proton and its bound nitrogen atom. Thus, each amino acid is described as one peak in the spectrum as every amino acid contains one HN in the peptide bound, except for proline. The side chains of asparagine and glutamine appear as doublets. Initial HSQC spectroscopy was carried out on the constructs ceSLX4⁴⁴³⁻⁶⁰⁶ and ceMUS81¹⁻¹⁵⁷. The proteins were expressed in M9 minimal medium containing 15N - ammonium sulfate as the sole nitrogen source and purified in a three

Table 2.2: Protein constructs of ceSLX4^{SAP} and ceMUS81^N for NMR experiments. More truncated SLX4^{SAP} and ceMUS81^N constructs were designed and tested for their expression (exp), solubility (sol), degradation (deg) and DNA binding. The success is indicated by + (positive) and - (negative). DNA binding is indicated by + (weak binding) to +++ (binding comparable to reference construct) and - (no binding). The reference constructs are for ceSLX4 443-606 and for ceMUS81 1-157. n.d. not determined

Construct No.	Protein	Construct	exp	sol	deg	DNA binding
48	ceMUS81	1 - 157	+++	+++	-	+++
49		1 - 140	+++	+++	++	
50		1 - 89	-			
51		1 - 80	-			
52		10 - 89	-			
53		17 - 140	++	++	+++	n.d.
54		17 - 80	++	++	+++	n.d.
36		ceSLX4	443 - 606	+++	+++	-
55	443 - 588		++	+++	+	n. d.
56	443 - 515		++	+++	+	-
57	467 - 560		+++	+++	-	+++
58	467 - 530		+++	+++	+	-
59	470 - 606		+++	+++	+++	+++
60	470 - 515		++	+++	-	-
61	488 - 606		++	+++	+++	n.d.
62	488 - 588		+++	+++	++	+++
63	535 - 606		++	++	+	-

step purification approach. The ¹⁵N labeled protein yield from 1 liter expression was 3.4 mg for ceSLX4⁴⁴³⁻⁶⁰⁶ and 3.6 mg for ceMUS81¹⁻¹⁵⁷. The proteins were measured at a concentration of 110 μM supplemented with 10% D₂O.

The recorded ¹H-¹⁵N HSQC spectra for ceSLX4⁴⁴³⁻⁶⁰⁶ (Figure 2.18 A, blue-green) and ceMUS81¹⁻¹⁵⁷ (Figure 2.18 B, blue-green) showed good spectra with round defined peaks for both proteins. However with a size of about 18 kDa, they are still too large for simple NMR measurement as many peaks overlap and thus can not be resolved for structural assignments. In addition, NMR chemical shift perturbation experiments of ceSLX4⁴⁴³⁻⁶⁰⁶ (Figure 2.18 A, red-orange) and ceMUS81¹⁻¹⁵⁷ (Figure 2.18 B, red-orange) were carried out with the respective interaction partner. Usually in chemical shift titration, the peaks from residues involved in the interaction with the binding partner move their position in the spectrum upon change of the chemical environment. Here, bound resonances are unobservable because of line broadening as a complex of 36 kDa has been formed and can not be detected within the relaxation time. Upon comparison with the spectrum without an interaction partner, it is visible that only a few resonances were not detected (Figure 2.18 black boxes). Most of the peaks in the spectrum of ¹H-¹⁵N-ceSLX4⁴⁴³⁻⁶⁰⁶ in presence of ceMUS81¹⁻¹⁵⁷ (darkred-orange) are still detected, indicating that their environment does not change. Thus, most amino acids in the used protein constructs are not involved in the interaction. The same observation was made for ¹H-¹⁵N-ceMUS81¹⁻¹⁵⁷. Taken together, the constructs of ceSLX4^{SAP} as well as the N-terminal domain of ceMUS81 had to be shortened in order to obtain a simpler spectrum and to decrease line broadening. Based on structural predictions further truncated versions were designed, expressed, purified and tested for DNA binding (Table 2.2).

The tested construct ceMUS81¹⁻¹⁵⁷ was designed according to limited proteolysis results. From the same limited proteolysis experiment, another N-terminal construct 1-105 was designed, but did not express. Therefore, other truncations were designed according to more refined structural

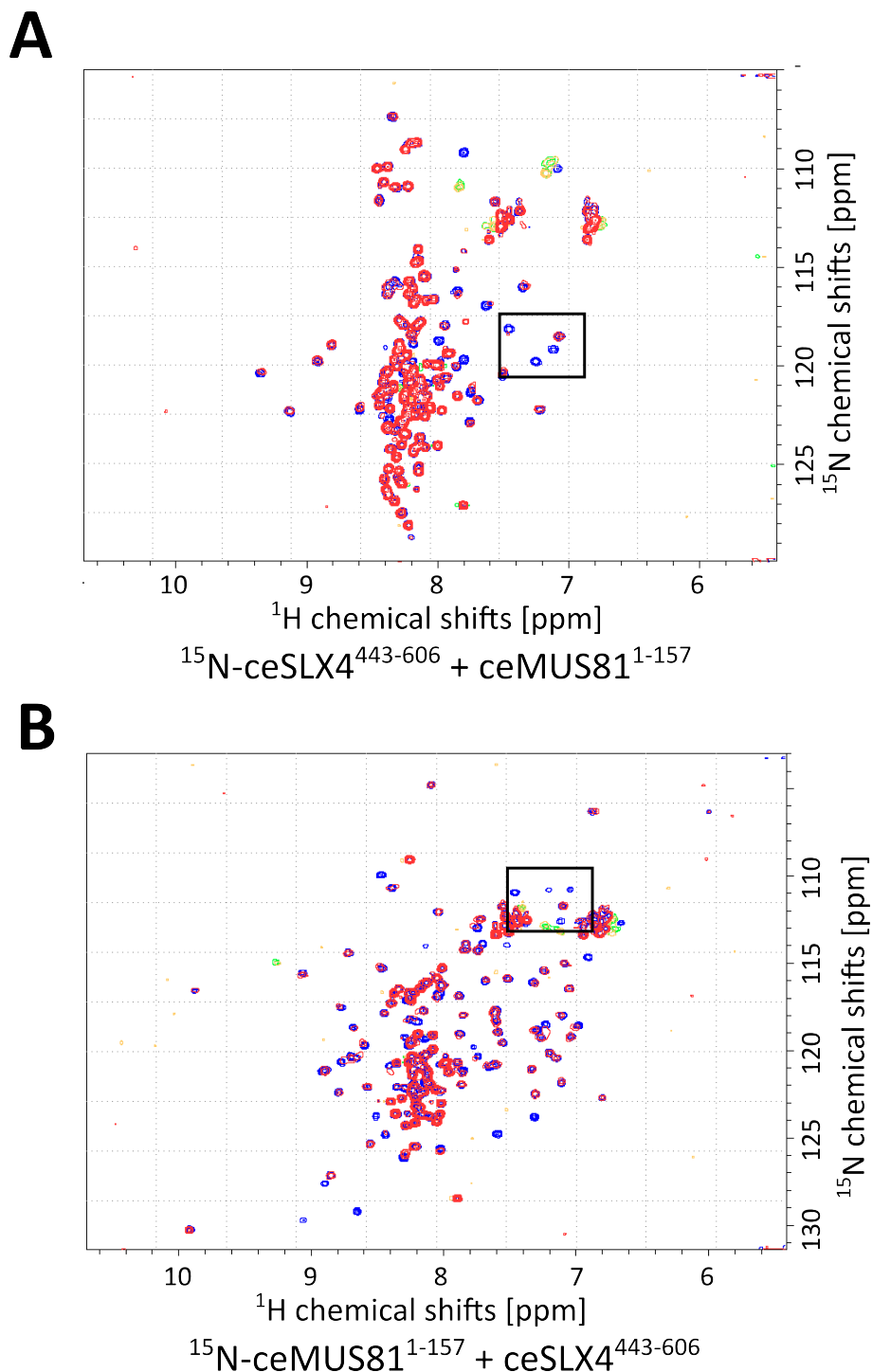


Figure 2.18: ^1H - ^{15}N HSQC NMR spectra of ceSLX4^{SAP} and ceMUS81^N (A) A HSQC spectrum of ^{15}N labeled ceSLX4 construct 443-606 was recorded alone (blue/green) and with a 1:1 ratio of the interaction partner ceMUS81 1-157 (red/orange). (B) A HSQC spectrum of ^{15}N labeled ceMUS81 construct 1-157 was recorded alone (blue/green) and with a 1:1 ratio of the interaction partner ceSLX4 443-606 (red/orange). The spectra were recorded on a 110 μM protein sample in 10% D_2O . The peaks in the black boxes are the peaks of single proteins and where not detected in complex with the interaction partner.

prediction and taking the chemical shift experiments into account. Many of the smaller constructs of ceMUS81 could either not be expressed or showed strong degradation. Construct ceMUS81¹⁻¹⁴⁰ was expressed and purified comparable to the reference ceMUS81¹⁻¹⁵⁷. A ¹H-¹⁵N HSQC spectrum was recorded for ceMUS81¹⁻¹⁴⁰ but this construct was still slightly too large and was not stable during the measurement as it showed precipitation. In summary a stable and shorter ceMUS81 N-terminal construct could not be found.

The further truncated ceSLX4^{SAP} constructs were expressed in comparable amounts to the reference construct ceSLX4⁴⁴³⁻⁶⁰⁶ and except for few constructs they were stable towards degradation during purification. Some constructs, e.g. ceSLX4⁴⁸⁸⁻⁵⁸⁸, showed degradation in ESI-TOF mass spectrometry and SDS-PAGE after one freeze-thaw cycle. The stable constructs allowed further recordings of nuclear magnetic resonance spectra.

2.4.2 NMR-Model of ceSLX4^{SAP}

The small size of 10.4 kDa and the stability of the construct ceSLX4⁴⁶⁷⁻⁵⁶⁰ allowed to record a high quality spectrum of the ceSLX4 SAP domain. The construct ceSLX4⁴⁶⁷⁻⁵⁶⁰ was suitable for further NMR experiments due to its stability and its reduction in size (94 aa compared to 164 aa ceSLX4⁴⁴³⁻⁶⁰⁶). The construct was expressed in M9 minimal medium supplemented with ¹⁵N-ammonium sulfate as sole nitrogen source and ¹³C-glucose as carbon source. ¹³C-¹⁵N double labeled ceSLX4⁴⁶⁷⁻⁵⁶⁰ was purified by Ni TED column, cation exchange chromatography and SEC.

From the construct ceSLX4⁴⁶⁷⁻⁵⁶⁰, hereafter referred to as ceSLX4^{SAP}, a model based on backbone assignments could be obtained. The last 37 residues were quite flexible and could not be resolved by the NMR structural model. Thus the model comprises residues 468-523 of ceSLX4 in an arrangement of two short α -helices and one longer α -helix (Figure 2.19 A, B). The three α -helices encompass residues M473 - S478 (helix 1), E481 - I491 (helix 2) and K499 - L512 (helix 3). Figure 2.19 A showed the ten best NMR models of ceSLX4^{SAP}. Especially the last ten C-terminal residues (514 - 523) showed a very high flexibility in the model. The first five amino acids are flexible in a loop, followed by a short α -helix 1. This helix is connected via two amino acids to a second short helix, which form an angle of 126° to each other. The second helix is connected via a seven residue loop to the third and longest helix in the structure, both forming an angle of 62°. Helix 1 and 3 are parallel to each other, but in opposite directions. In general ceSLX4^{SAP} forms a small helical bundle reminiscent of the shape of a seven. With the two α -helices separated by a turn the overall shape of ceSLX4^{SAP} resembles the common fold of SAP domains [168].

In general, the SAP (SAF-A/B, Acinus, PIAS) domain is found in proteins involved in chromosomal organization, like the SAF-A/B proteins [274], or in DNA repair (Ku70, STAT inhibitor PIAS) [168]. Usually, it is located at the N-terminus of proteins and contains strongly conserved hydrophobic, polar and bulky amino acids, which are separated by an invariant glycine [168]. The invariant glycine of ceSLX4^{SAP} (G492) is located in the loop region connecting α -helix 2 and α -helix 3 (marked by an orange sphere in Figure 2.19 B and E). ceSLX4^{SAP} also contains a hydrophobic core with the following residues: M473, F476, M479, L484, M488, I491, M493, M501, L505, Y509 and I510 (indicated in pale-orange in Figure 2.19 C and E). The extensive hydrophobic core is formed with nearly all conserved hydrophobic residues. The conserved hydrophobic residues are likely important for the structural integrity of the protein. Furthermore, the SAP domains of

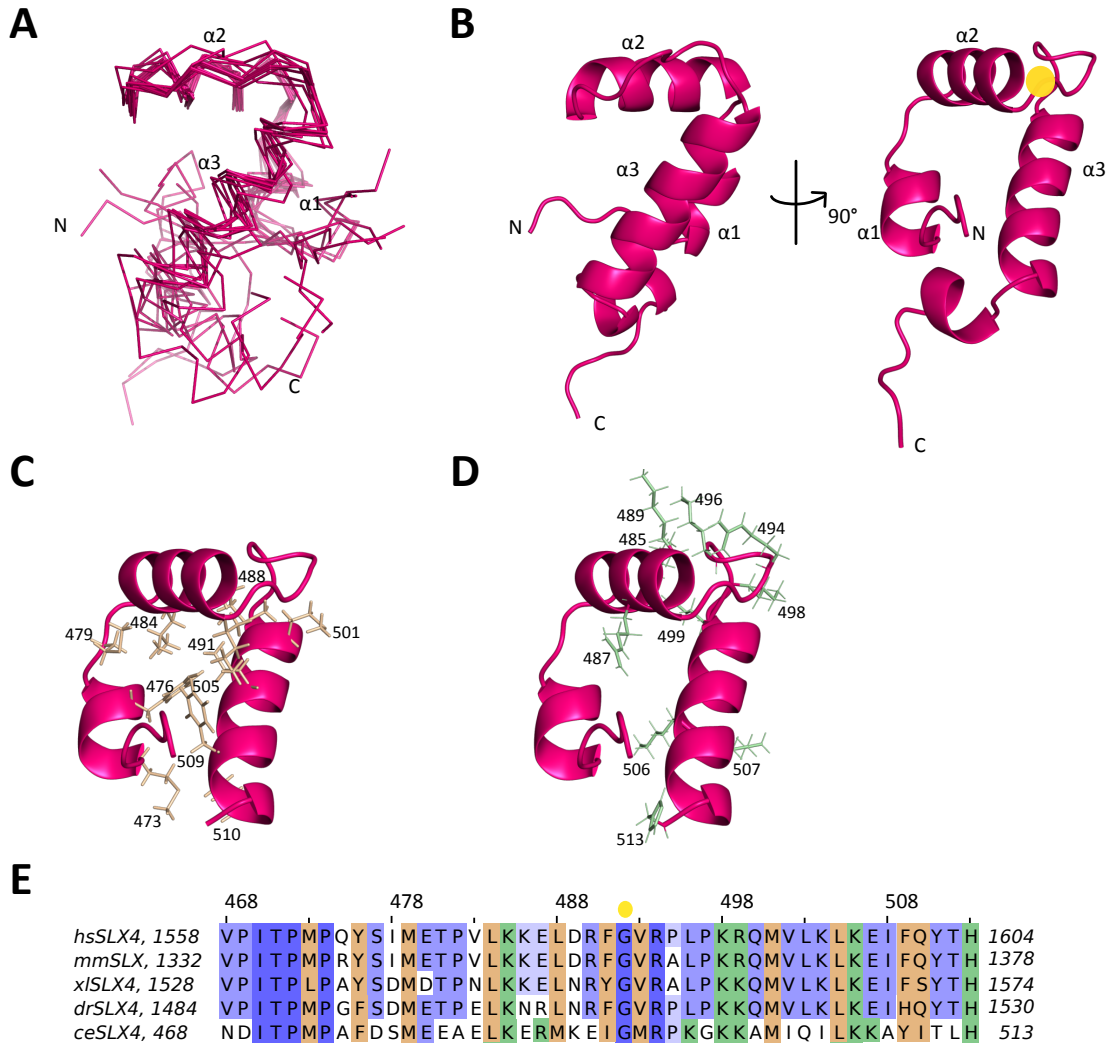


Figure 2.19: Structure of ceSLX4^{SAP} (A) Ensemble of the ten final NMR structures of ceSLX4⁴⁶⁸⁻⁵²³. The last ten residues show a high flexibility. (B) Cartoon representation of the the lowest energy model of (A). Left: shape of a seven, Right: 90° rotation around y-axis. The orange sphere marks the invariant glycine located in the loop region, a hallmark of SAP domains. (C) ceSLX4^{SAP} depicted with the hydrophobic core of the structured region (residues 468-513). (D) ceSLX4^{SAP} with the positively charged amino acids in the loop area. (E) Sequence alignment of the SAP domain of SLX4 from various model organisms, carried out with MUSCLE and visualized with Jalview [215, 216]. The hydrophobic residues are colored in pale orange, the positive residues in light green. The invariant glycine is marked by a yellow sphere. Uniprot accession numbers [217] for these proteins are as follows: Q8IY92 (human), Q6P1D7 (mouse), A0A068EUV4 (claw frog), F1QPU2 (zebra fish) and Q22145 (*C. elegans*).

other proteins contain a positively charged patch [168]. This positive patch is also existent in ceSLX4^{SAP} (Figure 2.19 D and E). Most of the positively charged residues are located on α -helix 2 and the loop region (K495, K489, R494, K496, K498, K499). They are facing the surface of the SAP domain.

Next, the ceSLX4^{SAP} model was subjected to three-dimensional structural comparison, in order to find other proteins with the same or similar fold of a SAP domain. The DALI (distance matrix alignment) server was applied for structural comparison and structure database searching [276]. Dali uses 3D Cartesian coordinates of C α atoms of each protein in order to calculate residue-residue distance matrices [277]. The Dali search found 147 entries and 58 of them were unique structural homologs (see Appendix). Among those entries was also PIAS (protein inhibitor of activated STAT), one of the eponyms of the SAP domain, but it was not picked for further analysis here, because it was ranked low in comparison to other proteins. Surprisingly, the Dali search did not give a hit for the yeast Holliday junction resolvase Ydc2/Cce1, which requires the SAP domain for the stabilization of the whole CCE1 protein on the HJ and thus plays a critical role for junction resolution [127]. From the database analysis eight entries were picked for direct structural comparison based on similar function to ceSLX4^{SAP} or on their high similarity (Figure 2.20 A, B). Additionally, to the superimposition by Dali, the structural alignment was further optimized by Pymol for five of the eight entries. The eight proteins can be separated into two groups, one group is also involved in DNA repair and the proteins in the other group are involved in various other DNA or RNA metabolic processes. For both groups the root mean square deviation is low and the structures superimpose well (Figure 2.20 A, B). Although all eight proteins share only a sequence identity between 10 and 20%, they all share same the invariant glycine, which is located in the loop region between helix 2 and helix 3 (Figure 2.20 C, D). For six of them the glycine is located in the superimposition at the same position as the ceSLX4^{SAP} glycine, right after helix 2. In Fan1 and Eri1 is the characteristic glycine closer to helix 3 (Figure 2.20 C, right structure). Among the reported entries the structure of the transcription factor rho had the highest Z-score of 4.5, which indicated the highest structural homology according to Dali. However, as the alignment between ceSLX4^{SAP} and some other structural homologs was further optimized using Pymol, the rmsd is still a bit higher compare to, e.g. Tho1. Rho is a transcriptional terminator and functions as a hexamer, binding either single-stranded RNA or DNA [278]. Rho contains a RNA-binding domain, whose α -helical domain shares structural similarity with ceSLX4^{SAP}. Furthermore, the SAP domain could be found in Tho1, a factor involved in mRNA biogenesis [279], in the 3'-5' exoribonuclease Eri1 (PDB Code 1zbn) as well as in the nuclear protein Hcc1 (PDB code 2do1). Eri1 performs 3' end processing of the 5.8S ribosomal RNA and can negatively regulate RNAi [280]. The structure of the *S. cerevisiae* Tho1 SAP domain has been solved by NMR experiments and dsDNA binding was analyzed by chemical shift perturbations [279].

Some DNA repair proteins showed a similar fold to ceSLX4^{SAP} (Figure 2.20 A, left). The Dali search revealed close resemblance to the SAP domain in FAN1 (FANCD2-associated nuclease 1) DNA structure-specific nuclease [281], Holliday Junction resolvase T4 endonuclease VII [147], DNA double-strand binding protein Ku70 [282] and *E. coli* exonuclease I [283]. The structural data of the four proteins was derived from crystal diffraction experiments.

Comparison of ceSLX4^{SAP} with exonuclease I showed that the SAP-like fold of ExoI (residues 163 - 200) is located in its exonuclease domain (residues 1 - 201) [283]. Another crystal structure

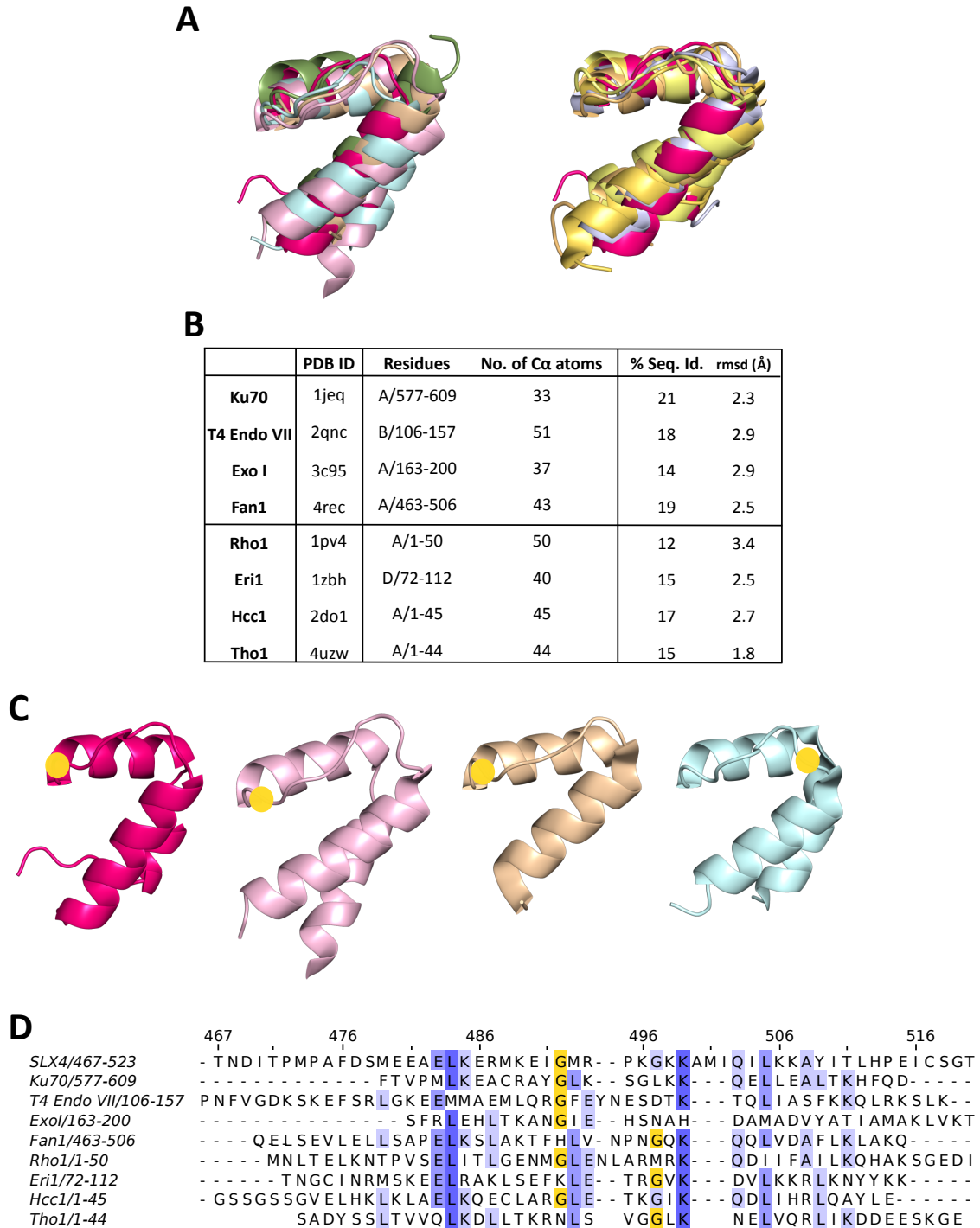


Figure 2.20: Superimposition of ceSLX4^{SAP} with other SAP domains. Continued on the next page.

Superimposition of ceSLX4^{SAP} with other SAP domains. (A) Superposition of ceSLX4^{SAP} (pink). Amino acids 514 to 523 diverge after the helix and are omitted for clarity. Left: Structural overlay with other proteins involved in DNA repair: Ku70 (1jeg, wheat), T4 Endonuclease VII (2qnc, lightpink), Exonuclease I (3c95, olive green) and Fan1 (4rec, lightcyan). Right: Structural comparison with other proteins containing a SAP domain: transcription factor Rho (1pv4, lightorange), 3'-5' Exonuclease Eri (1zbh, paleyellow), Hcc1 (2do1, bluewhite) and Tho1 (4uzw, yelloworange). (B) Table summarizing the structural comparison sequence alignment between ceSLX4^{SAP} and SAP domains from other proteins. The first four proteins are involved in DNA repair and the lower four proteins are involved in other DNA or RNA related processes. For each protein is the Protein Data Bank code (PDB ID) as well as the chain and residue numbers as well as number of C α atoms annotated. Furthermore, for sequence alignment is the percentage of sequence identity (% Sec. Id.) and the root mean square deviation for C α atoms (Å) for structural comparison depicted. (C) Comparison between the SAP domains of ceSLX4, T4 endonuclease VII, Ku70 and Fan1 for the location of the invariant glycine residue (marked by a yellow sphere). Colors are identical to (A). (D) Sequence alignment between the SAP domains of the different proteins from the structural alignment. The alignment was carried out with the multiple sequence alignment tool MUSCLE and visualized using Jalview [215, 216].

of Exo I was solved in presence of ssDNA and could show how the SAP-like fold is involved in the cleavage process of ssDNA [284]. Some residues in the SAP-fold of Exo I are involved binding and activation of Mg²⁺ within the active site, whereas others, e.g. Arg 142, Arg 165, Leu 166 and Phe 164, are involved in electrostatic interactions with the phosphates of the DNA [284].

Ku70 forms a heterodimer with Ku80 and encloses DNA ends from a DSB in a ring shape with high affinity, thus allowing the repair by NHEJ [35, 37, 282]. The SAP domain of Ku70 is located at the C-terminus of the protein and only present in Ku70, but not in Ku80 [168, 282]. The crystal structure of Ku70/Ku80 heterodimer with Ku70^{SAP} was solved in absence of a DNA substrate, but there is also a structure bound to DNA available. The SAP domain in Ku70 is located far away from the ring formed for DNA binding and exposed to the solvent in the crystal structure [171, 282]. Biochemical studies indicate that Ku70^{SAP} is responsible for the high affinity DNA binding of the heterodimer [285]. The Ku70^{SAP} structure has also been solved as a NMR structure and a structural comparison has been carried out as well, showing its similarity to transcription factor rho and T4 endonuclease VII [286]. Similar to ceSLX4^{SAP} within Ku70^{SAP} are conserved Lys residues, which confer DNA binding for Ku70^{SAP} [286].

For T4 endonuclease VII and FAN1 the crystal structures were solved in complex with a DNA substrate [147, 281] (Figure 2.21 A). Phage T4 endonuclease VII is a Holliday Junction resolvase and its crystal structure has been solved in complex with a HJ substrate [147]. The DNA-bound structure shows that the HJ substrate is opened as a parallelogram and thus interrupting the coaxial base stacking [147]. The three-helical bundle located between the arms of a HJ exhibited a fold similar to ceSLX4^{SAP} and has not been described as a SAP domain so far [147] (Figure 2.21 A left). These bundles form the tips of the S-shaped dimeric protein, which interact with an arm of the HJ [147]. The depiction of molecular shape and electrostatic surface potential shows how the SAP domain in T4 endonuclease VII is located between the arms of the HJ and forming an interaction platform with DNA (Figure 2.21 B left). The positive charge of the SAP fold in T4 endonuclease VII and its position between the arms indicates that the SAP domain is important for the stabilization of the protein-DNA complex (Figure 2.21 B left). Here, positively charged residues from helix 1 and 3 are contributing to the DNA interaction. The loop connecting helix 2 and 3 is in this structure turned away from the DNA substrate. The SAP fold could through its interactions make sure that the arms of the junctions are separated and could facilitate the

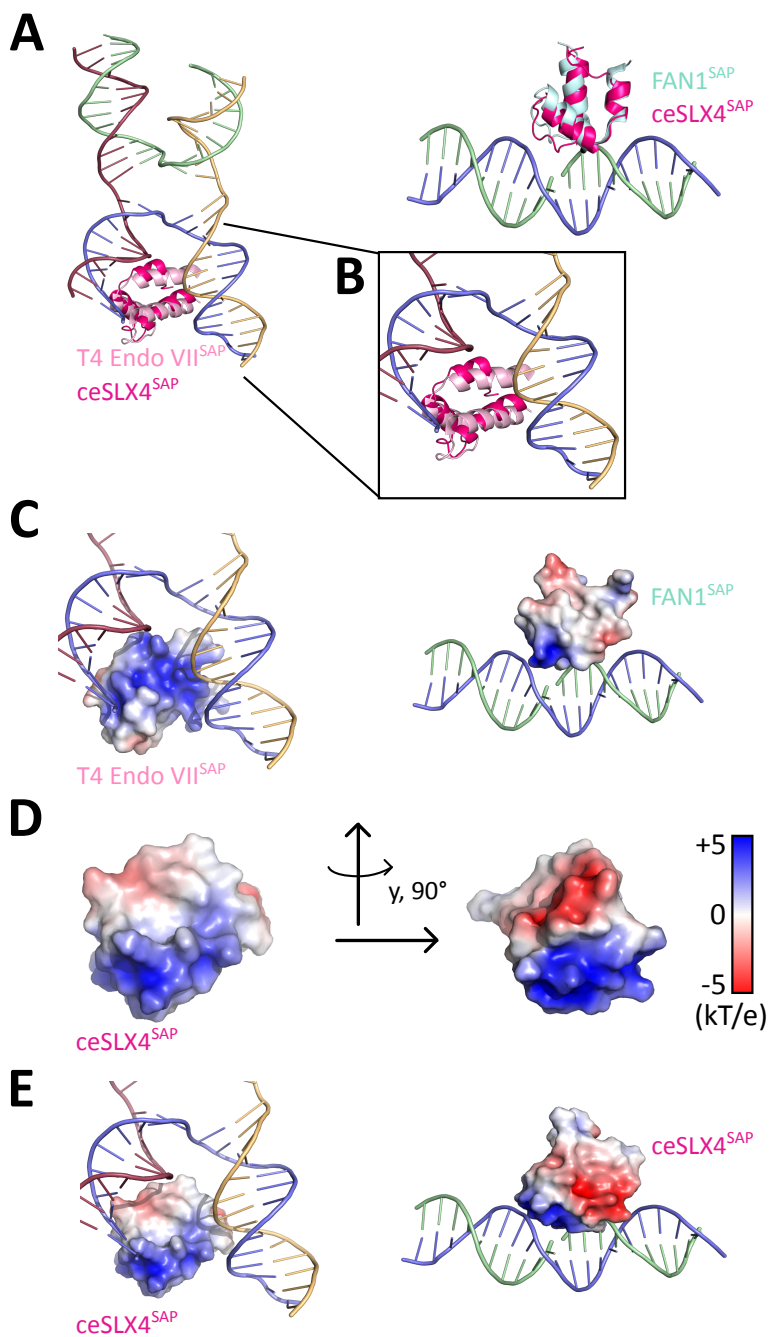


Figure 2.21: Structural comparison of ceSLX4^{SAP} with SAP domains from other protein in complex with DNA substrate. (A) Left: Superposition of ceSLX4^{SAP} (pink) with T4 Endonuclease VII^{SAP} (T4 endo VII, 2qnc, lightpink) in complex with HJ substrate (left) and with Fan1^{SAP} (4rec, lightcyan) and flap DNA substrate (right). (B) Zoomed presentation of the interaction site of T4 Endonuclease VII^{SAP} with HJ substrate (C) Molecular shape and electrostatic surface potential of T4 Endonuclease VII^{SAP} in same orientation as (B left) (2qnc, left) and Fan1^{SAP} (4rec, right) in same orientation as (B right). (D) Surface representation with electrostatic charge distribution of the NMR model of ceSLX4^{SAP} (blue: positive, red: negative). The left molecule is in same orientation as in (B). The right molecule is turned around y axis by 90° (E) Superimposition of ceSLX4^{SAP} shown colored by electrostatic surface potential with the DNA substrate from T4 Endonuclease VII (2qnc) and FAN1 (4rec). ceSLX4^{SAP} is in the same orientation as the respective other SAP domains from (C).

opening of the HJ into a parallelogram. Biochemical studies on T4 endonuclease VII showed that the last three amino-acids of the protein are dispensable for nucleolytical cleavage [148]. However, removal of additional four amino acids (150-153) in the SAP domain causes loss of activity [148]. This loss of activity is due to the loss of DNA binding, as the T4 endonuclease VII molecules with shorter SAP domain showed decreased DNA binding [148].

Human FAN1 is a DNA structure-specific nuclease involved in the processing of DNA inter-strand crosslinks [281]. The FAN1 molecule contains a SAP domain as a primary DNA-binding-domain with a canonical SAP fold and a SAP' region, that connects to the next domain (tetra-tricopeptide repeat (TPR) domain) [281]. FAN1 forms a 'head to tail' dimer, where one FAN1 molecule is carrying out the catalysis and the other auxiliary FAN1 molecule facilitates substrate orientation and unwinding of the flap structure [281]. The SAP domain has in this dimer two roles. In the catalytically active FAN1 molecule the canonical SAP domain interacts with the large groove of the DNA substrate (Figure 2.21 A, right). Inspecting the shape and charge distribution of the molecule, it is apparent that most of the domain is not charged and the site closely to the large groove is positively charged thereby stabilizing the interaction with the DNA substrate (Figure 2.21 B, right). Lys 478 and Lys 482 are located on helix 2 of the SAP domain and are in very close proximity to the DNA backbone of the substrate, visualized by the positively charged patch in the surface representation (Figure 2.21 B, right). The lysines belong to the most conserved region of FAN1, which achieves DNA binding through extensive electrostatic attraction [281]. The SAP domain of the auxiliary FAN1 serves as an interaction module between the TPR and the nuclease domain of the other molecule [281]. The SAP domain thus allows DNA-induced dimerization [281]. In FAN1 the SAP domain fulfills two functions, first it confers dimerization of the molecule by electrostatic interaction and second it is essential for nuclease activity.

Comparing the surface representation with electrostatic surface potential of ceSLX4^{SAP} two large charged patches are existent, one is positively charged and the other forms a small hutch with negative charges (Figure 2.21 C). The positively charged patch is created by lysine residues in helix 2 and the loop connecting helix 2 and 3. This strong separation is not seen in the FAN1 or T4 endonuclease VII. Superposition of ceSLX4^{SAP} with the SAP position of FAN1 and T4 endonuclease VII and comparing ceSLX4^{SAP} in a modeled DNA binding state suggests an indirect interaction with the HJ (Figure 2.21 D). However, when ceSLX4^{SAP} is modeled with the flap substrate of FAN1, the positive charges reach into the large groove of the DNA and interact with the phosphate backbone. The negatively charged patch is not in contact with the DNA. There is the possibility that the two charged patches on ceSLX4^{SAP} carry out different tasks. The positively charged patch in ceSLX4^{SAP} confers DNA binding and the negatively charged patch could be involved in ceMUS81^N interaction. ceMUS81^N is highly positively charged and could interact with the negatively charged patch on the SAP domain.

2.5 Biochemical Characterization of MUS81-EME1 and SLX4^{SAP}

2.5.1 DNA Cleavage Activity of MUS81-EME1

The heterodimer MUS81-EME1 is a XPF family member interacting with SLX4 and cleaves together with SLX1, associated with SLX4^{CCD}, HJs in a nick and counter-nick mechanism [141].

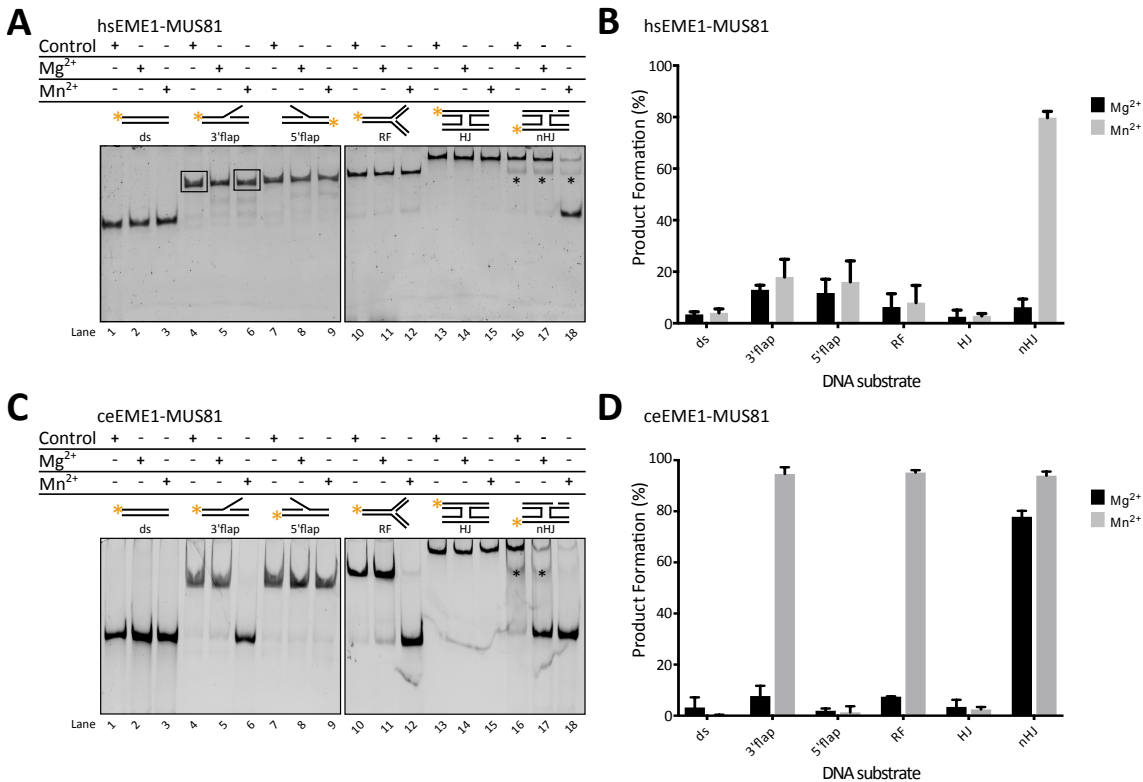


Figure 2.22: Cleavage activity of MUS81-EME1 complexes. (A + C) Nuclease activity of hsEME1-MUS81 and ceEME1-MUS81 with different DNA substrates and Mg²⁺ or Mn²⁺ as cofactor. 40 nM 5'FAM labeled substrate (marked with yellow asterisk) was mixed with 640 nM enzyme and incubated for 30 min at 37°C. The reaction products were separated on a native PAGE and analyzed with a phosphoimager. (B + D) Quantification of the nuclease assays carried out with different DNA substrates and either Mg²⁺ or Mn²⁺ as cofactor. The bar chart shows the percentage of product formation quantified by using ImageJ. The leftover uncut DNA substrate was normalized against the control substrate and product formation was calculated from the difference. The error bars depict the standard deviation from three independent experiments.

MUS81-EME1 is able to cleave nicked HJs alone as well as replication forks and 3'flaps [141, 174]. The Mg²⁺ ion binding site in the crystal structure of hsMUS81-EME1 is similar to the metal on the nucleophile site in other nucleases or polymerases, which employ a two-metal-ion catalysis [214, 223]. The crystal structure in complex with DNA substrate showed electron density only for one Mg²⁺ but not for the second. So far, published studies covering the cleavage activity Mg²⁺ was used as a cofactor [141, 213]. In order to test the cleavage activity of the purified MUS81-EME1 constructs, nuclease assays were carried out with Mg²⁺ as cofactor and additionally with Mn²⁺. The reason behind testing nuclease activity with Mn²⁺ as cofactor is based on its similar ligand preferences and being common surrogate for Mg²⁺ in the majority of enzymes [287]. In the assay different DNA substrates were tested resembling different intermediates occurring during DNA damage repair. For example the flap substrate resembles a replication intermediate created by the displacement of the 5' end of the previous Okazaki fragment during the extension of the nascent Okazaki fragment by polymerase δ .

The full-length human and *C. elegans* MUS81-EME1 nuclease complexes were purified as described previously (see section 2.1.1 and 2.3) and tested for their activity in a nuclease assay

(Figure 2.22). For this, the DNA : enzyme ratio was 1:15 applying single-turnover conditions. Substrate gel bands from three independent measurements were quantified and product formation was calculated from the difference between the control substrate and the substrate after cleavage.

By choosing single-turnover conditions with high enzyme concentrations the catalytic rate at the active site of the enzyme is being measured as all active sites are occupied. By single-turnover measurements the usage of the divalent cation becomes an important factor and events at the active site of the enzyme are isolated. Single-turnover condition might compensate for lacking phosphorylation of the MUS81-EME1 complex after expression in *E. coli* as phosphorylation of the EME1 ortholog Mms4 has shown to be important for nuclease activity in yeast [118].

Human MUS81-EME1 consisted of full-length EME1 and two truncated MUS81 fragments after purification. The complex cleaved in presence of Mg^{2+} only weakly 3' and 5' flaps with maximal 14% product formation (compare lanes 5 & 8 in Figure 2.22 A). The cleavage of 5'flaps occurred in the same extent as compared to 3'flaps and was surprising as cleavage of this substrate was not reported so far for hsMUS81-EME1. With the cofactor Mn^{2+} the cleavage of the flap substrates increased slightly (compare lanes 6 & 9). Product formation after nHJ cleavage reached nearly 80% with Mn^{2+} (lane 18).

In difference to the human MUS81-EME1 complex, *C. elegans* MUS81-EME1 complex with cofactor Mn^{2+} 3'flap, replication fork and nHJ nearly completely (compare lanes 6, 12 & 18 in Figure 2.22 C). Very weak product formation is seen for these substrates when Mg^{2+} is added, but nHJ is cleaved to 80% (lanes 5, 11 & 17).

The unexpected cleavage behavior in presence of Mg^{2+} was puzzling and the question was raised whether a co-factor needed for cleavage was needed. A co-factor, influencing cleavage activity of ceMUS81-EME1 could be ceSLX4^{SAP}. Therefore, the same cleavage assay was carried out in presence of the SAP domain, but the same result was obtained (see Appendix Figure 4.9). Information about the cleavage activity for hsMUS81-EME1 in presence of hsSLX4^{SAP} could not be acquired as constructs comprising hsSLX4 SAP domain posed to be degradation prone.

In the end, limited nuclease activity could be detected on the preferred substrate 3'flap and a clear difference between human and *C. elegans* proteins was detected. Cleavage tests of N-terminally truncated constructs of MUS81-EME1 from both organisms showed expected activity towards the substrates.

2.5.2 Interaction between ceSLX4^{SAP} and ceMUS81^N

Further biochemical characterizations were solely carried out on the *C. elegans* proteins as they were more stable and gave higher yield than the human proteins. For the *C. elegans* SLX4 ortholog the interaction with MUS81 has been shown by yeast-two-hybrid assay [241]. The involved domains were examined for human proteins by yeast two-hybrid screen and immunoprecipitation. The first 106 amino acids of hsMUS81, containing the HhH domain, suffice for the interaction with hsSLX4 [201]. Interaction experiments were all carried out either by the usage of homologous expressed protein and immunoprecipitation or yeast two hybrid screen [201, 202]. However, the interaction between MUS81 and SLX4 has not been studied *in vitro* with heterologous expressed protein until now.

First, the interaction of ceSLX4^{SAP} with ceMUS81^N was assessed by applying a direct method

Table 2.3: Interaction assays applied for the ceSLX4^{SAP}-ceMUS81^N complex. Molecular weights of ceSLX4^{SAP}, ceMUS81^N and ceSLX4^{SAP}-ceMUS81^N complex were determined theoretically. SEC size exclusion chromatography, SLS static light scattering, AUC analytical ultracentrifugation

Protein Complex	Exp. MW ^{SLS} (kDa)	Exp. MW ^{AUC} (kDa)	Theoretical MW (kDa)
ceSLX4 ^{SAP}	25 / 41.7	23.3	18.05
ceMUS81 ^N	27 / 57	21.2	18.5
ceSLX4 ^{SAP} -ceMUS81 ^N	37.8	19.3 / 28.5	36.55

like analytical gel filtration. For SEC interaction assays different pH (8.0, 7.5 and 6.5) and salt concentrations (50, 100 and 150 mM KCl) were tested. However, a shift of the elution volume was not detected in any of the experiments. The two proteins eluted in one peak consisting of the single peaks of each protein alone (Figure 2.23 and Table 2.3). Additionally, static light scattering (SLS) was performed to quantify the complex formation and determine the molecular weights of the proteins (Figure 2.24). During an SLS experiment the protein sample is injected on a size exclusion chromatography column and intensity of scattered light is measured to calculate the average molecular weight. The measurements were conducted by Dr. Claire Basquin. Depending on the concentration of ceSLX4^{SAP} or ceMUS81^N SLS indicated slightly higher molecular weights or even formation of polymers. The obtained molecular weight for the complex indicated dimer formation but as the single proteins showed deviating masses, it is not so solid to rely on these values. Thus, either a different method had to be applied to get a measure for molecular weight or the interaction partners had to be modified.

Next, it was examined whether post-translational modification like phosphorylation was required for complex assembly. One possible kinase is CDK1, which is important for interaction of MUS81 and SLX4 in human cells [200]. Another possible kinase involved in the phosphorylation of SLX4 is PLK1 (Polo-like kinase 1), which has been shown to interact with SLX4 and to phosphorylate it [114]. As for human PLK1 a purification protocol available, it was expressed and purified as a hyperactive mutant (T210D R337A I340A) and applied to ceSLX4^{SAP} in a phosphorylation assay. ceMUS81^N could not be phosphorylated as the assay conditions lead to strong precipitation of the protein. The successful phosphorylation of ceSLX4^{SAP} was assessed by total mass analysis and phosphorylation site mapping by proteomics. hsPLK1 phosphorylated ceSLX4^{SAP} at positions Ser443, Ser460, Thr467, Ser478, Thr511, Thr522, Thr578, Ser578, Ser583, Thr598 and Ser602. Phosphorylated ceSLX4^{SAP} and unphosphorylated ceMUS81^N were subjected to gel filtration analysis alone and in complex with each other (Figure 2.23). However, a peak shift could not be observed. The phosphorylation of the target protein by PLK1 did not enhance the binding between ceSLX4^{SAP} and ceMUS81^N, even though it was shown for human SLX4 [114]. Taken together, in order to analyze the interaction between the two domains, size exclusion chromatography is not suited. Here it seems, that the protein-protein interaction must have either weak binding affinity or the complex has a high off rate.

Next, complex formation was studied by sedimentation velocity experiment, carried out by Dr. Stephan Uebel and Susanne von Gronau in the biochemistry core facility (MPI of Biochemistry). Unmodified ceSLX4^{SAP} and ceMUS81^N were subjected alone and in 1:1 ratio to analytical ultracentrifugation (Figure 2.25 A). For the individual proteins the masses of the monomeric proteins could be retrieved, each of them is a monomer. The sample with the 1:1 mixture of the two

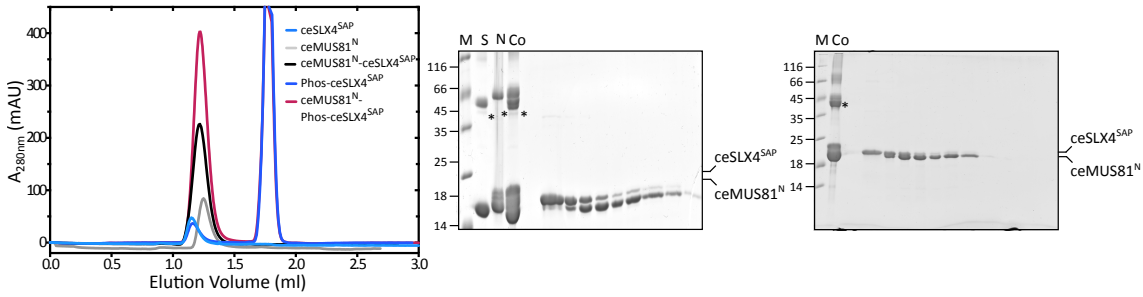


Figure 2.23: SEC studies for the ceSLX4^{SAP}-ceMUS81^N complex. SEC Analysis of unphosphorylated ceSLX4^{SAP} (lightblue), phosphorylated ceSLX4^{SAP} (blue), ceMUS81^N (lightgrey), unphosphorylated ceSLX4^{SAP}-ceMUS81^N (black) and phosphorylated ceSLX4^{SAP}-ceMUS81^N (magenta) in 1:1 ratio at pH 8.0 and 100mM KCl. The elution profile was visualized by UV absorbance at 280nm. Peak fractions are marked by lines and were analyzed by SDS-PAGE with Coomassie staining (right). (M) Protein Marker in kDa, (S) Input sample of ceSLX4^{SAP} (N) Input sample of ceMUS81^N (Co) Input sample for complex of complex. The asterisks mark artifact bands after long exposure to heat.

proteins exhibited two peaks, corresponding to 19.3kDa and 28.5kDa. The calculated size of the complex was with 36.5kDa 28% higher than the experimentally determined complex size. This result indicates a low binding affinity between ceSLX4^{SAP} and ceMUS81^N and an equilibrium between the monomeric proteins and the complex. Through the frictional ratio, the analytical ultracentrifugation experiment can give information about the shape of the molecule or complex. A frictional ratio of 1 indicates a perfect sphere, 1.2 resembles a globular protein and 2.0 a coiled-coil. The sedimentation velocity experiments revealed a frictional ratio of 2.315 for ceSLX4^{SAP} and 1.598 for ceMUS81^N, indicating elongated proteins. The elongation of ceSLX4^{SAP} was partially also seen in the NMR experiments. The last 37 residues of the molecule could not be resolved by NMR and residues 514 to 523 were highly flexible. The flexibility indicates an unstructured region, which is responsible for the high frictional ratio. The frictional ratio of protein complex was determined to 1.9, which indicates again an elongated shape of the proteins. Still, sedimentation velocity experiments could not give a definite answer whether the SAP domain of SLX4 and N-terminus of MUS81 interact with each other, which lead to further search for a method to gain information about the interaction.

As an indirect method to determine protein-protein interactions a thermal shift assay was applied to measure the complex formation between ceSLX4^{SAP} and ceMUS81^N. Thermal shift assays can be applied for quantification of protein-protein interaction and the calculation of binding constants [288–290]. In the ThermoFluor assay the melting temperature of the protein-ligand complex is measured over different ligand concentrations. The resulting curve is then fitted to the Boltzmann model and the K_D is determined. Initially the melting temperature of was ceSLX4^{SAP} (100 μ M), ceMUS81^N (100 μ M) and ceSLX4^{SAP}-ceMUS81^N (500 μ M + 100 μ M) was determined. For ceSLX4^{SAP} no fluorescence signal could be detected. A reason for this could be that the SAP domain does not expose hydrophobic residues, which are in general necessary for association with the SYPR Orange dye [291]. Thus, any signal seen for the complex stems from the fluorescence of SYPR Orange bound to ceMUS81^N. For measuring the interaction of the two proteins, the concentration of ceMUS81^N was kept constant at 100 μ M and ceSLX4^{SAP} was titrated in a concentration range from 5 μ M to 500 μ M. After incubation for 1 h on ice, the melting temperature

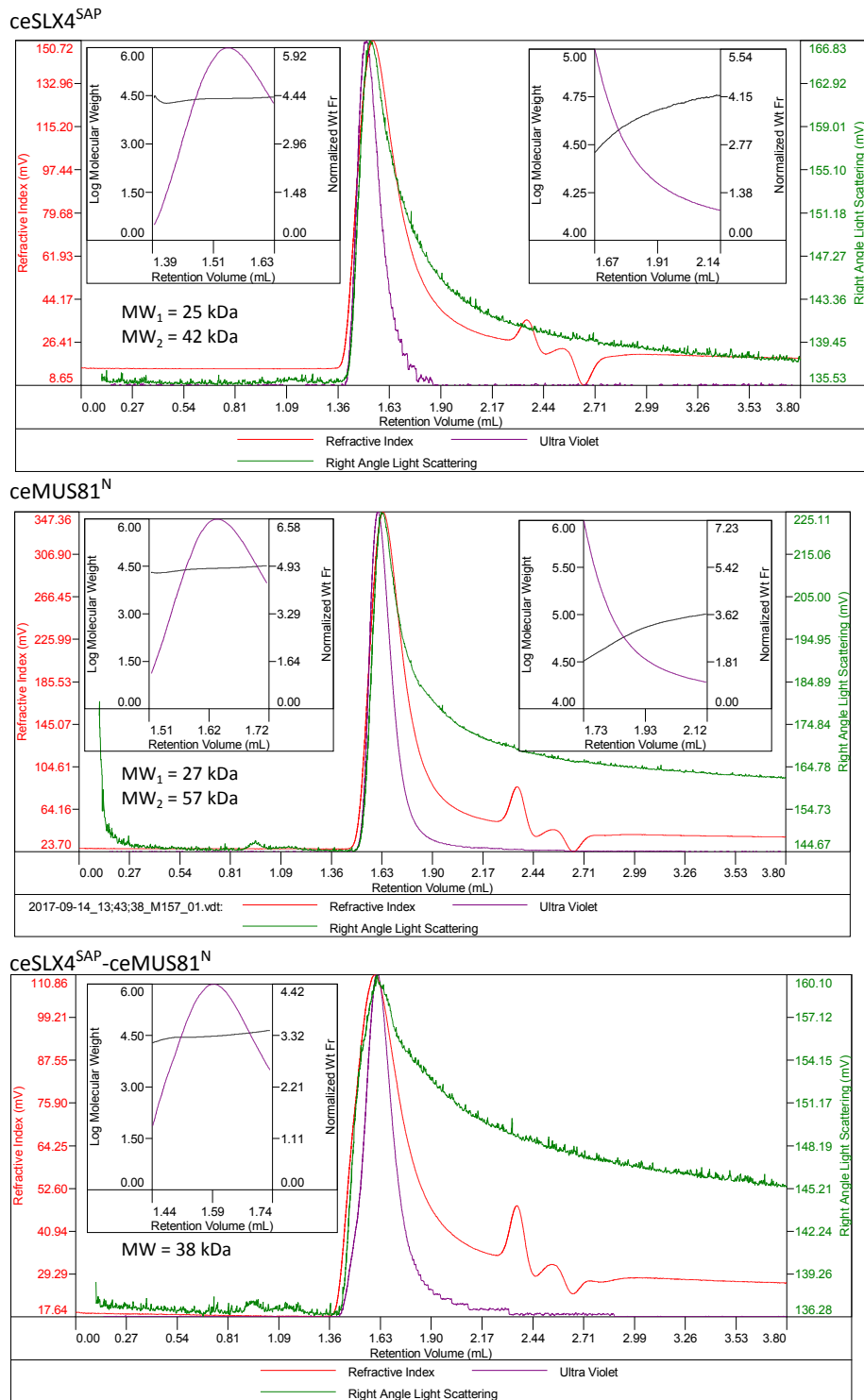


Figure 2.24: SLS studies for the ceSLX4^{SAP}-ceMUS81^N complex. Static light scattering was applied to determine the molecular weight of ceSLX4^{SAP}, ceMUS81^N and ceSLX4^{SAP}-ceMUS81^N. The curves are UV absorbance (violet), light scattering (green) and refractive index (red) for the indicated protein complex solution. The top panels in each spectrum show the average molecular weight values in the population of the peaks shown in logarithmic scale.

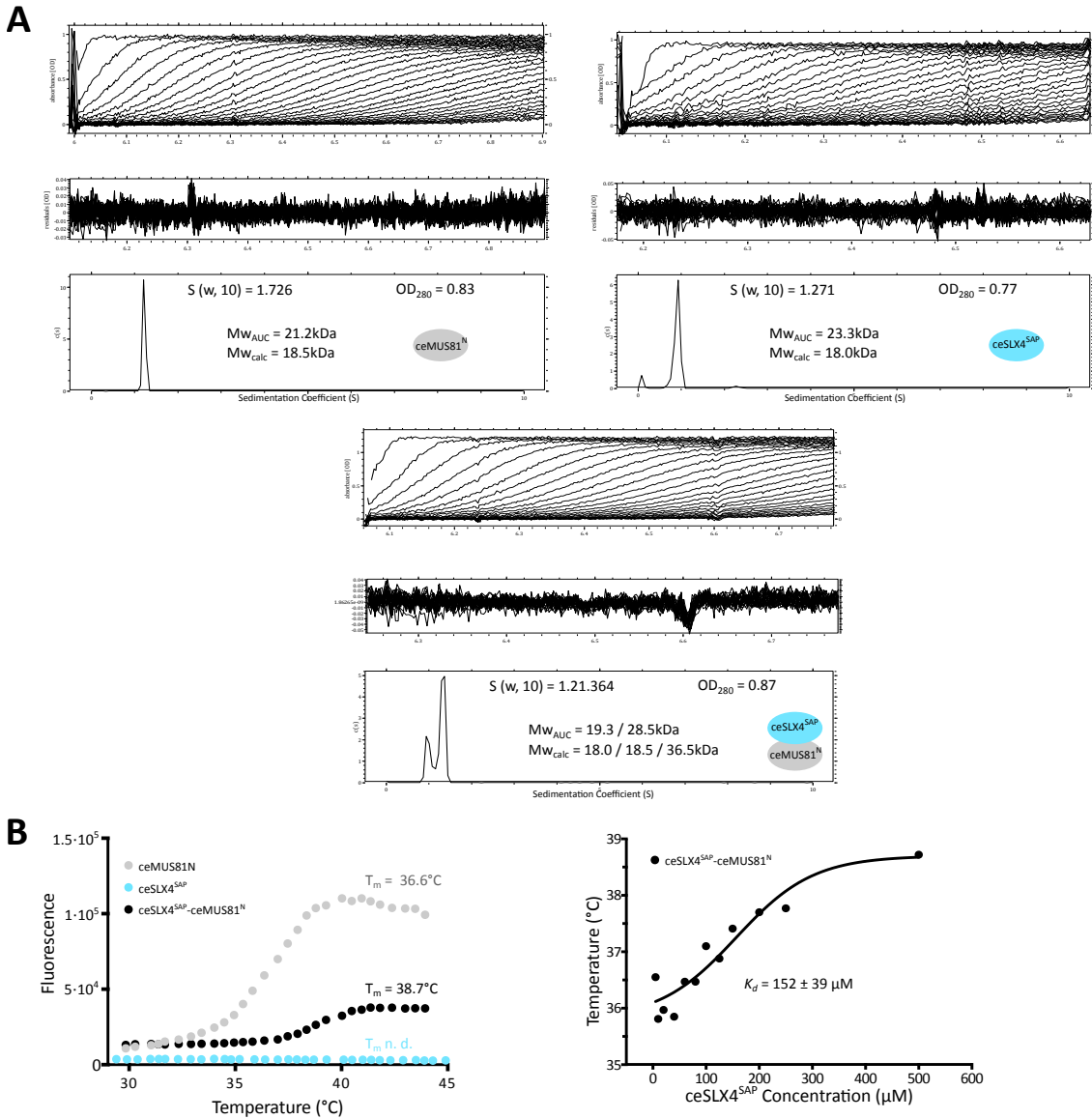


Figure 2.25: Interaction studies for the ceSLX4^{SAP}-ceMUS81^N complex. (A) Sedimentation velocity experiment for ceMUS81^N (left), ceSLX4^{SAP} (right) and the complex ceSLX4^{SAP}-ceMUS81^N (bottom). The determined molecular weights are depicted in each panel and discussed in the text. (B) Thermofluor measurements of ceMUS81^N and ceSLX4^{SAP}. Left: Fluorescence of the control samples depending on the temperature. The single proteins of ceMUS81^N and ceSLX4^{SAP} were measured at 100 μ M. The complex was measured with 100 μ M ceMUS81^N and 500 μ M ceSLX4^{SAP}. The fluorescence of Sypro Orange ceSLX4^{SAP} could not be detected. Right: The concentration of ceMUS81^N was kept constant at 100 μ M and ceSLX4^{SAP} was titrated in a concentration range of 5 μ M to 500 μ M. Analysed is the dependence of the melting temperature on the ceSLX4^{SAP} concentration. The data points were fitted by an Boltzmann equation. The binding isotherm is representative for the quadruplet measurement of binding isotherms. The K_D of ceSLX4^{SAP}-ceMUS81^N complex was found to be 152 μ M \pm 39 under these conditions.

for each titration point was measured (Figure 2.25 B). Applying a Boltzmann model for non-linear regression, the K_D values were calculated for each binding isotherm separately and then averaged. For the interaction between ceSLX4^{SAP} and ceMUS81^N the dissociation constant was determined to $152 \pm 39 \mu\text{M}$. The interaction between the two proteins is in moderate scale, which explains that it could not be detected by size exclusion chromatography. For the first time the interaction between SAP domain and N-terminus could be quantified and a dissociation constant could be obtained.

2.5.3 DNA Binding Properties of ceSLX4^{SAP} and ceMUS81^N

SLX4 has been described to contain a SAP domain [24, 113, 115]. This domain interacts with the MUS81-EME1 complex [201] and upon homology search it has been found that other proteins harboring a SAP domain can have DNA binding properties [168], e.g. Ku70, FAN1 or the yeast Cc1 / Ydc2 resolvase [127, 170, 281]. The SAP domain is typically located at the N- or C-terminus and contains an invariant glycine between two amphipathic helices with additional positively charged patches [168]. Hence, ceSLX4^{SAP} was probed for DNA binding using electrophoretic mobility shift assay (EMSA). Different 5'FAM-labeled DNA substrates mimicking several DNA repair intermediates were added to ceSLX4^{SAP} concentrations ranging from 0 to 2.5 μM (Fig. 2.26).

All tested substrates lead to ceSLX4^{SAP}-DNA complex formation at the lowest protein concentration. However, the ratio between bound and unbound substrate varied between the tested DNA structures. Quantification of the EMSA results showed that the interaction of ceSLX4^{SAP} with dsDNA is weaker compared to HJ substrates, especially at low protein concentrations. HJ resulted in 57% ceSLX4^{SAP}-DNA complexes at 0.25 μM whereas dsDNA has only 10% complex formation (compare lanes 2 and 47 in Figure 2.26 A). In general, ceSLX4^{SAP} does not show any preference for a certain substrate, it is a promiscuous DNA binding domain. For all substrates except for ssDNA higher order complexes are formed at protein concentrations starting between 0.75 and 1 μM , leading to a saturation of the binding. For 5'flap, dsDNA and ssDNA substrate up to 9% of the substrate at highest protein concentration remains unbound (see lanes 45, 54 and 63 in Fig. 2.26 A and C). The binding data was fitted with non-linear regression and applying a fit for cooperative binding mode, thermodynamic parameters were calculated from the EMSAs (Figure 2.26 C). As statistical measurement for agreement between model and measured data R^2 values are shown and except for ssDNA higher than 0.96. The high R^2 values and the low residuals show a good description of the data by the model. The dissociation constants K_D of the different DNA substrates are in the same overall range with slight differences between the substrates. It can be observed that with increasing complexity of the substrate the K_D decreases, with strongest binding towards nHJ. Compared to the others substrates the binding towards dsDNA is weakest with a difference in K_D of 300 nM to HJ. The Hill coefficient is for most of the substrates at 2.2, except for HJ with 3.3. The Hill coefficients of the fitted binding curves greater than 1.0 as well as the sigmoidal shape of the binding curve point towards positive cooperativity.

At the lowest protein concentrations the EMSA for nHJ and HJ showed several shifts indicating the formation of different complexes. For the nHJ substrate only two defined shifts and at higher protein concentrations higher order complexes were formed (lane 2 in Figure 2.26 A). HJ substrate presents three different shifts, which likely corresponds to four SAP molecules bound to one HJ,

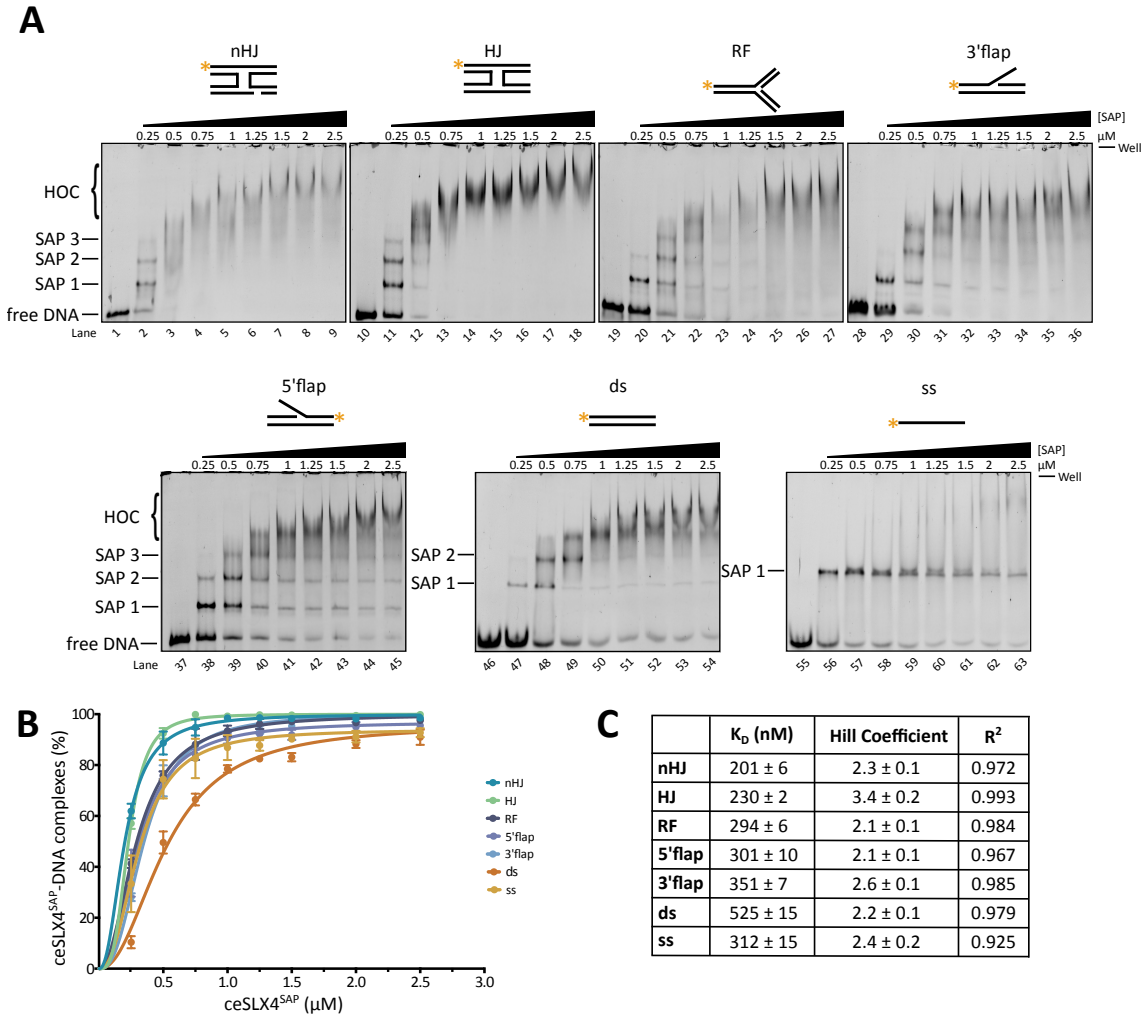


Figure 2.26: DNA binding properties of ceSLX4^{SAP}. (A) Electrophoretic mobility shift assays (EMSAs) of ceSLX4^{SAP} (construct 443-606) with various 5'FAM-labeled DNA substrates. The armlength of HJ, nHJ, RF and flaps was 16 bp, ds DNA had 32 bp and ssDNA was 32 bases long. Increasing amounts of added protein concentration are indicated by a triangle (total concentrations 0.25, 0.50, 0.75, 1, 1.25, 1.50, 2, 2.5 μM). The DNA concentration was kept constant at 40 nM. Different shifts of substrate-DNA complex were observed and are marked with SAP 1-3. Higher order complexes (HOC) were detected at larger protein concentrations. Unbound DNA is marked as free DNA. Seen is a fluorescence image of the native gel. (B) Quantitative analysis of the bound substrate using ImageJ and non-linear regression carried out with GraphPad Prism assuming cooperative binding. Error bars depict the standard deviation calculated from three independent experiments. For some data points the error bar is not shown as it was smaller than the symbol depicting the mean value. (C) Thermodynamic Parameters for ceSLX4^{SAP} binding to different DNA substrates calculated from non-linear regression with cooperative binding mode applying GraphPad Prism [292].

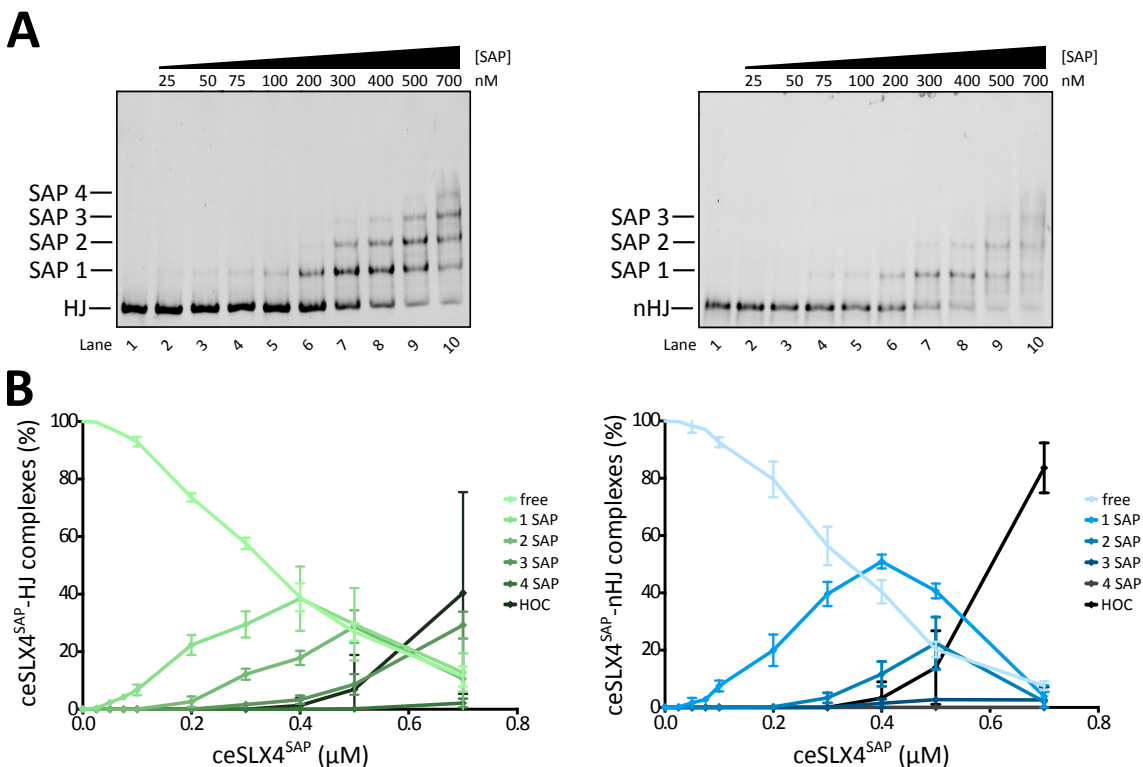


Figure 2.27: Analysis of cooperative binding of ceSLX4^{SAP}. (A) Electrophoretic mobility shift assays (EMSA) of ceSLX4^{SAP} with 5'FAM-labeled HJ and nHJ substrate. The armlength of HJs and nHJs was 16bp. The used protein concentrations ranged from 0, 25, 50, 75, 100, 200, 300, 400, 500 to 700 nM. Seen are fluorescence images of the native gel. (B) Quantitative analysis of the bound substrate using ImageJ, graphical representation with GraphPad Prism. Error bars depict the standard deviation calculated from at least three independent experiments. HOC = higher order complexes

one SAP per arm (see Figure 2.26 lane 11). Also for flap structures maximal three shifts are seen (lane 30 and 39), which would correspond to the binding of SAP to the two arms and the single-stranded overhang. As the HJ and nHJ substrate showed several shifts in the native gel, the EMSA was repeated with lower protein concentrations (Figure 2.27).

With a narrower concentration range it is visible that the critical concentration for forming ceSLX4^{SAP}-HJ complexes was 100 nM (Figure 2.27 A). At this concentration complex SAP 1 is formed, which most likely refers to the binding of one SAP molecule to the Holliday junction (SAP 1). At higher concentrations more shifts were visible, which could refer to two SAP (SAP 2), three SAP (SAP 3) and four SAP molecules (SAP 4) bound to the arms of the HJ. Thereby it is assumed that each molecule bound to one arm of the junction. For example, at a protein concentration of 300 nM a second shift in the native gel was visible, which indicated the binding of two SAP molecules to the HJ. At highest protein concentration few higher order complexes are formed, but additionally four shifts can be seen. The binding towards nHJ showed maximum three shifts on the native gel, indicating a maximal association of three SAP molecules with the nHJ (Figure 2.27 A right). Comparing the binding of ceSLX4^{SAP} between HJ and nHJ, it is visible that more SAP 1 complexes are formed at 400 nM protein concentration (Figure 2.27). Especially for nHJ the formation of higher order complexes is much higher than for HJ.

The formation of different shifts in electrophoretic band shift assay has been detected also for other SAP-domain containing proteins. SAP domains show cooperative binding in general. For spCCE1 the formation of different complexes in EMSA has also been reported [127]. Furthermore, EMSAs of T4 endonuclease VII show several shifts from different complexes as well [148]. There, SAP largely contributes to the DNA binding ability of T4 endonuclease VII [148].

Before the NMR backbone model was generated by Dr. André Mourão, several residues, which were highly conserved in SLX4^{SAP} from different organisms, were mutated to alanine in order to remove charged residues. For this, residues F476, K485, R494, K498 and H513 were mutated to alanine and hereafter named ceSLX4^{SAP-5A}. The mutant ceSLX4^{SAP-5A} was expressed and purified like the wild type protein. DNA binding was tested with different DNA Substrates (Figure 2.28 A - C). The ceSLX4^{SAP-5A} mutant displayed a weaker binding especially towards most of the substrates compared to the wild type protein. At protein concentrations of 2 or 2.5 μM not much higher order complexes are formed, but the protein-DNA mixture is retained in the wells of the native PAGE (Figure 2.28 A, e.g. lanes 26 and 27). The obtained K_D values assuming cooperative binding show in general a weaker DNA binding affinity and in addition the calculated R^2 values are lower than for the wild type protein indicating a higher scatter around the regression line (Figure 2.28 B, C). As another statistical measurement residual plots were inspected and showed similar agreement with the model as the wild type protein. Thus, the cooperative model describes the DNA binding for ceSLX4^{SAP-5A} well.

Although ceSLX4^{SAP-5A} showed a weaker DNA binding for all tested DNA substrates, the promiscuous DNA binding is retained (Figure 2.28 A - D). In direct comparison the overall DNA binding of ceSLX4^{SAP-5A} was weakened compared to the ceSLX4^{SAP} (Figure 2.28 D). The interaction with nHJ was strongly reduced when K_D values were compared, e.g. 201 nM for ceSLX4^{SAP} and 685 nM for ceSLX4^{SAP-5A}. This is an reduction of approximately 3.5-fold. Still the DNA interaction was only reduced but not abolished as there are enough positively charged residues left, which can interplay with the phosphate backbone. Next, the interaction of ceSLX4^{SAP-5A} with ceMUS81^N was tested by thermal shift assay (Figure 2.28 E). A melting curve curve could not be obtained, thus indicating that ceSLX4^{SAP-5A} does not associate with ceMUS81^N. ceSLX4^{SAP-5A} concentrations higher than 125 μM could not be measured as the high protein content lead to aggregation. Through the mutation of the five charged residues the interaction was abolished. The construct ceSLX4⁴⁴³⁻⁶⁰⁶ or ceSLX4^{SAP} might contain two sites, one site for DNA binding and the other site for interaction with ceMUS81^N.

Next, the location of the mutated residues were compared with the NMR model and so far published data. Residue F476 faces in our model inwards to the hydrophobic pocket. The residue K485 is located in α -helix 1 and most likely faces to the solution. The residues R494 and K498 are located in the loop region connecting helix 2 and helix 3. Histidine 513 is located at the end of helix 3, where the unstructured region of the SAP domain starts. Those four residues could give a contribution to DNA binding as well as interaction with ceMUS81^N. Castor et al. published an interaction study based on co-immunoprecipitation with several mutated residues in mouse SLX4^{SAP} and interaction with MUS81 was abolished when Y1340 in mmSLX4 was mutated to alanine [202]. This residue corresponds to F476 in *C. elegans* and mutation to alanine lead to a blockage in the interaction with SLX4. The other residues mutated in ceSLX4^{SAP-5A} are possibly involved in DNA binding and lead to the reduced DNA affinity.

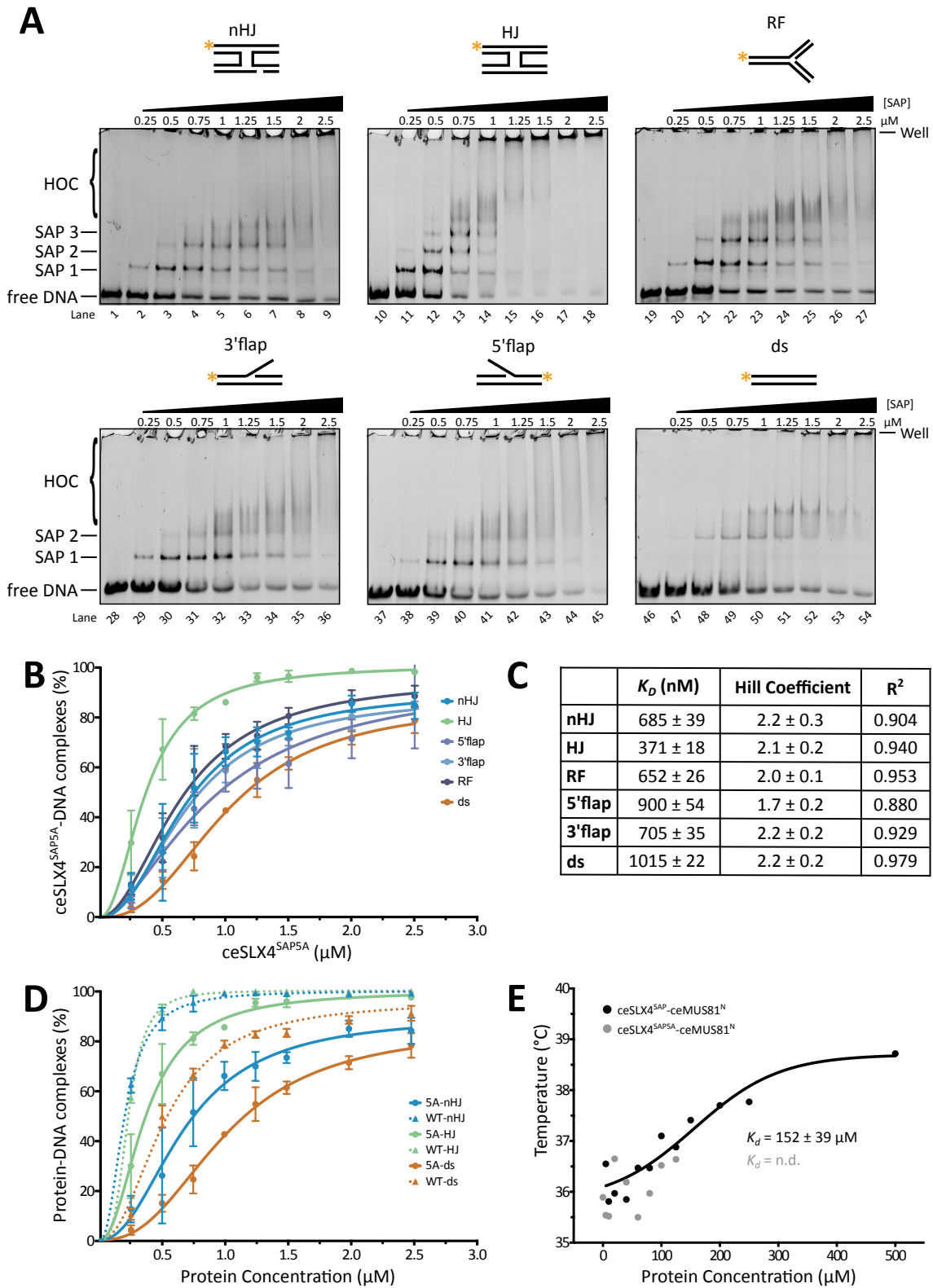


Figure 2.28: Analysis of cooperative binding of ceSLX4^{SAP-5A} and interaction with ceMUS81^N. Continued on the following page

Analysis of cooperative binding of ceSLX4^{SAP-5A} and interaction with ceMUS81^N. (A) Electrophoretic mobility shift assays (EMSAs) of ceSLX4^{SAP-F476A, K485A, R494A, K498A, H513A} with various 5'FAM-labeled DNA substrates as used before. Increasing amounts of added protein concentration are indicated by a triangle (total concentrations 0.25, 0.50, 0.75, 1, 1.25, 1.50, 2, 2.5 μ M). The DNA concentration was kept constant at 40 nM. Different shifts of substrate-DNA complex were observed and are marked with SAP 1-3. Higher order complexes (HOC) were detected at larger protein concentrations. Seen are fluorescence images of the native gel. (B) Quantitative analysis of the bound substrate using ImageJ and non-linear regression carried out with GraphPad Prism assuming cooperative binding. Error bars depict the standard deviation calculated from three independent experiments. For some data points the error bar is not shown as it was smaller than the symbol depicting the mean value. (C) Thermodynamic Parameters for ceSLX4^{SAP-5A} binding to different DNA substrates calculated from non-linear regression with cooperative binding mode applying GraphPad Prism [292]. (D) Comparison of the quantitative analysis of the bound substrate for the binding of ceSLX4^{SAP} and ceSLX4^{SAP-5A}. The wild type ceSLX4^{SAP}-DNA binding is depicted by dotted lines with triangle symbol and the mutant ceSLX4^{SAP-5A}-DNA interaction is depicted by the continuous line with circle symbol. (E) Thermofluor measurements of ceMUS81^N with either ceSLX4^{SAP} (black) or ceSLX4^{SAP-5A} (grey). The concentration of ceMUS81^N was kept constant at 100 μ M and either wild type or 5A mutant ceSLX4^{SAP} was titrated in a concentration range of 5 μ M to 500 μ M. Analyzed is the dependence of the melting temperature on the ceSLX4^{SAP} concentration. The data points of ceMUS81^N with ceSLX4^{SAP} were fitted by an Boltzmann equation. The data points for ceMUS81^N with ceSLX4^{SAP-5A} could not be fitted due to lacking complex formation.

The human MUS81 contains two domains at the N-terminus, the HhH and the WH domain (Figure 1.6), which have been described as DNA binding domains [143, 219]. In the case of ceMUS81 only the N-terminal HhH domain is predicted to be present in the residues 5 to 116. For the residues spanning from 117 to 206 no secondary structure is predicted (see Appendix 4.3). From residue 207 to the end of the protein a large structured region is predicted by the secondary structure prediction tool Jpred. This structured region corresponds to the ERCC4 domain. Comparison with the other species in the alignment shows that ceMUS81 is the shortest ortholog of the compared species and doesn't seem to contain a WH domain (see Appendix 4.3). To determine whether ceMUS81^N binds DNA, electrophoretic mobility shift assays were used to study its interaction with various DNA substrates.

All tested DNA substrates formed complexes with ceMUS81^N (residues 1 - 157) with the highest binding affinity with a K_D of 269 nM towards nHJ (Figure 2.29). The weakest binding is visible towards ssDNA with only 54% DNA bound at 5 μ M protein. In similar K_D range of 800 to 815 nM was the binding towards dsDNA (Figure 2.29). The binding constant towards RF, flaps and HJ were between 500 and 600 nM. The strongest binding of MUS81^N was visible with nHJ, showing a clear preference for binding towards nHJ over other, especially non-branched DNA substrates. For all tested substrates it can be seen that a saturation of the binding is not reached. Although the concentration of protein was in excess (125times higher), 10 - 50% of substrate is unbound depending on the substrate. The Hill coefficients of the non-linear regression showed values >1 , but remaining free substrate at excess protein amount indicates presence of unlimited low binding sites.

Previously, it was shown *in vivo* that the N-terminus of human MUS81 interacted with SLX4-SAP domain [201]. I could demonstrate *in vitro* that both domains interact with each other in absence of DNA substrate at low affinity. Thus, DNA binding was tested for the complex of ceSLX4^{SAP} and ceMUS81^N in a band shift assay as well.

Comparing the band shift assays of the complex among the substrates HJ and nHJ, binding is in the same range (Figure 2.30 A, B). Here, the EMSAs show much more formation of higher

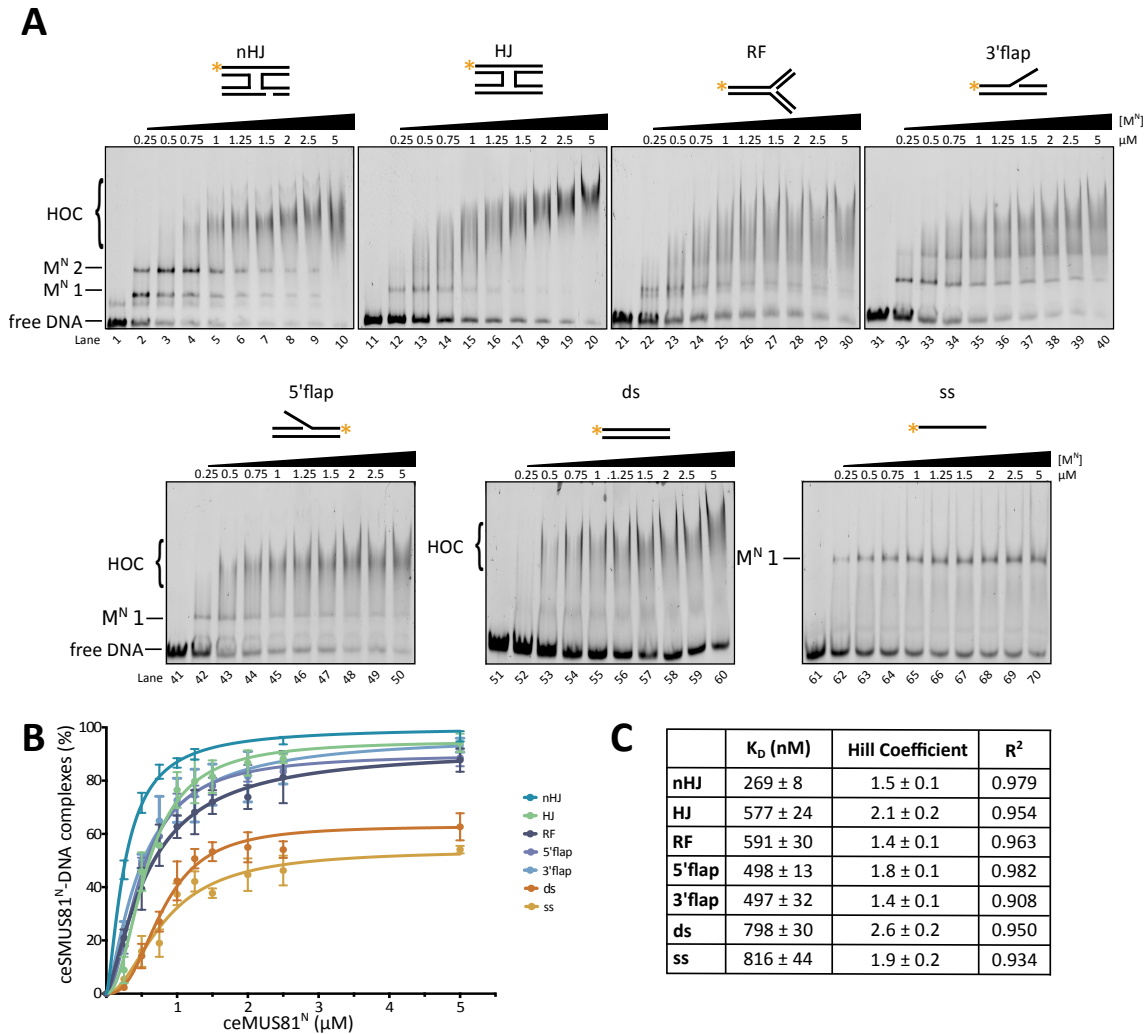


Figure 2.29: DNA binding properties of ceMUS81^N(A) Electrophoretic mobility shift assays (EMSAs) of ceMUS81^N (construct 1-157) with various 5'FAM-labeled DNA substrates. The armlength of HJ, nHJ, RF and flaps is 16 bp, ds DNA has 32 bp and ssDNA is 32 bases long. Increasing amounts of added protein concentration are indicated by a triangle (total concentrations 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2, 2.5, 5 μ M). The DNA concentration was kept constant at 40 nM. Different shifts depending on substrate were observed and are marked with M-N 1-2. Higher order complexes (HOC) were detected at larger protein concentrations. Unbound DNA is marked as free DNA. Seen is a fluorescence image of the native gel. (B) Quantitative analysis of the bound substrate using ImageJ and GraphPad Prism applying non-linear regression with cooperative binding mode. Error bars depict the standard deviation calculated from three independent experiments. For some data points the error bar is not shown as it smaller than the symbol depicting the mean value. (C) Thermodynamic parameters for ceMUS81^N binding to different DNA substrates calculated from non-linear regression with cooperative binding mode applying GraphPad Prism [292].

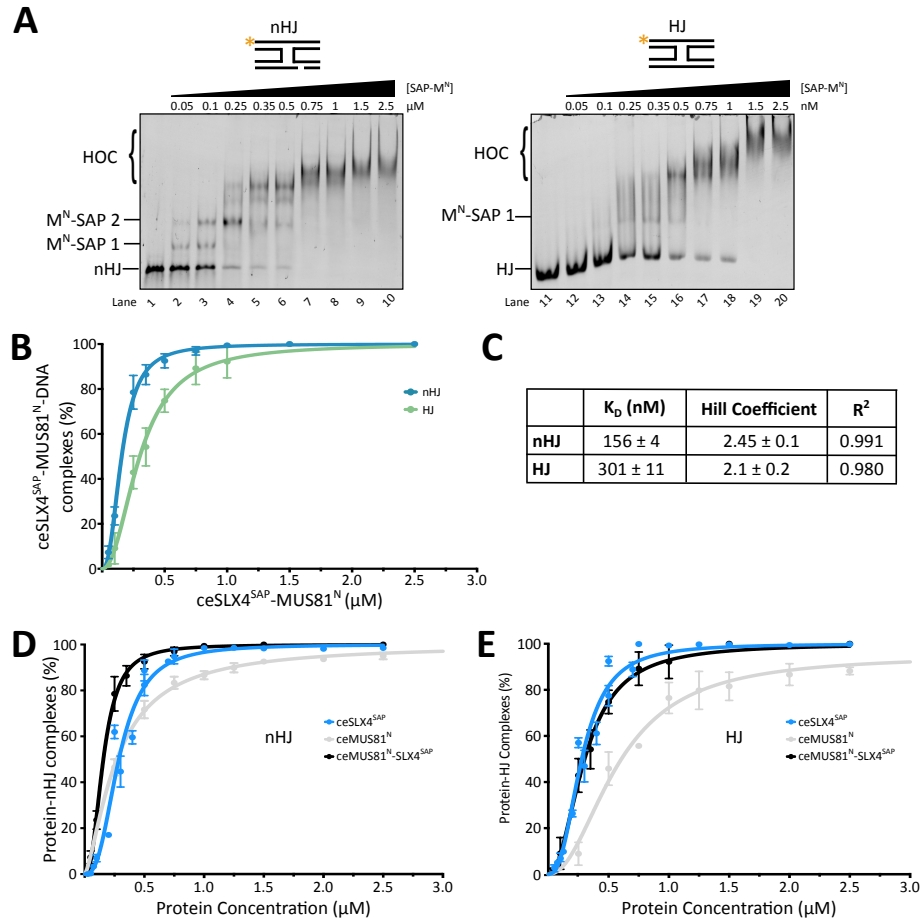


Figure 2.30: DNA binding properties of ceSLX4^{SAP}-ceMUS81^N complex. (A) Electrophoretic mobility shift assays (EMSAs) of ceSLX4^{SAP}-ceMUS81^N with 5'FAM-labeled nicked HJ and HJ. The armlength of nHJ and HJ was 16bp. Increasing amounts of added protein concentration are indicated by a triangle (total concentrations 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 2.00 and 2.5 μ M). The DNA concentration was kept constant at 40 nM. Different shifts depending on substrate were observed and are marked with M^N-SAP 1-2. Higher order complexes (HOC) were detected at larger protein concentrations. Unbound DNA is marked as either nHJ or HJ. Seen is a fluorescence image of the native gel. (B) Quantitative analysis of the bound substrate using ImageJ and non-linear regression was carried out with GraphPad Prism assuming cooperative binding. (C) Thermodynamic parameters of complex binding to nHJ or HJ. (D +E) Comparison between the nHJ (D) and HJ (E) binding of the different proteins. Error bars depict the standard deviation calculated from three independent experiments. For some data points the error bar is not shown as it smaller than the symbol depicting the mean value.

order complexes than for the single proteins and furthermore maximal two band shifts are visible, indicating the formation of two complexes. Furthermore, the distance between the shifts is larger than for the single proteins, indicating that two molecules bind at the same time. Comparing the quantifications of the single proteins ceSLX4^{SAP} and ceMUS81^N and the complex shows that binding of nHJ substrate gives a higher binding affinity than of the single proteins (Figure 2.30 D, E). The curve shape resembles a hyperbolic shape leading to a saturation. The binding towards intact HJ is for ceMUS81^N much weaker than for the SAP domain alone or for the complex (Figure 2.30 E). Binding of the complex to HJ is slightly stronger than of ceSLX4^{SAP} alone. Overall, the quantification of the band shift assays for the complex of ceSLX4^{SAP} and ceMUS81^N can not rule out the possibility that the two domains interact individually with the substrate. Furthermore, the protein concentration of 1 μ M is 150-fold lower than the before calculated K_D for complex formation.

Additionally to the shift assays, DNA binding of ceSLX4^{SAP}, ceMUS81^N and ceSLX4^{SAP}-ceMUS81^N was quantified by applying fluorescence anisotropy (Figure 2.31). In this method the 5'FAM-labelled DNA substrate is excited with vertically polarized light by a wavelength of 470 nm that selectively excites a subset of fluorophores according to the relative orientation of the incoming beam [293, 294]. The emission has a polarization as well and the extent of the polarization is described as anisotropy. Anisotropy originates from within the fluorophore structure, in which the transition moments for absorption and emission lie along specific directions. But the emission can become depolarized by a number of causes. One cause is rotational diffusion, which changes the direction of the transition moments. Fluorescence anisotropy measurements reveal the average angular displacement of the fluorophore that occurs between absorption and emission [294]. The displacement is dependent on the rate and extent of rotational diffusion. The rotational diffusion in turn depends on the size and shape of the molecule. With increasing size of the complex, the rotational diffusion is further decreased and the fluorescence anisotropy increases.

Binding isotherms showed that single ceSLX4^{SAP} binds to 32 bp dsDNA with a K_D of 149.7 \pm 15 nM, to 16 bp armlength HJ with a K_D of 135.7 \pm 17 nM and to 16 bp armlength nHJ with a K_D of 141.6 \pm 8.9 nM (Figure 2.31A & D). A clear difference in the binding towards dsDNA like in the EMSA can not be detected in anisotropy measurements. The DNA interaction is for all three substrates in the same range. However, the Hill coefficient is lowest for dsDNA and highest for nHJ, giving rise to higher cooperative binding when nHJ is given as a substrate. Analysis of the residual plots confirms that the chosen cooperative binding model for ceSLX4^{SAP} describes the measured data quite well, indicated by low residuals. However, the residual plot showed heteroscedasticity. Heteroscedasticity is a statistical phenomenon, which occurred at high protein concentrations when the standard error of the anisotropy increased. This leads also to higher residuals as the predicted values represent the measured values weaker, when they show variance. Taken together the findings from bands shift assays and fluorescence anisotropy demonstrate the binding of ceSLX4^{SAP} in a cooperative manner. The structure and the branching of the DNA substrate facilitates the binding of several SAP domains to one DNA substrate until saturation is reached and in the case of HJ until every arm is occupied by the SAP domain.

Titration of the three different 5'FAM-labeled DNA substrates (HJ, nHJ and dsDNA) with MUS81^N showed the formation of a small plateau between 400 nM and 1 μ M depending on the substrate (Figure 2.31 B). However, a saturation of the binding could not be reached, even at

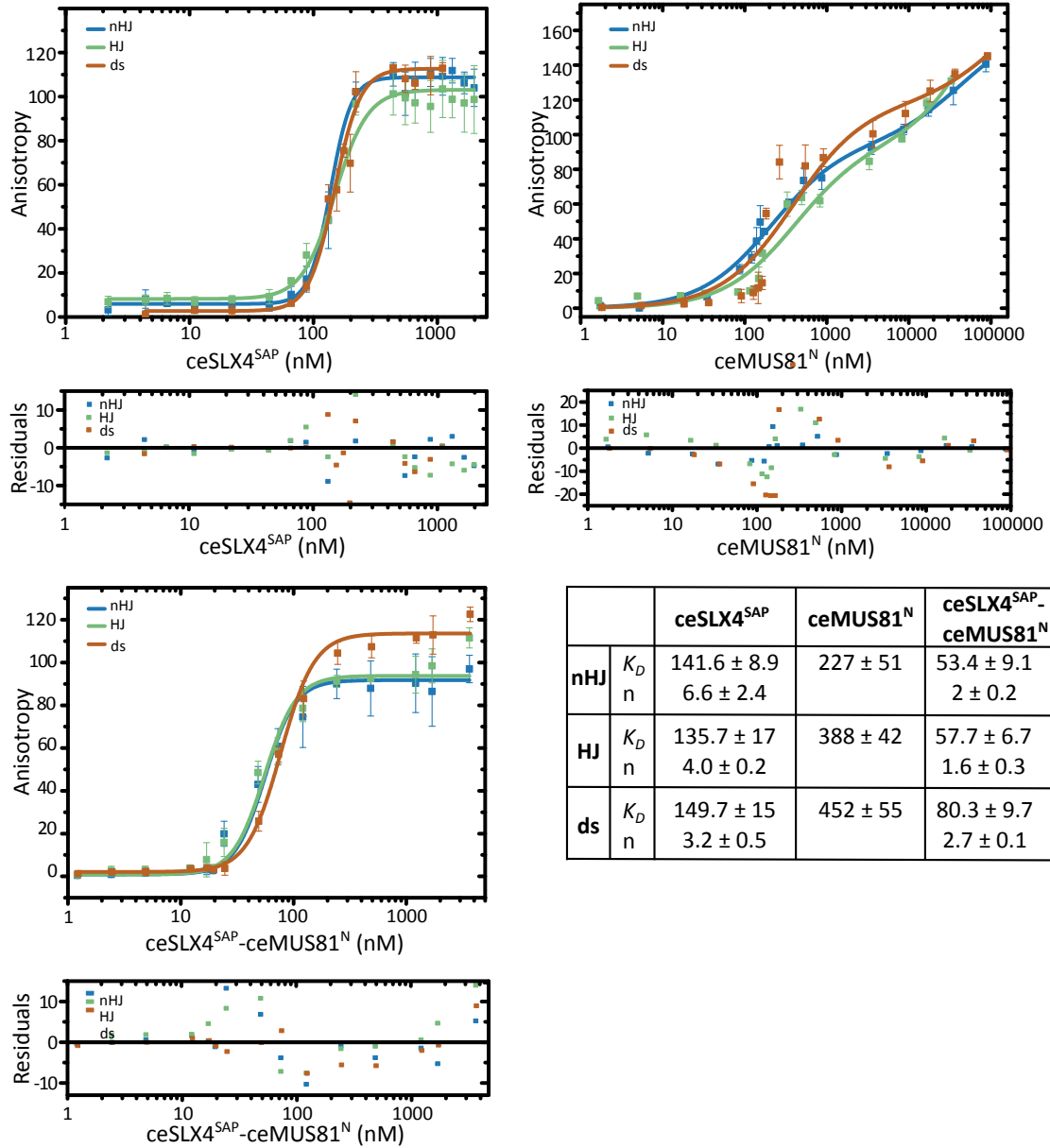


Figure 2.31: Fluorescence anisotropy measurements. Representative binding isotherms for binding of ceSLX4^{SAP} (A), ceMUS81^N (B) and ceSLX4^{SAP}-ceMUS81^N (C) to 16bp armlength HJ, nHJ and 32bp dsDNA (11.2nM) with 5'-FAM-label analyzed by fluorescence anisotropy. The anisotropy is shown as a function of protein concentration in logarithmic scale. (D) Summary of the calculated K_D in nM and Hill coefficients (n). For ceSLX4^{SAP} and ceSLX4^{SAP}-ceMUS81^N the data were calculated by using a nonlinear fit to Hill equation as described in Section 4.3.15. For ceMUS81^N data were fitted to a binding equation describing a two-site binding model described in Section 4.3.15. The error bars represent the standard deviation of three independent measurements. At the bottom of each anisotropy plot is the respective residual plot for the curve fit (difference between the calculated anisotropy of the curve fit and the measured anisotropy).

protein concentrations of more than 100 μM (10300-fold excess of protein over DNA). To analyze the DNA binding of ceMUS81^N a two-site binding model was applied. At the first inflection point measured data and fitted curve deviate most for dsDNA, which can be seen at the high residuals in the plot. For nHJ the two-site binding mode is best described leading to low residuals in the residual plot. The calculated K_{DS} have a high deviation. Still, a clear preference of ceMUS81^N towards nicked Holiday Junction can be detected, as the dissociation constant is significant smaller than for the other tested residues. Band shift assays and fluorescence anisotropy showed that DNA binding of ceMUS81^N does not reach 100% DNA binding at very high protein concentration. Especially in fluorescence anisotropy measurements it is visible that binding between ceMUS81^N and DNA substrates shows at low protein concentrations a cooperative binding mode and at concentrations exceeding 120 nM a second binding interaction is observed. This second binding interaction can be dependent on the positively charged ceMUS81^N, as it consists of 20.4% positively charged residues (Lys 13.4%, Arg 7%) and a pI of 9.28 resulting in a positively charged protein in the assay buffer (pH 6.5). Thus, the binding isotherms above 120 nM protein concentrations might result from electrostatic interactions between the already formed ceMUS81^N-DNA complex and excess ceMUS81^N.

Fluorescence anisotropy was also measured for the complex ceSLX4^{SAP}-ceMUS81^N. The calculated K_{DS} are below 100 nM and up to 60% lower than for ceSLX4^{SAP} alone. The binding towards HJ and nHJ is in the same range, but the binding towards dsDNA is weaker. This confirms the result of the band shift assays, showing a slight preference for branched DNA. Analysis of the residual plots confirms that the chosen cooperative binding model for ceSLX4^{SAP}-ceMUS81^N describes the measured data quite well, indicated by low residuals and no bias.

2.5.4 Mapping of the ceSLX4^{SAP} DNA Binding Region

The standard construct applied for biochemical investigations on the ceSLX4 SAP domain ranged from amino acid 443 to 606 and comprised 164 amino acids. This construct was also used for the first HSQC-spectrum and was too large for chemical shift perturbation experiments (Figure 2.18). Hereafter, several shorter SAP constructs were tested for expression, purification and DNA binding (Table 2.2). Figure 2.32 A shows the sequence alignment of several SAP domains from different organisms based on the ceSLX4⁴⁴³⁻⁶⁰⁶. Band shift assays of various shorter SAP constructs were performed with nHJ as DNA substrate and a protein concentration range between 250 nM and 5 μM (Figure 2.32 B, C). Constructs ending at residue 515, after the highly conserved region in the SAP domain, did not bind nHJ. At protein concentrations higher than 1.5 μM a smear was detected in the native gel. The region from Ala 530 to Arg 606, which contains a conserved area, revealed a similar binding behavior in the EMSA (Figure 2.32 B right). Both constructs alone did not bind to DNA like the standard construct. The construct ceSLX4⁴⁶⁷⁻⁵³⁰, comprising the highly conserved part of the SAP domain and the flexible region after helix-3, did not even show smear in the band shift assays at high protein concentrations (Figure 2.32 C). In this study the minimal DNA binding capability of ceSLX4^{SAP} was defined between the residues 467 and 560.

However, the combination of ceSLX4⁴⁶⁷⁻⁵³⁰ and ceMUS81^N displayed DNA binding. A quantitative analysis of the protein-DNA complexes was performed (Figure 2.32 D) and the complex of ceSLX4⁴⁶⁷⁻⁵³⁰ with ceMUS81^N showed similar binding affinity as the ceSLX4⁴⁴³⁻⁶⁰⁶ alone and

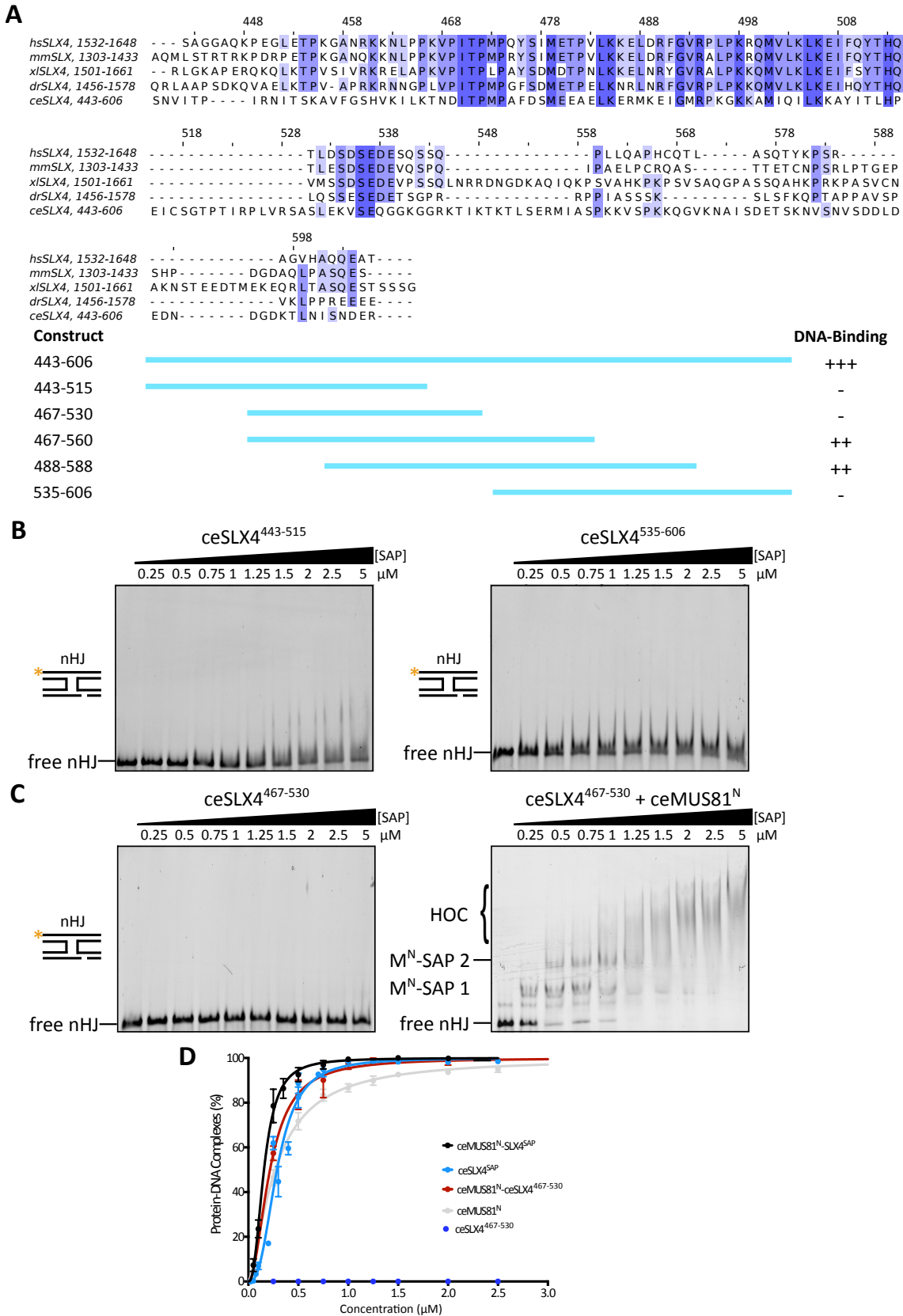


Figure 2.32: Determination of the DNA binding site in ceSLX4^{SAP} Continued on the following page.

Determination of the DNA binding site in ceSLX4^{SAP} (A) Sequence alignment of different SAP domains from different species compared with the longest SAP construct of ceSLX4. The panel below depicts the constructs tested for DNA binding. +++ indicates strong binding like seen for the standard construct ceSLX4⁴⁴³⁻⁶⁰⁶, ++ indicates DNA binding, but with lower affinity than the standard construct. (B) EMSAs of two constructs showing very weak to no DNA binding activity. For DNA binding a 5'FAM-labelled 16bp armlength nicked HJ substrate was used. DNA binding was tested for protein concentrations from 0 to 5 μ M. (C) The construct ceSLX4⁴⁶⁷⁻⁵³⁰ was tested for DNA binding and did not show DNA interaction. In complex with the N-terminal domain of ceMUS81 the binding towards nHJ was tested as well. (D) Quantification of the bound substrate using ImageJ and non-linear regression carried out with GraphPad Prism assuming cooperative binding. The quantification is based on triplicate measurements except for ceSLX4⁴⁶⁷⁻⁵³⁰.

higher affinity than ceMUS81^N alone. As the binding of the two proteins together is higher than ceMUS81^N alone, one can exclude that the DNA binding activity originates only from MUS81^N. However, the same binding affinity of ceSLX4⁴⁴³⁻⁶⁰⁶ in complex with ceMUS81^N is not reached.

The bands shifts assays give information about two different characteristics of the SAP domain. First, the highly conserved part of the SAP domain alone is not sufficient for DNA binding and second the combination of the two DNA binding domains has higher DNA binding affinity than the individual domains. For this process it can be possible that ceSLX4⁴⁶⁷⁻⁵³⁰ undergoes a conformational change upon interaction with ceMUS81^N and this increases binding affinity. Or it could be possible that interaction of ceMUS81^N with ceSLX4⁴⁶⁷⁻⁵³⁰ increases DNA binding affinity of ceMUS81^N.

3. Discussion

The DNA repair factor SLX4 is considered as a scaffold protein that integrates many different functions through diverse interactions. SLX4 interacts with the structure-selective endonucleases SLX1 and the MUS81-EME1 complex, forming the SM complex [141, 143]. So far, it is not known how the active sites of SLX1 and MUS81-EME1 are coordinated during HJ cleavage. The involvement of two active sites from different heterodimeric enzymes is unique to the SM complex and represents a novel mechanism of HJ resolution in eukaryotes [236]. My thesis addresses the open question how SLX4 is involved in DNA interaction during DNA substrate processing. The presented data shows a structural model for the SAP domain of SLX4 and that SLX4^{SAP} is an additional element in the DNA cleavage process. Based on the presented biochemical data I could show, that the SAP domain is a promiscuous DNA binding domain and its interaction with the N-terminus of MUS81 was analyzed *in vitro*. Interestingly, SLX4 expressed from HEK293 cells was highly phosphorylated.

3.1 Reconstitution of the SM-Complex *in vitro*

In order to gain structural and biochemical information about the Holliday-Junction resolvase complex SLX1-SLX4-MUS81-EME1, expression and purification of the four proteins was attempted in a first step. For this human and *C. elegans* orthologs were first expressed in *E. coli* and human SLX4-SLX1 was expressed in mammalian cells.

In general it was challenging to express the SLX1 orthologs. Expression of ceSLX1 full-length and the N-terminally truncated construct could be accomplished in bacterial cells. However, size-exclusion chromatography indicated soluble aggregates by early elution and in a cleavage test in complex with ceSLX4 activity was not detected. Attempts to purify human SLX1 alone failed, however it was possible to obtain functional protein from HEK293 co-expression with its interaction partner SLX4. The aggregation after over-expression could arise from the reported homodimerization of Slx1, when in inactive state [156]. Upon interaction with SLX4^{CCD} SLX1 becomes active. Thus, the co-expression with SLX4 constructs in HEK cells was carried out and SLX1 could be expressed soluble and showed nuclease activity after purification. Furthermore, bacterial expressed hsSLX4 constructs showed stability issues and were therefore challenging for crystallization as well. Moreover, *C. elegans* SLX4 constructs were over expressed in *E. coli* and was tested for DNA binding and interaction with MUS81^N. For ceSLX4^{SAP} and ceMUS81^N large amounts of pure proteins could be purified.

3.2 Post-Translational Modification of hsSLX4

Human SLX4 constructs were purified after expression in mammalian cell culture and samples showed a complex phosphorylation pattern. Phosphorylation site mapping detected six phosphorylation sites in hsSLX4¹³⁰⁰⁻¹⁸³⁴ alone and 18 in co-expression with hsSLX1¹¹⁻²⁷⁵. The here described sites, were found in all biological triplicates. But a batch to batch variation was seen with few sites detected from one expression but not in another one. The amount of phosphorylation and the distribution are interesting to note. The distribution of phosphorylation is different between hsSLX4¹³⁰⁰⁻¹⁸³⁴ expressed alone compared to co-expression with hsSLX1¹¹⁻²⁷⁵. When hsSLX4¹³⁰⁰⁻¹⁸³⁴ is expressed alone, the phosphorylation is reduced by approximately 3-fold. The phosphorylation sites were mostly found at the SIM3 domain and between SAP and CCD for SLX4¹³⁰⁰⁻¹⁸³⁴ expressed alone. hsSLX4¹³⁰⁰⁻¹⁸³⁴ co-expressed with SLX1¹¹⁻²⁷⁵ was much stronger phosphorylated between the SIM3 and SAP domain. This difference in distribution between the expression conditions is very interesting and needs further examination.

The comparison of the herein identified phosphorylation sites with the database Phosphosite Plus showed that only a third of the SLX4¹³⁰⁰⁻¹⁸³⁴ phosphorylation sites were also listed in the database. Most entries were found for T1315 and S1329, but the biological significance is not know. The database contains a list of phosphorylation sites found in large proteomic studies for hundreds to thousands of proteins. The relevance of all the SLX4 phosphorylation sites and stoichiometry is not described in the database. Further intricate experiments are needed in order to obtain quantitative information about the detected phosphorylation sites.

Additionally, phosphorylation sites were mapped after DNA damage using different DNA damaging agents in order to gain a more consistent distribution of the phosphorylation. The protocols for the DNA damage treatment were adopted from literature and the cells were not checked for the induction of the DNA damage or the DNA damage response. In order to confirm the induction and extent of DNA damage additional experiments have to be carried out, e.g. metaphase spreads, analysis of histone γ -H2AX content or 2D-electrophoresis. The phosphorylation pattern under the influence of DNA damaging agents showed only limited differences, however some sites were not changed under the influence of certain agents. Hydroxyurea showed the lowest difference in phosphorylation intensity compared to control cells. Hydroxyurea causes a long chain of reactions leading ultimately to DSB and it is possible that the incubation time or the concentration used was not enough in order to elicit DSBs as the DNA damage state of the cells was not analyzed.

In general, the analysis and interpretation of phosphorylation sites is challenging as samples show heterogenous phosphorylation pattern and it has been shown before that phosphorylation is not uniform and happens at several neighboring sites [295]. This could be easily seen, when comparing the detected phosphorylation sites from the analysis of the co-expression of hsSLX4¹³⁰⁰⁻¹⁸³⁴ with hsSLX1¹¹⁻²⁷⁵ with the untreated control cells in the experiment with DNA damaging agents. Detected phosphorylation sites always showed subtle differences in the expression batches, which requires enough sample replicas and a thorough analysis would involve more tests. Another difficulty for analysis is that the identification of the exact phosphorylated residue is not possible as one peptide mass fingerprinting does not yield direct sequence information [295] and on some peptides phosphorylation can occur on several residues. Furthermore, depending on the expression batch, different phosphorylation sites could be detected, which indicates a variation in

kinase activity. This variation and the high phosphorylation state might also be influenced by the expression conditions. After transfection of HEK cells they were shaken at 32°C to increase the protein expression [296, 297]. This sub-physiological temperature and the large amount of protein could lead to an activation of kinases, which keep the expressed protein soluble by adding negatively charged modifications. Although the system applied for detecting phosphorylation is an artificial system, it was still beneficial for the here described experiments. For phosphosite mapping large amounts of proteins for high coverage in MS are needed and SLX4 has in general low expression levels. Therefore, your expression system is a compromise to get enough coverage for MS analysis. Furthermore, it would also be interesting to measure the kinetics of phosphorylation after DNA damage in order to see if certain phosphorylation sites appear or disappear at a certain time point.

In the depicted experimental setup, hsSLX4¹³⁰⁰⁻¹⁸³⁴ was purified from unsynchronized cells. Thus, SLX4 comes from a mixture of different stages of the cell cycle. As phosphorylation of scSlx4 is regulated in a cell cycle dependent manner [193, 194], one can assume that human SLX4 is cell cycle regulated as well. Here, purified SLX4 consists of a mixture of molecules originated from cells at different cell cycle stages representing different phosphorylations. At a given step in cell cycle a single peptide can be unphosphorylated, weakly phosphorylated or highly phosphorylated. Future analysis of the phosphorylation state of human SLX4 could elucidate the phosphorylation status of different sites by synchronously grown cells and their arrest at different cell cycle stages. A systematic approach would give more information about the modification state at the respective cell cycle stage and the influence of the various DNA damaging agents.

In order to get a deeper understanding of the phosphorylation of human SLX4 it would be more beneficial to analyze SLX4 from expression in a stable cell line. In general, stable cell lines show less heterogeneity of the protein product and high productivity. However, to establish a stable cell line overexpressing SLX4 is time-consuming and it would involve an inducible expression system as SLX4 is toxic for the cells. Transient expression leads to high amount of expressed protein in short time and is flexible towards changes in the construct and trying different crystallization conditions. The problem is that the cell has to deal with high amount of DNA after transfection and then large protein amounts, which have to be kept in an soluble state. Thereby phosphorylation of the protein is also used as it keeps the highly expressed protein soluble. A stable cell line expression SLX4 is more beneficial as it has low batch to batch variability and lower heterogeneity of the product. Furthermore, a stable cell line produces protein amounts closer to the physiological level, which are lower than in transient expression. As SLX4 tends to degrade, an expression level closer to the physiological expression level could be beneficial for protein stability. With a stable cell line, the influence of SLX4 on the cell cycle could be better understood and it would be more straightforward to synchronize the cells with low SLX4 expression levels than with high expression levels. Nowadays, the establishment of stable cell lines has advanced and is even possible applying CRISPR/Cas9 technology [298]. In this method, the recombinant protein of interest is integrated into specific genomic loci using CRISPR/Cas9 and the gene expression is driven by endogenous promoters [298].

In SLX4 the exact physiological role of most phosphorylation sites is unclear. In this thesis I could show that the phosphorylation pattern of hsSLX4¹³⁰⁰⁻¹⁸³⁴ is very different from hsSLX4¹³⁰⁰⁻¹⁸³⁴ co-expressed with hsSLX1¹¹⁻²⁷⁵. One can so far only speculate about the reason for this experi-

mental finding. It was shown earlier that the disruption of murine *Slx4* gene lead to absence of SLX1 in Western Blot, suggesting that SLX4 regulates SLX1 protein stability [202]. Furthermore, SLX1 is not active and stable in absence of SLX4 [156]. Through phosphorylation SLX4 activates itself for SLX1 interaction. The phosphorylation between the domains is increased and causes conformational change to bind the SLX1 nuclease and the MUS81-EME1 nuclease complex [200]. It might be that a specific phosphorylation pattern indicates a certain role for SLX4 as it is as a scaffold protein involved in other DNA repair pathways.

An example for the importance of certain phosphorylation sites is the abolished SLX4-MUS81 complex formation after mutation of phosphorylated residues around the SAP domain [200]. Similar data for the interaction between SLX1 and SLX4 is not available and so far it is unknown how the interaction between human SLX4 and SLX1 is regulated. But comparison of the detected sites with the Phosphosite Plus database showed that nearly half of the sites were also found *in vivo* in large scale proteomic approaches. Furthermore, based on this data it would also be compelling to detect the phosphorylation sites of SLX4 when co-expressed with MUS81-EME1 and when co-expressed with both nucleases. A strong difference could already been seen between the phosphorylation of SLX4 expressed alone and expressed with SLX1, then there might be a difference to phosphorylation in co-expression with MUS81-EME1. The co-expression with another nuclease could change the phosphorylations around the SAP domain. The proteomic data could give a deeper insight into the regulation of SLX4 and how certain functions could be achieved by post-translational modification. The shown study and future studies have to be taken carefully when discussing the biological relevance of phosphorylation sites. There is no information about the stoichiometry of the phosphorylation site available as phosphorylated and unphosphorylated peptide are chemically distinct and behave differently during MS. The detection methods are also so sensitive, that phosphorylation sites might be detected, which are at very low occupancy [264].

It would be that the many detected phosphorylation sites, asynchronously grown cells and a possible hyperactivity of kinases during protein over-expression, resulted in a heterogenous protein sample. An applicable way to reduce the heterogeneity of the samples was not found. For future crystallization studies of hsSLX4¹³⁰⁰⁻¹⁸³⁴ and hsSLX1¹¹⁻²⁷⁵ the protein samples have to be either stabilized by addition of another interaction partner or reduced in size. Another option could be the expression in insect cells, which was not tested in the course of the presented research. Insect cell expression offers high levels of protein expression and introduces similar post-translational modifications as mammalian cell expression. Crystallization of the same constructs purified from HEK cells with the background of the phosphorylation sites might be challenging even after phosphatase treatment .

3.3 The ERCC4 Typ Nuclease MUS81-EME1

MUS81-EME1 belongs to the XPF family of endonucleases, which is diversified in different substrate preferences, thereby playing different physiological roles [142, 174]. Published assays for substrate preferences showed enzymatic activity on nicked HJs, replication forks and 3'flaps for human MUS81-EME1 [141, 174]. Full-length ceMUS81-EME1 was expressed and purified from bacterial cells and cleavage assays revealed that ceMUS81-EME1 had the expected activity towards 3'flap, RF and nHJ when the reaction was supplemented with Mn^{2+} (Figure 2.22). The

low activity of ceMUS81-EME1 in presence of Mg^{2+} towards its usual substrates was surprising. It was expected that ceMUS81-EME1 would behave like the human or yeast complex, which have been shown to cleave substrates in presence of Mg^{2+} in the reaction buffer [135, 141]. A possible reason for this could be lacking post-translational modification.

Cleavage tests for human MUS81-EME1 deviated even stronger from published results as hsMUS81-EME1 exhibited cleavage activity on nHJ only in presence of Mn^{2+} . There are several possible explanation for this. One could be the lack of the one Hhh domain at MUS81 as human full-length EME1 could be purified but human MUS81 was lacking either N- or C-terminally 11 kDa. The 11kDa could correspond to the hairpin-helix-hairpin domains located at each end. A recent publication attributed an influence of the N-terminal HhH domain to the cleavage reaction and specificity [143]. Therefore, it is possible that the existence of the HhH domain controls the cleavage activity and thus human MUS81-EME1 shows a divergence to before published nuclease specificity. A lacking C-terminal HhH domain can reduce the interaction of MUS81 with EME1, which in turn is needed for cleavage activity. Furthermore, an additional effect on the activity could also be lacking post-translational modification of the hsMUS81-EME1 complex. From studies in yeast it is known that the activity of the Holliday Junction resolvases is controlled in a cell cycle dependent manner [117, 118]. The yeast ortholog of EME1, Mms4, is being phosphorylated by Cdc5 leading to the disentanglement of joint molecules by the Mus81-Mms4 nuclease [117, 118]. The phosphorylation dependent regulation is carried out at the G_2/M transition to eliminate all chromosomal interaction for chromosome separation [118]. For human MUS81-EME1 the cleavage of HJs is cell-cycle regulated and increases upon SLX1-SLX4-binding [117, 141, 202], but the cleavage of RFs is not inhibited prior to mitosis nor globally stimulated by SLX1-SLX4 interaction [200]. Due to expression of the human MUS81-EME1 complex in *E. coli* phosphorylation is not introduced. In other publications the human MUS81-EME1 complex was purified from insect cell expression and is therefore phosphorylated. Thus, lacking post-translational modification could be another reason for the difference in cleavage activity. So far, it was proposed, that the SAP domain of SLX4 might be important for the cleavage activity and especially on substrate relaxation of MUS81-EME1 [143]. Under the tested conditions with *C. elegans* protein an influence of the SAP domain was not detected.

Here, I showed for the first time the cleavage activity of MUS81-EME1 complexes in presence of the two cations Mn^{2+} and Mg^{2+} . Until now, cleavage assays of MUS81-EME1 complexes have not been reported in the presence of Mn^{2+} . The expected activity towards 3'flap, RF and nHJ was only reached for the ceMUS81-EME1 complex with Mn^{2+} as cofactor. In general, the properties of Mn^{2+} are quite similar to Mg^{2+} . However, Mn^{2+} leads to relaxation in substrate specificity, which has been shown for other proteins, e.g. helicase / nuclease Hef [211], translesion DNA polymerase Dpo4 [299] and RNase H [300]. In the case of the here presented MUS81-EME1 complexes, Mn^{2+} is needed for proper activity. Future studies, could use human MUS81-EME1 complex purified from homologous source in order to analyze, which bivalent cation is bound after purification. Furthermore, stability studies could be carried out in presence of either of the divalent cation to find the one cation, which increases protein stability. So far, it has not been studied whether Mg^{2+} or Mn^{2+} is the natural ligand for MUS81-EME1. Both published crystal structures contained Mg^{2+} either in the crystallization buffer [213] or for soaking [214]. Reported biochemical assays were carried out solely with Mg^{2+} as cofactor for the cleavage reaction. The

cleavage assays published so far used only magnesium as co-factor in cleavage buffer. In the here presented cleavage tests the MUS81-EME1 complex displayed a strong difference in activity between the cations. Such a difference in activity between Mg^{2+} and Mn^{2+} has been observed for RNaseH [301]. The replacement of Mg^{2+} with Mn^{2+} does not only relax the coordination geometry but also activates the D192N mutant of RNase H and reduces substrate specificity of retroviral RNases H in order to cleave dsRNA [301]. Mn^{2+} in the D192N mutant compensates the distance shortening between the two-metal ions in D192N RNaseH as Mn^{2+} shifts the first metal binding site a bit further away from the second binding site and thus leads to a cleavage reaction [301]. Another study about RNase H proposed the binding of one Mn^{2+} in the active site, which is required for the activation of the protein [302]. Earlier it was also shown, that the binding affinity towards Mn^{2+} is higher than towards Mg^{2+} [303]. Thus, a similar mechanism could also play a role for the MUS81-EME1 complex. Instead of Mg^{2+} the actual cation is Mn^{2+} and the cleavage activity is enhanced. However, more investigations have to be carried out in order to determine the divalent cation involved in the active site. It would also be interesting, whether the activity of active site mutant changes upon interaction with Mn^{2+} , which was seen for RNase HI [301].

3.4 Interaction between ceMUS81^N and ceSLX4^{SAP}

Earlier studies on SLX4 and MUS81 showed by yeast-two-hybrid assay and co-immunoprecipitation that SLX4^{SAP} and MUS81^N interact with each other [201, 202]. So far, the complex between the two domains has not been reconstituted *in vitro* and furthermore it has not been quantified until now. Here, the interaction between the two domains from *C. elegans* was tested *in vitro* by direct measurements, size-exclusion chromatography, SLS and AUC, and by an indirect method, thermofluor. ceMUS81^N and ceSLX4^{SAP} interaction could not be detected by direct measurements including size exclusion chromatography (analytical SEC and SLS). In order to simulate the protein modification of homologous expressed protein, phosphorylation of the proteins was attempted but in a gel filtration assay the interaction of the two domains was not detected. It could be possible that the phosphorylation by PLK1 did not give the correct phosphorylation pattern for the interaction to happen. Sedimentation velocity experiments did not give clear data about the complex formation, either. Finally, a thermal shift assay could show and quantify the interaction to 152 μ M (Figure 2.25). Other quantitative methods, e.g. isothermal titration calorimetry and microscale thermophoresis, were applied as well. However, due to various technical problems during the experimental procedure, they did not give a quantitative reproducible measure in the case of these two protein constructs. The direct measurements could not show interaction between the two domains, as either no shift was seen in gel filtration or the single domains showed much higher masses than expected. This findings indicated that the affinity between the domains is weak and the complex falls apart during gel filtration.

The low affinity of the two proteins measured *in vitro* could also be an effect of missing post-translational modification. It remains to be tested whether phosphorylation with a kinase other than PLK1, e.g. CDK1, influences interaction *in vitro*. A recent study showed by coimmunoprecipitation, that CDK1 phosphorylation at the SAP domain is important for interaction with MUS81 [200]. The dissociation constant of ceSLX4^{SAP} and ceMUS81^N gives a weak binding between the two domains. It also means that at a concentration of 152 μ M half of the proteins are present as a

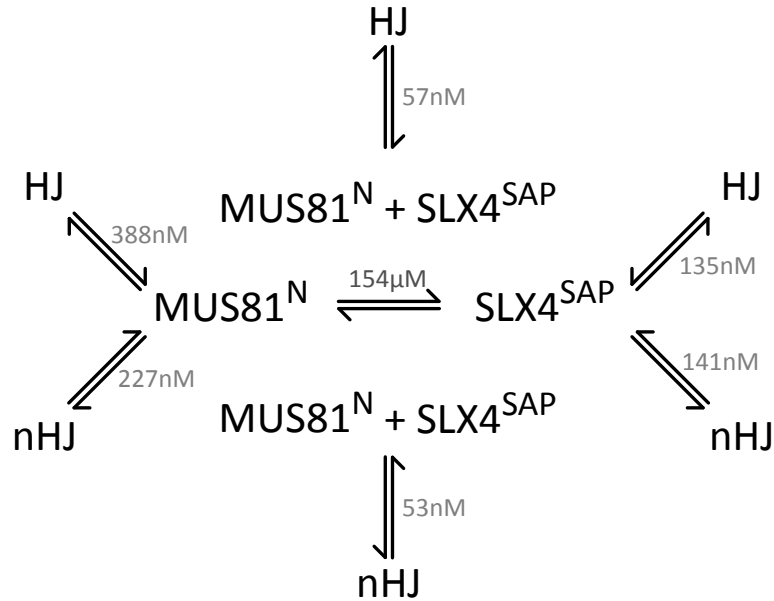


Figure 3.1: Thermodynamic cycle of DNA binding for ceSLX4^{SAP}, ceMUS81^N and ceSLX4^{SAP}-ceMUS81^N complex. Interaction Scheme of ceMUS81^N and ceSLX4^{SAP} with each other, alone with nHJ and HJ as well as both together with HJ and nHJ. The binding affinities with DNA were calculated from fluorescence anisotropy measurement (Figure 2.31, the binding affinity between was calculated from Thermofluor measurements (Figure 2.25))

complex. This kind of high protein concentration might be biologically only reached in the nucleus in a very confined region. MUS81 and SLX4 are both located in the nucleus and form foci after induction of DNA damage [114, 209, 304]. Deep proteome mapping of HeLa cells identified the copy number per cell for MUS81, EME1 and SLX4 at 2700, 1700 and 220, respectively [305]. Given that HeLa cells have an average nucleus size of 690 fl (690 μm³) [306], the average concentrations of MUS81, EME1 and SLX4 are 6.49 nM, 4.1 nM and 528.95 pM, respectively. The concentration of MUS81 in the cell is thus more than 20,000-fold lower than the measured K_D of the interaction between ceMUS81^N and ceSLX4^{SAP} measured *in vitro*. Thus, there must be the possibility in the cell to lower the high binding constant by additional measures, e.g. post-translational modification or phase separation. The weak binding between the complexes in absence of DNA can be explained by the fact that the two proteins might not strongly interact without the presence of DNA damage, without DNA or additional binding factors in general. Upon DNA damage SLX4 will be recruited to the DNA lesion and then recruit the nucleases [14, 158]. EME1 in the MUS81 complex will be phosphorylated and by this activates MUS81. Depending on a certain phosphorylation MUS81 could undergo a conformational change and interact with SLX4. The before mentioned studies showing the interaction between the two proteins used protein from one step purification coming from expression of stable transfected cell lines.

3.5 Biological Implications for the DNA binding of ceSLX4 and ceMUS81

SLX4 contains a DNA binding domain, the SAP domain, which has been found in other proteins being involved in DNA metabolism. Typical examples are chromatin-associated proteins, such as the scaffold attachment factors SAF-A and SAF-B, PIAS family members, DNA repair proteins, like the non-homologous end joining protein Ku70 [168]. The typical characterization of the SAP, two amphipatic helices stabilized by hydrophobic residues in their interface and polar residues facing to the solvent. The two helices are usually separated by a conserved glycine in the loop region and could also be found in ceSLX4^{SAP}. Direct evidence that the SAP motif binds DNA has come from studies of SAF-A and Ku70 [169, 170]. The SAP domain can either be used as the main DNA binding component of a protein (e.g. SAF-A) [169] or as an additional stabilizing element (e.g. Ku70 and SpCCE1) [127, 171]. A single SAP domain forms a relatively weak interaction with DNA, which can be undetectable in solution in the case of SAF-Box [274]. When many SAP domains are brought into close proximity, cooperative effects lead to high affinity DNA binding as in the case of SAF-A [274].

Characteristically, the SAP domain contains many positively charged residues, which undergo ionic and hydrogen bonding with the phosphate backbone of the DNA backbone. DNA binding of ceSLX4^{SAP} was determined by band shift assays and fluorescence anisotropy, both measurements revealed K_{DS} in the range from 140 to 200 nM. Band shift assays of ceSLX4^{SAP} showed a cooperative behavior with higher level shifts depending on the branching of the substrate. As SAP domains show cooperative binding [274], it is possible that each arm of the HJ is occupied by one SAP molecule (Figure 3.3). At low protein concentrations one SAP binds to one arm of the junction (Figure 3.3). With increasing protein concentration more HJs are occupied by one SAP, but at the same time the ratio of HJs attracting a second or even third SAP is increasing. If non-cooperative binding was assumed then all HJs would be occupied first by one SAP and at increasing protein concentrations a second SAP would bind to the arms. In an EMSA, before the second shift would appear all free DNA would be bound by one SAP and with increasing protein concentrations a second shift would be visible.

In the case of the binding to nHJ, the band shift assay shows hardly any nHJ-complexes with four SAP. A possible reason for this could be that the SAP does not interact strongly with the arm containing the nick owing to the higher flexibility of the arm. In the band shift assays with lower protein concentrations of ceSLX4^{SAP} several shifts owing to the formation of different complexes were visible (Figure 2.27). This pattern of shift positions is also seen for T4 endonuclease VII [148]. Band shifts assays with HJ substrate showed up to four shifts with increasing amounts of protein. Golz et al. assumed that each shift represents a different complex between T4 endonuclease VII and HJ with all binding sites being saturated in the highest shift position. Stepwise loading reflects one physical unit (monomer, dimer and multimer) of the protein binding first and attract three further units to the same DNA molecule. For ceSLX4^{SAP} the molecular mechanism and the stoichiometry of the recognition reaction between SAP and HJ could be similar. A possible way to further examine the stoichiometry is the introduction of mutations for amino acids, which are involved in the SAP-SAP interface. Mutation of these sites could lead reduction

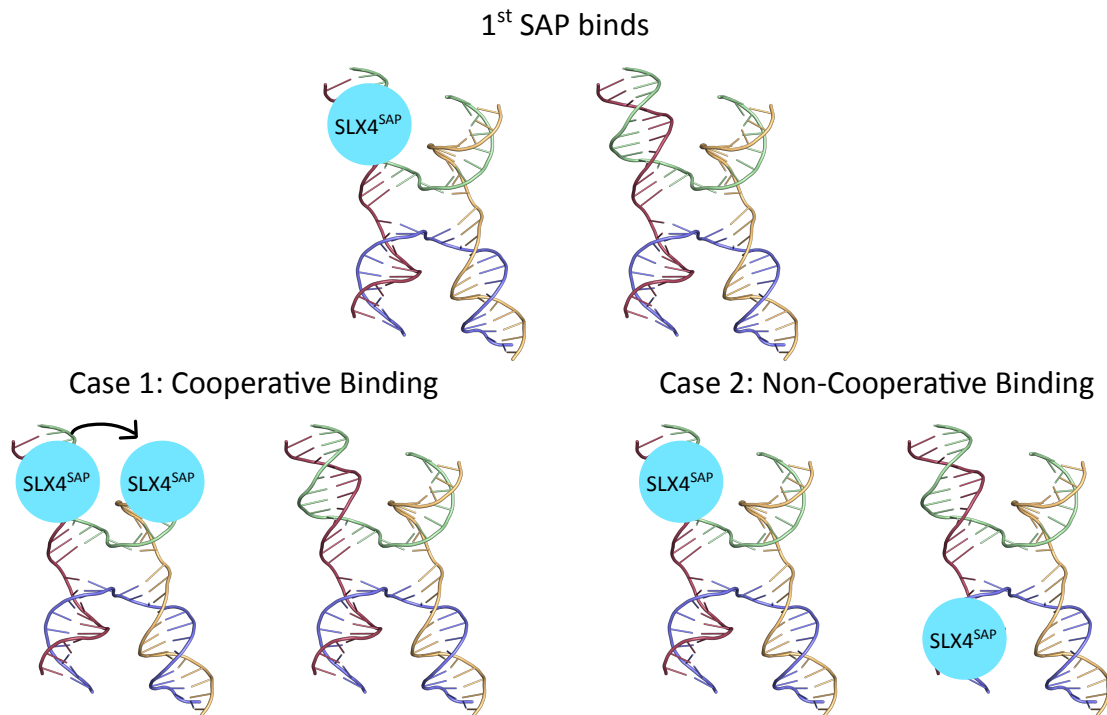


Figure 3.2: Model for cooperative binding of ceSLX4^{SAP}. After the binding of the first SAP domain to the Holliday Junction, there are two possible scenarios for the binding of a second SAP domain. First, assuming cooperative binding, the second SAP domain will interact cooperatively with the already bound SAP domain and bind to a neighboring arm of the same HJ. Second, assuming non-cooperative binding, the second SAP domain will interact with any arm of another HJ.

or abolishment of the several shifts.

Another hint towards cooperative binding comes from the comparison with the SAP domain of SAF-A. SAF-A^{SAP} shows only strong DNA binding with protein immobilized either on a membrane or on Sepharose beads [169]. Testing the DNA binding in solution leads to only weak DNA binding. The same is true for ceSLX4^{SAP}, DNA interaction in solution during gel filtration could not be detected. A possible explanation for this effect could be that protein-protein interactions are necessary for a specific binding. During gel filtration experiments the proteins are too much diluted and lose contact with each other. However, as there is no structural information of ceSLX4^{SAP} in complex with HJ available, it is very speculative if, how and where SAP would interact with each other.

The cooperative binding towards DNA seen with ceSLX4^{SAP} alone and in complex with ceMUS81^N could result either from alterations in DNA structure or can be mediated by direct protein-protein interactions. For the first case, DNA structure can change after protein binding to one site, which in turn facilitates binding to the others sites. From biophysical and crystallization studies of the HJ two main conformations of the HJ are known, open square-planar or stacked-X conformation [307, 308]. The square planar shape is adopted in the absence of cations, when the repulsion of the unshielded negatively charged phosphates directs the four arms of the junction into a corner of a square [307]. The stacked X shape of the HJ exists in presence of divalent metal ions and the arms undergo pairwise coaxial stacking [307, 308]. Yeast SpCCE and *S. pombe* YDC2

bind the HJ in an open square shape. Whereas, the X-shape of the HJ has been seen in complex with *E. coli* RuvC or *Sulfolobus sulfataricus* Hjc [307]. Interestingly, in the crystal structure of T4 endonuclease VII the HJ is a hybrid of the square-planar and stacked X-conformation of the HJ [147]. T4 endonuclease VII protrudes into the junction point and opens it to a parallelogram [147]. Taking the different HJ conformations into account, it can be possible that either ceSLX4^{SAP} and ceMUS81^N alone or in complex stabilize a certain conformation of the HJ in order to facilitate the cleavage mechanism.

Secondly, the cooperative binding could arise from direct protein-protein interactions. Then the protein bound to one arm of the junction could interact with a protein bound to the other arm which requires the DNA to undergo conformational changes. To distinguish between these possibilities, further experimental evidence is needed. To check for alterations in DNA structure after protein binding, further bands shift assays employing Holliday Junctions with more diverse arm lengths, different symmetries (2-fold or no symmetry instead of 4-fold symmetry) and more nicks could lead to greater understanding in protein-DNA complex formation on the Holliday Junction. A change in DNA conformation induced by protein binding could also be detected by DNase I footprinting as DNase I cuts preferentially sequences in widened minor grooves [309]. Thus, from the available experimental data it can be assumed here that for HJ substrate up to four ceSLX4^{SAP} can bind it, one per arm. At very high protein concentrations, the formation of higher order complexes is more prevalent, due to possible non-specific interactions.

Cooperative binding of ceSLX4^{SAP} started at concentrations of 100 nM in the refined band shift assay with HJ (Figure 2.27 A). But the SLX4 concentration in the nucleus could be 0.5 nM estimated from the proteomic analysis of HeLa cells [305]. Upon DNA damage induction SLX4 was seen to form foci [114, 304]. In these foci, SLX4 concentration could locally exceed the theoretical concentration of 0.5 nM. DNA lesions could lead through the DNA damage signaling to an increase of SLX4, which in turn attracts more SLX4 proteins due to the cooperative behavior of the SAP domain. Furthermore, SLX4 contains a dimerization domain [157], which could further enhance the cooperative binding of two SAP domains. The dimerization and the cooperative binding mode could help to immobilize the HJ for the activity of the nucleases, which interact with SLX4. It remains speculative, whether the SAP domain has also a control function towards the start of the nuclease recruitment. Only after reaching a certain protein concentration threshold the SAP domain binds DNA and in complex with associated nuclease complex formation with DNA happens already at lower concentrations. In general, SLX4 is involved in many different cellular pathways, thus cooperative binding to HJ could be of considerable importance to different ways of controlling and bringing it into activity at certain cell stages.

The SAP domain seems to work as an auxiliary DNA binding domain in general. In T4 endonuclease VII the SAP domain is located between the arms of the HJ and positive charged residues undergo hydrogen bonding with the HJ arm [147]. In yeast SpCCE1, another HJ resolvase, the SAP domain is required for the stabilization of the interaction with the HJ [127]. As SpCCE1 undergoes branch migration the stable junction binding is necessary to move the resolvase to the cleavable sites [127]. The promiscuity of the SAP domain could be seen as a versatile tool by SLX4 to be able to scan the genome for damage while piggybacking the nucleases. Upon the encounter of a cleavable substrate the affinity of SLX4^{SAP}, maybe also directed by MUS81^N, increases and the two nicks in the HJ are introduced.

Fluorescence anisotropy data of ceMUS81^N (residues 1 - 157) showed a two-site binding mode. However, the N-terminus of ceMUS81 shows in the alignment with MUS81 orthologs from other species only one structured domain, instead of two domains. Thus, it is unlikely that the used ceMUS81 constructs contains two DNA binding sites. The band shift assays especially with dsDNA and ssDNA show that saturation is not perceived at the tested conditions. First, this could suggest negative cooperativity or multiple classes of binding sites (also discussed in [310]). Negative cooperativity can be excluded as the curve shape of the anisotropy measurement and the band shift assays point clearly towards positive cooperativity. The presence of two binding sites can neither be excluded nor included. Additional experiments would be needed in order to find out about a second binding site on this construct, e.g. cross-linking experiments. However, it is more likely that the cooperative binding of ceMUS81^N is overlaid by non-specific binding at high protein concentrations, which can be described by a linear function. As suggested by Mendel & Mendel [310], in such a case of non-specific binding, the total binding should be measured and the binding data is separated into two components, one of which represents the non-specific binding. Thus, at low ceMUS81^N concentrations the prevalent binding mode is mostly represented by cooperative binding to the DNA. However, it is already influenced by non-specific binding. At protein concentrations higher than 120 nM, the binding isotherm is mainly related to non-specific binding. Biologically, this case will not occur as the MUS81 concentration is unlikely to jump to a concentration of up to 10 μ M in the nucleus. Anisotropy was measured at pH 6.5 resulting in a positively charged protein (pI 9.28). The excess protein undergoes electrostatic interactions at protein concentration higher than 120 nM. Taken together, the non-specific interaction is an artifact of the measurement with no biological relevance. As the non-specific binding is influencing the DNA binding, the calculated K_d values are an estimate for ceMUS81^N.

In previous studies on human MUS81, fluorescence anisotropy measurements of the N-terminal HhH domain (residues 10 - 90) [143] and the WH domain (residues 128 - 230) [219] were carried out as well. The WH domain binds dsDNA with a K_D of 6.6 μ M showing a plateau in the binding reaction at 60 μ M [219]. The DNA binding of the HhH domain was tested in presence of flap structures, single and double-stranded DNA [143]. Wyatt et al. could not reach a plateau anisotropy value for dsDNA, but could detect a preference for ssDNA in comparison to dsDNA and additionally, binding towards flap structures, all in low nM range. As the ceMUS81 N-terminus does likely not contain the WH domain, the DNA binding property can be attributed to the HhH domain. The ceMUS81^N construct shows a similar behavior towards dsDNA in anisotropy as the human HhH domain.

ceMUS81^N favors the binding towards nHJ over dsDNA shown by the lower dissociation constants in band shift assays and fluorescence anisotropy. In contrast to hsMUS81^{HhH} [143], ceMUS81^N band shift assays show no strong difference between binding to ssDNA and dsDNA. The band shift assays show like the fluorescence anisotropy measurements non-specific binding towards the tested DNA substrates, which can be seen from unbound substrate even at high protein concentrations. The weakest binding is visible towards ssDNA with only 54% DNA bound at 5 μ M protein. In similar K_D range of 800 to 815 nM was the binding towards dsDNA (Figure 2.29). The binding constant towards RF, flaps and HJ was between 500 and 600 nM. The strongest binding of the MUS81^N was visible with nHJ, showing a clear preference for binding towards nHJ over other, especially non-branched DNA substrates. For all tested substrates it can be seen that at excess

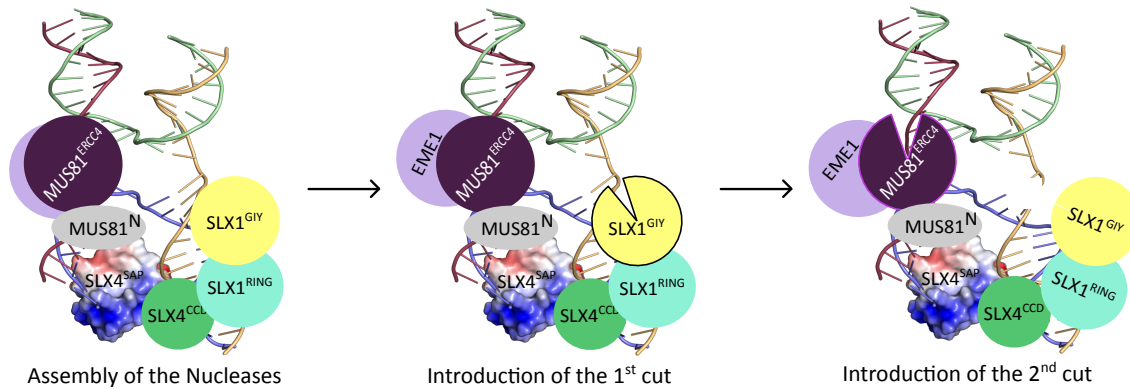


Figure 3.3: Model for HJ cleavage. Through the interaction of SLX4^{SAP} (displayed with electrostatic surface potential) with the HJ substrate, the interacting nucleases might be placed and coordinated by SLX4 at the substrate. SLX1 will then introduce the first cut at the HJ substrate. This leads to a conformational change in MUS81 and MUS81 will introduce the second cut at the HJ.

of protein there is still unbound substrate available. Although the concentration of protein was 125times higher than the substrate concentration (5000 nM - 40 nM), substrate is not fully bound at the highest concentration of ceMUS81^N. Taken together, the shape of the anisotropy curve for ceMUS81^N shows positive cooperativity, which is overlaid by non-specific binding. therefore, the binding isotherm gives rise to a more linear shape after the first inflection point.

3.6 A Modified HJ Resolution Model

The following section will give an updated model about the resolution of Holliday Junctions by an SM-complex in the context of the full length proteins based on the NMR model of ceSLX4^{SAP}, the here presented biochemical analysis of the single domains as well as published results.

SLX4^{SAP} and MUS81^N are alone and in complex promiscuous DNA binding domains. Based on this promiscuity it can be possible that the proteins scan alone or in form of the SM-complex the genome until they encounter a suitable substrate, e.g. a Holliday Junction.

In a first step the SAP domain of SLX4 would then bind to the HJ substrate in a cooperative manner (Figure 3.3). This binding could occur similar to T4 endonuclease VII, where the SAP domains is located between the arms of the open-planar HJ and stabilizes the interaction [147]. SLX4 brings the MUS81-EME1 nuclease through interaction with the N-terminal domain MUS81 into position. Upon DNA-binding of SAP at the same time a conformational change in the CCD domain of SLX4 could bring the associated SLX1 nuclease into position. It has been proposed that SLX1 forms an inactive homodimer and upon formation of the SLX1-SLX4 complex the SLX1 nuclease becomes active [156]. Another possibility is SLX1 only exists in a complex with SLX4 and gets activated by post-translational modifications similar to the interaction between SLX4 and MUS81 [200]. After the activation, SLX1 nuclease recognizes the HJ substrate and introduces a nick in the HJ. The performance of the cleavage reaction puts SLX1 in a post-nick state and might lead to a conformational change in SLX1. SLX1 in post-nick state could subsequently trigger a movement within SLX4, resulting in the activation of MUS81. In this relay the N-terminus of

MUS81 could be the link between MUS81 activity after the first incision. MUS81 recognizes the substrate and introduces the second cut on the HJ. The motions across the C-terminal part of SLX4 and N-terminus of MUS81 could relay the catalytic residues in MUS81 and then promote the positioning of the catalytic residues. Furthermore, the catalysis could be assisted by changes in the HJ structure itself after the initial incision. The introduction of the first nick transforms the relatively stiff HJ into a more bendable DNA substrate. This substrate is better being recognized by the MUS81-EME1 complex through its 5' end binding pocket [214]. In human MUS81 are two N-terminal domains involved in DNA binding, the HhH and the WH domain. The latter domain is not present in *C. elegans* MUS81. In the end, SLX4 has to coordinate the active site of two nucleases in order to introduce the two nicks within the lifetime of the complex. Additionally to the nucleases there are three more domains in the human complex, which interact with the DNA substrate, SLX4^{SAP}, MUS81^{HhH} and MUS81^{WH}. In total the remarkable number five DNA-interacting sites have to be coordinated within the SM-complex. This also includes the proper positioning of the active sites on the substrate. Each DNA-binding domain has to recognize the DNA and through several conformational changes in the whole complex the nuclease domains have to recognize their specific substrate and cleave in a spatial and timely coordinated manner.

This presented model can also be extended as SLX4 can dimerize via its BTB domain [157, 164]. Together with the dimerization multiple MUS81 molecules are bridged by SLX4, but it is not known whether dimerization is cell-cycle regulated or required for the bridging of MUS81 molecules [164, 200]. There is the possibility that the complex is activated through the proposed SUMO-E3 ligase activity of SLX4 [164]. SLX4 has been shown to SUMOylate XPF and itself, which is proposed to be important for activation of the endonuclease partners [164]. The nuclease XPF-ERCC1 works together with the complex of SLX1-SLX4-MUS81-EME1 by enhancing the cleavage activity of SLX1 and MUS81 [143]. In the context of a double HJ, which occurs at the end of homologous recombination, the dimerization of SLX4 and its associated endonucleases could be important for the coordinated introduction of four nicks on two HJs.

The outlined model explains a possible cleavage mode for HJs, but there are more DNA intermediate structures. During recombination other DNA substrates arise as intermediates and SLX4 has been shown to be involved in other repair functions, e.g. in ICL repair. ICL repair is initiated by signaling by the FA core complex and monoubiquitylation of FANCD2, which lead to recruitment of the nuclease XPF-ERCC1 by SLX4 via a so far not determined mechanism [108]. It is proposed the SLX4 does not only bring XPF but also other nucleases to the lesion [108], which suggests that the nucleases also recognize other substrates in the context of the SMX-complex. After the cleavage reaction of XPF-ERCC1 translesion synthesis and homologous recombination are promoted. In the course of homologous recombination SLX4 might be involved again with its nuclease set.

3.7 Open Questions and Outlook

In conclusion, the presented study describes for the first time a model for the structure of SLX4^{SAP} and shows that this domain is a promiscuous DNA binding domain. Interaction of SLX4 with DNA is central for its function to position the nucleases at the junction. Hydrophobic residues between the helices of the SAP fold are important for structural integrity. The positive patch on

ceSLX4^{SAP} is essential for DNA binding. The negatively charged patch could be an interaction site for the N-terminus of MUS81. The here presented studies on SLX4-SLX1 and MUS81-EME1 raise further structural and biochemical questions which are being addressed in future work. In the four-protein complex there are three DNA-binding sites and two active sites. Human SLX4 has been shown in literature to alter the substrate specificity of MUS81-EME1 [143], giving rise to the question on the conformational changes and repositioning of nucleases involved in complex formation. So far, it is not yet clear, how the four proteins will be positioned concretely at the HJ. There is also the question why the cells have evolved this multi-nuclease complex. Its activity has to be tightly regulated by post-translational modification and is constrained to a late phase of the cell cycle [141, 143, 200]. The temporal regulation prevents unrestrained cleavage leading to chromosome aberrations. NMR experiments showed how flexible the truncated versions of ceSLX4^{SAP} and ceMUS81^N still were. Based on this experiments much shorter SAP constructs were designed and tested for DNA binding. Now, optimized constructs for SLX4^{SAP} and MUS81^N are a good starting point for further structural studies. NMR experiments are a helpful tool for designing protein constructs which can not be identified by limited proteolysis anymore. Future structural studies on SLX4 with its nucleases and in complex with Holliday Junctions could give an deeper insight on the organization of these proteins at the Holliday Junction. As SLX4 contains a dimerization domain, it would be further interesting, how a dimer of SLX4 with associated nucleases is being coordinated and regulated especially in the context of double Holliday Junctions and how these enzymes facilitate different DNA repair functions.

Furthermore, the influence of SAP on the cleavage reaction is not yet fully investigated. Kinetic measurements could shed light to the extent of how SAP is making an impact on the cleavage of SLX1 and MUS81. So far, it has been investigated that mutations in the SAP domain lead to decreased survival in presence of DNA damaging agents and reduce interaction with MUS81 [202].

4. Materials and Methods

4.1 Materials

4.1.1 Consumables and Chemicals

Unless otherwise stated all chemicals were purchased from Roche Diagnostics (Mannheim, Germany), Merck (Darmstadt, Germany), Qiagen (Hilden, Germany), Roth (Karlsruhe, Germany), Sigma-Aldrich (München, Germany) and Serva (Heidelberg, Germany). Enzymes were ordered from New England Biolabs (NEB) (Frankfurt/Main, Germany), ThermoFisher Scientific (Waltham, MA, USA) or prepared by Biochemistry Core Facility of Max-Planck Institute of Biochemistry (Martinsried, Germany). The enzymes used for tag cleavage (Senp2, 3C, Senp2 mod) were prepared by Maren Klügel and Marcus Hammerl. Nucleotids were ordered from Eurofins/MWG (Ebersberg, Germany). For expression in HEK cells, the FreeStyle™ expression medium was ordered from ThermoFisher Scientific (Waltham, MA, USA) and the transfectant polyethyleneimide (PEI) was ordered from Polysciences (Warrington, PA, USA). Gel and plasmid extraction kits were purchased from Macherey-Nagel (Düren, Germany). DNA oligonucleotids were synthesized by Eurofins Genomics (Ebersberg, Germany). Chromatography material and columns were ordered from GE Healthcare (München, Germany), Macherey-Nagel (Düren, Germany), Qiagen (Hilden, Germany), IBA (Göttingen, Germany) and Roche (Basel, Switzerland). Crystallization Screens and cryo-loops were obtained from Hampton Research (Aliso Viejo, CA, USA), and Qiagen (Hilden, Germany).

4.1.2 Media

Table 4.1: Media for bacterial and mammalian cell expression.

Medium	Composition	Cell Type
LB	1% (w/v) Bacto Trypton, 0.5% (w/v) Yeast extract, 1% (w/v) NaCl, pH 7.2	<i>E. coli</i>
SOC	2% (w/v) Bacto Trypton, 0.5% (w/v) Yeast extract, 100 mM NaCl, 1 mM MgCl ₂ , 2.5 mM KCl, 10 mM MgSO ₄ , 0.4% glucose, pH 7.2	<i>E. coli</i>
TB	1.2% (w/v) Bacto Trypton, 2.4% (w/v) Yeast extract, 0.004% (v/v) glycerol, supplemented with phosphate buffer (10%(v/v), 0.17 M KH ₂ PO ₄ , 0.72M K ₂ HPO ₄	<i>E. coli</i>
M9	33.7 mM Na ₂ HPO ₄ , 22 mM KH ₂ PO ₄ , 8.55 mM NaCl, 9.35 mM ¹⁵ NH ₄ Cl, supplemented with 0.4% glucose or 13C-glucose, 1 mM MgSO ₄ , 0.3 mM CaCl ₂ , 1 µg biotin, 1 µg thiamin, 1x trace element solution	<i>E. coli</i>

Medium	Composition	Cell Type
FreeStyle	animal origin-free, chemically-defined, protein-free, supplemented with 1% FCS	Mammalian Cells

4.1.3 Bacterial Strains and Cell Line

Table 4.2: Bacterial strains

Bacterial Strain	Species	Genotype
XL1 blue	<i>E. coli</i>	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac</i> [F' <i>proAB lacI^qZΔM15 Tn10</i> (Tet ^r)]
DH5α	<i>E. coli</i>	F ⁻ <i>θ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17</i> (rK ⁻ mK ⁺) <i>phoA supE44 λ⁻ thi-1 gyrA96 relA1</i>
BL21 DE3 gold	<i>E. coli</i>	B F ⁻ <i>opmT hsdS</i> (r _B ⁻ m _B ⁻) <i>dcm⁺ Tet^r gal λ(DE3) endA Hte</i>
BL21 DE3 gold pRIL	<i>E. coli</i>	B F ⁻ <i>opmT hsdS</i> (r _B ⁻ m _B ⁻) <i>dcm⁺ Tet^r gal λ(DE3) endA Hte</i> [pRIL Cam ^r]
BL21 DE3 gold pRARE	<i>E. coli</i>	B F ⁻ <i>opmT hsdS</i> (r _B ⁻ m _B ⁻) <i>dcm⁺ Tet^r gal λ(DE3) endA Hte</i> [pRARE Cam ^r]
B834 (DE3) pLysS	<i>E. coli</i>	F ⁻ <i>opmT hsdS</i> (r _B ⁻ m _B ⁻) <i>gal dcm met⁻ λ(DE3) pLysS</i> (Cam ^R)

Table 4.3: Cell line

Cell line	Description
HEK293S	Human embryonic kidney cells (<i>GntI⁻</i>) cells containing Adenovirus 5 DNA, used for transient expression of recombinant protein with PEI, cultivated in suspension

4.1.4 Antibiotic Solutions

Table 4.4: Antibiotic stock solutions

Antibiotic	Stock Solution	End Concentration
Ampicillin	100 mg/ml	0.1 mg/ml
Kanamycin	50 mg/ml	50 μg/ml
Chloramphenicol	34 mg/ml	34 μg/ml
Tetracycline	10 mg/ml	0.01 mg/ml

4.1.5 Vectors and Plasmids

Table 4.5: Plasmids used for expression. bax vectors were used for expression in *E. coli* and hex vectors for expression in HEK293S cell culture. GFP (green fluorescent protein) and mVenus both contained the A206K mutation to ensure monomeric state of the fluorescent dyes.

Name	Application	Tag	Tag cleavage
pCB-bax04	Bacterial expression of proteins	N-terminal 8x His-SUMO-Tag	Senp2
pCB-bax10	Bacterial expression of proteins	C-terminal CPD-8xHis-Tag	CPD

Name	Application	Tag	Tag cleavage
pCB-bax25	Bacterial expression of proteins	N-terminal 8x His-SUMO-Tag, C-terminal CPD-Twinstrep-tag	Senp2, CPD
pCB-hex12	Mammalian expression of proteins	N-terminal-Kozak sequence-8xHis-eGFP(A206K)-SUMO (R63T, R70E)	Senp2
pCB-hex28	Mammalian expression of proteins	N-terminal hTee-658, C-terminal 3C-mVenus(A206K)-SSGS-linker-rho1d4	3C protease
pCB-hex28L21	Mammalian expression of proteins	N-terminal L21 leader-Kozak-sequence, C-terminal-3C-mVenus(A206K)-SSGS-linker-rho1d4	3C protease
pCB-hex29L21	Mammalian expression of proteins	N-terminal-hTee-658-L21leader-Kozak sequence-eGFP(A206K)-3C	3C protease

Table 4.6: Cloned constructs. The insertion site LIC-SUMO contains the SUMO tag always at the N-terminus. The insertion site LIC-CPD contains the CPD tag at the C-terminus. LIC ligation independent cloning, SUMO small ubiquitin like modifier, CDP cysteine protease domain.

No.	Construct	Vector	Primer	Insertion site
1a	hsEME1syn 1-570-RBS-MUS81syn 1-551	pCB-bax04	CBp1243, CBp1204	LIC SUMO
1b	hsEME1syn 1-570-RBS-MUS81syn 1-551	pCB-bax25	CBp1243, CBp1331	LIC SUMO CPD
2a	hsEME1syn 178-570-RBS-MUS81syn 241-551	pCB-bax04	CBp1213, CBp1204	LIC SUMO
2b	hsEME1syn 178-570-RBS-MUS81syn 241-551	pCB-bax25	CBp1213, CBp1331	LIC SUMO CPD
3	hsEME1syn 246-570-RBS-MUS81syn 241-551	pCB-bax04	CBp1206, CBp1204	LIC SUMO
4a	hsEME1syn 210-570-RBS-MUS81syn 133-551	pCB-bax04	CBp1205, CBp1204	LIC SUMO
4b	hsEME1syn 210-570-RBS-MUS81syn 133-551	pCB-bax25	CBp1205, CBp1331	LIC SUMO CPD
5a	hsEME1syn 233-570-RBS-MUS81syn 256-551	pCB-bax04	CBp1214, CBp1204	LIC SUMO
5b	hsEME1syn 233-570-RBS-MUS81syn 256-551	pCB-bax25	CBp1214, CBp1331	LIC SUMO CPD
6a	hsMUS81syn 2-225	pCB-bax04	CBp1333, CBp1335	LIC SUMO
6b	hsMUS81syn 1-225	pCB-bax10	CBp1334, CBp1336	LIC CPD
7a	hsSLX4 1-1834	pCB-bax04	CBp083, CBp084	LIC SUMO
7b	hsSLX4 1-1834	pCB-bax10	Cbp495, CBp498	LIC SUMO
8	hsSLX4 1309-1824	pCB-bax10	CBp670, CBp676	LIC CPD
9	hsSLX4 1492-1824	pCB-bax10	CBp671, CBp676	LIC CPD
10	hsSLX4 1516-1824	pCB-bax10	CBp672, CBp676	LIC CPD
11	hsSLX4 1647-1824	pCB-bax10	CBp673, CBp676	LIC CPD
12	hsSLX4 1706-1824	pCB-bax10	CBp674, CBp676	LIC CPD
13	hsSLX4 1748-1824	pCB-bax10	CBp1171, CBp676	LIC CPD
14a	hsSLX4 1492-1648	pCB-bax04	CBp1373, CBp1375	LIC SUMO
14b	hsSLX4 1492-1648	pCB-bax10	CBp671, CBp675	LIC CPD
15a	hsSLX4 1516-1648	pCB-bax04	CBp1374, CBp1375	LIC SUMO
15b	hsSLX4 1516-1648	pCB-bax10	CBp672, CBp675	LIC CPD
16	hsSLX4 1565-1648	pCB-bax04	CBp150, CBp675	LIC SUMO
17a	hsSLX1 1-275	pCB-bax04	CBp085, CBp086	LIC SUMO

No.	Construct	Vector	Primer	Insertion site
17b	hsSLX1 1-275	pCB-bax10	CBp501, CBp796	LIC CPD
18	hsSLX1 11-258 + hsSLX4 1300-1824	pCB-bax10, pCB-bax10	CBp490, CBP491, CBp497, CBp676	LIC CPD
19	hsSLX1 11-258 + hsSLX4 1706-1824	pCB-bax10, pCB-bax10	CBp490, CBP491, CB674, CBp676	LIC CPD
20	hsSLX4 1-1834	pCB-hex12	CBp083, CBp084	LIC SUMO
21	hsSLX4 667-1834	pCB-hex12	CBp354, CBp084	LIC SUMO
22	hsSLX4 1300-1934	pCB-hex12	CBp149, CBp084	LIC SUMO
23a	hsSLX1 1-275	pCB-hex12	CBp085, CBp086	LIC SUMO
23b	hsSLX1 1-275	pCB-hex29	CBp085, CBp086	LIC SUMO
24	hsSLX1 11-275, hsSLX4 1300-1834	pCB-hex12, hex28L21	CBp200, CBp086, CBp729, CBp738	LIC SUMO, LIC 3C
25	hsSLX1 11-258, hsSLX4 1300-1834	pCB-hex12, hex28L21	CBp200, CBp203, CBp729, CBp738	LIC SUMO, LIC 3C
26	hsSLX1 11-275, hsSLX4 1489-1818	pCB-hex12, hex28L21	CBp200, CBp086, CBp823, CBp827	LIC SUMO
27	hsSLX1 11-275, hsSLX4 1647-1818	pCB-hex12, hex28L21	CBp200, CBp086, CBp816, CBp827	LIC SUMO
28	hsSLX1 11-258, hsSLX4 1647-1818	pCB-hex12, hex28L21	CBp200, CBp203, CBp816, CBp827	LIC SUMO
29	hsSLX1 11-275, hsSLX4 1706-1818	pCB-hex12, hex28L21	CBp200, CBp086, CBp817, CBp827	LIC SUMO
30	hsSLX4 1706-1834-P2A- hsSLX1 11-258	hex28L21	CBp817, CBp216, CBp595, CBp933	LIC 3C
31	ceSLX4 1-718	pCB-bax04	CBp472, CBp473	LIC SUMO
32	ceSLX4 225-718	pCB-bax04	CBp685, CBp473	LIC SUMO
33	ceSLX4 277-718	pCB-bax04	CBp686, CBp473	LIC SUMO
34	ceSLX4 467-718	pCB-bax04	CBp687, CBp473	LIC SUMO
35	ceSLX4 535-718	pCB-bax04	CBp682, CBp473	LIC SUMO
36	ceSLX4 443-606	pCB-bax04	CBp1350, CBp683	LIC SUMO
37	ceSLX4 467-606	pCB-bax04	CBp687, CBp683	LIC SUMO
38	ceSLX1 1-443	pCB-bax04	CBp474, CBp475	LIC SUMO
39	ceSLX1 168-443	pCB-bax04	CBp550, CBp475	LIC SUMO
40	ceSLX4 1-718-RBS-ceSLX1 1-443	pCB-bax04	CBp472, CBp636, CBp637, CBp475	LIC SUMO
41	ceSLX4 1-718-RBS-ceSLX1 168-443	pCB-bax04	CBp472, CBp636, CBp638, CBp475	LIC SUMO
42	ceMUS81 1-445-RBS-ceEME1 1-454	pCB-bax04	CBp476, CBp633, CBp634, CBp527	LIC SUMO
43	ceMUS81 191-445-RBS-ceEME1 1-454	pCB-bax04	CBp574, CBp633, CBp634, CBp527	LIC SUMO
44	ceMUS81 1-445-RBS-ceEME1 161-454	pCB-bax04	CBp476, CBp633, CBp709, CBp527	LIC SUMO
45	ceEME1 1-454 -RBS-ceMUS81 1-445	pCB-bax04	CBp526, CBp635, CBp632, CBp477	LIC SUMO
46	ceEME1 161-454 -RBS-ceMUS81 1-445	pCB-bax04	CBp575, CBp635, CBp632, CBp477	LIC SUMO
47	ceEME1 231-454 -RBS-ceMUS81 198-445	pCB-bax04	CBp1183, CBp635, CBp1186, CBp477	LIC SUMO
48a	ceMUS81 1-157	pCB-bax04	CBp476, CBp1185	LIC SUMO
48b	ceMUS81 1-157	pCB-bax10	CBp611, CBp1285	LIC CPD
49	ceMUS81 1-140	pCB-bax04	CBp476, CBp1566	LIC SUMO

No.	Construct	Vector	Primer	Insertion site
50	ceMUS81 1-89	pCB-bax04	CBp476, CBp1818	LIC SUMO
51	ceMUS81 1-80	pCB-bax04	CBp476, CBp1565	LIC SUMO
52	ceMUS81 10-89	pCB-bax04	CBp1817, CBp1818	LIC SUMO
53	ceMUS81 17-140	pCB-bax04	CBp1564, CBp1566	LIC SUMO
54	ceMUS81 17-80	pCB-bax04	CBp1564, CBp1565	LIC SUMO
55	ceSLX4 443-588	pCB-bax04	CBp1350, CBp1563	LIC SUMO
56	ceSLX4 443-515	pCB-bax04	CBp1350, CBp1562	LIC SUMO
57	ceSLX4 467-560	pCB-bax04	CBp687, CBp1693	LIC SUMO
58	ceSLX4 467-530	pCB-bax04	CBp687, CBp1691	LIC SUMO
59	ceSLX4 470-606	pCB-bax04	CBp1561, CBp683	LIC SUMO
60	ceSLX4 470-515	pCB-bax04	CBp1561, CBp1562	LIC SUMO
61	ceSLX4 488-606	pCB-bax04	CBp681, CBp683	LIC SUMO
62	ceSLX4 488-588	pCB-bax04	CBp681, CBp1563	LIC SUMO
63	ceSLX4 535-606	pCB-bax04	CBp682, CBp683	LIC SUMO
64	pCB-hex15-hsPLK1 R337A I340A (DAA)	T210D pCB-hex15	CBp468, CbBp469	LIC SUMO

4.1.6 Oligonucleotides

Table 4.7: Oligonucleotides for cloning. The small letters correspond to the LIC overhangs. Capital letters correspond to the sequence of the gene of interest.

Name	Sequence (5' → 3')
CBp083	accaggaacaaaccggcgccgctcgatgAAACTGAGTGTGAATGAGGCTCAGCTAGGCTTCTAC
CBp084	gcaaagcaccggcctcgtaGTTCCGCTCCACCTTCTTCTTGCCCC
CBp085	accaggaacaaaccggcgccgctcgatgGTTCCCGCGGGGGTCGCG
CBp086	gcaaagcaccggcctcgtaGGTCTCCAGCAGGTCTGTCCAGTGTG
CBp126	gcaaagcaccggcctcgtaGTTCCGCTCCACCTTCTTCTTGCCCCGAGGC
CBp149	accaggaacaaaccggcgccgctcgGAAGGGAACGAAGTCGCACAGAAGTTTTCTGTCATC
CBp185	ccagggcctcgGAAGGGAACGAAGTCGCACAGAAGTTTTCTG
CBp200	accaggaacaaaccggcgccgctcgGGGCGCTTTTTTCGGCGTCTACC
CBp203	gcaaagcaccggcctcgtaTTCTTTCTCAGTGTCCATCTGGCACAG
CBp216	ggaacagcactctcGTTCCGCTCCACCTTCTTCTTGCCCCG
CBp217	gaggagaacccggcctatgGTTCCCGCGGGGGTCGCG
CBp354	ccagggcctcgatGGCGCCGCACCTTGCTCTCCC
CBp468	accaggaacaaaccggcggcATGAGTGCTGCAGTACTGCAGG
CBp469	gcaaagcaccggcTTAGGAGGCCTTGAGACGGTTGCTGG
CBp472	accaggaacaaaccggcggcATGAATGATTCATCACTAAAAAAGGCAC
CBp473	gcaaagcaccggcTTATAGTCTTCTTTGTAGCCGCCACG
CBp474	accaggaacaaaccggcggcATGGAGACATTCATATTGTCGTCTGATTC
CBp475	gcaaagcaccggcTTATTTTTCGACGAAATTGGACTGATTTCTCG
CBp476	accaggaacaaaccggcggcATGTGCAATCAATCAAAAACGGGTACAG
CBp477	gcaaagcaccggcTTATTGAACAAAAAATTTGAAAAGGTTACGAGTA
CBp490	aagaaggagatatacatATGGGGCGCTTTTTTCGGCGTCTACC
CBp491	accgccccgagTTCTTTCTCAGTGTCCATCTGGCACAG
CBp495	aagaaggagatatacatATGAACTGAGTGTGAATGAGGCTCAGC
CBp497	aagaaggagatatacatATGGAAGGGAACGAAGTCGCACAGAAGTT
CBp498	accgccccgagGTTCCGCTCCACCTTCTTCTTGCC
CBp501	cgcgccgccaccATGGGTCCCGCGGGGGTCGCG
CBp526	accaggaacaaaccggcggcATGGACGAGGCGATTGTAGTCTCTC
CBp527	gcaaagcaccggcTTATTCGACAACAGAATTTCCAGTTTCATCAGTG
CBp550	accaggaacaaaccggcggcGTTCAAATGAGTTTTACGGAGTTTATTG
CBp574	accaggaacaaaccggcggcTCAGATCCATTATCTTTCAGACTTCTCAC
CBp575	accaggaacaaaccggcggcACCAATTCGAAATGCGAGCTCTACAC
CBp595	gaggagaacccggcctGGGCGCTTTTTTCGGCGTCTACC
CBp611	aagaaggagatatacatATGTGCAATCAATCAAAAACGGGTACAG
CBp632	ataatttgttaacttaagaaggagatatacatATGTGCAATCAATCAAAAACGGGTACAG
CBp633	gttaaacaaaattattactagtttaTTGAACAAAAAATTTGAAAAGGTTACGAGTA

Name	Sequence (5' → 3')
CBp634	ataatTTgtTTaactTTaagaaggagatatacatATGGACGAGGCGATTGTAGTCCTCTC
CBp635	gttaaacaaaattattactagtttaTTTCGACAACAGAAATTTCCAGTTTC
CBp636	gttaaacaaaattattactagttTATAGTCTTCTTTTGTAGCCGCCACG
CBp637	ataatTTgtTTaactTTaagaaggagatatacatATGGAGACATTTCATATTGTGCTGTGATTC
CBp638	ataatTTgtTTaactTTaagaaggagatatacatATGGTTCAAAATGAGTTTTTACGGAGTTTA
CBp670	aagaaggagatatacatATGTCTGTTCATCAGGCCCCAGACACCAC
CBp671	aagaaggagatatacatATGTCTCGGGCGCGGGCTCC
CBp672	aagaaggagatatacatATGGGGGAAGAGCAGAGGCCTCCAG
CBp673	aagaaggagatatacatATGGCCACCACAGGACCTGGGGCCC
CBp674	aagaaggagatatacatATGTCTGTGGATGGCAGTGACAGCTCC
CBp675	accgcccgcgagGGTGGCCTCCTGCTGGGCATG
CBp676	accgcccgcgagCTGCCGCTCCTGCCCTGG
CBp681	accaggaacaaaccggcggcATGAAAGAGATCGGGATGCGGCC
CBp682	accaggaacaaaccggcggcGTTTCGGAACAAGGAGGAAAGGGTG
CBp683	gcaaagcaccggccTTATCGCTCGTCATTTGATATATTC AATG
CBp685	accaggaacaaaccggcggcTACGAACA ACTCTCATCTGACATGC
CBp686	accaggaacaaaccggcggcTCTGAAAATATTCAAGTCGAGCACAC
CBp687	accaggaacaaaccggcggcACCAATGATATAACTCCAATGCCTGC
CBp709	ataatTTgtTTaactTTaagaaggagatatacatATGACCAATTCGAAATGCGAGCTC
CBp729	aaaaaaccgccaccATGGAAGGGAACGAAGTCGCACAGAAGTT
CBp738	ggaacagaaccaccagGTTCCGCTCCACCTTCTTCTTGCC
CBp796	accgcccgcgagGGTCTCCAGCAGGTCTGTCCAGTG
CBp816	aaaaaaccgccaccATGGCCACCACAGGACCTGGGGCCC
CBp817	aaaaaaccgccaccATGTCTGTGGATGGCAGTGACAGCTCC
CBp823	aaaaaaccgccaccATGCAAGAGAAGTCCTCGGGCGCG
CBp827	ggaacagaaccaccagGAGCTTCTCCCTGCGGGTGGC
CBp933	ggaacagcacctccagTTCTTTCTCAGTGTCCATCTGGCACAG
CBp1171	aagaaggagatatacatATGCAGGCGGGGACACAGACGA
CBp1183	accaggaacaaaccggcggcATGTTTGAATATTCTTCAACACAAA ACTTGTTC
CBp1185	gcaaagcaccggccTTATTTCTGTTT TAGTGAGTGGAACACTTGGTTG
CBp1186	ataatTTgtTTaactTTaagaaggagatatacatATGCTTCTCACTTGTGCGACCTTTTGAG
CBp1204	gcaaagcaccggccTTAGGTGAGGACGACCATAGCTAC
CBp1205	accaggaacaaaccggcggcAGCCATGGTTGTGTCAGCAGC
CBp1206	accaggaacaaaccggcggcGAAGAATGTCTGAAACATATTATTGTTGT
CBp1213	accaggaacaaaccggcggcGGTCAGAGCAGCAGCCTGGC
CBp1214	accaggaacaaaccggcggcAATGCAGCACTGGTTACCCGTATGAAAG
CBp1243	aagaaggagatatacatATGcaccatcaccatcaccatcaccat
CBp1285	accgcccgcgagTTTTCGTTTTAGTGAGTGGAACACTTGGTTG
CBp1331	accgcccgcgagGGTCAGCGGACCATAGCTACAATACAG
CBp1333	accaggaacaaaccggcggcGCAGCACCGGTTTCGCCTGG
CBp1334	aagaaggagatatacatATGGCAGCACCGGTTTCGCCTG
CBp1335	gcaaagcaccggccTTAAGGTCAGGCTCAGACCTTCGC
CBp1336	accgcccgcgagCAGGCTCAGACCTTCGCCTTC
CBp1350	accaggaacaaaccggcggcTCCAATGTCATTACTCCAATCCGCAACAT
CBp1373	accaggaacaaaccggcggcTCCTCGGGCGCGGGCTCC
CBp1374	accaggaacaaaccggcggcGGGGAAGAGCAGAGGCCTCCAG
CBp1375	gcaaagcaccggccTTAGGTGGCCTCCTGCTGGGCATG
CBp1561	accaggaacaaaccggcggcATAACTCCAATGCCTGCATTTCGATTC
CBp1562	gcaaagcaccggccTTATTTCTGGATGTAATGTGATATATGCTTTCT
CBp1563	gcaaagcaccggccTTAATCATCGGAAACATTTGGATACATTTTTTGG
CBp1564	accaggaacaaaccggcggcAACACTTTTTTTCGAAAAAGTTTTAAAAATCCAAC
CBp1565	gcaaagcaccggccTTAGACGTTCCACGTTTTCTTCCAATTTTAG
CBp1566	gcaaagcaccggccTTATTCGAATTCATTTTCTATATCAACAATGGTT
CBp1691	gcaaagcaccggccTTATGCAGATCTGACAAGCGGACG
CBp1693	gcaaagcaccggccTTACGGCGATGCAATCATTCTTTTCC
CBp1817	accaggaacaaaccggcggcGTGAGAATTGAGCATTCGAAGAACACT
CBp1818	gcaaagcaccggccTTAATCCTGGCCATTGTTCTCTCGGC

Table 4.8: Oligonucleotides used for mutagenesis.

Name	Sequence (5' → 3')	Description
CBp1366	GACTCAGCACGAGCCCAGGGCCTCAG GAAAC	TAG- forward mutagenesis primer for hsSLX4 (T1290E)

Name	Sequence (5' → 3')	Description
CBp1367	CGTTCCTTCCCTGTTTCCTACTG	reverse mutagenesis primer for hsSLX4 (T1290)
CBp1368	AGTGCCCATAGAGCCGATGCCACAGT CATTATG	ATTC- forward mutagenesis primer for hsSLX4 (T1561E)
CBp1369	CACCGGCGTCTCCATAATGGAATAC	reverse mutagenesis primer for hsSLX4 (T1561)
CBp1370	CCATTATGGAGGAGCCGGTGCTGAA GAACTG	GAAG- forward mutagenesis primer for hsSLX4 (T1571E)
CBp1371	CTCCAAACCTATCCAGTTCCTTCTTCAG	reverse mutagenesis primer for hsSLX4 (T1571)
CBp1372	GTGCCCATAGAGCCGATGCCACAGTATT CCATTATGGAGGAGCCGGTGCTGAA GAAGG	forward mutagenesis primer for hsSLX4 (T1561E, T1571E)

Table 4.9: Oligonucleotides used for sequencing.

Name	Sequence (5' → 3')	Description
T7fw	TAATACGACTCACTATAGG	forward sequencing primer for pCB-bax vectors
T7rev	GCTAGTTATTGCTCAGCGG	reverse sequencing primer for pCB-bax vectors
CBp099	CGCGCCACCAGACATAATAG	forward sequencing primer for pCB-hex vectors
CBp079	CGCGCCACCAGACATAATAG	reverse sequencing primer for pCB-hex vectors
CBp107	AGCCTCCTCCTTCCTGTTTG	forward sequencing primer for human slx4 - 1
CBp108	AGAGGTCTGAAACGGAGAGG	forward sequencing primer for human slx4 - 2
CBp109	TCAGAGCTGAGCGAGCGAAG	forward sequencing primer for human slx4 - 3
CBp110	GACTCAGAGGGCAAACCATG	forward sequencing primer for human slx4 - 4
CBp111	TCAGAGCCGTCCCAAATAAC	forward sequencing primer for human slx4 - 5
CBp112	TCCCAGAGCCCACCAAGAAG	forward sequencing primer for human slx4 - 6
CBp113	AGTCGGAGACAGTGACGATG	forward sequencing primer for human slx4 - 7
CBp114	ACCAGACCCTGGACTCAGAC	forward sequencing primer for human slx4 - 8
CBp214	GGTGCTGTTCCAGGGACCAGGCG	forward primer for amplifying 3C-P2A LIC cassette
CBp215	GGGCCGGGGTTCTCCTCAACGTC	reverse primer for amplifying 3C-P2A LIC cassette
CBp257	TGTCAAAGACAGGGTGTTC	forward sequencing primer for hex-sumo
CBp510	GCGGGTTTACCACCATCTG	reverse sequencing primer for Ccpd vectors
CBp579	GTCCAGCTCGACCAGGATG	reverse sequencing primer hex-gfp-rev (pEGFPN1rev)
CBp657	TCCCGATACGAACAACCTCTC	forward sequencing primer for ceSLX4 (central part)
CBp1196	ATTGCTCTAAAATTGGAAGAAACG	forward sequencing primer for ceMUS81 (start @ 69)
CBp1277	CTGCAGTTCACGCAGCAGTT	reverse sequencing primer of hsMUS81syn (ending @ 292)

4.1.7 Buffer for Large Scale Plasmid DNA Preparation

Table 4.10: Buffers for large scale preparation of vector DNA. H₃PO₄: buffer pH was adjusted using phosphoric acid; EDTA ethylenediaminetetraacetic acid; SDS sodium dodecyl sulfate.

Buffer	Composition
Res+ (Resuspension)	50 mM Tris-H ₃ PO ₄ pH 8.0 (4°C), 10 mM EDTA, 100 µg/ml RNase A
Lys + (Lysis)	200 mM Sodium hydroxide, 1% (w/v) SDS
Neu+ (Neutralization)	2.8 M Potassium acetate pH 5.1, set pH with acetic acid
Equ+ (equilibration)	100 mM Tris-H ₃ PO ₄ pH 6.5, 0.9 M KCl, 10% v/v ethanol, 0.6% (v/v) Triton X-114

Buffer	Composition
Endo+ (Endotoxin Wash)	100 mM Tris-H ₃ PO ₄ pH 6.5, 1.15 M KCl, 10% (v/v) ethanol, 0.6% (v/v) Triton X-114
Fil+ (Filter Wash)	100 mM Tris-H ₃ PO ₄ pH 6.5, 1.15 M KCl
Wash+ (Column Wash)	100 mM Tris-H ₃ PO ₄ pH 7.0, 1.15 M KCl, 15% (v/v) ethanol
Elu+ (Elution)	100 mM Tris-H ₃ PO ₄ pH 8.9, 1 M KCl, 10% (v/v) isopropanol

4.1.8 Buffers for Protein Purification

Table 4.11: Buffers used for Protein Purification. *Lys 2 - 7 were supplemented with 1mM AEBSF, 1 μ M Leupeptin, 1 μ M Pepstatin A, 1 μ M Aprotinin. FL full length; IMAC immobilized metal affinity chromatography; β -ME β -Mercaptoethanol; DTT dithiothreitol; EDTA ethylenediaminetetraacetic acid; CW Chaperone wash; ATP adenosine triphosphate; IP6 Inositol-6-phosphate; HABA 2-[4 -hydroxy-benzeneazo]benzoic acid; GFP-B GFP affinity chromatography; Strep affinity purification with recombinant Strep-tag[®]; IEX ion exchange chromatography; SEC size exclusion chromatography; TCEP Tris(2-carboxyethyl)phosphine; Phosphatase inhibitor cocktail see Table 4.12

Buffer	Composition	Protein	Application
Lys 1	50 mM NaH ₂ PO ₄ pH 7.5, 500 mM NaCl, 25 mM Imidazole, 10 mM MgCl ₂	Bacterial expression test	Cell lysis
Lys 2*	20 mM Hepes pH 8.0, 500 mM NaCl, 5 mM β -ME, 30 mM Imidazole	hsEME1-MUS81 FL & truncations, ceEME1-MUS81, ceMUS81 ^N	Cell lysis
Lys 3*	50 mM Tris pH 7.5, 450 mM NaCl, 5 mM DTT, 0.5% NP40, 3750 U <i>Serratia marcescens</i> DNase, 1 mM EDTA	hsSLX4 and hsSLX1 constructs	Cell lysis
Lys 4*	20 mM Tris pH 7.5, 1 M NaCl, 5 mM DTT, 1 mM EDTA, 5 mM Imidazole	hsSLX4 SAP constructs, ceSLX4 constructs	Cell lysis
Lys 5*	1x PBS, 750 mM NaCl, 5 mM DTT, 0.5% NP4, 3 mM EDTA	hsPLK1 DAA	Cell Lysis
Lys 6*	20 mM Hepes pH 8.0, 500 mM NaCl, 3 mM DTT, 5 mM Imidazole, 5 mM EDTA	ceMUS81 N-terminal constructs	Cell lysis
Lys 7*	20 mM Tris pH 7.5, 1 M KCl, 5 mM β -ME, 10 mM Imidazole	hsSLX4 CPD constructs	Cell lysis
Lys 8*	50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT, 0.5% NP40, 1 mM EDTA, 75 mM urea, phosphatase inhibitor cocktail	hsSLX4 and hsSLX1 constructs	Cell lysis
His 1	20 mM Hepes pH 8.0, 500 mM NaCl, 5% (w/v) glycerol, 5 mM β -ME, 30 mM Imidazole	hsEME1-MUS81, ceEME1-MUS81, ceMUS81 ^N	IMAC
His 2	20 mM Hepes pH 8.0, 1 M NaCl, 5% (w/v) glycerol, 5 mM β -ME, 30 mM Imidazole	hsEME1-MUS81 truncations	IMAC
His 3	20 mM Tris pH 7.5, 1 M NaCl, 10% (w/v) glycerol, 5 mM DTT, 5 mM Imidazole	hsSLX4 SAP constructs, ceSLX4	IMAC
His 4	20 mM Hepes pH 8.0, 500 mM NaCl, 5 mM DTT, 5 mM EDTA, 5 mM Imidazole	ceMUS81 N-terminal constructs	IMAC
His 5	20 mM Tris pH 7.5, 1 M KCl, 5mM β -ME, 5% (w/v) glycerol, 30 mM Imidazole	hsSLX4 CPD constructs	IMAC
CW 1	20 mM Hepes pH 8.0, 500 mM NaCl, 5% (w/v) glycerol, 5 mM β -ME, 2 mM ATP, 5 mM MgSO ₄	hsEME1-MUS81 FL & truncations, ceEME1-MUS81	IMAC
CW 2	20 mM Hepes pH 8.0, 1 M NaCl, 5% (w/v) glycerol, 5 mM β -ME, 2 mM ATP, 5 mM MgSO ₄	hsMUS81 ²⁻²²⁵	IMAC

Buffer	Composition	Protein	Application
CW 3	20 mM Tris pH 7.5, 1 M NaCl, 5 mM DTT, 2 mM ATP, 5 mM MgCl ₂	hsSLX4 SAP constructs, ceSLX4	IMAC
CW 4	50 mM Tris pH 7.5, 500 mM NaCl, 3 mM DTT, 2 mM ATP, 5 mM MgSO ₄	hsSLX4 construct, hsPLK1 DAA	IMAC
CW 5	20 mM Hepes pH 8.0, 500 mM NaCl, 3 mM DTT, 2 mM ATP, 5 mM MgSO ₄	ceMUS81 N-terminal constructs	IMAC
CW 6	20 mM Tris pH 7.5, 1 M KCl, 5mM β-ME, 2 mM ATP, 5 mM MgSO ₄	hsSLX4 CPD constructs	IMAC
His Elu 1	50 mM NaH ₂ PO ₄ pH 7.5, 250 mM NaCl, 700 mM Imidazole	Bacterial expression test	IMAC
His Elu 2	20 mM Hepes pH 8.0, 100 mM NaCl, 5% (w/v) glycerol, 600 mM Imidazole, 5 mM β-Mercaptoethanol	hsEME1-MUS81FL & truncations, ceEME1-MUS81	IMAC
His Elu 3	20 mM Tris pH 7.5, 100 mM NaCl, 10% (w/v) glycerol, 300 mM Imidazole, 5 mM DTT	ceSLX4, ceMUS81 ^N	IMAC
His Elu 4	20 mM Hepes pH 8.0, 100 mM NaCl, 5% (w/v) glycerol, 1 mM IP6, 5 mM β-ME	ceMUS81 ^N	IMAC
His Elu 5	20 mM Tris pH 7.5, 150 mM KCl, 5 mM β-ME, 300 mM Imidazole, 5% (w/v) glycerol	hsSLX4 CPD constructs	IMAC
His Elu 6	50 mM Tris pH 8.0, 100 mM NaCl, 5% (w/v) glycerol, 350 mM Imidazole, 5 mM β-Mercaptoethanol	hsEME1-MUS81 truncations	
Strep 1	50 mM Tris pH 8.0, 500 mM NaCl, 1 mM EDTA	hsEME1-MUS81 truncations	Strep
Strep 2	50 mM Tris pH 8.0, 150 mM NaCl, 10 mM IP6	hsEME1-MUS81 truncations	Strep
Strep 3	50 mM Tris pH 8.0, 150 mM NaCl, 0.1 mM EDTA, 2.5 mM desthiobiotin	hsEME1-MUS81 truncations	Strep
Strep 4	50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM HABA	hsEME1-MUS81 truncations	Strep
Strep 5	50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA	hsEME1-MUS81 truncations	Strep
GFP 1	50 mM Tris pH 7.5, 1 M NaCl, 5 mM DTT	hsSLX4 and hsSLX1 constructs, hsPLK1 DAA	GFP-B
GFP 2	50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT	hsSLX4 and hsSLX1 constructs, hsPLK1 DAA	GFP-B
IEX 1 A	20 mM Hepes pH 8.0, 3 mM DTT, 5% (w/v) glycerol	hsEME1-MUS81 FL & truncations, hsSLX4 SAP constructs, ceSLX4 ^{SAP} , ceMUS81 ^N	IEX
IEX 1 B	20 mM Hepes pH 8.0, 2 M NaCl, 3 mM DTT, 5% (w/v) glycerol	hsEME1-MUS81 FL & truncations, hsSLX4 SAP constructs, ceSLX4 ^{SAP} , ceMUS81 ^N	IEX
IEX 2 A	20 mM MES pH 6.0, 3 mM DTT, 10% (w/v) glycerol	ceSLX4	IEX
IEX 2 B	20 mM MES pH 6.0, 2 M NaCl, 3 mM DTT, 10% (w/v) glycerol	ceSLX4	IEX
IEX 3 A	20 mM Bis-Tris pH 6.5, 5% (w/v) glycerol, 3 mM DTT, 1 mM EDTA	hsSLX4 CPD constructs	CEX
IEX 3 B	20 mM Bis-Tris pH 6.5, 2 M NaCl, 5% (w/v) glycerol, 3 mM DTT, 1 mM EDTA	hsSLX4 CPD constructs	CEX
SEC 1	20 mM Hepes pH 8.0, 150 mM NaCl, 3 mM DTT	hsEME1-MUS81 FL	SEC
SEC 2	20 mM Hepes pH 8.0, 200 mM NaCl, 3 mM DTT	hsEME1-MUS81 truncations, ceSLX4 SAP constructs	SEC
SEC 3	20 mM Hepes pH 8.0, 300 mM NaCl, 1 mM TCEP	hsMUS81 ²⁻²²⁵	SEC
SEC 4	20 mM Bis-Tris-Propane pH 7.5, 200 mM KCl, 5% w/v glycerol, 3 mM DTT	hsSLX4 and hsSLX1 constructs	SEC

Buffer	Composition	Protein	Application
SEC 5	20 mM Hepes pH 7.5, 150 mM NaCl, 3 mM DTT	ceSLX4	SEC
SEC 6	20 mM Bis-Tris-Propane pH 7.5, 200 mM KCl, 1 mM TCEP	hsSLX4 SAP constructs, ceSLX4 SAP constructs, hsPLK1 DAA, ceMUS81 ^N	SEC
SEC 7	20 mM Hepes pH 8.0, 100 mM NaCl, 5% (w/v) glycerol, 3 mM DTT	ceEME1-MUS81	SEC
SEC 8	20 mM Potassium phosphate pH 8.0, 100 mM KCl, 1mM TCEP	Interaction assay on SEC	SEC
SEC 9	20 mM Potassium phosphate pH 6.5, 100 mM KCl, 1 mM TCEP	ceMUS81 ^N , Interaction assay on SEC	SEC
SEC 10	20 mM Bis-Tris-Propane, 200 mM KCl, 5% (w/v) glycerol, 3 mM DTT	hsSLX4 CPD constructs	SEC

Table 4.12: Phosphatase inhibitors. Phosphatase Inhibitor Cocktail A is derived from Merck Phosphatase Inhibitor Cocktail II, Phosphatase Inhibitor Cocktail B is derived from Merck Phosphatase Inhibitor Cocktail III.

Phosphatase Cocktail	Concentration	Components	Target Phosphatases
Cocktail A	100x		
	200 mM	Imidazole	Alkaline Phosphatase
	100 mM	Sodium Fluoride	Acid Phosphatase
	115 mM	Sodium Molybdate	Acid Phosphatase
	100 mM	Sodium Orthovanadate	Protein Tyrosine phosphatases, Alkaline Phosphatase
	400 mM	Sodium Tartrate	Acid Phosphatase
Cocktail B	10x		
	50 mM	Sodium Fluoride	Acid Phosphatase
	20 mM	β -Glycerophosphate	Ser/Thr Phosphatase
	20 mM	Sodium Pyrophosphate	Ser/Thr Phosphatase
	2 mM	Sodium Orthovanadate	Protein Tyrosine phosphatases, Alkaline Phosphatase

4.1.9 Equipment

Table 4.13: Equipment

Instrument	Supplier
Ultrasonic homogenizer	Bandelin electronics, Berlin, Germany
Dounce homogenizer	B. Braun, Melsungen, Germany
ÄKTA purification systems	GE Healthcare, München, Germany
NanoDrop (ND-1000) spectrophotometer	PeqLab, Erlangen, Germany
Typhoon TM FL 7000 PhosphoImager	GE Healthcare, München, Germany
Genius Pro Fluorescence reader	Tecan, Männedorf, Switzerland
Cellometer Auto T4 Bright Field Cell Counter	Nexcelom Bioscience, Lawrence, MA, USA
JANSi UVEX-m	JANSi, Seattle, WA, USA
Prometheus NT.48 nanoDSF	Nanotemper Technologies GmbH, Munich, Germany

4.1.10 Bioinformatic tools and software

Table 4.14: Software

Software	Developer / Supplier
APBS	[311]
ApE v2.0.44	M. Wayne Davis
BLASTp	[312]
coot	[313]
Dali server	[276, 277]
GlobPlot	[270]
GPS 3.0	[265]
GraphPad Prism 7.0c	GraphPad Software, La Jolla, CA, USA
HHpred	Söding et al. [314]
i-Control	Tecan, Männedorf, Switzerland
Inkscape	Version 0.48.2 [315]
Jalview	Waterhouse et al. [216]
MaxQuant	Cox and Mann, 2008 [316]
MUSCLE	Edgar [215]
Origin 8.1	Origin Lab, Northampton, MA, USA
PDB2PQR	[317]
Phaser Version	[318]
Phenix	[319]
Phyre ²	[320]
Protparam	[321]
Pymol	[322]
SEDFIT package, version 12.1b	Schuck [323]
UNICORN	GE Healthcare
XDS	[324]

Alignments

The homologous sequences were collected using the program Jalview and the alignment of the sequences was carried out with the algorithm MUSCLE (Multiple Sequence comparison by Log-Expectation) [215, 216]. Jalview was used for the visualization of the alignment.

Secondary Structure Prediction

Secondary structure prediction and homology detection was carried out with HHpred, JPred and Phyre² [314, 320, 325]. The tool GlopPlot was used for globular domain prediction [270].

Calculation of Protein Properties

Molecular weight, absorption coefficient and the isoelectric point (pI) of the protein were calculated with the ProtParam tool from the ExPASy homepage (www.expasy.org/tool/protaram.html). The calculation of the pI was important for the design of purification strategy. In order to quantify the purified protein amounts, absorption coefficients and molecular weights were obtained from the ProtParam tool.

Figures

The molecular figures of ceSLX4^{SAP} were created using the program pymol. Graphical representations were made with the vector graphics editor Inkscape [315].

4.2 Methods

4.2.1 Cloning Procedures

DNA Templates

Human SLX4 was provided by Wade Harper (HMS, Boston, USA). SLX1 was obtained from Mammalian Gene Collection (MGC, IMAGE clone 2823474). hs-MUS81 and EME1 were synthesized in two pieces as codon optimized GeneArt String DNA fragments by Invitrogen™ and cloned as a bicistronic construct into bax04 by T4 processing. *C. elegans* cDNA was provided by Dr. Karsten Klage (MPIB) and genes were directly amplified by polymerase chain reaction.

Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) was applied for the amplification of the DNA sequences. The used constructs were amplified from template plasmids, cDNA or synthesized DNA. The reactions were carried out according to standard protocol and standard program (see tables 4.15 and 4.16) using Phusion™ High-Fidelity DNA Polymerase Master Mix (NEB). The 2x Phusion Master Mix contained already the Phusion polymerase, Phusion DNA polymerase buffer and dNTPs. The annealing temperature was adjusted to the respective primers and ranged between 58 and 63°C. The extension time was adjusted to the size of the gene of interest with 30s / kb.

Table 4.15: PCR reaction mixture.

Component	Stock Concentration	Final Concentration	Amount
Template	~ 50-200 ng/ μ l	5-20 ng	1 μ l
Forward Primer	10 μ M	0.5 μ M	1 μ l
Reverse Primer	10 μ M	0.5 μ M	1 μ l
Phusion Master Mix	2x	1x	10 μ l
DMSO	100%	3%	0.6 μ l
ddH ₂ O	add to 20 μ l		6.4 μ l

Table 4.16: PCR program for amplification

Step	Temperature (°C)	Duration	Cycles
Initial Denaturation	98	30 s	
Denaturation	98	20 s	
Annealing	58 - 63	30 s	30
Extension	72	30 s / kb	
Final Extension	72	5 min	
Cooling	10	∞	

Constructs containing two or more proteins were generated in a two-step PCR. First, the individual proteins were amplified by PCR in a way that the reverse primer and the forward primer of the open reading frame in first position and second position, respectively, contained overhangs with a ribosome binding site. After gel extraction the fragments were mixed in 1:1

molar ratio for the second PCR step. In this PCR the primer of the first open reading frame and the reverse primer of the second open reading frame were used.

The PCR product was afterwards analyzed by agarose gel electrophoresis (0.8 % w/v) in 1x TAE buffer (40mM Tris, 20mM Acetic acid, 1mM EDTA pH 8.0). The correctly sized PCR product was cut out of the gel and isolated using the NucleoSpin™ Gel and PCR Clean-up Kit (MN).

Cloning into bax and hex Vectors

The amplified PCR product was cloned into various pCB vectors, being generated by the lab of Christian Biertümpfel, by Ligase independent cloning (LIC). This method relies on the generation of complementary single-stranded overhangs by T4 DNA polymerase, which contains a 3' - 5' exonuclease activity [326]. Vector and PCR product can circularize based on their cohesive ends without ligase activity. Remaining nicks on the plasmid are repaired by *E. coli* RecA [326].

First, the vector was linearized with restriction enzyme *BspQI* (NEB) for 1h at 50°C as described in Table 4.17. To separate between cut and uncut vector the digested vector was analyzed on a 0.8% agarose gel loading the equivalent volume of 250ng vector per lane. The corresponding bands of linearized vector were cut and the DNA was isolated with NucleoSpin™ Gel and PCR Clean-up Kit (MN).

Table 4.17: Mixture for vector linearization in order to prepare it for T4 processing

Component	Stock Concentration	Amount
Vector		2 µg
Cutsmart buffer	10x	4 µl
<i>BspQI</i>	10,000 U/ml	8 U
ddH ₂ O		to 40 µl

In order to generate the overhangs for annealing with the construct the linearized vector was processed with T4 DNA Polymerase (NEB) (see Table 4.18).

Table 4.18: Mixture for T4 processing of the vector in order to prepare it for LIC cloning

Component	Stock Concentration	Amount
linearized vector		450 ng
NEB buffer 2.1	10x	3 µl
T4 DNA Polymerase	3,000 U/ml	0.6 µl
dTTP	25 mM	3 µl
DTT	100 mM	1.5 µl
ddH ₂ O		to 30 µl

Likewise, the gel-purified PCR product (see section 4.2.1) was processed with T4 DNA polymerase by mixing according to table 4.19. Insert and vector processing mixtures were incubated

at room temperature for 30 min. The enzyme was heat inactivated by incubation for 20 min at 75°C .

Table 4.19: Mixture for T4 processing of the insert in order to prepare it for T4 processing.

Component	Stock Concentration	Amount
linearized vector		600 ng
NEB buffer 2.1	10x	2 μ l
T4 DNA Polymerase	3,000 U/ml	0.4 μ l
dATP	25 mM	2 μ l
DTT	100 mM	1 μ l
ddH ₂ O		to 20 μ l

The annealing reaction was performed by mixing 4 μ l T4 processed insert with 1 μ l T4 processed vector (Table 4.19) and incubating 20 min at room temperature. After addition of 1 μ l EDTA (4.16 mM final concentration), the reaction mix was incubated for 10 min at room temperature. For transformation, chemical competent *E. coli* cells were mixed with 4 μ l of annealing reaction and proceeded as in section 4.2.1. Maximal five colonies were picked and tested for false-positives by colony-PCR (see section 4.2.1) or restriction digest. Two positive clones were amplified and plasmids were prepared using NucleoSpinTM Plasmid Kit (Macherey-Nagel, Düren, Germany). To verify the correct sequence the clones were sent for sequencing (Eurofins Genomics, Ebersberg, Germany or Biochemistry Core Facility (MPIB)).

Screening of Constructs

In order to identify positive clones, maximal five colonies were picked and either subjected to colony PCR (cPCR) or grown over night, isolated and subjected to restriction digest. cPCR was carried out in a 10 μ l reaction volume according to Table 4.20 and 4.21. The extension time depended on the length of the inserted gene of interest and was usually calculated to 1 min per kb. The cPCR products were analyzed by 0.8% w/v agarose gel electrophoresis in 1x TAE.

Table 4.20: Reaction mixture for colony PCR

Component	Stock Concentration	Final Concentration	Amount
Forward T7 Primer	10 μ M	0.3 μ M	0.3 μ l
Reverse T7 Primer	10 μ M	0.3 μ M	0.3 μ l
dNTPs	2mM	0.2 mM	1 μ l
ThermoPol buffer	10x	1x	1 μ l
Taq polymerase	5000 U/ml	1 U	0.2 μ l
ddH ₂ O		7.2 μ l	

Table 4.21: PCR program for colony screening

Step	Temperature (°C)	Duration	Cycles
Initial Denaturation	95	5 min	

Step	Temperature (°C)	Duration	Cycles
Denaturation	95	30 s	25
Annealing	52	30 s	
Extension	72	1 min / kb	
Final Extension	72	5 min	
Cooling	10	∞	

Alternatively to colony PCR, restriction digest was carried to check for positive clones. The prepared plasmid was subjected to restriction in 10 μ l reaction volume containing 500 ng plasmid, 1 μ l 10x CutSmart buffer and 4 U of the respective restriction enzymes. The reactions were incubated at 37°C for 1 h and analyzed on a 0.8% w/v agarose gel. The sequence of one to two positive clones was further verified by sequencing (Eurofins, Ebersberg, Germany or Biochemistry Core Facility, MPIB, Martinsried, Germany).

Cloning of Mutants by Site-Directed Mutagenesis

Mutations were inserted into the gene of interest via PCR with partially overlapping primers. The forwards primer contained the mutation and is complementary with the reverse primer for 10 - 15 nucleotides at the 3' end. The mutation in the forward primer was flanked by 15 - 20 nucleotides. Both primers had a length between 30 - 40 nucleotides. For insertion of the mutation the standard PCR reaction mixture was used (see Table 4.15). The PCR program for site-directed mutagenesis was applied (see Table 4.22). Afterwards, the PCR product was digested with 5 U DpnI for 1 h at 37°C. For transformation, 4 μ l of the digested PCR product was mixed with XL1 blue competent *E. coli* cells. A maximum of five colonies was subjected to amplification and plasmids were prepared with NucleospinTM Plasmid Kit (Macherey-Nagel, Düren, Germany). To assess the successful insertion of the mutation the plasmids were sent to sequencing (Eurofins, Ebersberg, Germany or Biochemistry Core Facility, MPIB, Martinsried, Germany).

Table 4.22: PCR program for site-directed mutagenesis

Step	Temperature (°C)	Duration	Cycles
Initial Denaturation	98	30 s	
Denaturation	98	20 s	30
Annealing	58 - 63	60 s	
Extension	72	30 s / kb	
Final Extension	72	5min	
Cooling	10	∞	

The construct ceSLX4⁴⁴³⁻⁶⁰⁶ F476A K485A R494A K498A H513A was synthesized by Eurofins (Ebersberg, Germany) and resuspended in water to a concentration of 30 ng/ μ l. The DNA fragment was subjected to T4 processing (see Table 4.19), annealing with N-terminal His8-SUMO vector (pCB-bax04) and cloning into XL1 blue bacteria.

Large Scale Preparation of DNA Plasmids

For transient transfection of HEK cells, large amounts of vector were needed. Thus, the desired vector was transformed into XL1 blue cells (for procedure see following subsections) and plated on agar plates containing the respective antibiotics. Next day, one colony was picked and a 25 ml LB pre-culture with the antibiotics was shaken in the incubator at 250 rpm and 37°C over night. Then a 600 ml LB culture with the respective antibiotics was inoculated with 600 µl pre-culture and incubated for 16 - 18 h at 37°C and 220 rpm in 2 l Erlenmeyer flasks with baffles reaching final OD₆₀₀ values higher than 4.5. The bacterial cells were harvested at 8000 g for 10 min. Pellets were either flash frozen in liquid nitrogen and stored at -80°C or used directly. For large scale plasmid preparation the buffers from table 4.10 were applied according to the steps written in the manual MN NucleoBondTMXtra Maxi for low-copy plasmid preparation. DNA binding columns were used from the Kit and regenerated after usage extensively washing with 1 M HCl, incubation over night and neutralizing with water. Folded paper filters (diameter 185mm) for loading the lysate on the columns were supplied by WhatmanTM/ GE Healthcare. After precipitation of the eluted plasmid DNA by isopropanol, the pellet was centrifuged at 3220 g for 90 min. The supernatant was removed and after washing with 70% ethanol the pellet was spun down at 3220 g for another 60 min. After this the plasmid pellet was transferred under the sterile hood and dried under sterile conditions. The pellet was dissolved completely in 1 ml sterile ddH₂O shaking at 37°C and 165 rpm. The purity of the plasmid solution was assessed by measuring 260/280 and 260/230 ratios. The concentration of the plasmid was measured by Nanodrop.

Preparation of Competent Cells

Chemical competent cells were prepared according to the following protocol. The bacterial strain for cloning or expression was plated on agar plates containing the respective antibiotics. One single colony was picked for over night pre-culture. 100 ml LB media were inoculated with 2ml pre-culture in a 500 ml flask and the culture was grown at 37°C to OD₆₀₀ 0.5. Cells were cooled down to 4°C and centrifuged at 3000 rpm for 15 min in an Eppendorf Centrifuge 5810 R (Hamburg, Germany). Bacterial pellet was resuspended in 10 ml TSS buffer (1% (w/v) tryptone, 0.5% (w/v) Yeast extract, 100 mM NaCl, 10 % (w/v) PEG3000, 5% (v/v) DMSO, 50 mM MgCl₂, pH 6.5) and 2.5 ml 86% glycerol was added. Cells were aliquoted to 50 µl and flash frozen in liquid nitrogen. Competence was preserved by long-term storage at -80°C.

Transformation of Competent Cells

Chemical-competent cells were thawed from -80°C for 5min on ice. For re-transformation of plasmid ~10 ng and for cloning 4 µl annealing reaction were mixed with 50 µl cells and incubated for 25 min on ice. The heat shock was applied at 42°C for 45 s and the cells were cooled down on ice for 5 min. After adding 200 µl SOC-medium, the cells were shaken at 37°C 1000 rpm for maximal one hour. The bacteria were plated on agar-plates containing the respective antibiotic (Table 4.4) and incubated over night at 37°C.

4.2.2 Protein Expression

Pull-Down Experiments for Expression and Solubility Testing for Expression in *E. coli*

In order to assess the expression and solubility of cloned constructs pull-down experiments were performed by testing different bacterial strains, media and Isopropyl- β -D-thiogalacto-pyranoside (IPTG) concentrations. The different bacterial strains were transformed with the respective plasmid and grown over night on agar plates with respective antibiotics, Afterwards a 3 ml overnight LB-culture was inoculated with one colony. Small Scale expression tests were carried out in 4 ml media inoculated with 40 μ l over night culture in 24 deep-well plates with round bottom (Whatman, Maidstone, UK). Bacteria were grown for 6 h at 37°C shaking at 800 rpm. Cultures were induced with respective IPTG concentration or auto-induction and shaken at 16°C or room temperature over night. Next day, cultures were harvested at 3130 g for 10 min and the pellet was resuspended in Lysis buffer 1 supplemented with 1 μ l benzonase / 100 ml lysis buffer. Sonication and pull down steps were carried out on a Tecan Workstation, Tecan Freedom Evo 150. Bacterial cells were lysed by sonication with a Vibra-Cell™ VC750 Processor (Sonics, Newton, CT, USA) complemented with 24 Tip Horn from Qsonica, LLC (Newton, CT, USA). Cells were sonicated for 4 min at 50% with pulse on/off 1 s/1 s and again for 2 min with the same settings. Lysates were spun down at 13200 rpm for 30 min at 4°C and applied to 50 μ l MediaScout RoboColumns (Atoll GmbH, Ravensburg; now: OPUS RoboColumns by Repligen, Waltham, MA) with Chelating Sepharose FF (GE Healthcare). Proteins were eluted with elution buffer His Elu 1. Samples of cells before induction, induced cells, lysate, insoluble fraction and eluate were analyzed by SDS-PAGE. The steps from pellet resuspension to elution were carried out by Ariane Fischer (Crystallization Facility, MPIB).

Expression in *E. coli*

Large scale protein expression was typically carried out in 2l Erlenmeyer flasks with baffles (2l flasks were manufactured from Klimax; Glasgertebau Ochs introduced 3 baffles up to 30% of total flask height) containing 500 ml medium supplemented with appropriate antibiotics (Table 4.4). The medium was inoculated with 500 μ l small-scale pre-culture (1:100 dilution). For the pre-culture, 20 ml LB medium with the appropriate antibiotics were inoculated with colonies of freshly-transformed *E. coli* cells and grown for 1 h at 37°C shaking at 220 rpm. After inoculation the large scale bacterial cultures were raised at 37°C and 120 rpm to an optical density of OD₆₀₀ 1.5 - 1.7. The temperature of the shaker was then reduced to 16°C and cells were cooled down at 4°C. Protein expression was induced with respective amount of IPTG for 16 - 18 h at 16°C. The cells were harvested by centrifugation (8000 g, 10 min) and either freshly used for protein purification or stored at -80°C.

Table 4.23: Expression conditions for large scale protein production in *E. coli* cells. All expressions were carried out at 16°C for 16 - 18 h. FL full length

No.	Expressed construct	<i>E. coli</i> strain	Condition
1	hsEME1-MUS81 FL	BL21 DE3 gold	TB 0.5 mM IPTG

No.	Expressed construct	<i>E. coli</i> strain	Media
2 - 5	hsEME1-MUS81 truncations	BL21 (DE3) gold	TB 0.5 mM IPTG
6	hsMUS81 2-225	BL21 DE3 gold	TB 0.5 mM IPTG
7	hsSLX4 1-1834 (FL)	BL21 DE3 gold pRARE	TB 0.5 mM IPTG
8	hsSLX4 1309- 1824	B834 pLysS	ZY
9	hsSLX4 1492 - 1824	BL21 DE gold pRARE	TB 0.5 mM IPTG
10	hsSLX4 1516 - 1824	BL21 DE gold pRARE	TB 0.5 mM IPTG
11	hsSLX4 1647 - 1824	B834 pLysS	TB 0.5 mM IPTG
12	hsSLX4 1706-1824	B834 pLysS	TB 0.5 mM IPTG
13	hsSLX4 1748-1824	B834 pLysS	TB 0.5 mM IPTG
14 - 16	hsSLX4 ^{SAP} truncations	BL21 DE3 gold pRARE	TB 0.5mM IPTG
17	hsSLX1 1-275	BL21 DE gold pRIL	ZY
18	hsSLX4 1300-1824 + hsSLX1 11-258	BL21 DE gold, B834 pLysS	TB 0.5 mM IPTG, ZY
19	hsSLX4 1706-1824 + hsSLX1 11-258	BL21 DE3 gold pRARE, B834 pLysS	TB 0.5 mM IPTG, ZY
31	ceSLX4 1-1834 (FL)	BL21 DE gold pRIL	TB 1 mM IPTG
32 - 35	ceSLX4 N-terminal truncations	BL21 DE gold pRARE	TB 0.5 mM IPTG
36 - 37	ceSLX4 ^{SAP} constructs	BL21 DE gold pRARE	TB 0.5 mM IPTG
38 + 39	ceSLX1 FL & 168-443	B834 pLysS	TB 0.5 mM IPTG
40 + 41	ceSLX4-ceSLX1 FL & 168-443	BL21 DE gold pRARE	TB 0.5 mM IPTG
42 - 44	ceMUS81-EME1 constructs	BL21 DE gold pRARE	TB 0.5 mM IPTG
45 - 47	ceEME1-MUS81 constructs	BL21 DE gold pRARE	TB 0.5 mM IPTG
48 - 54	ceMUS81 N-terminal constructs	BL21 DE gold pRARE	TB 0.5 mM IPTG
55 - 63	ceSLX4 ^{SAP} constructs	BL21 DE gold pRARE	TB 0.5 mM IPTG

Expression in *E. coli* in Minimal Medium

Expression of ¹⁵N- and ¹⁵N-¹³C-labeled proteins was carried out in M9 minimal medium supplemented with ¹⁵N ammonium chloride, 0.4% glucose or 2 g ¹³C-glucose, 1 mM MgSO₄, 0.3 mM CaCl₂, 1 μg biotin, 1 μg thiamin and 1x trace elements solution (0.134 mM EDTA, 31 μM FeCl₃·6H₂O, 0.62 μM ZnCl₂, 0.76 μM CuCl₂·2H₂O) according to the protocol by Dr. Arie Geerloff, Helmholtz Center Munich. Bacterial cells transfected with the respective plasmid were plated on agar plates containing appropriate antibiotics. A 20 ml overnight pre-culture in LB medium was grown and cells were pelleted for 10 min at 3220 g. The supernatant was discarded and the cell pellet was washed with 5 ml sterile water and resuspended in M9 media. The bacterial cells were grown until OD₆₀₀ 0.8 at 37° C. Afterwards, the temperature was reduced to 20° C and the bacterial cells were induced with the respective amount of IPTG. Cells were harvested like described before.

Expression Tests for HEK Culture

For initial vector and construct testing different ratios between transfectant (Polyethyleneimide, PEI) and DNA as well as expression times were tested in HEK293 cells. Proteins were tagged with GFP in order to facilitate protein detection. For example, in a 1:1 setup 1 μg vector and 1 μg PEI per 1 ml of cell culture were mixed with 500 μl Freestyle media and incubated for 8 min at 37°C. The transfection mix was added to 1.5×10^6 cells/ml in 10 ml Freestyle™ expression media (ThermoFisher Scientific) supplemented with 1% FCS. Small scale expression tests were carried out in TubeSpin® Bioreactors 50 (TPP, Trasadingen, Switzerland) and grown for up to three days with every day sampling. Whole cell samples were spun down at 1200 g for 3 min. The cell pellet was dissolved with 6x GFP loading dye (150 mM Tris / HCl pH 7.5, 15 mM EDTA, 15% (w/v) glycerol, 0.02% bromphenol blue, 150 mM DTT, 5% SDS according to [257]) and an equivalent of 0.2×10^6 cells/ml was loaded on a SDS-PAGE. In gel Fluorescence was detected with Typhoon™ FLA 7000 PhosphoImager (GE Healthcare) with excitation at 473 nm and filter at 520 nm.

Expression in HEK Cells

HEK293 cells were adapted to suspension culture and grown in FreeStyle™ expression media (ThermoFisher Scientific) supplemented with 1% fetal calf serum (FCS). The cells were cultivated in wide neck DURAN™ GLS 80™ laboratory glass bottles with permeable membrane cap (DWK Life Sciences, Mainz, Germany) and raised at 37°C, 7.5% CO₂, 63% humidity, and 165rpm to high cell density. One day before expression set-up the cells were split to 0.8×10^6 cells/ml. On the day of the expression the HEK293 cells were grown to a cells density of 1.5×10^6 cells / ml. The cells were transiently transfected with a 4:2 mixture of polyethylenimid (PEI) and DNA in FreeStyle™ expression media. For each ml of cell culture 2 μg DNA was mixed in a 50 ml conical centrifuge tube with 4 μg PEI in a volume of 5% of the total expression volume. The mixture was incubated at room temperature for 8 min and added directly to the target cell culture. To slow down cell growth valproic acid (VPA) was added to an end concentration of 3.75 mM and cells were cultivated at 32°C. 24 h after setting up the expression the cells were harvested by centrifugation at 2000 g for 20 min. The cell pellet was then washed with 1x phosphate buffered saline (PBS) and directly used for protein purification.

Table 4.24: Expression conditions for large scale protein production in suspension culture of HEK cells.

No.	Expressed construct	Condition (PEI:DNA)
20	hsSLX4 1-1834	3:1
21	hsSLX4 667-1834	3:1
22	hsSLX4 1300-1834	3:1
23	hsSLX1 1-275	2:1
24	hsSLX1 11-275 + hsSLX4 1300-1834	4:2
25	hsSLX1 11-258 + hsSLX4 1300-1834	4:2
26	hsSLX1 11-275 + hsSLX4 1489-1834	4:2
27 - 29	hsSLX1 constructs with hsSLX4 CCD constructs	4:2
30	hsSLX4 1706-1834 - P2A-hsSLX1 11-258	4:2

4.2.3 Protein Purification

In general, all purification steps were carried out at 4°C. The detailed buffer composition for the purification is listed in Table 4.11. Loading of the cleared lysate on the Ni and ion exchange (IEX) columns and washing steps were performed using a peristaltic pump (Reglo ICC, Ismatec, Cole-Parmer GmbH, Wertheim, Germany). Linear gradients for elution were administered by ÄKTA Prime systems (GE Healthcare) by using mixtures of IEX A and B buffers. Gel filtration was carried out on ÄKTA Purifier system.

Purification of hsEME1-MUS81

Recombinantly and bicistronically expressed HIS-SUMO-hsEME1-MUS81 was purified from freshly harvested BL21 gold cells. The cell pellet was resuspended at 4°C with lysisbuffer 2 (Lys2) and lysed by sonication (Bandelin Sonoplus, tip VS70, pulse ON/OFF 0.5/1 s, 40% amplitude, 10 min). The lysate was clarified by centrifugation at 75600 g for 30 min. The soluble fraction was loaded on a Ni-NTA column (5 ml, Protino™ Ni-NTA, Macherey-Nagel) and washed with His-1 buffer (His 1) followed by a chaperone-wash (CW 1) and another wash with His-1 buffer. The immobilized protein complex was eluted with high imidazole buffer (His Elu 2). The eluted protein complex was directly loaded onto a Heparin column (5 ml, HiTrap Heparin HP, GE Healthcare), which was directly connected to the lower end of the Ni column. For washing and elution the Heparin column was connected to a ÄKTA prime system and washed with 5% IEX B buffer (100 mM NaCl). For elution a gradient from 5% to 60% IEX B (1.2 M NaCl) in 60 ml was applied. The His-SUMO tag on hsEME1 was cleaved with SUMO-protease (final concentration 0.01mg/ml) for 30 min at 4°C. The cleavage mix was concentrated by ultra-filtration and subjected to gel filtration (Superdex 200 16/600 column, GE Healthcare) pre-equilibrated with SEC buffer 1 (SEC 1). In the end, the purified proteins were concentrated by ultra-filtration and used for further analysis or crystallization. To store the protein, 86% glycerol was added to a final concentration of 10% (v/v), aliquoted, flash frozen in liquid nitrogen and frozen at -80°C.

Purification of hsEME1-MUS81 Truncations from bax25

hsEME1-MUS81 constructs in vector bax25 with N-terminal SUMO tag on EME1 and C-terminal CPD-TwinStrep tag on MUS81 were bacterially expressed in BL21 gold cells. After harvest of bacteria the bacterial pellet was resuspended in lysis buffer (Lys 2) and lysed by sonication like the full length proteins of hsEME1-MUS81. After clarification of the lysate, the soluble protein fraction was loaded on Ni column (5 ml, Protino™ Ni-NTA, Macherey-Nagel). Afterwards the column was washed with His washbuffer (His 2) and chaperon wash buffer (CW 1). Proteins were eluted with His elutionbuffer (His Elu 6). In order to reduce the imidazole concentration to 100mM, the Ni pool was diluted with 3.5x volume of His Elu 6 buffer without imidazole. The diluted Ni pool was loaded on Strep-Tactin® Superflow® column (1 ml, IBA GmbH, Göttingen, Germany), equilibrated with Strep 1 buffer. The bound protein was eluted by activation of CPD by washing with Strep 2 buffer. The remaining CPD-Twinstrep-tag was displaced from the column by washing with Strep 3 buffer. The Strep-Tactin® Superflow® column was regenerated by subsequently washing with Strep 4 and Strep 5 buffer. The remaining SUMO-tag on hsEME1 construct was cleaved by the addition of Senp2 (0.01 mg/ml).

Purification of hsEME1-MUS81 truncations - N-terminal SUMO tag

hsEME1-MUS81 constructs were expressed in a bicistronic vector and purified from either fresh pellet after harvest or stored bacterial pellets at -80°C . Lysis was carried out in the same way as for full-length hsEME1-MUS81 applying Lys 2 buffer. After application of cleared lysate on Ni-NTA column (5 ml, ProtinoTM Ni-NTA, Macherey-Nagel) and washed with His-2 buffer (His 2) followed by a chaperone-wash (CW 1). The immobilized protein complex was eluted with His Elu 2 directly into Heparin column (5 ml, HiTrap Heparin HP, GE Healthcare). Wash and elution of Heparin column was done in the same order as for hsEME1-MUS81 FL with IEX A and IEX B buffers, except for the washing step with 10% IEX B. Cleavage and concentration of the proteins were performed according to the protocol for FL proteins. The cleaved mix was subjected to gel filtration (Superdex 75 16/600 column, GE Healthcare) pre-equilibrated with SEC buffer 2 (SEC 2). The purified proteins were used for further analysis and crystallization. Proteins were stored in 10% (v/v) glycerol at -80°C .

Purification of hsMUS81²⁻²²⁵

Recombinantly expressed hsMUS81 2-225 with N-terminal His8-SUMO solubility tag was purified from BL21 gold cells. Lysis was carried out in the same way as for full-length hsEME1-MUS81 applying Lys 2 buffer. After application of cleared lysate on Ni-NTA column (5 ml, ProtinoTM Ni-NTA, Macherey-Nagel) the column was washed with His-2 buffer (His 2) followed by a chaperone-wash (CW 2). The immobilized protein complex was eluted with His Elu 2 directly into Heparin column (5 ml, HiTrap Heparin HP, GE Healthcare). Wash and elution of Heparin column was done in the same order as for hsEME1-MUS81 FL with IEX A and IEX B buffers. The Heparin pool was cleaved for 1 h at 4°C on the rocker and after dilution to 100 mM NaCl re-applied to Heparin column in order to separate protein from tag. The pool of the second Heparin column run was concentrated and injected into gel filtration column (Superdex 75 16/600 column, GE Healthcare) equilibrated with SEC buffer 3 (SEC 3). The purified proteins were either used for further analysis and crystallization or were stored in 10% (v/v) glycerol at -80°C .

Purification of hsSLX4¹³⁰⁰⁻¹⁸³⁴ + hsSLX1¹¹⁻²⁷⁵ from HEK Cell Expression

Transiently co-expressed hsSLX4¹³⁰⁰⁻¹⁸³⁴ with C-terminal 3C-mVENUS-1d4 tag and hsSLX1¹¹⁻²⁷⁵ with N-terminal His8-eGFP-SUMO tag was purified from HEK293 cells directly after harvest. The cell pellet was homogenized in 50 ml lysisbuffer 3 (Lys 3) by a Dounce homogenizer (B. Braun, Melsungen), moving the pestle 30 times up and down, and subjected to sonication (Bandelin Sonoplus, tip VS70, pulse ON/OFF 0.5/1.5 s, 20% amplitude, 5 min). The lysate was clarified by centrifugation at 75600 g for 1 h. The soluble fraction was filtered by 0.22 μm CorningTM bottle top filters (Corning Incorporated Life Sciences, Tewksbury, MA, USA) and then applied to GFP binder column (1 ml, HiTrap NHS-Activated HP affinity column (GE Healthcare) coupled with anti-GFP V_HH purified protein). The column was washed extensively with High salt wash buffer (GFP 1) and then with Low Salt buffer (GFP 2). For elution 100 μl 1 mg/ml Senp2 mod (modified for application in mammalian cell culture) was mixed with 10 μl 1 mg/ml 3C protease in a volume of 1 ml GFP 2 buffer, injected into the GFP binder column and incubated at 4°C over night. The proteins were eluted with 2.5 ml High salt buffer (GFP 1) and the GFP binder column was

regenerated with 0.1 M Citrate. The proteins were applied to gel filtration column S200 16/600 (GE Healthcare) equilibrated with SEC buffer (SEC 4). The purified proteins were used for further analysis or stored in 10% (v/v) glycerol at -80°C .

Purification of CPD-Tagged hsSLX4 Constructs

Bacterially expressed hsSLX4 constructs with C-terminal CPD-His8 tag were purified by resuspension in Lys 7 buffer and the cells were sonicated (Bandelin Sonoplus, tip VS70T, pulse ON/OFF 0.5/1 s, 40% amplitude, 15 min). After separating soluble and insoluble fraction by centrifugation for 45 min at 75600 g, the soluble fraction was loaded on two IMAC columns (1CV = 10ml, His-Trap HP, GE Healthcare) and washed with His 5 buffer. The ATP buffer wash was carried out with CW 6 buffer. The protein was eluted with His Elu 5 buffer, fractions containing the tagged protein were pooled and tag cleavage was induced by addition of 0.1 M IP6 (end concentration 40 μM) for 30 min at 4°C . Salt concentration and pH were adjusted by dilution with buffer IEX 3 A to 50 mM salt and pH 6.5 and the sample was loaded on cation exchange column (5 ml, HiTrapTM SP HP, GE Healthcare). After washing with 3% IEX 3 B (60 mM NaCl) the protein was eluted by applying a gradient from 3% to 60% B. The peak fractions were pooled and concentrated via ultrafiltration. At last the protein was purified by SEC (Superdex 75 10/300, GE Healthcare) pre-equilibrated with SEC buffer (SEC 10), concentrated, flash-frozen and stored at -80°C .

Purification of hsSLX4 SAP Constructs

Constructs of human SLX4 SAP domain were expressed with N-terminal His8-SUMO tag in *E. coli* and cell pellet was directly used after harvest. The cell paste was resuspended in lysis buffer Lys 4 supplemented with 10mM EDTA and 2 μM Pepstatin and lysed by sonication (Bandelin Sonoplus, tip VS70T, pulse ON/OFF 0.5/1 s, 40% amplitude, 12 min). The cleared lysate (centrifugation at 75600 g for 3 0min) was loaded on a Ni-column (5ml, cOmpleteTM His-Tag purification column, Roche) and subsequently washed with buffer His 3, CW 3 and eluted with His Elu 3 supplemented with 3 mM DTT. The Ni pool was diluted with IEX 1 buffer A to 44 mM NaCl and loaded on anion exchange column (Q Sepharose Fast Flow, GE Healthcare). The protein was eluted by a linear gradient from 4% to 60% IEX 1 B buffer. The peak fractions were pooled and the SUMO tag was cleaved over night at 4°C by addition of Senp2 (end concentration 0.014 mg/ml). After concentrating by ultrafiltration the cleavage mix was injected on gel filtration column (Superdex 75 16/600, GE Healthcare) pre-equilibrated with SEC buffer (SEC 6). The peak fractions were pooled, concentrated, flash-frozen and stored at -80°C .

Purification of ceSLX4

Full-length ceSLX4 was expressed with N-terminal His8-SUMO tag in BL21 gold pRIL cells. The cell pellet was resuspended in Lys 4 buffer and lysed by sonication (Bandelin Sonoplus, tip VS70T, pulse ON/OFF 0.5/1 s, 40% amplitude, 8 min). The lysate was clarified by centrifugation at 75600 g for 45 min. The soluble fraction was loaded on a Ni-column (5 ml, cOmpleteTM His-Tag purification column, Roche) and washed with His-3 buffer (His 3) followed by a chaperone-wash (CW 3) and elution with Elu 3 buffer (His Elu 3). The His8-SUMO-tag was cleaved over night by

addition of Senp2 to the pooled fractions. The cleavage mixture was loaded on to a SP-sepharose column (5 ml, HiTrapTM SP HP, GE Healthcare), washed with 5% IEX 2 B (100 mM NaCl), and eluted from cation exchange column using a gradient to 60% IEX 2 B. The protein was purified in a last step by SEC (Superose 6 16/600 column, GE Healthcare) pre-equilibrated with SEC buffer (SEC 5), concentrated by ultra-filtration, flash-frozen and stored at -80°C.

Purification ceSLX4^{SAP} Constructs

This purification protocol applies to all ceSLX4 SAP constructs. His8-SUMO tagged SAP constructs of ceSLX4 were expressed in *E. coli*. The cell pellets were resuspended in Lys 4 buffer supplemented with 10 mM EDTA and 2 µM Pepstatin and lysed by sonication (Bandelin Sonoplus, tip VS70T, pulse ON/OFF 0.5/1 s, 40% amplitude, 10 min). After centrifugation (75600 g 30 min) the clarified fraction was loaded on His-cOmplete Ni column (5 ml, cOmpleteTM His-Tag purification column, Roche) and the resin was first washed with His buffer 3 (His 3), then with chaperon-wash buffer 3 (CW 3). The His-tagged protein was directly eluted with His Elu 3 into a subsequent connected Heparin column (5 ml, HiTrapTM Heparin HP, GE Healthcare). After washing with 5% IEX 1 buffer B (IEX 1 B), the protein was eluted with an linear gradient to 60% IEX 1 buffer B (IEX 1 B). The His8-SUMO-tag was cleaved off by incubation with Senp2 protease (final concentration 0.02 mg/ml) over night at 4°C. The cleavage mix was desalted to 5% IEX 1 buffer B (IEX 1 B, 100 mM NaCl) and applied to SP-Sepharose column (5 ml, HiTrapTM SP HP, GE Healthcare), washed with 5% IEX 2 B (100 mM NaCl), and eluted from cation exchange column using a linear gradient to 60% IEX 2 B. Finally, the protein was purified by SEC (Superdex 75 16/600, GE Healthcare) using SEC buffer 6 (SEC 6), concentrated, supplemented with 10% glycerol for storage, flash frozen and stored at -80°C.

Purification of ceEME1-MUS81

The bicistronically expressed protein complex consisting of ceEME1 and ceMUS81 was purified after expression in *E. coli*. Cells were resuspended in lysisbuffer 2 (Lys 2) and lysed by sonication (Bandelin Sonoplus, tip VS70T, pulse ON/OFF 0.5/1s, 40% amplitude, 10min). The clarified lysate was loaded onto Ni-NTA Superflow cartridges (2x 5 ml, Qiagen GmbH, Hilden, Germany). After washing with washbuffer His 1 and CW 1, the proteins were directly eluted with His Elu 2 into one connected Heparin column (5 ml, HiTrapTM Heparin HP, GE Healthcare). Subsequently, the proteins were washed with 5% IEX 1 B (100 mM NaCl) and eluted by a linear gradient from 5% to 60% IEX 1 B. The His8-SUMO tag was cleaved off from ceEME1 by incubation with Senp2 (final concentration 0.01 mg/ml) during concentrating of the proteins. The cleavage mixture was subjected to gel filtration (Superdex 200 16/600, GE Healthcare) with buffer SEC 7, concentrated, flash frozen and stored at -80°C.

Purification of Truncated ceEME1-MUS81 Constructs

Truncated versions of the bicistronically expressed ceEME1-MUS81 complex were purified in the same way as outlined for ceEME1-MUS81 full length proteins (Section ?? with the exception that instead of Qiagen Ni-NTA Superflow cartridges ProtinoTM Ni-NTA columns were applied because they had higher binding capacity for the better expressed truncated ceEME1-MUS81 constructs.

Purification of ceMUS81^N

ceMUS81^N construct 1-157 was purified either with C-terminal CPD tag or with N-terminal His8-SUMO tag. This paragraph describes the purification with the CPD-tag and the subsequent paragraph explains the purification of the SUMO-tagged ceMUS81^N and more truncated N-terminal ceMUS81 constructs. ceMUS81^N was expressed in *E. coli* and resuspended in Lys 2 buffer. Cells were lysed by sonication (Bandelin Sonoplus, tip VS70T, pulse ON/OFF 0.5/1 s, 40% amplitude, 10 min) and after clarification was the lysate loaded on two Ni-NTA columns (10ml, Ni-NTA superflow cartridges, Qiagen). The resin was washed with His 1 buffer, followed by a chaperon wash with buffer CW 1. Tag cleavage was induced by washing with His Elu 4. The remaining His8-CPD-tag was removed from Ni column by washing with His Elu 1 buffer. The eluted protein after His Elu 3 was loaded on two Heparin columns (10 ml, HiTrap Heparin HP, GE Healthcare), and after removing unbound contaminants with 5% buffer IEX 1 B (100 mM NaCl) eluted by a gradient from 100 mM NaCl to 1.2 M NaCl. The pooled fractions were subjected to gel filtration (S75 16/600, GE Healthcare) with SEC 6 buffer. The fractions containing the clean protein were pooled, flash frozen and stored at -80°C.

Purification of ceMUS81 N-terminal Truncations

Further truncations of ceMUS81^N expressed with N-terminal-His8-SUMO tag were expressed in BL21 (DE3) pRARE *E. coli* cells and lysed by sonication, as outlined before, in Lys 6 buffer. The lysate was clarified by centrifugation at 75600g for 30min at 10°C. The clarified fraction was loaded on His-cOmplete Ni column (5ml, cOmpleteTM His-Tag purification column, Roche) and washed with His 4 buffer followed by chaperon wash (CW 5) and elution with His Elu 2 buffer, supplemented with 3mM DTT. Tag cleavage was induced by addition of Senp2 (final concentration 0.0025mg/ml) and incubation for 2h on ice. The cleavage mix was loaded at low pump speed on an IEX column (HiTrapTM SP HP, GE Healthcare). After washing with 5% IEX 1 B buffer (100mM NaCl), the protein was eluted in a gradient to 60% IEX 1 B buffer (1.2M NaCl). The SP pool was pooled, concentrated by ultra-filtration and subjected to SEC (S75 16/600, GE Healthcare) pre-equilibrated with SEC 9 buffer. In the end, the purified protein was concentrated by ultra-filtration, flash frozen after supplementing with 10% (v/v) glycerol and stored at -80°C.

Purification of hsPLK1 DAA

The hyperactive mutant of hsPLK1 (T210D R337A I340A) was lysed by douncing with 30 strokes and sonication for 2 min in Lys buffer 5 after three day expression in HEK293 suspension culture. Lysate was loaded on GFP-binder column (1ml, HiTrap NHS-Activated HP affinity column (GE Healthcare) coupled with anti-GFP V_HH purified protein) and washed subsequently with GFP 1 buffer, CW 4 and GFP 2 buffer. Cleavage of the tag was carried out by incubation over night at 4°C with 1 ml 0.1 mg/ml Senp2 (modified for application with proteins expressed in human cell culture) supplemented with GFP 2 buffer. Next day, PLK1 was eluted by applying 4 ml GFP 1 buffer and further purified by gel filtration (S200 16/600, GE Healthcare) in buffer SEC 6. The SEC pool was supplemented with 10% glycerol (w/v) and stored at -80°C.

4.3 Biochemical Assays

4.3.1 Determination of Protein Concentration and Purity

Protein purity was assessed by SDS-Gel electrophoresis, by monitoring the absorption ratio A_{260}/A_{280} during size exclusion chromatography and at Nanodrop measurements. Protein concentrations were measured by NanoDrop spectrophotometer (Thermo Scientific, UK) at absorbance 280nm and 260nm. The extinction coefficients were calculated using the ProtParam tool (ExpASy, SIB bioinformatics resource portal) [321]. Protein concentrations for the ceSLX4^{SAP} constructs were not measured with NanoDrop as the extinction coefficient contained already an error of 10% due to the presence of only one tyrosine in the protein. The protein concentration was assessed using the Bicinchoninic acid (BCA) Protein Assay Kit (NovagenTM, EMD Millipore Comp, Billerica, MA USA) according to manufacturer proceedings for micro-scale assay.

Protein concentrations for MUS81, EME1 and SLX4 in one HeLa cell [305] were estimated by applying the following formulas: The estimated volume of one HeLa nucleus $690 \mu^3$ [306], which corresponds to 690 fl ($690 \cdot 10^{-15}$ l).

$$n = \frac{N}{N_A} \quad (4.1)$$

with n being the substance amount, N the particle number and N_A the Avogadro constant ($6.02214086 \cdot 10^{23} \text{ mol}^{-1}$). The molar concentration c is then calculated by the following formula:

$$c = \frac{n}{V} \quad (4.2)$$

with V being the volume of the cell nucleus.

4.3.2 Polyacrylamide Gel Electrophoresis

Purity of protein was assessed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) according to Laemmli [327] using 12.5 to 18% gels. SDS-Gels were stained with Coomassie brilliant blue R-250.

4.3.3 Native Gel Electrophoresis

To analyze the DNA-binding properties of the used constructs, EMSA was carried out with self-made Tris-Borate gels (0.5x Tris-Borate, 6% 19:1 Acrylamide:Bisacrylamide, 0.2% TEMED, 0.08% APS). The gels were pre-equilibrated for 120 - 180 min in 0.5x TB buffer at 100 V and 4°C. After loading the samples, the gels were incubated for 5 min at 4°C to let the sample diffuse slowly into the gel. Gels were then ran for 10 min at 200 V and for another 30 to 60 min at 100 V. The gels were analyzed by fluorescence imaging (see section ??).

4.3.4 Limited Proteolysis

For limited proteolysis, 10 μ l protein was used at 0.6 mg/ml in respective gel filtration buffer, mixed with 3 μ l protease solution and incubated for 30 min on ice. Trypsin, Subtilisin, Elastase, GluC and Chymotrypsin were diluted from 1 mg/ml stock using protease dilution buffer (20 mM

Hepes pH 7.5, 50 mM NaCl, 10 mM MgSO₄) and were applied at three different concentrations (0.1 mg/ml, 0.01 mg/ml, 0.001 mg/ml). The reaction was stopped by adding 3 μ l 6x LAEMMLI buffer and incubation for 5 min at 95°C. The degradation products were separated by SDS-PAGE and stained with Coomassie for visualization. Peptide finger printing was carried out by Nagarjuna Nagaraj, PhD (MPIB Biochemistry Core Facility).

4.3.5 Mass Spectrometry Analysis

For peptide fingerprinting after limited proteolysis, protein samples were cut out from SDS-PAGE. Sample processing and peptide analysis was carried out by Dr. Nagarjuna Nagaraj and his team (Biochemistry Core Facility, MPIB). Prior to protein digest, the protein sequence was subjected to *in silico* enzymatic digest with MS-Digest of ProteinProspector [328]. Samples were digested with Trypsin or a combination of Trypsin and GluC for human SLX4 constructs [329] in order to increase sequence coverage. The peptides were analyzed by Orbitrap mass spectrometry [330] and identified using Max Quant software [316]. To determine the total mass of proteins or protein fragments, 40 μ l of a 1 mg/ml protein solution was analyzed by electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS). The measurement was carried out with a BRUKER microTOF mass spectrometer by Elisabeth Weyher-Stingl, Antonio Piras or Victoria Sanchez (MPIB Biochemistry Core Facility, MPIB).

4.3.6 Treatment with DNA Damaging Agents

HEK293 cells were raised to 1.5×10^6 cells per ml and transiently transfected with C-terminally mVenus tagged hsSLX4¹³⁰⁰⁻¹⁸³⁴ and N-terminally eGFP-tagged hsSLX1¹¹⁻²⁷⁵ according to the HEK expression protocol (see section 4.2.2). Next day the cells were separated into 50 ml aliquots, DNA damaging agents were added and the mixture was incubated according to table 4.25.

Table 4.25: DNA damaging agents. Application scheme of DNA damaging agents to HEK suspension culture. The reference gives credit to the publication of the experimental conditions used.

DNA damaging Agent	Stock concentration	End concentration	Incubation time	Reference
Hydroxyurea	150 mM	2 mM	3h	[267]
Mitomycin C	500 μ g/ml	0.1 μ g/ml	1h	[112]
Methylmethanesulfonate	1 M	2 mM	1h	[113]
Etoposide	50 mM	5 μ M	1h	[268]

The cells were harvested by centrifugation at 2000 g for 10 min and resuspended in lysis buffer (Lys 8). This lysis buffer was supplemented with Phosphatase inhibitor cocktail A & B. Cells were homogenized in a 5 ml dounce homogenizer, incubated for 30 min on ice and sonicated for 2 min (VS70T tip, 20%, 0.5/1.5 s on/off). The lysate was transferred into 2 ml reaction tubes and the lysate was clarified by centrifugation at 21130 g for 15 min at 4°C. The soluble fraction was then incubated for 1 h at 4°C with GFP-binder beads (200 μ l 50% slurry, HiTrap NHS-Activated HP affinity beads (GE Healthcare) coupled with anti-GFP V_{HH} purified protein), equilibrated with GFP 2 buffer. For separation of bound and unbound protein, the beads were centrifuged at 2500 g for 2 min. Beads were washed three times with GFP 1 buffer, two times with CW 4 buffer followed by washing in GFP 2 buffer. The tag was cleaved off by incubation for 1 h on rotating

wheel with 3C protease (end concentration 0.06 mg/ml) in GFP 2 buffer. Proteins were separated from the beads by centrifugation at 2500 g for 2 min and subsequently subjected to phosphosite analysis by Dr. Nagarjuna Nagaraj (MPIB).

4.3.7 Phosphosite Identification by MS

Sample Preparation and Processing

hsSLX4¹³⁰⁰⁻¹⁸³⁴ and hsSLX4¹³⁰⁰⁻¹⁸³⁴-hsSLX1¹¹⁻²⁷⁵ purified by anti-GFP-affinity were dephosphorylated by different phosphatases. For this, 40 µg protein was incubated with 7 U of the respective phosphatase and reaction buffer in a 70 µl reaction volume for 1 h at 37°C (see Table 4.26). The reactions were stopped by cooling down on ice.

Table 4.26: Dephosphorylation of hsSLX4¹³⁰⁰⁻¹⁸³⁴. Different phosphatases and their respective buffer were applied for dephosphorylation of hsSLX4¹³⁰⁰⁻¹⁸³⁴.

Phosphatase	Reaction buffer	Supplier
λ-Phosphatase	10 mM MnCl ₂	NEB
Shrimp phosphatase	Cutsmart	NEB
Fast Alkaline Phosphatase	FastAP buffer	ThermoFisher
Antarctic Phosphatase	Antarctic PPT buffer	NEB

The samples were further handled by Dr. Nagarjuna Nagaraj (Biochemistry Core Facility, MPIB) and co-workers. According to their protocol 40 µl sample were added to 40 µl guanidium denaturation buffer (6 M guanidium hydrochloride, 10 mM TCEP and 40 mM Chloroacetamide in 25 mM Tris pH 8.5) and the mixture was heated at 70°C for 2 minutes. The samples were then diluted by adding 400 µl of 50 mM ammonium bicarbonate solution and vortexed well. Each sample was divided into two aliquots. One aliquot was digested with 2 µg of sequencing grade modified trypsin (Promega, Madison, WI, USA) and the other half with 2 µg trypsin followed by 2 µg of sequencing grade Glu-C (Roche) The digestion was carried out overnight at 25°C and 37°C for Glu-C and trypsin respectively. After digestion the peptides were purified using C18 StageTips and loaded for LC-MSMS analysis. The purified peptides were loaded on 15 cm long pulled emitters packed with 1.9 µm repositil C18 beads (Dr Maisch GmbH, Ammerbuch-Entringen, Germany) and eluted over a 140 minute gradient directly into a Q Exactive HF mass spectrometer via a nano-electrospray source. The mass spectrometer was operated in a data-dependent manner and up to top 15 peptides like features were selected for HCD fragmentation. Dynamic exclusion was enabled to minimize repeat sequencing of same peptides. All raw data were processed using MaxQuant computational platform. The peak lists were searched against a custom database containing only the relevant proteins present in the sample using Andromeda search engine that is built into Maxquant platform.

Analysis of Phosphorylation Sites

The analysis of the data was carried but myself. For this two parameters were taken into account, the localization probability and the peak intensity. The probability for one phosphorylation at a specific site had to be higher than 75%. This is important, when one peptide contained more than

one detected phosphorylation site. Phosphorylation sites were only taken into account when the peak intensity in the spectrum was significantly above the background, usually ranging from $1 \times E^7$ to $3 \times E^{11}$. For the Phosphosite detection of hsSLX4¹³⁰⁰⁻¹⁸³⁴ alone three different independent phosphorylation site detections from three different expression batches were analyzed. For the detection of the phosphosites in hsSLX4¹³⁰⁰⁻¹⁸³⁴ in co-expression with hsSLX1 and in presence of DNA-damaging agents, three independent measurements were performed. The intensities were grouped by treatment and the average intensity given in the three measurements for one site was taken. For graphical presentation in a heat map, logarithmized (\log_2) intensities were clustered.

4.3.8 *In vitro* Kinase assay

The hyperactive mutant hsPLK1 T210D R337A I340A was applied for phosphorylation of ceSLX4^{SAP}. The reaction was carried out in 50 mM Tris pH 7.5, 100 mM KCl, 10 mM MgCl₂, 5 mM DTT, 200 μ M ATP according to [331]. 500 μ M ceSLX4^{SAP} was incubated with 109 nM hsPLK1 T210D R337A I340A for 30 min at 30°C. The resulting phosphorylation mix was either subjected to gel filtration interaction assays or sent for phosphosite detection by mass spectrometry and ESI-TOF mass spectrometry.

4.3.9 Cleavage Assay

DNA substrates were synthesized by Eurofins (Ebersberg, Germany) and resuspended in annealing buffer (20 mM Hepes/KOH pH 7.9, 50 mM KCl, 0.1 mM EDTA). The substrates resemble different DNA intermediates during DNA damage repair and were designed based on 30 bp arm length HJ (see Table 4.27). Oligonucleotide CB209 was 5' 6FAM labeled. Cleavage assays were carried out in cleavage buffer (50 mM Tris pH 8.0, 1 mM DTT, 100 μ g/ml BSA, 5 mM MgCl₂ and 150 mM NaCl) using 40 nM DNA substrate and 640 nM protein. The reaction mixture was incubated for 30 min at 37°C and the reaction was stopped by adding 1 μ l Proteinase K, 0.5 μ l 0.5 M EDTA and 0.3 μ l 10%SDS. The cleavage products were separated by native gel electrophoresis with 10% TB gels at room temperature for 30 min running constant at 2 W. Gels were analyzed by fluorescence imaging (excitation 473 nm, filter 520 nm, TyphoonTM FL 7000 PhosphoImager, GE Healthcare). The DNA substrate bands were quantified using ImageJ. The uncut DNA substrate bands were normalized against the control substrate bands and product formation was calculated from the difference.

Table 4.27: DNA substrates used for nuclease activity tests

DNA Substrate	Oligo Name	Sequence
ds	CB209	5'FAM-ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGA CATCTTTGCCACCTGCAGGTTACCC
	CB215	GGGTGAACCTGCAGGTGGGCAAAGATGTCCTAG CAAG- GCACTGGTAGAATTCGGCAGCGT
3'flap	CB209	5'FAM-ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGA CATCTTTGCCACCTGCAGGTTACCC
	CB212	CGATAGTCCGATCCTCTAGACAGCTCCATGTAGCAAG GCACTGGTAGAATTCGGCAGCGT
	CB217	CATGGAGCTGTCTAGAGGATCCGACTATCG

DNA Substrate	Oligo Name	Sequence
5'flap	CB209	5'FAM-ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGA CATCTTTGCCACCTGCAGGTTACCC
	CB212	CGATAGTCGGATCCTCTAGACAGCTCCATGTAGCAAG GCACTGGTAGAATTCGGCAGCGT
	CB218	GGGTGAACCTGCAGGTGGGCAAAGATGTCC
RF	CB209	5'FAM-ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGA CATCTTTGCCACCTGCAGGTTACCC
	CB212	CGATAGTCGGATCCTCTAGACAGCTCCATGTAGCAAG GCACTGGTAGAATTCGGCAGCGT
	CB217	CATGGAGCTGTCTAGAGGATCCGACTATCG
	CB218	GGGTGAACCTGCAGGTGGGCAAAGATGTCC
HJ	CB209	5'FAM-ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGA CATCTTTGCCACCTGCAGGTTACCC
	CB210	GGGTGAACCTGCAGGTGGGCAAAGATGTCCATCT GTTG- TAATCGTCAAGCTTTATGCCGT
	CB211	ACGGCATAAAGCTTGACGATTACAACAGATCATG GAGCT- GTCTAGAGGATCCGACTATCG
	CB212	CGATAGTCGGATCCTCTAGACAGCTCCATGTAGCAAG GCACTGGTAGAATTCGGCAGCGT
nHJ	CB209	5'FAM-ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGA CATCTTTGCCACCTGCAGGTTACCC
	CB210	GGGTGAACCTGCAGGTGGGCAAAGATGTCCATC TGTG- TAATCGTCAAGCTTTATGCCGT
	CB213	ACGGCATAAAGCTTGACGATTACAACAGATC
	CB214	P-ATGGAGCTGTCTAGAGGATCCGACTATCG
	CB212	CGATAGTCGGATCCTCTAGACAGCTCCATGTAGCAAG GCACTGGTAGAATTCGGCAGCGT

4.3.10 Electrophoretic Shift Assays

For EMSA analysis, proteins were incubated with DNA substrates and the formed complexes were analyzed by native PAGE. EMSAs were run like described in section 4.3.3. For first analysis DNA binding was analyzed in a protein concentration range from 0.25 to 2.5 or 5 μ M. Gels were analyzed by fluorescence imaging (excitation 473 nm, filter 520 nm, TyphoonTM FL 7000 PhosphoImager, GE Healthcare).

Table 4.28: DNA substrates used for Fluorescence Anisotropy measurements and bands shift assays

DNA Substrate	Oligo Name	Sequence
ss	4w1616-1	5'FAM-CTCGAAGAATTCCGGA TTAGGGATGCCGTCTG
ds	4w1616-1	5'FAM-CTCGAAGAATTCCGGA TTAGGGATGCCGTCTG
	ds1616-2	CAGACGGCATCCCTAA TCCGGAATTCTTCGAG
3'flap	4w1616-1	5'FAM-CTCGAAGAATTCCGGA TTAGGGATGCCGTCTG
	4w1616-4	CGTTCTGAGCCTAGCG TCCGGAATTCTTCGAG
	4w1616-3n6	pCGCTAGGCTCAGAACG
5'flap	4w1616-1	5'FAM-CTCGAAGAATTCCGGA TTAGGGATGCCGTCTG
	4w1616-4	CGTTCTGAGCCTAGCG TCCGGAATTCTTCGAG

DNA Substrate	Oligo Name	Sequence
	4w1616-2n5	CAGACGGCATCCCCTAA
RF	4w1616-1	5'FAM-CTCGAAGAATTCCGGA TTAGGGATGCCGTCTG
	4w1616-2n5	CAGACGGCATCCCCTAA
	4w1616-3n6	pCGCTAGGCTCAGAACG
	4w1616-4	CGTTCTGAGCCTAGCG TCCGGAATTCTTCGAG
HJ	4w1616-1	5'FAM-CTCGAAGAATTCCGGA TTAGGGATGCCGTCTG
	4w1616-2	CAGACGGCATCCCCTAA GCTCCATCGTGGCGGA
	4w1616-3	TCCGCCACGATGGAGC CGCTAGGCTCAGAACG
	4w1616-4	CGTTCTGAGCCTAGCG TCCGGAATTCTTCGAG
nHJ	4w1616-1	5'FAM-CTCGAAGAATTCCGGA TTAGGGATGCCGTCTG
	4w1616-2n3	CAGACGGCATCCCCTAA GC
	4w1616-2n4	pTCCATCGTGGCGGA
	4w1616-3	TCCGCCACGATGGAGC CGCTAGGCTCAGAACG
	4w1616-4	CGTTCTGAGCCTAGCG TCCGGAATTCTTCGAG

The K_D was calculated using Graphpad Prism 7 by applying non-linear regression and specific binding with Hill Slope (Equation 4.3).

$$y = B_{max} \cdot \frac{x^h}{K_D^h + X^h} \quad (4.3)$$

y is the Protein-DNA complex in percent, x is the protein DNA concentration. B_{max} is the maximum specific in percent. K_D is the protein concentration needed to achieve half-maximum binding at equilibrium. h is the Hill slope. If h is 1.0, then a monomer binds with no cooperativity to one site. When h is greater than 1.0, then the binding graph shows sigmoidal shape, meaning the protein has multiple binding sites with positive cooperativity. The quality of the non-linear regression was assessed with R squared and residual plot analysis and the agreement between observed and calculated binding isotherm was evaluated applying a confidence interval of 95%.

4.3.11 SEC-based Protein-Protein Interaction

For the analysis of interaction between ceSLX4^{SAP} and ceMUS81^N, both proteins were mixed in a 1:1 molar ratio. Both proteins were mixed together at a concentration of 600 μ M of each protein. The mixture was incubated for 30 min on ice. Finally, a 40 μ l sample at a final concentration of 150 μ M, supplemented with SEC 8 buffer, was injected on a Superdex 75 3.2/30 column (GE Healthcare) on ÄKTAmicro, equilibrated with SEC 8 buffer. The single proteins were subjected to analytical gel filtration at 1.11 μ M (ceMUS81^N) and 1.29 μ M (ceSLX4^{SAP}).

4.3.12 Static Light Scattering

In order to determine molecular weight and monodispersity of a protein sample static light scattering experiments were carried out. Static light scattering experiments were performed by Dr. Claire Basquin at the MPI for Biochemistry. In principle, the sample passes through a size exclusion chromatography column, which is directly connected to a detector for light scattering (LS), ultraviolet absorbance (UV) and refractive index (RI). Here, ÄKTA micro (GE Healthcare)

was used as a chromatography system coupled to a TDA 302 detector (Viscotek, now Malvern Pananalytical, Kassel, Germany) monitoring refractive index and right angle scattering. The protein samples (xx) were either injected onto Superdex S200 5/150 or Superdex S75 5/150 GL (3 ml column volume, GE Healthcare) column. As a standard bovine serum albumine was used and the refractive index increment (dn/dc) was set to 0.185 ml/g for the calculations.

4.3.13 Analytical Ultracentrifugation

Sedimentation velocity experiments were performed in the Biochemistry Core Facility at Max-Planck-Institute of Biochemistry by Susanne von Gronau and Dr. Stephan Uebel on an Optima XL-I analytical ultracentrifuge (Beckman Inc., Palo Alto, CA, U.S.A.) using an An-60 Ti rotor and double-sector epon centerpieces. The single protein constructs ceSLX4^{SAP} and ceMUS81^N were submitted in 20 mM Bis-tris-Propane pH 7.5, 200 mM KCl, 1 mM TCEP (SEC 6) at concentrations of 25 mg/ml (OD₂₈₀ 0.77) and 1.75 mg/ml (OD₂₈₀ 0.8), respectively. The complex of both proteins was submitted at 3.2 mg/ml (OD₂₈₀ 0.87). Buffer density and viscosity was measured with a DMA 5000 densitometer and an AMVn viscosimeter, respectively (both Anton Paar, Graz, Austria). Protein concentration distribution was monitored at 280 nm at 50,000 rpm and 10°C. Time-derivative analysis was computed with the SEDFIT package, version 12.1b [323], resulting in a c(s) distribution and an estimate for the molecular weight Mf from the sedimentation coefficient and the diffusion coefficient, as inferred from the broadening of the sedimentation boundary, assuming that all observed species share the same frictional coefficient f/f₀.

4.3.14 Thermal Shift Assay

For Thermofluor 100 μM ceMUS81^N was mixed with 5 - 500 μM ceSLX4^{SAP} in a total volume of 50 μl of SEC 9 buffer supplemented with 3 mM TCEP. The reaction mixtures were incubated for 1h on ice. Afterwards the protein complexes were mixed with 1 μl 1:10 Sypro Orange dye (5000x, Invitrogen) in a 96-well PCR plate (Eppendorf). The plate was sealed and heated in a real-time PCR system (Eppendorf) from 20°C to 80°C applying increments of 0.5°C. Changes in fluorescence were observed simultaneously with excitation and emission wavelengths of 470 nm and 550 nm, respectively. To obtain the binding constant, a Boltzmann model was used to fit the fluorescence data (Equation 4.4).

$$y = Top + \frac{Bottom - Top}{1 + e^{\frac{x - K_D}{dx}}} \quad (4.4)$$

The thermal shift measurements were carried out in quadruplets by Dr. Claire Basquin. The dissociation constant was calculated from the average of the K_{DS} from each of the four binding isotherms.

4.3.15 Fluorescence Anisotropy

Fluorescence Anisotropy is a method to monitor the binding interaction of two species applying equilibrium titration experiments and relies on the change in anisotropy when two species are interacting. The anisotropy is measured as the difference in polarization along different axes of

the fluorophore, applying 5'-6-carboxy-fluorescein (5'-FAM)-labeled Holliday Junction, (HJ with 16bp armlengths), nicked Holliday Junction (HJ with 16 bp armlength and nick at 2 bp from the junction point on the exchanging strand) and 32 bp dsDNA (Table 4.28). Fluorescence anisotropy measurements were carried out at room temperature with Infinite M1000 Pro (Tecan, Männedorf, Switzerland) using Tecan i-Control software. The DNA substrates were used at a final concentration of 11.2 nM and were incubated at room temperature for 10 min with varying concentrations of ceSLX4^{SAP} and ceMUS81^N alone as well as in complex using assay buffer (20 mM Potassium chloride pH 6.5, 100 mM KCl and 1 mM TCEP). The excitation and emission wavelengths were 470 nm and 535 nm, respectively. Each titration point was measured three times with 10 reads at an integration time of 40 μ s. Measurements were made in triplicate. The data for ceSLX4^{SAP} and ceSLX4^{SAP}-ceMUS81^N were analyzed by non-linear regression using the Hill function (Equation 4.5) included in the program Origin (Origin Lab; <http://www.originlab.com>).

$$A_{total} = A_0 + (A_{max} - A_0) \frac{x^n}{k^n + x^n} \quad (4.5)$$

A_{total} is the measured anisotropy, A_0 is the intrinsic anisotropy of the DNA substrate, A_{max} is the anisotropy of the saturated protein-DNA complex, n is the Hill coefficient and x is the concentration of the protein. For ceMUS81^N a two-site binding mode was assumed and curves were fit with equation 4.6.

$$A_{total} = \frac{A_{max1} \cdot x}{k_1 + x} + \frac{A_{max2} \cdot x}{k_2 + x} \quad (4.6)$$

A_{total} is the measured anisotropy, A_{max1} is the maximal anisotropy for the first binding site, A_{max2} is the maximal anisotropy for the second binding class and x is the concentration of the protein. To assess the quality of the non-linear regression, residual plot analysis was performed and the agreement between the observed and calculated binding isotherms was evaluated.

4.3.16 Nano Differential Scanning Fluorimetry

Nano differential scanning fluorimetry (nanoDSF) is a method to determine protein stability by testing different buffers using the intrinsic fluorescence of tryptophan and tyrosine residues in a protein. The fluorescence of Trp and Tyr residues is used as a measure for protein stability. For the experiment, the protein with an end concentration of 0.5 mg/ml was mixed with the buffer in a volume of 10 μ l. Different buffer substances, pH values and salt concentrations were tested. Samples were taken up into Prometheus NT.48 glass Capillaries (Nanotemper Technologies GmbH, Munich, Germany). Measurements were carried out on a Prometheus NT.48 by Nanotemper TEchnologies GmbH (Munich, Germany) with a laser intensity of 30 % scanning a temperature range from 20 - 90°C with a ramp of 1°C / min. The ratio of fluorescence at 350 and 330 nm was plotted against the temperature using the program PR.ThermControl (Nanotemper Technologies GmbH). The melting temperature of the proteins was determined for each tested condition by the minimum or maximum of the peak of the first derivative. For the crystallization trials of the ceSLX4^{SAP} constructs buffer conditions with low melting temperature and thus low protein stability were applied.

4.3.17 Crystallization Procedures

Initial Screening

Protein samples were purified and subjected alone or in complex with various DNA substrates for crystallization. The design of the DNA substrates was based on the 16bp arm length Holliday Junction as used for EMSAs (see Table 4.28). For the screening of the HJ arm length the length of the single oligos was varied keeping either four-fold symmetry or two-fold symmetry. Initial crystallization experiments were performed in 96 well sitting drop plates using commercially available or self-made screens. For initial screening, 100 nl protein solution was mixed with 100 nl reservoir solution using a Phenix nanodispenser robot (Art Robbins Instrument, Sunnyvale, CA, USA). Initial hits were optimized in 96-well sitting drop plates or in 72-well Terasaki microbatch plates (Molecular Dimension, Newmarket, Suffolk, UK). The crystallization plates except for the microbatch plates were automatically imaged by the imaging system of the crystallization facility (Max-Planck Institute for Biochemistry, Martinsried). The initial screening plates as well as deep well plates with screening buffer conditions were set up by Dr. Karina Valer-Saldaña or Sabine Pleyer from the Crystallization Facility (MPIB). Crystal hits were checked for protein content with the UV fluorescence microscope JANSi UVEX-m scanning in near UV (300 - 400 nm) zone.

Microbatch under oil crystallization experiments were carried out for example for the complex of ceEMEI¹⁶¹⁻⁴⁵⁴-ceMUS81^{K270M}-ceSLX4⁴⁶⁷⁻⁶⁰⁶ with various DNA substrates tested. Drops of 1 μ l protein sample with 1 μ l reservoir were setup under a layer of 50% paraffin oil and 50% silicone oil in 72-well Terasaki microbatch plates (Molecular Dimensions, Newmarket, UK).

To reduce vapor diffusion, oil was layered on the reservoir of 96-well sitting drop crystallization trials. For each reservoir, 70 μ l of a mixture of 80% paraffin oil and 20% silicone oil was put on the reservoir.

Matrix-Microseeding

Matrix-Microseeding was carried out according to protocol from Hampton Research Seed Bead Kit and drops were setup with 50 nl seed stock, 150 nl protein and 200 nl reservoir. The primary screening temperature was 20°C, infrequently plates were set up at 4°C.

Formulation Buffer Screening

The *C. elegans* SLX4 constructs of the C-terminus expressed very well and the proteins were highly soluble after purification. Thus, high protein concentrations could be achieved for crystallization which lead to difficulties with pipetting when setting up crystallization trials. Therefore different formulation buffers were tested in a nanoDSF experiment (see Section 4.3.16).

Cryo-Protection

For cryo-protection, 0.5 μ l cryo-protectant (list of cryo-protectant used, see Table 4.29) was added to the drop with the corresponding crystals and in subsequent steps another 1 and 2 μ l cryoprotectant was added. In the end, the crystals were in the respective cryo-protectant and were picked up from the cryo-protectant solution using nylon cryo-loops with suitable size (Hampton Research, Aliso Viejo, CA, USA) and plunged into liquid nitrogen.

Table 4.29: List of the respective cryoprotectants used for the protecting the crystals at the synchrotron.

Protein-DNA Complex	Cryo-Protectant Solution
hsEMEI ¹⁷⁸⁻⁵⁷⁰ -MUS81 ²⁴¹⁻⁵⁵¹ + 17-5'flap	30% ethylene glycol, 0.1 M Mes pH 6.5, 5.5% ethanol
ceEMEI ¹⁶¹⁻⁴⁵⁴ -MUS81 ^{K270M} + ceSLX4 ⁴⁶⁷⁻⁶⁰⁶ + 4w4xn16n1	20 % PEG 3350, 0.2 M Calcium acetate, 34% glycerol
ceEMEI ¹⁶¹⁻⁴⁵⁴ -MUS81 ^{D268A} + ceSLX4 ⁴⁶⁷⁻⁶⁰⁶ + 4w4xn17n3	1.4 M Sodium/Potassium Phosphate, 0.1 M Sodium Malonate, 25% glycerol
ceEMEI ¹⁶¹⁻⁴⁵⁴ -MUS81 ^{D268A} + ceSLX4 ⁴⁶⁷⁻⁶⁰⁶ + 4w4x17n3	27% PEG MME500 0.1 M Bis-Tris pH 5.5, 0.05 M Magnesium chloride, 17.5% glycerol
ceMUS81 ¹⁻¹⁴⁰	50mM Tris pH 8, 0.2M Magnesium chloride, 35% PEG3350, 15% glycerol

Data Collection

Data of the plunged crystals were collected at 100°K at the Swiss Light Source (Villingen, Switzerland) by Dr. Jerome Basquin using the PXIII beamline and a MarCCD detector.

Molecular Replacement

The data obtained from PXIII beamline was processed using XDS [324]. Molecular replacement was carried out using Phaser [318] and the published structure of hsMUS81-EMEI bound to DNA substrate (protein data base entry 4P0P) as a search model. Refinement was carried out using Phenix [319]. The model was visualized and compared to the published structure applying coot [313].

4.3.18 NMR Spectroscopy

Sample Preparation

Proteins were expressed in M9 minimal medium supplemented with ¹⁵N ammonium chloride, 0.4% glucose, 1 mM MgSO₄, 0.3 mM CaCl₂, 1 μg biotin, 1 μg thiamin and 1x trace elements solution (0.134 mM EDTA, 31 μM FeCl₃-6h₂O, 0.62 μM ZnCl₂, 0.76 μM CuCl₂-2H₂O) according to the protocol by Dr. Arie Geerlof, Helmholtz Center Munich. Bacterial cells transfected with the respective plasmid were plated on agar plates containing appropriate antibiotics. A 20 ml overnight pre-culture in LB medium was grown and cells were pelleted for 10 min at 3220 g. The supernatant was discarded and the cell pellet was washed with 5 ml sterile water and resuspended in M9 media. The bacterial cells were grown until OD₆₀₀ 0.8 at 37° C. Afterwards, the temperature was reduced to 20° C and the bacterial cells were induced with the respective amount of IPTG. Cells were harvested like in section 4.2.2 and purification was carried out like described for ceSLX4^{SAP} and ceMUS81^N, except after Ni TED column the elution fraction was cleaved over night and then subjected to cation exchange on a SP-Sepharose column and gel filtration with buffer SEC 9. The labelled proteins were concentrated to amounts of higher than 300 μM and supplemented with 10% heavy water D₂O. Proteins with ¹⁵N-¹³C label were expressed in M9 medium as well, but in addition glucose was replaced by ¹³C-glucose.

NMR Experiments

NMR experiments were carried by Dr. André Mourão at the Helmholtz Center Munich. NMR spectra were acquired at 298 K on a Bruker Avance III 600 and 800-MHz equipped with a cryoprobe. The protein samples were measured in 3 mm NMR tubes. Data analysis was performed using NMRPipe. ^1H , ^{15}N and ^{13}C backbone resonance assignments were obtained from the following experiments: HNCO, HNCA, HNCACB, CBCACONH and ^{15}N -NOESY experiments were carried out.

Chemical Shift Perturbations

Chemical shift perturbations were performed with 110 μM ^{15}N -labelled ceSLX4^{SAP} or ^{15}N -labelled ceMUS81^N in a two dimensional ^{15}N - ^1H - HSQC (heteronuclear single quantum coherence) experiment. In this experimental setup each amino acid residue (except for proline) is described by one peak in the spectrum. Each peak is defined as amide ^{15}N shift (ω_1) and ^1H shift (ω_2). The chemical shift of a peak depends on the chemical surrounding of the amino acid amide group. Furthermore, the side chain amides of glutamate and asparagine are represented in the spectrum. Ideally, a folded protein will generate a ^{15}N - ^1H - HSQC spectrum with a broad distribution of well separated signals. When a ligand binds to the protein, the environment of a spin will change and this change gives rise to chemical shift changes in the NMR spectrum. Residues involved in ligand binding are expected to display a larger chemical shift than residues more far away. Through these changes the interface of a protein can be mapped.

Chemical shift perturbations obtained from ^1H - ^{15}N HSQC titrations were calculated based on the equation:

$$\Delta\delta = \sqrt{(\Delta\delta^1H)^2 + \left(\frac{\Delta\delta^{15N}}{6}\right)^2} \quad (4.7)$$

$\Delta\delta^1\text{H}$ and $\Delta\delta^{15}\text{N}$ are the ^1H and ^{15}N chemical shift in parts per millions.

4.3.19 Structural Model Calculation

The calculation of the structural model was carried out by Dr. André Mourão at the Helmholtz Center Munich similar to an already published approach [332]. A structural model for ceSLX4^{SAP} was calculated using the Chemical-Shift-ROSETTA (CS-ROSETTA Version 4.8) server from the Biological Magnetic Resonance Data Bank Rosetta. CS-ROSETTA is a tool for *de novo* protein structure generation, using ^1H , ^{15}N and ^{13}C chemical shifts as input. CS-ROSETTA employs SPARTA-based selection of protein fragments from the PDB, in conjunction with a regular ROSETTA Monte Carlo assembly and relaxation procedure, to generate structures of minimized energies. The NMR model was visualized and analyzed using pymol [322].

Structure Alignment

For structural comparison the Dali tool was applied and the ceSLX4^{SAP} NMR model was searched against all known structures [276, 277]. A DALI search [276] for all 10 models retrieved in total 147 entries with 58 unique entries. Eight entries with Z-scores higher than 3 and similar biological function were analyzed closer. Structural superimpositions and sequence alignments were carried

out with Dali [276]. For the molecules 1jeg, 4rec, 1zbh, 2do1 and 4uzw the structural alignment was further optimized by Pymol using the command *align*. Molecular representations were generated using the Pymol software [322]. Electrostatic surface potentials were calculated with PDB2PQR [317] and APBS [311].

Acknowledgements

Firstly, I would like to express my sincerest gratitude to my supervisor Dr. Christian Biertümpfel for his expertise, enthusiasm and motivation. I thank him for giving me the chance to work on the SLX4 project, for supervising me and trusting me. Thanks a lot, Christian, for all your scientific and personal support and your constructive feedback on my work. Moreover, I would like to particularly thank my head of the thesis committee, Prof. Dr. Elena Conti, for her constant feedback, support and help on my project. I thank both Prof. Dr. Elena and Christian for giving me the opportunity to work in such an outstanding department of Structural Cell Biology at the Max-Planck Institute of Biochemistry and being my TAC members. I am grateful for the structural and positive feedback I received from my other TAC members, Prof. Dr. Andreas Ladurner and Dr. Stephan Gruber. I want to thank all my TAC members for their informative and helpful comments and encouragement.

I am very thankful to Dr. André Mourão and our fruitful collaboration. He showed me for the first time some signals of my protein and pointed out that NMR is a good option for my proteins. Furthermore, he provided me with a first model of the SAP domain and I am very grateful for all his suggestions for further experiments based on our initial NMR measurements. I also want to acknowledge Prof. Dr. Wade Harper (Harvard Medical School) and Dr. Karsten Klage (MPIB) for sharing the initial human SLX4 plasmid and *C. elegans* cDNA, respectively.

I have benefited a lot from the resources and infrastructure of the Biochemistry Core Facility. The people in the core facility do not only offer a broad range of methods and reliable analysis, but they are always open for professional discussion and help. Here, I want to thank especially Dr. Stephan Uebel, who is an endless resource of benefit for the whole institute and a giver of very good scientific advice. Thanks for discussing model fitting with me. I highly appreciate the expertise of Lissy Weyher in spectroscopy and mass spectrometry. I want to thank Dr. Nagarjuna Nagaraj and Victoria Sanchez for peptide fingerprinting and countless total mass analysis.

Furthermore, I was very happy to be a part of the International Max-Planck Research School for Molecular Life Sciences (IMPRS-LS). I appreciate the support I received from the coordination office, especially from Dr. Ingrid Wolf. I benefitted a lot from the offered workshops and activities.

My thanks extend to all present and past members of the Department of Structural Cell Biology for the great scientific support and the enjoyable lab atmosphere. Without the encouragement of Dr. Claire Basquin and her comprehensive biophysical knowledge, I would have definitely missed some methods to try. I enjoyed discussing with you and I'm very grateful for the time you spent on my project. I want to thank Dr. Jérôme Basquin, Dr. Karina Valer-Saldaña and Sabine Pleyer for running the crystallization Facility in such a smooth and efficient manner. Thanks for setting up plenty crystallization drops and discussing special needs. I thank Jérôme for testing my crystals

on various synchrotron trips. My gratitude goes to Ariane Fischer for running many expression tests and for helpful discussions on different expression conditions. I'm very happy to have you also as a friend and I enjoyed lunch time so much with you. I appreciate the careful and critical reading of my manuscript as well as the scientific discussions with Dr. Sebastian Falk a lot. I also thank Dr. Iuliia Iermak and Dr. Julia Adam for reading my manuscript. I am grateful for the technical support I got from Walter Erhardt and Peter Reichelt. Walter, without constant battery exchange on my pipettes the whole assays would have been impossible. Thank you for all the IT support on various computer problems. You are missed badly. Peter, your patient help and maintenance on chromatographic columns and purification system is invaluable. Both of you are the good-hearted soul of the Conti department. Petra Lee and Ulrike Goldschmitt were great helpers in all organizational issues. A big thank you goes to Judith Ebert, who is an incredible lab manager and supported me in many ways. Without our cleaning kitchen I would have spent hours on cleaning glass ware. I'm very grateful for the ladies, who are doing such an important job for us. I'm very happy to have met Sylvia, who is a very positive person and I enjoyed talking to her in the early morning hours.

Moreover, I want to thank Dr. Shun-Hsiao Lee and Giulia Chapparini for being a great lab member and for all the scientific discussions we had. My gratitude goes to our former lab technician Maren Klüegel, who helped building up the lab, and our present technician Marcus Hammerl, who both assisted me a lot in all aspects of cloning, expression and purification. My gratitude extends to the interns Jerome Schnittger, Katharina Bayer, Christopher Kuhn, Beata Kaczmarek and Sophia Beslmüller, who joined my project for a limited time. It was a great pleasure not only to teach you but also to learn from you. Furthermore, I want to thank Dr. Naoko Mizuno for excellent scientific input during our lunchmeetings with her group. I thank the Mizuno group for very helpful discussions.

Very special thanks go to the former members of the Lorentzen and Wolf labs for welcoming us in a great way and creating a unique atmosphere on our floor. I enjoyed working a lot with Dr. Kristina Weber and Dr. Melanie Vetter and sharing instruments was a pleasure and fun. Dr. Ira Schmalen and Dr. Justine Witosch were a great source of experience and great advisors for junior PhD students. I want to thank all the four of you for the lasting friendship and still being great companions to date.

Julia, we started at the same day and we have been friends ever since (Chaka!). We had a great time through all the ups and downs of PhD student life. You were for four years my fellow lab member and I thank you a lot for cheering me up and listening.

I express my sincere gratitude to Dr. Rajan Prabu, who is a great advisor and friend. I want to thank you for all your support in mastering the challenges of my PhD.

Ein großer Dank geht auch an die Weber Familie für kulinarischen Support und das Mitfiebern. Ohne meine nicht-wissenschaftlichen Freunde, Rebecca, Anja, Eike-Kristin mit Familien, wäre das Leben so eintönig geworden. Vielen Dank, dass ihr mir abseits vom Labor geholfen habt. Ein großer Dank geht an Thomas Ries für seine Seelsorge.

Zu guter letzt möchte ich mich bei meiner geliebten Omi, meinem Vater und meiner Tante Susanne für ihre große Unterstützung und ihr Zuhören bedanken. Ihr standet mir in all den Jahren des Studiums und der Promotion immer mit Rat zur Seite. Dieter, dir möchte ich ganz besonders für deine konstante Unterstützung und Aufmunterung danken. Mein Dank an dich ist

grenzenlos. Ohne euren Glauben an mich und die Unterstützung, die ihr mir zuteil habt werden lassen, hätte ich es nicht bisher geschafft.

Abbreviations

53BP1	p53-binding protein 1
A	Alanine
aa	Amino acid
AESBF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
AEX	Anion exchange chromatography
AP	Alkaline Phosphatase
APLF	Aprataxin and polynucleotide kinase/phosphatase-like factor
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
ATR	ATM and Rad3-related
ATRIP	ATR interacting protein
BER	Base excision repair
BIR	Break-induced repair
BLM	Boom syndrome RecQ-like helicase
β-ME	β -mercaptoethanol
BRCA	breast cancer susceptibility protein 1
BSA	Bovine serum albumin
BTB	Bric a brac, Tramtrack, Broad-complex
BTP	1,3-bis(tris(hydroxymethyl)methylamino)propane
CCD	conserved C-terminal domain
Cdc	Cell division cycle
CDK	Cyclin-dependent kinase
ce	<i>Caenorhabditis elegans</i>
<i>C. elegans</i>	<i>Caenorhabditi elegans</i>
CHK2	Checkpoint kinase 2
CPD	Cysteine protease domain
CtBP	CtBP C-terminal binding protein
CtIP	CtBP-interacting protein
DDR	DNA damage response
D-loop	Displacement loop
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
Dpb11	DNA polymerase B II
ds	Double-stranded
DSB	Double-strand break
DTT	Dithiothreitol
E	Glutamate
EDTA	Ethylenediaminetetraacetic acid

EME1	Essential meiotic endonuclease 1
EMSA	Electrophoretic mobility shift assay
ERCC	Excision repair cross complementation group
ESI-TOF MS	Electron spray ionization - time of flight mass spectrometry
FA	Fanconi anemia
FAAP24	FA-associated protein 24 kDa
FAM	Fluorescein amidite
FAN1	FANCD2-associated nuclease 1
FANCD2	Fanconi anemia complementation group D2
FANCM	Fanconi anemia complementation group M
Fast-AP	Fast alkaline phosphatase
FCS	Fetal calf serum
FEN1	Flap endonuclease
FL	Full length
GEN	Gap endonuclease
GFP	Green fluorescent protein
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HhH	Helix-hairpin-helix
HIM	High incidence of males
HR	Homologous recombination
ICL	Interstrand crosslink
IMAC	Immobilized metal ion affinity chromatography
IP	Interacting protein
IPTG	Isopropyl- β -D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
KCl	Potassium chloride
K_D	Dissociation constant
LIC	Ligase independent cloning
MEI	meiotic protein
MES	2-(N-morpholino)ethanesulfonic acid
MgCl₂	Magnesium chloride
MgSO₄	Magnesium sulfate
MLR	MUS312-MEI9 interaction-like region
MMEJ	Microhomology mediated end joining
MMR	Mismatch Repair
MMS	Methyl methane sulfonate
Mms4	methyl methane sulfonate sensitive 4
MPIB	Max-Planck Institute for Biochemistry
MRE11	Meiotic recombination 11
MRN	MRE11-RAD50-NBS1
MS	Mass spectrometry
MSH	MutS homolog
MUS81	Methyl methanesulfonate UV sensitive, clone 81
MutS	Mutator S
NaCl	Sodium chlorid
NBS1	Nijmegen breakage syndrome 1
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
Ni	Nickel

NP40	Nonidet P-40, octylphenoxypolyethoxyethanol
nt	Nucleotide
NTA	Nitrolotriactic acid
PAGE	Polyacrylamide gel electrophoresis
PARP	Poly(ADP-ribose) polymerase
RAD	Radiation sensitive
RIF1	Replication timing regulator factor 1
RING	Really interesting new gene
RPA	Replication protein A
PALB2	Partner and localizer of BRCA2
PBS	Phosphate buffered saline
PCNA	Proliferative cell nuclear antigen
PCR	Polymerase chain reaction
PEI	Polyethyleneimide
PEG	Polyethylene glycol
pI	isoelectric point
PKcs	Protein kinase catalytic subunit
PLK1	Polo like kinase 1
pol	Polymerase
PPT	Phosphatase
RAG	Recombination activating gene
RF	Replication fork
RING	Really interesting new gene
RMI	RecQ mediated genome instability protein
RPA	Replication protein A
rpm	Revolutions per minute
RTEL	Regulator of telomere length
Rtt107	Regulator of Ty1 transposition protein 107
Ruv	Resistance to UV light
SAF	Scaffold attachment factor
SAP	SAF-A/B, Acinus and PIAS domain
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i> , budding yeast
SDS	Sodium dodecyl sulfate
SDSA	Synthesis-depedent strand annealing
SEC	Size exclusion chromatography
Sgs	Slow growth suppressor
SIM	SUMO-interacting motif
SLX	Synthetic lethal of unknown function
SM	SLX4-SLX1-MUS81-EME1
<i>S. pombe</i>	<i>Schizosaccharomyces pombe</i> , fission yeast
ss	Single-stranded
SSA	Single-strand annealing
SSBR	Single-strand break repair
SUMO	Small ubiquitin modifier
TAE	Tris-acetate-EDTA
TLS	Translesion synthesis
Top	Topoisomerase
TOPBP1	DNA topoisomerase 2 binding protein

Tris	Tris(hydroxymethyl)aminomethane, 2-Amino-2-(hydroxymethyl)-propane-1,3-diol
TRF	Telomeric repeat-binding factor
UBZ	Ubiquitin-binding zinc finger domain
UV	Ultraviolet
VDJ	Variable diverse joining
VPA	Valproic acid
WH	Winged helix
XLF	XRCC4-like factor
XPF	Xeroderma pigmentosum complementation group F
XPG	Xeroderma pigmentosum complementation group G
X-ray	Röntgen radiation
XRCC	X-ray cross complementation chinese hamster gene

Appendix

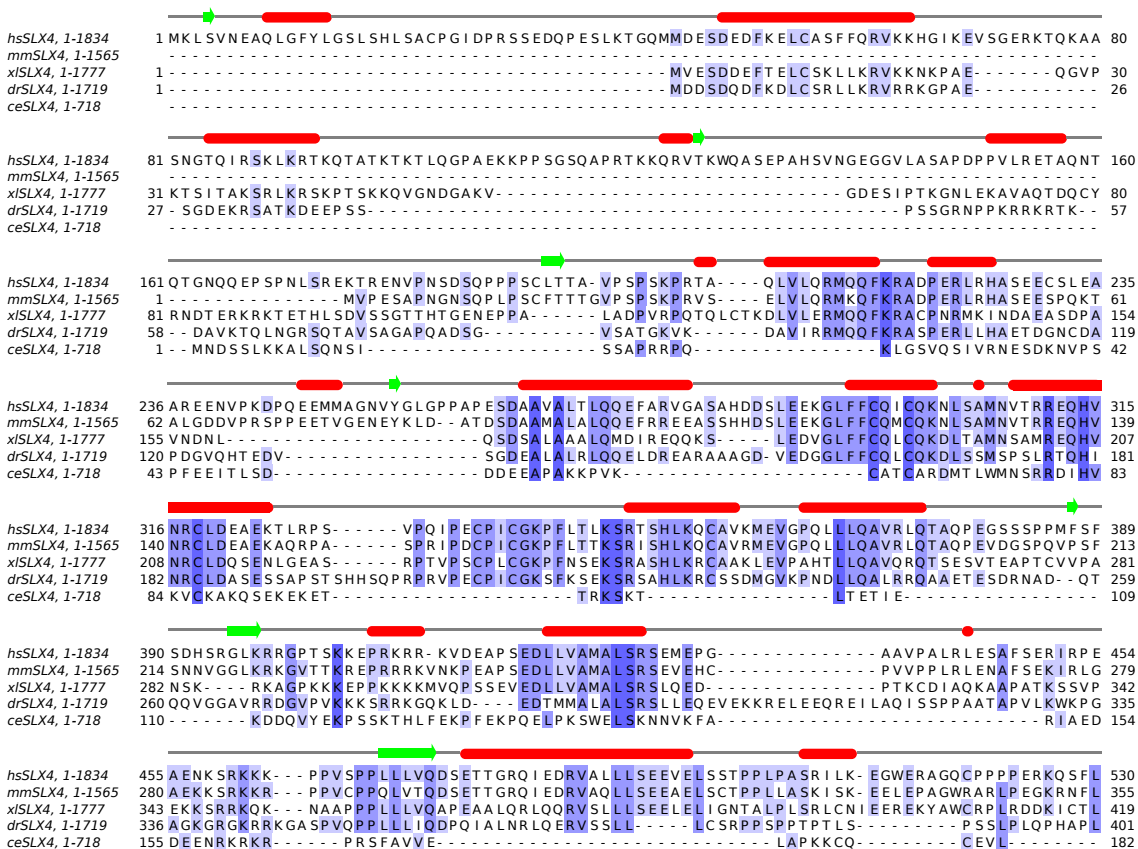


Figure 4.1: See next page for figure legend

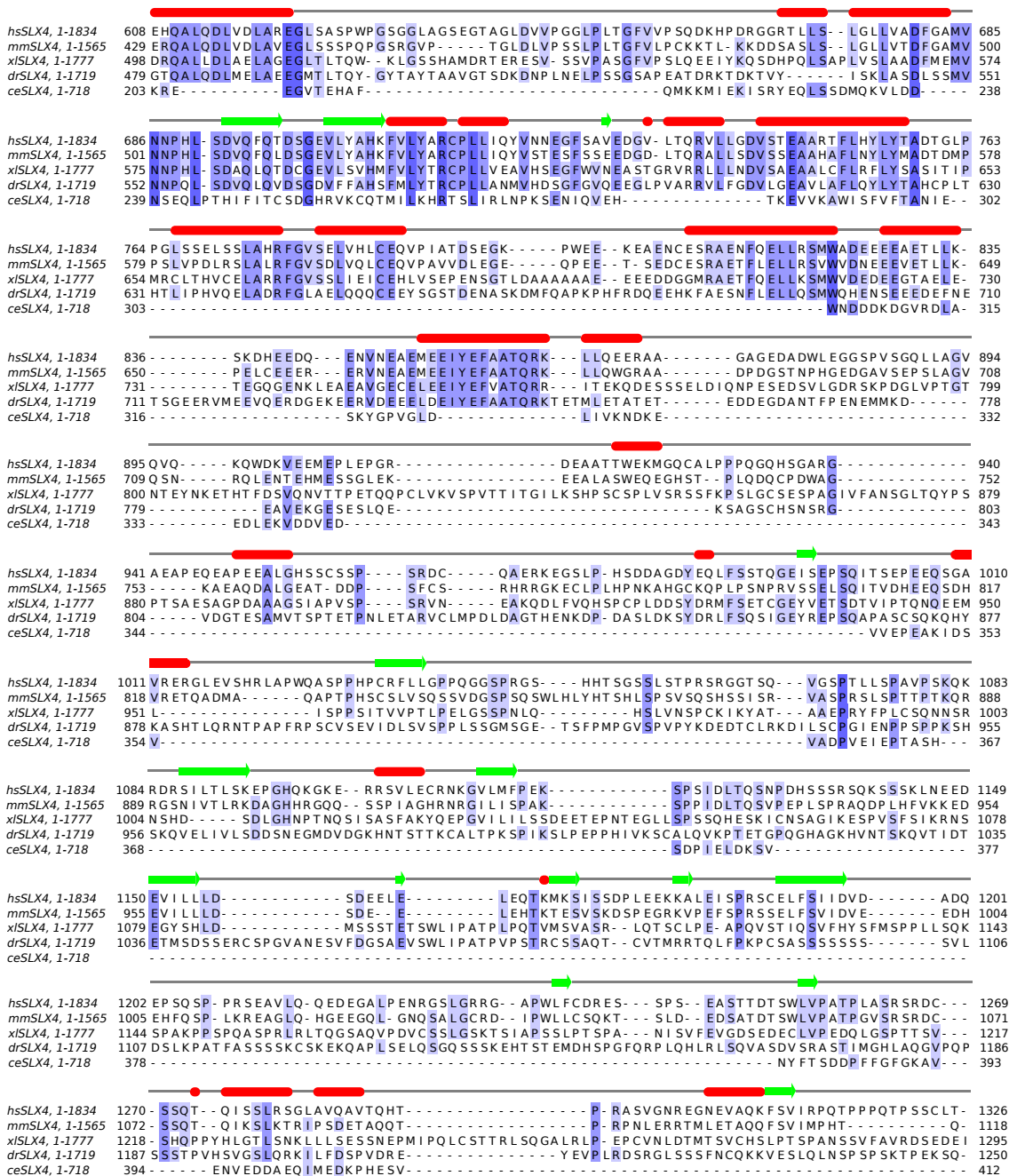
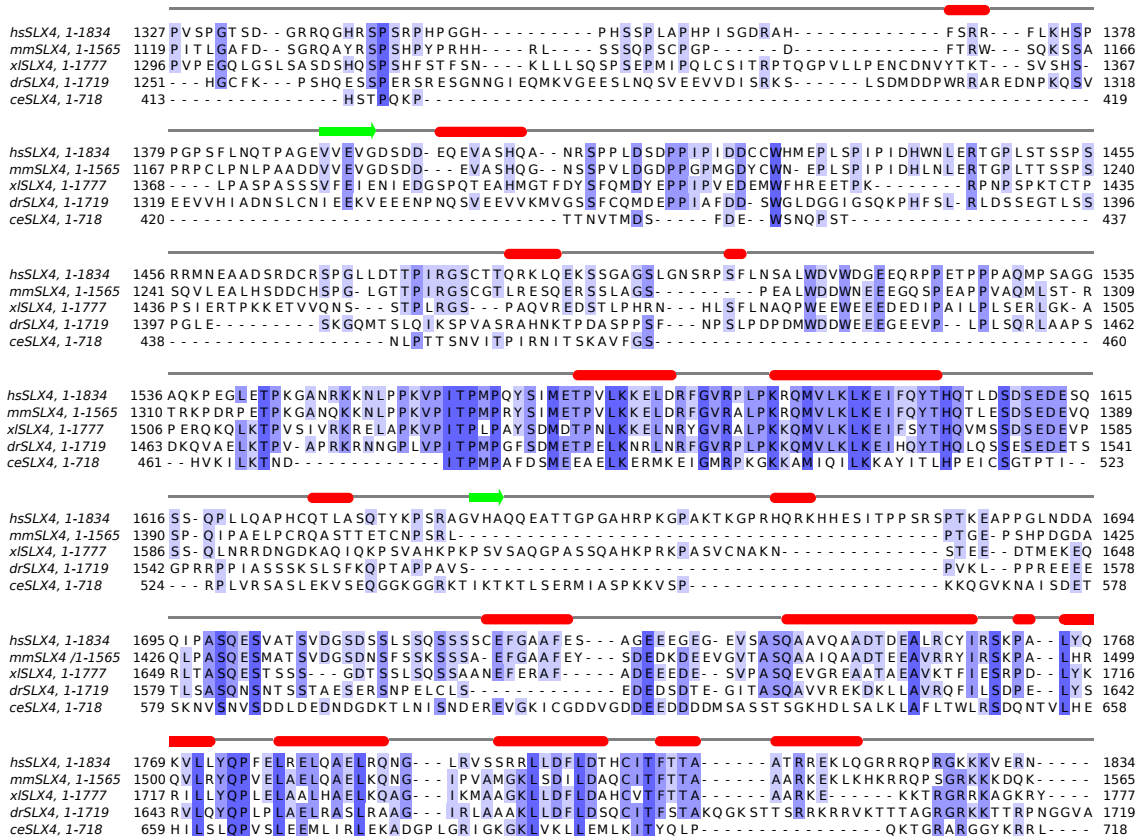


Figure 4.1: See next page for figure legend



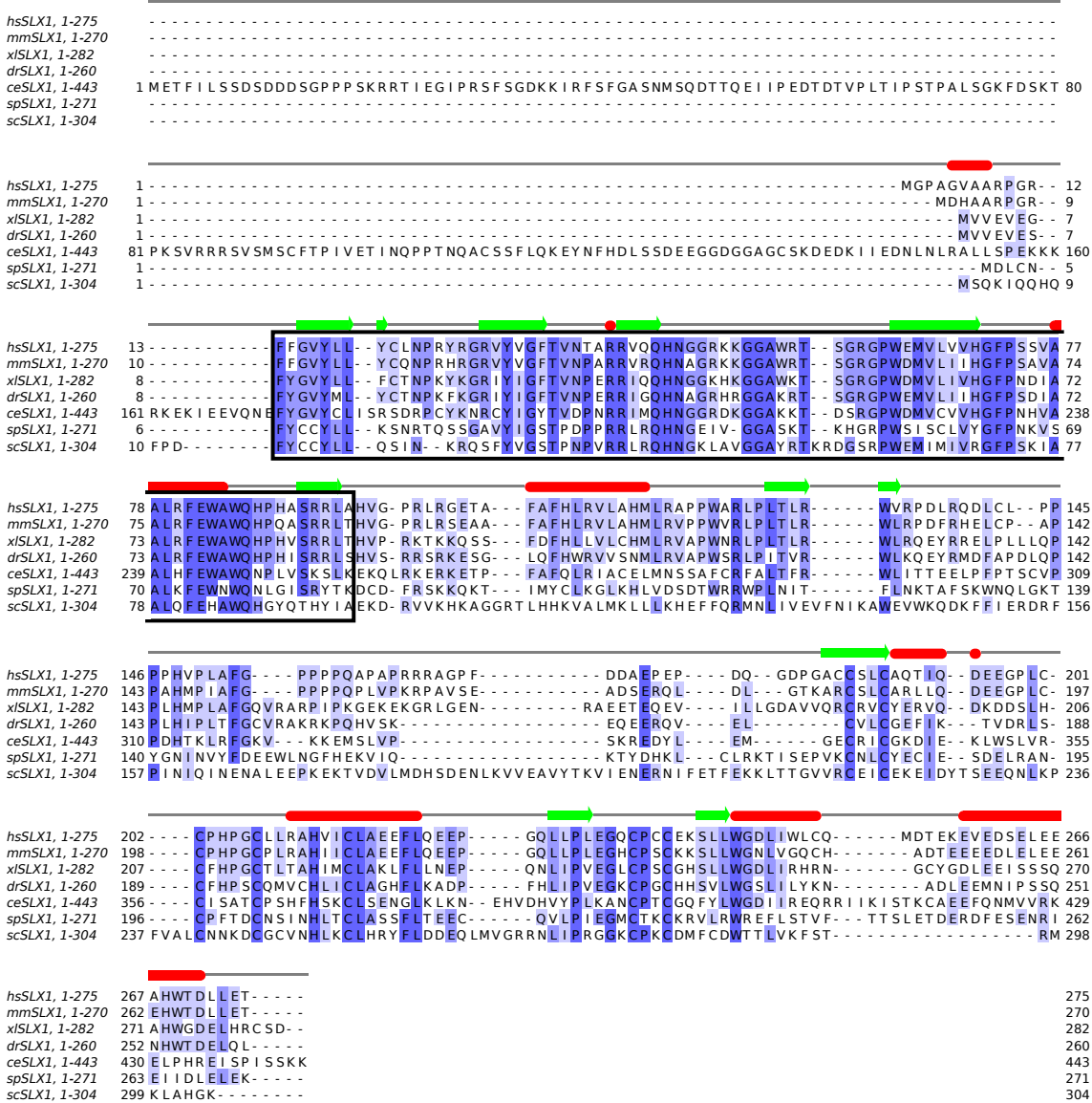


Figure 4.2: Multiple sequence alignment of SLX1 orthologs. Dark blue indicates strictly conserved residues, white indicates variable residues. The black box indicates the GIY-YIG domain. The orthologs were aligned with MUSCLE [215] and displayed using Jalview [216]. Secondary structure elements are indicated by red bars (alpha-helical) and green arrows (beta-sheets) and were calculated using Jpred [325]. The alignment shows the sequences of *Homo sapiens* (*hs*), *Mus musculus* (*mm*), *Xenopus leavis* (*xl*), *Danio rerio* (*dr*), *Caenorhabditis elegans* (*ce*), *Shizosaccharomyces pombe* (*sp*) and *Saccharomyces cerevisiae* (*sc*). Uniprot accession numbers [217] from top to bottom: Q9BQ83, Q8BX32, Q0IH86, A0A0R4IAZ1, P91351, Q9P7M3 and P38324.

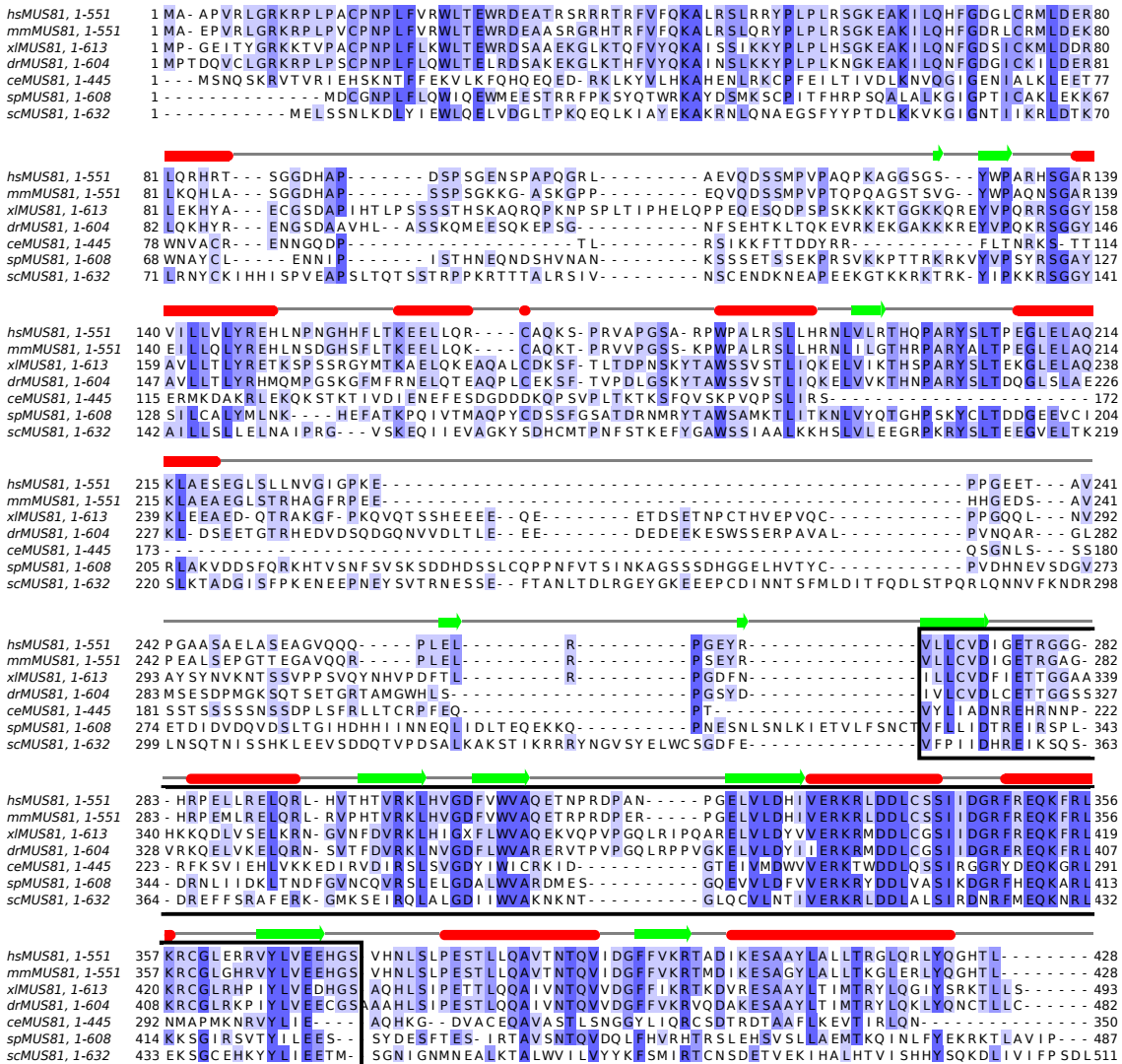


Figure 4.3: See next page for figure legend

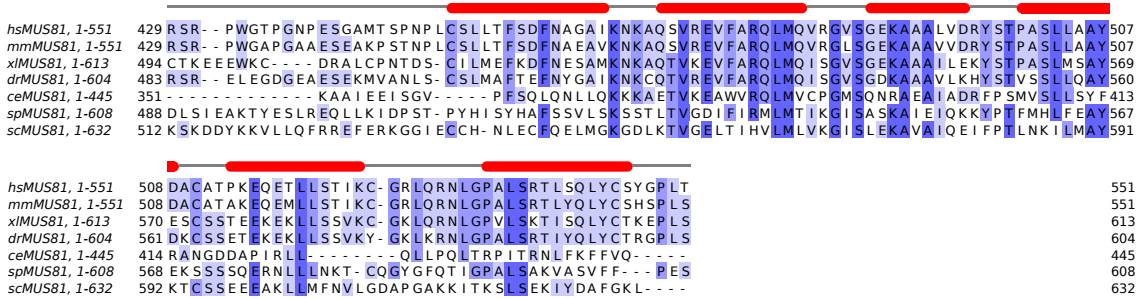


Figure 4.3: Multiple sequence alignment of MUS81 orthologs. Dark blue indicates strictly conserved residues, white indicates variable residues. The black box indicates the ERCC4 domain. The orthologs were aligned with MUSCLE [215] and displayed using Jalview [216]. Secondary structure elements for hsMUS81 are indicated by red bars (alpha-helical) and green arrows (beta-sheets) and were calculated using Jpred [325]. The alignment shows the sequences of *Homo sapiens* (*hs*), *Mus musculus* (*mm*), *Xenopus laevis* (*xl*), *Danio rerio* (*dr*), *Caenorhabditis elegans* (*ce*), *Shizosaccharomyces pombe* (*sp*) and *Saccharomyces cerevisiae* (*sc*). Uniprot accession numbers [217] from top to bottom: Q96NY9, Q91ZJ0, A0A068ER06, Q7SXA9, P91153, P87231 and Q04149.

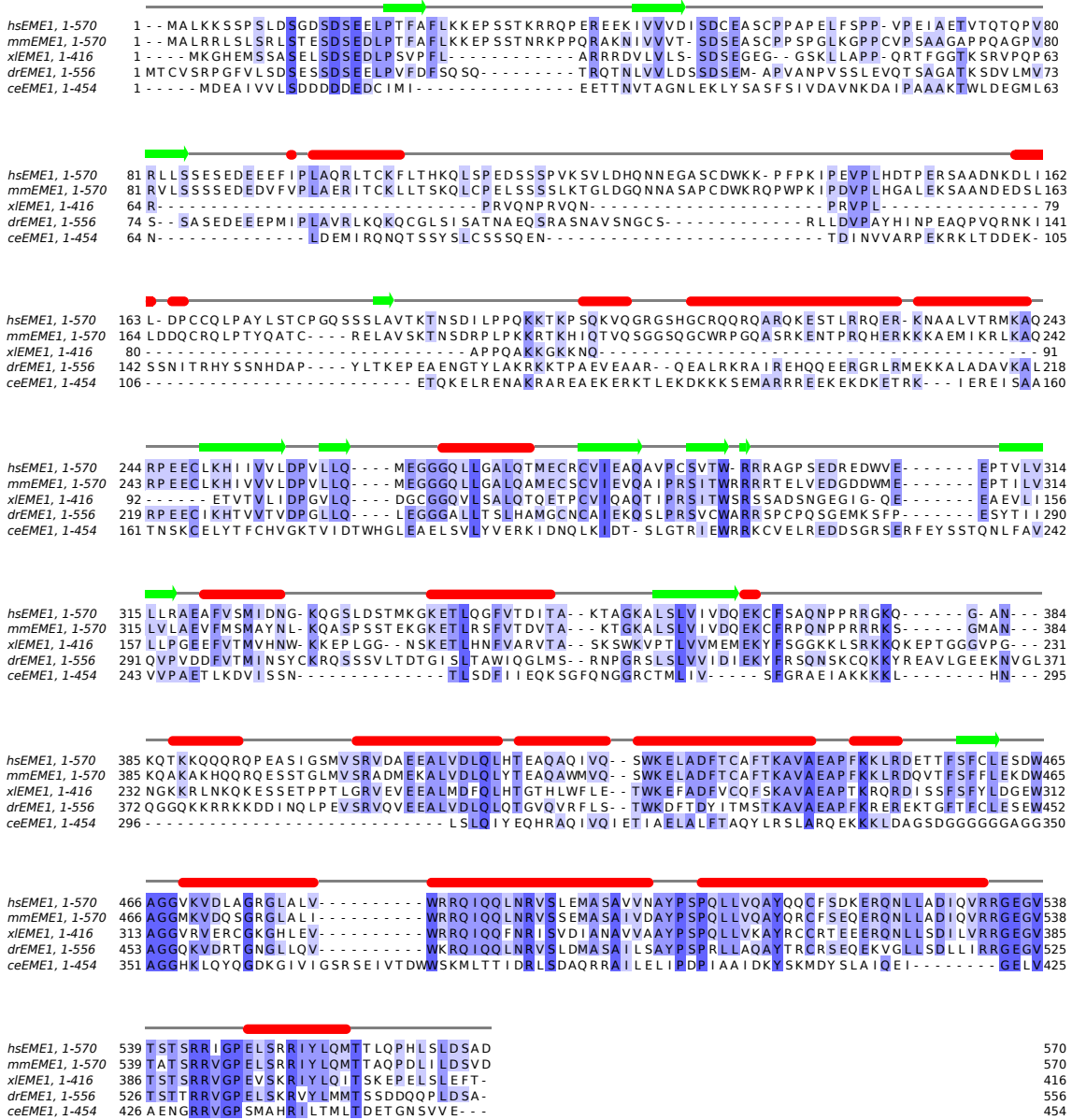


Figure 4.4: Multiple sequence alignment of EME1 orthologs. Dark blue indicates strictly conserved residues, white indicates variable residues. The orthologs were aligned with MUSCLE [215] and displayed using Jalview [216]. Secondary structure elements are indicated by red bars (alpha-helical) and green arrows (beta-sheets) and were calculated using Jpred [325]. The alignment shows the sequences of *Homo sapiens* (hs), *Mus musculus* (mm), *Xenopus laevis* (xl), *Danio rerio* (dr), *Caenorhabditis elegans* (ce), *Shizosaccharomyces pombe* (sp) and *Saccharomyces cerevisiae* (sc). Uniprot accession numbers [217] from top to bottom: Q96AY2, Q8BJW7, A9ULX4, E7F6N8, Q9ZG5, Q9C103 and P38257.

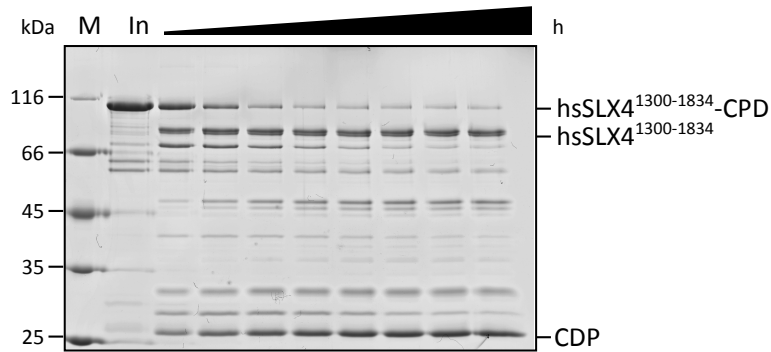


Figure 4.5: Degradation of hsSLX4¹³⁰⁰⁻¹⁸³⁴ after IP6 cleavage. hsSLX4¹³⁰⁰⁻¹⁸³⁴ was expressed with a CPD-tag in bacterial cells. After induction of CPD-cleavage protein samples of 18 µg were taken every half an hour (0.5, 1, 1.5, 2, 2.5, 3, 3.5 and 4 h). In Input sample

P2 P1 P1' P2' P3' P4'							P2 P1 P1' P2' P3' P4'								
X	L	A	G	G	K	Consensus	X	L	A	G	G	K	Consensus		
1324	C	L	T	P	V	S	1329	1594	K	L	K	E	I	F	1599
1356	P	L	A	P	H	P	1361	1605	T	L	D	S	D	S	1610
1373	F	L	K	H	S	P	1378	1619	P	L	L	Q	A	P	1624
1383	F	L	N	Q	T	P	1388	1620	L	L	Q	A	P	H	1625
1414	P	L	D	S	D	P	1419	1628	T	L	A	S	Q	T	1633
1432	P	L	S	P	I	P	1437	1689	G	L	N	D	D	A	1694
1442	N	L	E	R	T	G	1447	1713	S	L	S	S	Q	S	1718
1448	P	L	S	T	S	S	1453	1755	A	L	R	C	Y	I	1760
1471	G	L	L	D	T	T	1476	1765	A	L	Y	Q	K	V	1770
1472	L	L	D	T	T	P	1477	1770	V	L	L	Y	Q	P	1775
1486	K	L	Q	E	K	S	1492	1771	L	L	Y	Q	P	F	1776
1497	S	L	G	N	S	R	1502	1777	E	L	R	E	L	Q	1782
1505	F	L	N	S	A	L	1510	1780	E	L	Q	A	E	L	1785
1509	A	L	W	D	V	W	1514	1784	E	L	R	Q	N	G	1789
1541	G	L	E	G	P	K	1546	1789	G	L	S	R	R	L	1797
1553	N	L	P	P	K	V	1558	1796	R	L	L	D	F	L	1801
1573	V	L	K	K	E	L	1578	1797	L	L	D	F	L	D	1801
1577	E	L	D	R	F	G	1582	1800	F	L	D	T	H	C	1806
1585	P	L	P	K	R	Q	1590	1817	K	L	Q	G	R	R	1822
1592	V	L	K	L	K	E	1597								

Figure 4.6: Possible cleavage sites of the CPD protease on hsSLX4¹³⁰⁰⁻¹⁸³⁴ The cysteine protease domain (CPD) is activated upon addition of IP6 and undergoes cleavage at the P1 Leu site. As the construct hsSLX4¹³⁰⁰⁻¹⁸³⁴ showed degradation bands upon CPD activation, sites of Leu residues were counted to 39 and analyzed regarding the consensus sequence. The most common residues at P2 position (A, G, S or W [252]) were marked red and sum up to nine sites. The chemically similar residue to W was marked in orange. Residues which did not correspond the consensus sequence were not marked.

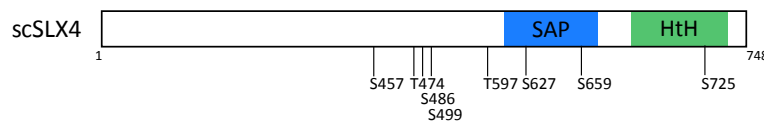


Figure 4.7: Phosphorylation sites in scSLX4. Phosphorylation site S486 is cell cycle regulated by CDK. The other phosphorylation sites are induced through DNA damage by Mec1. Adopted from [261].

Table 4.30: Structural comparison of ceSLX4^{SAP} using Dali server showed 147 entries for other proteins containing a SAP domain. The entries, which were used for further comparison, are marked in red. This table resembles the output retrieved from the Dali server. No. refers to the number in the Dali hit list. PDB and chain gives the protein database number and the chain of the respective molecule. The Z-score is a measure for the homology of the structures. A Z-score above means that the structures are definitely homologous, between 8 and 20 the two structures are probably homologous, between 2 and 8 there is a grey area. A Z-score below 2 is not significant [277]. The rmsd gives information about the root-mean-square deviation of the C α atoms in superposition. The number of the structurally equivalent positions is abbreviated by lali. Nres gives the number of total residues in the molecule chain. % id gives the percentage of amino acid identity in the aligned positions.

No.	PDB and Chain	Z-Score	rmsd (Å)	lali	nres	% id	Description
1	2a8v-C	4.5	3.7	52	118	13	RNA BINDING DOMAIN OF RHO TRANSCRIPTION
2	1xpu-D	4.4	3.7	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
3	1pvo-C	4.3	3.7	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
4	1xpu-E	4.3	3.7	51	407	14	RHO TRANSCRIPTION TERMINATION FACTOR
5	1xpu-B	4.3	3.8	51	408	14	RHO TRANSCRIPTION TERMINATION FACTOR
6	5jji-C	4.2	3.5	51	417	14	TRANSCRIPTION TERMINATION FACTOR RHO
7	5jji-L	4.1	3.8	52	415	13	TRANSCRIPTION TERMINATION FACTOR RHO
8	1xpr-E	4.1	3.8	50	407	12	RHO TRANSCRIPTION TERMINATION FACTOR
9	2do1-A	4.0	3.3	46	55	17	NUCLEAR PROTEIN HCC-1
10	1zbh-D	4.0	4.1	47	289	15	3'-5' EXONUCLEASE ERI1
11	1a62-A	4.0	3.5	52	125	12	RHO
12	1zbh-A	3.9	3.7	47	289	15	3'-5' EXONUCLEASE ERI1
13	1pv4-A	3.9	3.4	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
14	4uzw-A	3.8	5.1	46	50	15	PROTEIN THO1
15	3ice-D	3.8	3.6	48	410	15	TRANSCRIPTION TERMINATION FACTOR RHO
16	1pv4-C	3.8	3.7	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
17	1pv4-D	3.7	3.6	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
18	1xpu-A	3.7	3.4	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
19	3ice-C	3.6	3.6	51	413	10	TRANSCRIPTION TERMINATION FACTOR RHO
20	3ice-B	3.6	3.8	47	408	15	TRANSCRIPTION TERMINATION FACTOR RHO
21	1xpu-C	3.6	3.7	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
22	1a63-A	3.6	3.4	51	130	12	RHO
23	2hjq-A	3.4	4.0	49	111	12	HYPOTHETICAL PROTEIN YQBF
24	5jjk-C	3.4	3.9	48	411	15	TRANSCRIPTION TERMINATION FACTOR RHO
25	5jjk-A	3.4	4.0	47	405	15	TRANSCRIPTION TERMINATION FACTOR RHO
26	5jji-D	3.4	3.9	47	406	15	TRANSCRIPTION TERMINATION FACTOR RHO
27	5jjk-D	3.4	3.8	47	405	15	TRANSCRIPTION TERMINATION FACTOR RHO
28	5jji-D	3.3	3.9	47	406	15	TRANSCRIPTION TERMINATION FACTOR RHO
29	2kw9-A	3.3	4.0	51	75	20	MKL/MYOCARDIN-LIKE PROTEIN 1
30	4rec-A	3.3	2.8	42	593	19	FANCONI-ASSOCIATED NUCLEASE 1
31	1pvo-A	3.2	3.5	51	408	14	MOLECULE: 5'-R(P*UP*C)-3'
32	2kvu-A	3.2	2.5	45	75	22	MKL/MYOCARDIN-LIKE PROTEIN 1
33	4reb-H	3.2	2.7	42	584	19	FANCONI-ASSOCIATED NUCLEASE 1
34	4ria-B	3.2	2.4	43	613	12	FANCONI-ASSOCIATED NUCLEASE 1
35	1v66-A	3.1	1.9	40	65	13	PROTEIN INHIBITOR OF ACTIVATED STAT PRO- TEIN 1
36	1jeq-A	3.1	2.4	42	548	21	KU70
37	2wqg-A	3.1	2.9	41	51	20	THO1
38	2gwm-A	3.1	2.6	49	200	8	65 KDA VIRULENCE PROTEIN
39	4rea-A	3.1	2.5	41	590	20	FANCONI-ASSOCIATED NUCLEASE 1
40	3nts-A	3.0	2.7	49	185	2	VSDC
41	1xpr-F	3.0	3.9	49	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
42	1xpo-A	3.0	3.9	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
43	3d3q-B	3.0	2.4	42	303	7	TRNA DELTA(2)-ISOPENTENYLPYROPHOSPHATE TRANSFERAS
44	3d3q-A	3.0	2.4	42	307	7	TRNA DELTA(2)-ISOPENTENYLPYROPHOSPHATE TRANSFERAS
45	2do5-A	2.9	2.4	40	58	15	SPLICING FACTOR 3B SUBUNIT 2
46	4fml-B	2.9	2.6	49	196	2	VSDC
47	5v2o-E	2.9	2.8	42	60	14	TP2

No.	PDB and Chain	Z-Score	rmsd (Å)	lali	nres	% id	Description
48	3c95-A	2.9	2.9	42	446	14	EXODEOXYRIBONUCLEASE I
49	4fml-A	2.8	2.7	49	204	2	VSDC
50	3nts-B	2.8	2.6	49	189	2	VSDC
51	5hy7-D	2.8	3.3	43	55	14	PUTATIVE PRE-MRNA SPLICING PROTEIN
52	2yu0-A	2.8	2.7	47	94	2	INTERFERON-ACTIVABLE PROTEIN 205
53	4ric-B	2.8	2.4	43	628	12	FANCONI-ASSOCIATED NUCLEASE 1
54	4ric-A	2.8	2.4	43	615	12	FANCONI-ASSOCIATED NUCLEASE 1
55	4rid-A	2.8	2.6	40	615	20	FANCONI-ASSOCIATED NUCLEASE 1
56	4js4-A	2.8	2.9	42	468	14	DT16 OLIGONUCLEOTIDE
57	3hp9-A	2.8	2.9	42	452	14	EXODEOXYRIBONUCLEASE I
58	2qxf-A	2.8	2.9	42	433	14	EXODEOXYRIBONUCLEASE I
59	1fxx-A	2.8	3.0	42	459	14	EXONUCLEASE I
60	2gwl-A	2.7	2.5	49	200	8	65 KDA VIRULENCE PROTEIN
61	5hy7-C	2.7	3.3	43	55	14	PUTATIVE PRE-MRNA SPLICING PROTEIN
62	5ue8-A	2.7	2.9	43	847	7	PROTEIN UNC-13 HOMOLOG A
63	1zrj-A	2.7	3.0	40	50	18	E1B-55KDA-ASSOCIATED PROTEIN 5 ISOFORM C
64	1a8v-B	2.7	3.6	52	116	12	TRANSCRIPTION TERMINATION FACTOR RHO
65	2qnc-B	2.7	2.9	40	157	18	T4 ENDONUCLEASE VII
66	1e7d-B	2.7	3.3	44	157	16	RECOMBINATION ENDONUCLEASE VII
67	2qnf-A	2.7	2.9	40	157	18	RECOMBINATION ENDONUCLEASE VII
68	4rib-B	2.7	2.8	40	628	20	FANCONI-ASSOCIATED NUCLEASE 1
69	4rid-B	2.7	2.6	40	628	20	FANCONI-ASSOCIATED NUCLEASE 1
70	3c94-A	2.7	2.9	42	458	14	EXODEOXYRIBONUCLEASE I
71	1gjj-A	2.6	2.3	43	93	9	LAP2
72	1h9e-A	2.6	2.5	46	56	7	LAMINA-ASSOCIATED POLYPEPTIDE 2
73	1pvo-F	2.6	3.8	51	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
74	1pvo-E	2.6	3.6	51	407	12	RHO TRANSCRIPTION TERMINATION FACTOR
75	1xpr-D	2.6	3.7	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
76	2a8v-A	2.6	3.7	52	118	13	RNA BINDING DOMAIN OF RHO TRANSCRIPTION
77	1a8v-A	2.6	3.7	52	115	12	TRANSCRIPTION TERMINATION FACTOR RHO
78	1en7-A	2.6	2.8	40	157	18	RECOMBINATION ENDONUCLEASE VII
79	2qnc-A	2.6	2.9	40	157	18	T4 ENDONUCLEASE VII
80	1en7-B	2.6	2.6	39	157	18	RECOMBINATION ENDONUCLEASE VII
81	2qnf-B	2.6	2.9	40	157	18	RECOMBINATION ENDONUCLEASE VII
82	4rib-A	2.6	2.4	43	615	12	FANCONI-ASSOCIATED NUCLEASE 1
83	4jrq-B	2.6	2.4	39	459	15	5CY-DA13
84	1e7d-A	2.5	3.0	41	157	17	RECOMBINATION ENDONUCLEASE VII
85	2ld7-A	2.5	4.2	44	94	11	HISTONE DEACETYLASE COMPLEX SUBUNIT SAP30
86	5ife-B	2.5	3.2	43	66	16	SPLICING FACTOR 3B SUBUNIT 5
87	1e7l-A	2.5	3.3	42	157	17	RECOMBINATION ENDONUCLEASE VII
88	1e7l-B	2.5	3.4	42	157	17	RECOMBINATION ENDONUCLEASE VII
89	3l0o-A	2.5	3.8	43	413	21	TRANSCRIPTION TERMINATION FACTOR RHO
90	5muu-D	2.5	2.6	40	148	15	MAJOR INNER PROTEIN P1
91	5muu-L	2.5	2.6	40	148	15	MAJOR INNER PROTEIN P1
92	5muu-M	2.5	2.6	40	148	15	MAJOR INNER PROTEIN P1
93	5muu-F	2.5	2.6	40	148	15	MAJOR INNER PROTEIN P1
94	5muu-J	2.5	2.6	40	148	15	MAJOR INNER PROTEIN P1
95	5muu-I	2.5	2.6	40	148	15	MAJOR INNER PROTEIN P1
96	5muu-E	2.5	2.6	40	148	15	MAJOR INNER PROTEIN P1
97	5muu-H	2.5	2.6	40	148	15	MAJOR INNER PROTEIN P1
98	5muu-G	2.5	2.6	40	148	15	MAJOR INNER PROTEIN P1
99	2eap-A	2.4	3.2	45	90	9	LYMPHOCYTE CYTOSOLIC PROTEIN 2;
100	1su8-A	2.4	3.5	42	633	12	CARBON MONOXIDE DEHYDROGENASE 2
101	1jlr-A	2.4	2.3	38	54	18	THYROID AUTOANTIGEN
102	1suf-A	2.4	5.2	44	633	11	CARBON MONOXIDE DEHYDROGENASE 2
103	1pv4-F	2.4	3.7	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
104	1pvo-D	2.4	3.8	51	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
105	1jei-A	2.4	2.4	37	53	5	EMERIN
106	2a8v-B	2.4	3.5	52	118	12	RNA BINDING DOMAIN OF RHO TRANSCRIPTION
107	4rea-B	2.4	2.8	42	588	19	FANCONI-ASSOCIATED NUCLEASE 1
108	4ria-A	2.4	3.1	42	613	19	FANCONI-ASSOCIATED NUCLEASE 1

No.	PDB and Chain	Z-Score	rmsd (Å)	lali	nres	% id	Description
109	4ri8-B	2.4	3.1	42	615	19	FANCONI-ASSOCIATED NUCLEASE 1
110	4ri8-A	2.4	3.1	42	615	19	FANCONI-ASSOCIATED NUCLEASE 1
111	4ri9-B	2.4	3.1	42	615	19	FANCONI-ASSOCIATED NUCLEASE 1
112	4ri9-A	2.4	3.1	42	615	19	FANCONI-ASSOCIATED NUCLEASE 1
113	4jrq-A	2.4	3.0	40	460	15	5CY-DA13
114	4xrm-B	2.3	3.2	45	64	7	TRANSCRIPTION FACTOR MEIS1
115	1h1j-S	2.3	2.4	38	44	16	THO1 PROTEIN
116	5v2o-F	2.3	2.5	41	60	15	TP2
117	2ooc-B	2.3	4.3	51	105	6	HISTIDINE PHOSPHOTRANSFERASE
118	4f3l-A	2.3	4.2	40	319	15	BMAL1B
119	4up8-A	2.3	4.9	46	580	7	LYSINE-TRNA LIGASE
120	5uzu-B	2.2	2.7	41	68	10	UNCHARACTERISED PROTEIN OF <i>S. aureus</i>
121	1hwz-A	2.2	4.1	42	501	0	GLUTAMATE DEHYDROGENASE
122	1xpu-F	2.2	3.8	50	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
123	3iz6-Q	2.2	4.3	43	141	7	40S RIBOSOMAL PROTEIN RACK1
124	1xpo-C	2.2	3.6	51	408	12	RHO TRANSCRIPTION TERMINATION FACTOR
125	5sy5-D	2.2	4.1	40	283	25	ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR
126	1uyr-A	2.2	3.5	42	698	12	ACETYL-COA CARBOXYLASE
127	2xan-A	2.2	3.1	35	414	9	INOSITOL-PENTAKISPHOSPHATE 2-KINASE
128	2hm2-Q	2.2	3.2	42	89	12	PYRIN-ONLY PROTEIN 1
129	4aya-A	2.1	4.2	42	59	10	DNA-BINDING PROTEIN INHIBITOR ID-2
130	2yr4-A	2.1	3.8	41	674	7	PRO-ENZYME OF L-PHENYLALANINE OXIDASE
131	4myz-B	2.1	2.6	32	52	13	CURK, CURL FUSION PROTEIN
132	4ams-A	2.1	2.9	41	367	10	MG662
133	1zs4-A	2.1	3.2	44	82	9	bacteriophage lambda cII protein
134	2bbz-A	2.1	3.3	46	190	11	VIRAL CASP8 AND FADD-LIKE APOPTOSIS REGULATORY
135	2woo-A	2.1	2.6	43	303	5	ATPASE GET3
136	3u5v-A	2.1	4.8	40	62	8	PROTEIN MAX, TRANSCRIPTION FACTOR E2-ALPHA CHIMER
137	1lyl-A	2.1	3.2	39	482	13	LYSYL-TRNA SYNTHETASE (LYSU)
138	4lk1-F	2.0	3.7	44	542	5	DNA-DIRECTED RNA POLYMERASE SUBUNIT ALPHA
139	5jh0-D	2.0	3.2	39	157	5	ARS-BINDING FACTOR 2, MITOCHONDRIAL
140	2rno-A	2.0	3.1	43	110	21	PUTATIVE DNA-BINDING PROTEIN
141	1l6n-A	2.0	3.2	44	288	2	GAG POLYPROTEIN
142	2ko6-A	2.0	2.8	41	89	12	UNCHARACTERIZED PROTEIN YIHD
143	2zm5-A	2.0	2.6	40	306	10	TRNA DELTA(2)-ISOPENTENYLPYROPHOSPHATE
144	2do9-A	2.0	3.1	46	115	13	NACHT-, LRR- AND PYD-CONTAINING PROTEIN 10
145	2dbg-A	2.0	3.5	48	103	4	MYELOID CELL NUCLEAR DIFFERENTIATION ANTIGEN
146	4o7q-A	2.0	3.3	49	94	8	INTERFERON-INDUCIBLE PROTEIN AIM2
147	5jqe-A	2.0	3.1	43	544	9	SUGAR ABC TRANSPORTER SUBSTRATE-BINDING PROTEIN

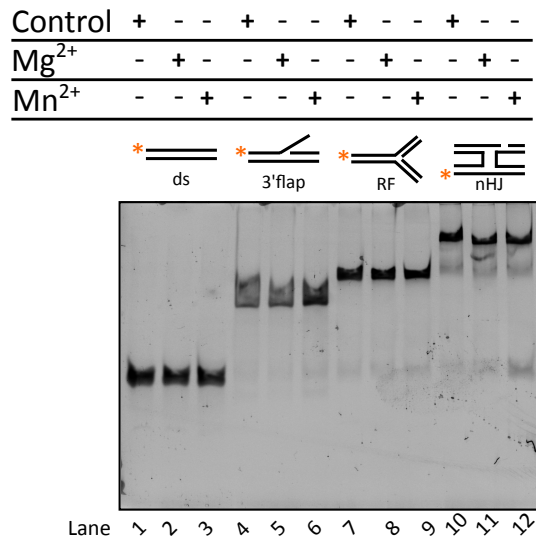


Figure 4.8: Cleavage activity of ceMUS81 towards DNA substrates. Nuclease activity of ceMUS81 expressed and purified alone was tested with different DNA substrates and Mg²⁺ or Mn²⁺ as cofactor. 40 nM 5'FAM labeled substrate was mixed with 640 nM enzyme and incubated for 30 min at 37°C. The reaction products were separated on a native 8% TB-PAGE and analyzed with a phosphoimager.

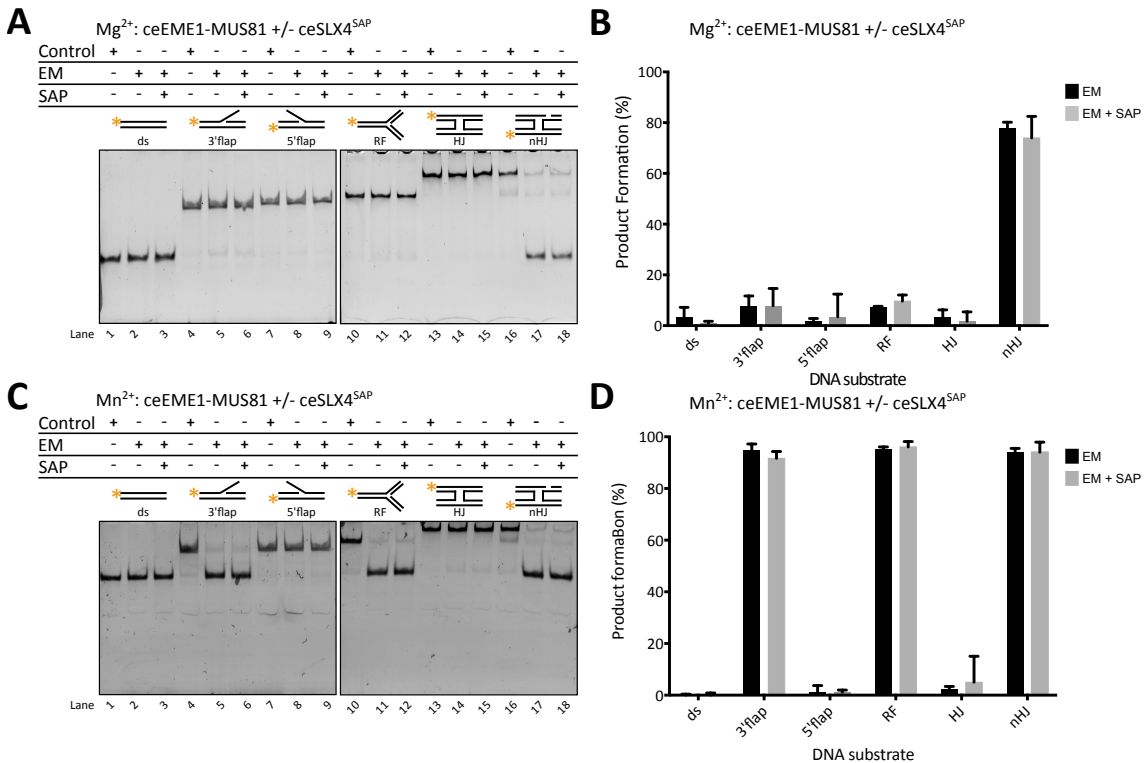


Figure 4.9: Cleavage activity of ceMUS81-ceEME1-ceSLX4^{SAP} complexes (A) Nuclease activity of ceMUS81-EME1 in presence and absence of ceSLX4^{SAP} in presence of Mg^{2+} . (B) Quantification of the nuclease assay carried out with different DNA substrates and Mg^{2+} as cofactor. (C) Nuclease activity of ceMUS81-EME1 in presence and absence of ceSLX4^{SAP} in presence of Mn^{2+} . (B) Quantification of the nuclease assay carried out with different DNA substrates and Mn^{2+} as cofactor. The bar chart shows the percentage of product formation quantified by using ImageJ. The leftover uncut DNA substrate was normalized against the control substrate and product formation was calculated from the difference. The error bars depict the standard deviation from three independent experiments.

List of Figures

1.1	The complexity of DNA damage repair	4
1.2	Pathways in double-strand-break repair	8
1.3	Processing of Holliday Junctions	11
1.4	The SLX4 scaffold protein and its interaction partners	16
1.5	The structure of SLX4-SLX1	21
1.6	Sequence similarities, domain architecture and crystal structure of MUS81-EME1	23
1.7	DNA substrates and cleavage properties of nucleases associated with SLX4	26
2.1	SEC analysis of hsEME1-MUS81 using a Superdex 200 16/600 column	34
2.2	Fusion protein purification systems	35
2.3	Schematic representation of hsEME1-MUS81 constructs	36
2.4	SEC analysis of hsMUS81 constructs based on limited proteolysis using a Superdex 75 16/600 column.	37
2.5	Expression test of different hsSLX4 constructs in HEK293 suspension culture.	38
2.6	SEC analysis of hsSLX4 ¹³⁰⁰⁻¹⁸³⁴ co-expressed with hsSLX1 ¹¹⁻²⁷⁵ using a Superdex 200 16/600 column.	39
2.7	Activity test and limited proteolysis of hsSLX4 ¹³⁰⁰⁻¹⁸³⁴ -hsSLX1 ¹¹⁻²⁷⁵	40
2.8	Schematic representation of hsSLX4 and hsSLX1 constructs	41
2.9	Electrophoretic shift analysis of hsSLX4 ¹³⁰⁰⁻¹⁸³⁴ -hsSLX1 ¹¹⁻²⁷⁵ after phosphatase treatment	43
2.10	Phosphorylation site mapping of hsSLX4 ¹³⁰⁰⁻¹⁸³⁴ alone and co-expressed with hsSLX1 ¹¹⁻²⁷⁵	45
2.11	Heat Map cluster of hsSLX4 ¹³⁰⁰⁻¹⁸³⁴ phosphopeptides in co-expression with hsSLX1 ¹¹⁻²⁷⁵ under influence of different DNA damaging agents	48
2.12	Schematic representation of ceSLX4, ceSLX1 and ceEME1-MUS81 constructs	51
2.13	SEC analysis of full-length ceSLX4, ceSLX1 and ceEME1-MUS81 proteins	53
2.14	Limited proteolysis experiments for ceSLX4 and ceEME1 ¹⁶¹⁻⁴⁴⁵ -ceMUS81	54
2.15	SEC analysis of ceMUS81 ¹⁻¹⁵⁷ and ceSLX4 ⁴⁴³⁻⁶⁰⁶	55
2.16	Stability test of ceSLX4 ⁴⁴³⁻⁷¹⁸	57
2.17	Crystallization of different ceSLX4 constructs	57
2.18	¹ H- ¹⁵ N HSQC NMR spectra of ceSLX4 ^{SAP} and ceMUS81 ^N	61
2.19	Structure of ceSLX4 ^{SAP}	63
2.20	Superimposition of ceSLX4 ^{SAP} with other SAP domains	65

2.21	Structural comparison of ceSLX4 ^{SAP} with SAP domains from other protein in complex with DNA substrate	67
2.22	Cleavage activity of MUS81-EME1 complexes	69
2.23	SEC studies for the ceSLX4 ^{SAP} -ceMUS81 ^N complex.	72
2.24	SLS studies for the ceSLX4 ^{SAP} -ceMUS81 ^N complex.	73
2.25	Interaction studies for the ceSLX4 ^{SAP} -ceMUS81 ^N complex.	74
2.26	DNA binding properties of ceSLX4 ^{SAP}	76
2.27	Analysis of cooperative binding of ceSLX4 ^{SAP}	77
2.28	Analysis of cooperative binding of ceSLX4 ^{SAP-5A} and interaction with ceMUS81 ^N	79
2.29	DNA binding properties of ceMUS81 ^N	81
2.30	DNA binding properties of ceSLX4 ^{SAP} -ceMUS81 ^N complex.	82
2.31	Fluorescence anisotropy measurements	84
2.32	Determination of the DNA binding site in ceSLX4 ^{SAP}	86
3.1	Thermodynamic cycle of DNA binding for ceSLX4 ^{SAP} , ceMUS81 ^N and ceSLX4 ^{SAP} -ceMUS81 ^N complex	95
3.2	Model for cooperative binding of ceSLX4 ^{SAP}	97
3.3	Model for HJ cleavage	100
4.1	Multiple sequence alignment of SLX4 orthologs	147
4.2	Multiple sequence alignment of SLX1 orthologs.	150
4.3	Multiple sequence alignment of MUS81 orthologs.	151
4.4	Multiple sequence alignment of EME1 orthologs.	153
4.5	Degradation of hsSLX4 ¹³⁰⁰⁻¹⁸³⁴ after IP6 cleavage	154
4.6	Possible cleavage sites of the CPD protease on hsSLX4 ¹³⁰⁰⁻¹⁸³⁴	154
4.7	Phosphorylation sites in scSLX4	154
4.8	Cleavage activity of ceMUS81 towards DNA substrates	158
4.9	Cleavage activity of ceMUS81-ceEME1-ceSLX4 ^{SAP} complexes	159

List of Tables

2.1	Comparison of experimentally found phosphorylation sites in hsSLX4 ¹³⁰⁰⁻¹⁸³⁴ with the database Phosphosite Plus	47
2.2	Protein constructs of ceSLX4 ^{SAP} and ceMUS81 ^N for NMR experiments	60
2.3	Interaction assays applied for the ceSLX4 ^{SAP} -ceMUS81 ^N complex	71
4.1	Media for bacterial and mammalian cell expression.	103
4.2	Bacterial strains	104
4.3	Cell line	104
4.4	Antibiotic stock solutions	104
4.5	Plasmids used for expression	104
4.6	Cloned constructs	105
4.7	Oligonucleotides for cloning	107
4.8	Oligonucleotides for mutagenesis.	108
4.9	Oligonucleotides for sequencing	109
4.10	Buffers for large scale preparation of vector DNA	109
4.11	Buffers for protein purification	110
4.12	Phosphatase inhibitors	112
4.13	Equipment	112
4.14	Software	113
4.15	PCR reaction mixture	114
4.16	PCR program for amplification	114
4.17	Mixture for vector linearization	115
4.18	Mixture for T4 processing of the vector	115
4.19	Mixture for T4 processing of the insert	116
4.20	colony PCR	116
4.21	PCR program for colony screening	116
4.22	PCR program for site-directed mutagenesis	117
4.23	Expression conditions in <i>E. coli</i> cells	119
4.24	Expression conditions for HEK suspension culture	121
4.25	DNA damaging agents	128
4.26	Dephosphorylation of hsSLX4 ¹³⁰⁰⁻¹⁸³⁴	129
4.27	DNA substrates for nuclease activity tests	130
4.28	DNA substrates used for Fluorescence Anisotropy measurements and bands shift assays . . .	131

4.29	List of the respective cryoprotectants used for the protecting the crystals at the synchrotron.	136
4.30	Structural comparison of ceSLX4 ^{SAP} using Dali server	155

Bibliography

- [1] O. D. Schärer. Chemistry and Biology of DNA Repair. *Angew. Chem. Int. Ed.*, 42:2946–2974, 2003.
- [2] J. H. Hoeijmakers. Genome maintenance mechanisms for preventing cancer. *Nature*, 411:366–374, 2001.
- [3] J. Stingele, R. Bellelli, and S. J. Boulton. Mechanisms of DNA-protein crosslink repair. *Nat Rev Mol Cell Biol.*, 18:563–573, 2017.
- [4] E. C. Friedberg. DNA damage and repair. *Nature*, 421:436–440, 2003.
- [5] J. San Filippo, P. Sung, and H. Klein. Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry*, 77:229–257, 2008.
- [6] T. Lindahl. The Intrinsic Fragility of DNA (Nobel Lecture). *Angew Chem Int Ed Engl.*, 55:8528–8534, 2016.
- [7] B. B. Zhou and S. J. Elledge. The DNA damage response: putting checkpoints in perspective. *Nature*, 408:433–439, 2000.
- [8] G. A. Cromie, J. C. Connelly, and D. R. F. Leach. Recombination at Double-Strand Breaks and DNA Ends: Conserved Mechanisms from Phage to Humans. *Mol Cell*, 8:1163–1174, 2001.
- [9] L. Aravind, D. R. Walker, and E. V. Koonin. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucl. Acids Res.*, 27:1223–1242, 1999.
- [10] J. M. Murray and A. M. Carr. Integrating DNA damage repair with the cell cycle. *Curr Op Cell Biol*, 52:120–125, 2018.
- [11] D. Branzei and M. Foiani. Regulation of DNA repair throughout the cell cycle. *Nature Rev Mol Cell Biol*, 9:297–308, 2009.
- [12] W.-D. Heyer, K. T. Ehmsen, and J. Liu. Regulation of Homologous Recombination in Eukaryotes. *Annu. Rev. Genetic*, 44:113–139, 2010.
- [13] R. Che, J. Zhang, M. Nepal, B. Han, and P. Fei. Multifaceted Fanconi Anemia Signaling. *Trends Genet.*, 34:171–183, 2018.
- [14] Y. Kim, F. P. Lach, R. Desetty, H. Hanenberg, A. D. Auerbach, and A. Smogorzewska. Mutations of the SLX4 gene in Fanconi anemia. *Nat. Genet.*, 43:142–146, 2011.

- [15] J. E. Cleaver. Defective repair replication of DNA in xeroderma pigmentosum. 1968. *DNA Repair*, 3:183–187, 2004.
- [16] A. Ciccia and S. J. Elledge. The DNA Damage Response: Making It Safe to Play with Knives. *Mol Cell*, 40:179–204, 2010.
- [17] V.A. Bohr. Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance. *Trends Biochem. Sci.*, 33:609–620, 2008.
- [18] S. Cory and J. M. Adams. The Bcl2 family: regulators of the cellular life-or-death switch. *Nat Rev Cancer*, 2:647–656, 2002.
- [19] L. C. Huang, K. C. Clarkin, and G. M. Wahl. Sensitivity and selectivity of the DNA damage sensor responsible for activating p53-dependent G1 arrest. *Proc Natl Acad Sci USA*, 93:4827–4832, 1996.
- [20] C. B. Bennett, A. L. Lewis, K. K. Baldwin, and M. A. Resnick. Lethality induced by a single site-specific double-strand break in a dispensable yeast plasmid. *Proc Natl Acad Sci USA*, 90:5613–5617, 1993.
- [21] M. Berti and A. Vindigni. Replication stress: getting back on track. *Nat Struct Mol Biol.*, 23:103–109, 2016.
- [22] A. H. Syeda, M. Hawkins, and P. McGlynn. Recombination and replication. *Cold Spring Harb Perspect Biol*, 6:a016550, 2014.
- [23] D. T. Long, M. Räschle, V. Joukov, and J. C. Walter. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science*, 333:84–87, 2011.
- [24] T.T. Saito, J.L. Youds, S.J. Boulton, and M.P. Colaiácovo. *Caenorhabditis elegans* HIM-18/SLX-4 interacts with SLX-1 and XPF-1 and maintains genomic integrity in the germline by processing recombination intermediates. *PLoS Genet*, 5:e1000735, 2009.
- [25] T. T. Saito, F. Mohideen, K. Meyer, J. W. Harper, and Colaiácovo M. P. SLX-1 is required for maintaining genomic integrity and promoting meiotic noncrossovers in the *Caenorhabditis elegans* germline. *PLoS Genet.*, 8:e1002888, 2012.
- [26] I. Lam and S. Keeney. Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harb Perspect Biol.*, 7:a016634, 2014.
- [27] M. J. Neale and S. Keeney. Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature*, 442:153–158, 2006.
- [28] L. Krejci, V. Altmannova, M. Spirek, and X. Zhao. Homologous recombination and its regulation. *Nucl. Ac. Res.*, 40:5795–5818, 2012.
- [29] T. Hassold and P. Hunt. To err (meiotically) is human: the genesis of human aneuploidy. *Nat. Rev. Genet.*, 2:280–291, 2001.
- [30] M. Gellert. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem.*, 71:101–132, 2002.

- [31] M. A. Oettinger, D. G. Schatz, C Gorka, and D Baltimore. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science*, 248:1517–1523, 1990.
- [32] K. Schwarz, G. H. Gauss, L. Ludwig, U. Pannicke, Z. Li, D. Lindner, W. Friedrich, R. A. Seger, T. E. Hansen-Hagge, S. Desiderio, M. R. Lieber, and C. R. Bartram. RAG Mutations in Human B Cell-Negative SCID. *Science*, 274:97–99, 1996.
- [33] S. D. Fugmann, A. I. Lee, P. E. Shockett, I. J. Villey, and D. G. Schatz. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.*, 18:495–527, 2000.
- [34] C. Boboila, F. W. Alt, and B. Schwer. Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. *Adv. Immunol.*, 116:1–49, 2012.
- [35] M. N. Lieber. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem.*, 79:181–211, 2010.
- [36] H. H. Y. Chang, N. R. Pannunzio, N. Adachi, and M. R. Lieber. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol.*, 18:495–506, 2017.
- [37] N. R. Pannunzio, G. Watanabe, and MR. Lieber. Nonhomologous DNA End Joining for Repair of DNA Double-Strand Breaks. *J Biol Chem.*, pii:jbc.TM117.000374., 2017.
- [38] W.-D Heyer. Regulation of Recombination and Genomic maintenance. *Cold Spring Harb Perspect Biol*, 7:a016501, 2015.
- [39] E. K. Schwartz and W.-D Heyer. Processing of joint molecule intermediates by structure-selective endonucleases during homologous recombination in eukaryotes. *Chromosoma*, 120:109–127, 2011.
- [40] M. McVey, V. Y. Khodaverdian, D. Meyer, P. A. Cerqueira, and W.-D. Heyer. Eukaryotic DNA polymerases in homologous recombination. *Annu. Rev Genet.*, 50:393–421, 2016.
- [41] S. E. Polo and S. P. Jackson. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes & Dev.*, 25:409–433, 2011.
- [42] J. W. Szostak, T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl. The double-strand-break repair model for recombination. *Cell*, 33:25–35, 1983.
- [43] E. P. Mimitou and L. S. Symington. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *EMBO J.*, 29:3358–3369, 2010.
- [44] C. Escribano-Díaz, A. Orthwein, A. Fradet-Turcotte, M. Xing, J. T. Young, J. Tkáč, M. A. Cook, A. P. Rosebrock, M. Munro, M. D. Canny, D. Xu, and D. Durocher. A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol Cell.*, 49:872–883, 2013.

- [45] A. Beucher, J. Birraux, L. Tchouandong, O. Barton, A. Shibata, S. Conrad, A. A. Goodarzi, A. Krempler, P. A. Jeggo, and M. Löbrich. ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *EMBO J.*, 28:3413–3427, 2009.
- [46] A. Marechal and L. Zou. DNA Damage Sensing by the ATM and ATR Kinases. *CSH Perspec Biol*, 5:a012716, 2013.
- [47] T. H. Stracker and J. H. J. Petrini. The MRE11 complex: starting from the ends. *Nat. Rev. Mol. Cell. Biol.*, 12:90–103, 2011.
- [48] M. Lisby, J. H. Barlow, R. C. Burgess, and R. Rothstein. Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell*, 118:699–713, 2004.
- [49] A. Lammens, A. Schele, and K. P. Hopfner. Structural biochemistry of ATP-driven dimerization and DNA-stimulated activation of SMC ATPases. *Curr Biol*, 14:1778–1782, 2004.
- [50] R. Bhargava, D. O. Onyango, and J. M. Stark. Regulation of single-strand annealing and its role in genome maintenance. *Trends in Genet*, 32:566–575, 2016.
- [51] T. M. Gottlieb and S. P. Jackson. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell*, 72:131–142, 1993.
- [52] K. Meek, V. Dang, and S. P. Lees-Miller. DNA-PK: the means to justify the ends? *Adv. Immunol.*, 99:33–58, 2008.
- [53] N. Jette and S. P. Lees-Miller. The DNA-dependent protein kinase: A multifunctional protein kinase with roles in DNA double strand break repair and mitosis. *Prog in Biophys and Mol. Biol.*, 117:194–205, 2015.
- [54] H. H. Chang and M. R. Lieber. Structure-specific nuclease activities of Artemis and the Artemis: DNA-PKcs complex. *Nucleic Acids Res.*, 44:4991–4997, 2016.
- [55] H. H. Chang, G Watanabe, and M. R. Lieber. Unifying the DNA end-processing roles of the artemis nuclease: Ku-dependent Artemis resection. *J. Biol. Chem.*, 290:24036–24050, 2015.
- [56] B.L. Mahaney, K. Meek, and S.P. Lees-Miller. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem. J.*, 417:639–650, 2009.
- [57] Y. Liu, J. R. Cussiol, D. Dibitto, J. R. Sims, S. Twayana, R. S. Weiss, R. Freire, F. Marini, A. Pellicoli, and M. B. Smolka. TOPBP1(Dpb11) plays a conserved role in homologous recombination DNA repair through the coordinated recruitment of 53BP1(Rad9). *J. Cell Biol.*, 216:623–639, 2017.
- [58] H. Gaillard, T. Garca-Muse, and A. Aguilera. Replication stress and cancer. *Nat Rev Cancer.*, 15:276–289, 2015.
- [59] C. Gelot, I. Magdalou, and B. S. Lopez. Replication stress in Mammalian cells and its consequences for mitosis. *Genes (Basel)*, 6:267–298, 2015.

- [60] C. X. Deng and R. H. Wang. Roles of BRCA1 in DNA damage repair: a link between development and cancer. *Hum. Mol. Genet.*, 12:R113–R123, 2003.
- [61] R. Prakash, Y. Zhang, W. Feng, and M. Jasin. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harb Perspect Biol*, 7:a016600, 2015.
- [62] J. H. Seol, E. Y. Shim, and S. E. Lee. Microhomology-mediated end joining: Good, bad and ugly. *Mutat Res.*, pii::S0027–5107(17)30041–6, 2017.
- [63] R. Holliday. A mechanism for gene conversion in fungi. *Genet. Res.*, 5:282–304, 1964.
- [64] M. E. Moynahan and M. Jasin. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.*, 11:196–207, 2010.
- [65] L. S. Symington. End resection at double-strand breaks: mechanism and regulation. *Cold Spring Harb Perspect Biol*, 6:a016436, 2014.
- [66] A. A. Sartori, C. Lukas, J. Coates, M. Mistrik, S. Fu, J. Bartek, R. Baer, J. Lukas, and Jackson S. P. Human CtIP promotes DNA end resection. *Nature*, 450:509–514, 2007.
- [67] L. Chen, C. J. Nievera, A. Y. Lee, and X. Wu. Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. *J Biol Chem.*, 282:7713–7720, 2008.
- [68] P. Huertas and S. P. Jackson. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J Biol Chem.*, 284:9558–9565, 2009.
- [69] E. P. Mimitou and L. S. Symington. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*, 455:770–774, 2008.
- [70] Z. Zhu, W. H. Chung, E. Y. Shim, S. E. Lee, and G. Ira. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell*, 134:981–994, 2008.
- [71] M.S. Wold. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem.*, 66:61–92, 1997.
- [72] P. Sung, L. Krejci, S. Van Komen, and M. G. Sehorn. Rad51 recombinase and recombination mediators. *J Biol Chem.*, 278:42729–42732, 2003.
- [73] P. Verma and R. A. Greenberg. Noncanonical views of homology-directed DNA repair. *Genes & Dev*, 30:1138–1154, 2016.
- [74] S. W. Morrical. DNA-pairing and annealing processes in homologous recombination and homology-directed repair. *Cold Spring Harb Perspect Biol*, 7:a016444, 2015.
- [75] N. Bennardo, A. Cheng, N. Huang, and J. M. Stark. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet.*, 4:e1000110, 2008.
- [76] G. M. Adair, R. L. Rolig, D. Moore-Faver, M. Zabelshansky, J. H. Wilson, and R. S. Nairn. Role of ERCC1 in removal of long non-homologous tails during targeted homologous recombination. *EMBO J.*, 19:5552–5561, 2000.

- [77] N. Sugawara, F. Pâques, M. Colaiácovo, and J. E. Haber. Role of *Saccharomyces cerevisiae* Msh2 and Msh3 repair proteins in double-strand break-induced recombination. *PNAS*, 94:9214–9219, 1997.
- [78] J. M. Svendsen and J. W. Harper. GEN1/Yen1 and the SLX4 complex: solutions to the problem of Holliday Junction resolution. *Genes & Dev*, 24:521–536, 2010.
- [79] S. Flott, C. Alabert, G. W. Toh, R. Toth, N. Sugawara, D. G. Campbell, J. E. Haber, P. Pasero, and J. Rouse. Phosphorylation of Slx4 by Mec1 and Tel1 regulates the single-strand annealing mode of DNA repair in budding yeast. *Mol Cell Biol.*, 27:6433–6445, 2007.
- [80] L. S. Symington. Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiol Mol Biol Rev.*, 6:630–670, 2002.
- [81] N. Nassif, J. Penney, S. Pal, W. R. Engels, and G.B. Gloor. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol. Cell. Biol.*, 14:1613–1625, 1994.
- [82] J. M. Daley, W. A. Gaines, Y. Kwon, and P. Sung. Regulation of DNA Pairing in Homologous Recombination. *CSH Perspect. Biol.*, 6:a017954, 2014.
- [83] M. Sebesta, P. Burkovics, S. Juhasz, S. Zhang, J. E. Szabo, M. Y. W. T. Lee, L. Haracska, and L. Krejci. Role of PCNA and TLS polymerases in D-loop extension during homologous recombination in humans. *DNA Rep*, 12:691–698, 2013.
- [84] L. Wu and I. D. Hickson. The Bloom’s syndrome helicase suppresses crossing over during homologous recombination. *Nature*, 426:870–874, 2003.
- [85] S. Paliwal, R. Kanagaraj, A. Sturzenegger, K. Burdova, and P. Janscak. Human RECQ5 helicase promotes repair of DNA double-strand breaks by synthesis-dependent strand annealing. *Nucleic Acids Res.*, 42:2380–2390, 2014.
- [86] R. Prakash, D. Satory, E. Dray, A. Papusha, J. Scheller, W. Kramer, L. Krejci, H. Klein, J. E. Haber, P. Sung, and G. Ira. Yeast Mph1 helicase dissociates Rad51-made D-loops: implications for crossover control in mitotic recombination. *Genes Dev*, 23:67–79, 2009.
- [87] L.J. Barber, J.L. Youds, J.D. Ward, M.J. McIlwraith, N.J. O’Neil, M.I. Petalcorin, J.S. Martin, S.J. Collis, S.B. Cantor, M. Auclair, H. Tissenbaum, S. C. West, A. M. Rose, and S. J. Boulton. RTEL1 maintains genomic stability by suppressing homologous recombination. *Cell*, 135:261–271, 2008.
- [88] K. Rodgers and M. McVey. Error-Prone Repair of DNA Double-Strand Breaks. *J. Cell Physiol.*, 231:15–24, 2016.
- [89] A. Mehta and J. E. Haber. Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. *CSH Pers Biol*, 6:a016428, 2014.
- [90] A. Malkova, E. L. Ivanov, and J. E. Haber. Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *PNAS*, 93:7131–7136, 1996.

- [91] A. Malkova and G. Ira. Break-induced replication: functions and molecular mechanism. *Curr Opin Genet Dev.*, 23:271–279, 2013.
- [92] C. J. Sakofsky and A. Malkova. Break induced replication in eukaryotes: mechanisms, functions, and consequences. *Crit. Rev. in Bioch. and Mol. Biol.*, 52:395–413, 2017.
- [93] M. A. Wilson, Y. Kwon, Y. Xu, W.-H. Chung, P. Chi, H. Niu, R. Mayle, X. Chen, A. Malkova, P. Sung, and G. Ira. Pif1 helicase and Pol δ promote recombination-coupled DNA synthesis via bubble migration. *Nature*, 502:393–396, 2013.
- [94] N. Saini, S. Ramakrishnan, R. Elango, S. Ayyar, Y. Zhang, A. Deem, G. Ira, J. E. Haber, K. S. Lobachev, and A. Malkova. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature*, 502:389–392, 2013.
- [95] A. Malkova, M. L. Naylor, M. Yamaguchi, G. Ira, and J. E. Haber. RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. *Mol Cell Biol.*, 25:933–944, 2005.
- [96] J. R. Lydeard, S. Jain, M. Yamaguchi, and J. E. Haber. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *2007*, 448:820–823, *Nature*.
- [97] B. Llorente, C. E. Smith, and L. S. Symington. Break-induced replication - What is it and what is it for? *Cell Cycle*, 7:859–864, 2008.
- [98] C. I. White and J. E. Haber. Intermediates of recombination during mating type switching in *Saccharomyces cerevisiae*. *EMBO J.*, 9:663–673, 1990.
- [99] M. J. McIlwraith, A. Vaisman, Y. Liu, E. Fanning, R. Woodgate, and S. C. West. Human DNA polymerase eta promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Mol Cell*, 20:783–792, 2005.
- [100] J. A. Solinger, K. Kiianitsa, and W.-D. Heyer. Rad54, a Swi2/Snf2-like recombinational repair protein, disassembles Rad51:dsDNA filaments. *Mol. Cell*, 10:859–864, 2002.
- [101] M. Bzymek, N. H. Thayer, S. D. Oh, N. Kleckner, and N. Hunter. Double Holliday junctions are intermediates of DNA break repair. *Nature*, 464:937–941, 2010.
- [102] A. V. Nimonkar, R. A. Sica, and S. C. Kowalczykowski. Rad52 promotes second-end DNA capture in double-stranded break repair to form complement-stabilized joint molecules. *Proc Natl Acad Sci U S A.*, 106:3077–3082, 2009.
- [103] M. J. McIlwraith and S. C. West. DNA repair synthesis facilitates RAD52-mediated second-end capture during DSB repair. *Mol Cell*, 29:510–516, 2008.
- [104] A. V. Nimonkar and S. C. Kowalczykowski. Second-end DNA capture in double-strand break repair. *Cell Cycle*, 8:1816–1817, 2009.
- [105] D. V. Bugreev, F. Hanaoka, and A. V. Mazin. Rad54 dissociates homologous recombination intermediates by branch migration. *Nat Struct Mol Biol.*, 14:146–153, 2007.

- [106] M.E. Robu, R.B. Inman, and M.M. Cox. RecA protein promotes the regression of stalled replication forks in vitro. *PNAS*, 98:8211–8218, 2001.
- [107] P. Cejka, J. L. Plank, C. Z. Bachrati, I. D. Hickson, and S. C. Kowalczykowski. Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3. *Nat. Struct. Mol. Biol.*, 17:1377–1382, 2010.
- [108] P. M. Dehé, S. Coulon, S. Scaglione, P. Shanahan, A. Takedachi, J. A. Wohlschlegel, J. R. 3rd Yates, B. Llorente, P. Russell, and P. H. Gaillard. Regulation of Mus81-Eme1 Holliday junction resolvase in response to DNA damage. *Nat Struct Mol Biol.*, 20:598–603, 2013.
- [109] Y. D. Tay and L. Wu. Overlapping roles for Yen1 and Mus81 in cellular Holliday junction processing. *J Biol Chem.*, 285:11427–11432, 2010.
- [110] E. K. Schwartz, W. D. Wright, K. T. Ehmsen, J. E. Evans, H. Stahlberg, and W. D. Heyer. Mus81-Mms4 Functions as a Single Heterodimer To Cleave Nicked Intermediates in Recombinational DNA Repair. *Mol. Cell. Biol.*, 32:3065–3080, 2012.
- [111] S.C.Y. Ip, U. Rass, M.G. Blanco, H.R. Flynn, J.M. Skehel, and S.C. West. Identification of Holliday junction resolvases from humans and yeast. *Nature*, 456:357–361, 2008.
- [112] E. Garner, Y. Kim, F. P. Lach, M. C. Kottemann, and A. Smogorzewska. Human GEN1 and the SLX4-Associated Nucleases MUS81 and SLX1 Are Essential for the Resolution of Replication-Induced Holliday Junctions. *Cell Reports*, 5:1–9, 2013.
- [113] S. Fekairi, S. Scaglione, C. Chahwan, E.R. Taylor, A. Tissier, S. Coulon, M.Q. Dong, C. Ruse, J.R. 3rd Yates, P. Russell, R.P. Fuchs, C.H. McGowan, and Gaillard P.H. Human SLX4 Is a Holliday Junction Resolvase Subunit that Binds Multiple DNA Repair/Recombination Endonucleases. *Cell*, 138:78–89, 2009.
- [114] J. M. Svendsen, A. Smogorzewska, M. E. Sowa, B. C. O’Connell, S. P. Gygi, S. J. Elledge, and J. W. Harper. Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell*, 138:63–77, 2009.
- [115] I. M. Muñoz, K. Hain, A. C. Déclais, M. Gardiner, G. W. Toh, L. Sanchez-Pulido, J. M. Heuckmann, R. Toth, T. Macartney, B. Eppink, R. Kanaar, C. P. Ponting, D. M. Lilley, and J. Rouse. Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. *Mol. Cell.*, 35:116–127, 2009.
- [116] A. H. Bizard and I. D. Hickson. The dissolution of double Holliday junctions. *CSH Perspect Biol*, 6:a016477, 2014.
- [117] J. Matos, M. G. Blanco, S. Maslen, J. M. Skehel, and S. C. West. Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis. *Cell*, 147:158–172, 2011.
- [118] J. Matos, M. G. Blanco, and West S. C. Cell-cycle kinases coordinate the resolution of recombination intermediates with chromosome segregation. *Cell Rep.*, 4:76–86, 2013.

- [119] M. Gallo-Fernández, I. Saugar, M. Á. Ortiz-Bazán, M. V. Vázquez, and J. A. Tercero. Cell cycle-dependent regulation of the nuclease activity of Mus81-Eme1/Mms4. *Nucleic Acids Res.*, 40:8325–8335., 2012.
- [120] I. Saugar, M. V. Vázquez, M. Gallo-Fernández, M. Á. Ortiz-Bazán, M. Segurado, A. Calzada, and J. A. Tercero. Temporal regulation of the Mus81-Mms4 endonuclease ensures cell survival under conditions of DNA damage. *Nucleic Acids Res.*, 41:8943–8958, 2013.
- [121] N. A. Ellis, J. Groden, T. Z. Ye, J. Straughen, D. J. Lennon, S. Ciocci, M. Proytcheva, and J. German. The Bloom’s syndrome gene product is homologous to RecQ helicases. *Cell*, 83:655–666, 1995.
- [122] L. Wu, S. L. Davies, P. S. North, H. Goulaouic, J. F. Riou, H. Turley, K. C. Gatter, and I. D. Hickson. The Bloom’s syndrome gene product interacts with topoisomerase III. *J. Biol. Chem.*, 275:9636–9644, 2000.
- [123] S. Gravel, J. R. Chapman, C. Magill, and S. P. Jackson. DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes Dev*, 22:2767–2772, 2008.
- [124] D.M Lilley and M.F. White. The junction-resolving enzymes. *Nature reviews. Molecular cell biology*, 2:433–443, 2001.
- [125] G.J. Sharples. The X philes: structure-specific endonucleases that resolve Holliday junctions. *Molecular microbiology*, 39:823–834, 2001.
- [126] P. M. Dehé and P. H. Gaillard. Control of structure-specific endonucleases to maintain genome stability. *Nat Rev Mol Cell Biol.*, 18:315–330, 2017.
- [127] J.S. Ahn and M.C. Whitby. The role of the SAP motif in promoting Holliday junction binding and resolution by SpCCE1. *The Journal of biological chemistry*, 278:29121–29129, 2003.
- [128] D.R. Duckett, A.I. Murchie, S. Diekmann, E. von Kitzing, B. Kemper, and D.M. Lilley. The structure of the Holliday junction, and its resolution. *Cell*, 55:79–89, 1988.
- [129] M. J. Giraud-Panis and D. M. Lilley. Structural recognition and distortion by the DNA junction-resolving enzyme RusA. *J Mol Biol.*, 278:117–133, 1998.
- [130] M. F. White and D. M. Lilley. Interaction of the resolving enzyme YDC2 with the four-way DNA junction. *Nucleic Acids Res.*, 26:5609–5616, 1998.
- [131] F. E. Benson and S. C. West. Substrate specificity of the Escherichia coli RuvC protein. Resolution of three- and four-stranded recombination intermediates. *J Biol Chem.*, 269:5195–5201, 1994.
- [132] R. Shah, R. J. Bennett, and S. C. West. Genetic recombination in E. coli: RuvC protein cleaves Holliday junctions at resolution hotspots in vitro. *Cell*, 79:853–864, 1994.
- [133] J. M. Fogg, M. J. Schofield, M. F. White, and D. M. Lilley. Sequence and functional-group specificity for cleavage of DNA junctions by RuvC of Escherichia coli. *Biochemistry*, 38:11349–11358, 1999.

- [134] M. C. Whitby, E. L. Bolt, S. N. Chan, and R. G. Lloyd. Interactions between RuvA and RuvC at Holliday junctions: inhibition of junction cleavage and formation of a RuvA-RuvC-DNA complex. *J Mol Biol.*, 264:878–890, 1996.
- [135] M. N. Boddy, P. H. Gaillard, W. H. McDonald, P. Shanahan, J. R. 3rd Yates, and P. Russell. Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell*, 107:537–548, 2001.
- [136] S. C. West, M. G. Blanco, Y. W. Chan, J. Matos, S. Sarbajna, and H. D. M. Wyatt. Resolution of Recombination intermediates: Mechanisms and regulation. *CSH Symp Quant Biol*, 80:103–109, 2015.
- [137] L. D. et al. Wood. The genomic landscapes of human breast and colorectal cancers. *Science*, 318:1108–1113, 2007.
- [138] S. L. Andersen, H. K. Kuo, D. Savukoski, M. H. Brodsky, and J. Sekelsky. Three structure-selective endonucleases are essential in the absence of BLM helicase in *Drosophila*. *PLoS Genet.*, 7:e1002315, 2011.
- [139] M. G. Blanco, J. Matos, U. Rass, S. C. Ip, and S. C. West. Functional overlap between the structure-specific nucleases Yen1 and Mus81-Mms4 for DNA-damage repair in *S. cerevisiae*. *DNA Repair*, 9:394–402, 2010.
- [140] T. Wechsler, S. Newman, and S. C. West. Aberrant chromosome morphology in human cells defective for Holliday junction resolution. *Nature*, 471:642–646, 2011.
- [141] H.D. Wyatt, S. Sarbajna, J.A. Matos, and West S.C. Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells. *Mol. Cell*, 52 (2):234–247, 2013.
- [142] A. Ciccia, A. Constantinou, and S. C. West. Identification and characterization of the human Mus81-Eme1 Endonuclease. *J. Biol. Chem.*, 278:25172–25178, 2003.
- [143] H.D. Wyatt, R.C. Laister, S.R. Martin, C.H. Arrowsmith, and S.C. West. The SMX DNA Repair Tri-nuclease. *Mol. Cell*, 65 (5):848–860, 2017.
- [144] A. C. Déclais and D. M. Lilley. New insight into the recognition of branched DNA structure by junction-resolving enzymes. *Curr Opin Struct Biol.*, 18:86–95, 2008.
- [145] M. J. Dickman, S. M. Ingleston, S. E. Sedelnikova, J. B. Rafferty, R. G. Lloyd, J. A. Grasby, and D. P. Hornby. The RuvABC resolvosome. *Eur J Biochem.*, 269:5492–5501, 2002.
- [146] H. Iwasaki, M. Takahagi, T. Shiba, A. Nakata, and H. Shinagawa. *Escherichia coli* RuvC protein is an endonuclease that resolves the Holliday structure. *EMBO J.*, 10:4381–4389, 1991.
- [147] C. Biertümpfel, W. Yang, and D. Suck. Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature*, 449:616–620, 2007.

- [148] S. Golz, A. Christoph, K. Birkenkamp-Demtröder, and B. Kemper. Identification of amino acids of endonuclease VII essential for binding and cleavage of cruciform DNA. *Eur. J. Biochem.*, 245:573–580, 1997.
- [149] R. P. Birkenbihl and B. Kemper. Endonuclease VII has two DNA-binding sites each composed from one N- and one C-terminus provided by different subunits of the protein dimer. *EMBO J.*, 17:4527–4534, 1998.
- [150] Y. W. Chan, K. Fugger, and S. C. West. Unresolved recombination intermediates lead to ultra-fine anaphase bridges, chromosome breaks and aberrations. *Nat Cell Biol.*, 20:92–103, 2018.
- [151] J. R. Mullen, V. Kaliraman, S. S. Ibrahim, and S. J. Brill. Requirement for three novel protein complexes in the absence of the Sgs1 DNA helicase in *Saccharomyces cerevisiae*. *Genetics*, 157:103–118, 2001.
- [152] W. M. Fricke and S. J. Brill. Slx1-Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1-Top3. *Genes Dev.*, 17:1768–1778, 2003.
- [153] S. Coulon, P. H. Gaillard, C. Chahwan, W. H. McDonald, J. R. 3rd Yates, and P. Russell. Slx1-Slx4 are subunits of a structure-specific endonuclease that maintains ribosomal DNA in fission yeast. *Mol Biol Cell.*, 15:71–80, 2004.
- [154] S. L. Andersen, D. T. Bergstralh, K. P. Kohl, J. R. LaRocque, C. B. Moore, and J. Sekelsky. Drosophila MUS312 and the vertebrate ortholog BTBD12 interact with DNA structure-specific endonucleases in DNA repair and recombination. *Mol Cell*, 35:128–135, 2009.
- [155] Y. Kim. Nuclease delivery: versatile functions of SLX4/FANCP in genome maintenance. *Mol. Cells*, 37:569–574, 2014.
- [156] V. Gaur, H.D. Wyatt, W. Komorowska, Szczepanowski R.H., D. de Sanctis, K.M. Gorecka, S.C. West, and M. Nowotny. Structural and Mechanistic Analysis of the Slx1-Slx4 Endonuclease. *Cell Reports*, 10:1467–1476, 2015.
- [157] J. Yin, B. Wan, J. Sarkar, K. Horvath, J. Wu, Y. Chen, G. Cheng, K. Wan, P. Chin, M. Lei, and Y. Liu. Dimerization of SLX4 contributes to functioning of the SLX4-nuclease complex. *Nucleic Acids Res.*, 44:4871–4880, 2016.
- [158] Y. Kim, G. S. Spitz, U. Veturi, F. P. Lach, A. D. Auerbach, and A. Smogorzewska. Regulation of multiple DNA repair pathways by the Fanconi anemia protein SLX4. *Blood*, 121:154–63, 2013.
- [159] K. N. Yamamoto, S. Kobayashi, M. Tsuda, H. Kurumizaka, M. Takata, K. Kono, J. Jiricny, S. Takeda, and K. Hirota. Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proc Natl Acad Sci U S A.*, 108:6492–6496, 2011.
- [160] C. Lachaud, D. Castor, K. Hain, I. Muñoz, J. Wilson, T. J. MacArtney, D. Schindler, and J. Rouse. Distinct functional roles for the two SLX4 ubiquitin-binding UBZ domains mutated in Fanconi anemia. *J Cell Sci.*, 127:2811–2817, 2014.

- [161] D. Klein Douwel, R. A. Boonen, D. T. Long, A. A. Szypowska, M. Räschle, J. C. Walter, and P. Knipscheer. XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol Cell.*, 54:460–471, 2014.
- [162] J. J. Sekelsky, K. S. McKim, G. M. Chin, and Hawley R. S. The *Drosophila* meiotic recombination gene *mei-9* encodes a homologue of the yeast excision repair protein Rad1. *Genetics.*, 141:619–627, 1995.
- [163] P. J. Stogios, G. S. Downs, J. J. Jauhal, S. K. Nandra, and G. G. Priv. Sequence and structural analysis of BTB domain proteins. *Genome Biol.*, 6:R82, 2005.
- [164] J. H. Guervilly, A. Takedachi, V. Naim, S. Scaglione, C. Chawhan, Y. Lovera, E. Despras, I. Kuraoka, P. Kannouche, F. Rosselli, and P. H. Gaillard. The SLX4 complex is a SUMO E3 ligase that impacts on replication stress outcome and genome stability. *Mol Cell.*, 57:123–137, 2015.
- [165] E. Ouyang, J. and Garner, A. Hallet, H. D. Nguyen, K. A. Rickman, G. Gill, A. Smogorzewska, and L. Zou. Noncovalent interactions with SUMO and ubiquitin orchestrate distinct functions of the SLX4 complex in genome maintenance. *Mol Cell.*, 57:108–122, 2015.
- [166] J. R. Gareau and C. D. Lima. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol.*, 11:861–871, 2010.
- [167] P. Sarangi, Z. Bartosova, V. Altmannova, C. Holland, M. Chavdarova, S. E. Lee, L. Krejci, and X. Zhao. Sumoylation of the Rad1 nuclease promotes DNA repair and regulates its DNA association. *Nucleic Acids Res.*, 42:6393–6404, 2014.
- [168] L. Aravind and E.V. Koonin. SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends in biochemical sciences*, 25:112–114, 2000.
- [169] F. Göhring, B.L. Schwab, P. Nicotera, M. Leist, and F.O. Fackelmayer. The novel SAR-binding domain of scaffold attachment factor A (SAF-A) is a target in apoptotic nuclear breakdown. *The EMBO journal*, 16:7361–7371, 1997.
- [170] J. Wang, M. Satoh, C.H. Chou, and W.H. Reeves. Similar DNA binding properties of free P70 (KU) subunit and P70/P80 heterodimer. *FEBS letters*, 351:219–224, 1994.
- [171] C.H. Chou, J. Wang, M. W. Knuth, and W.H. Reeves. Role of a major autoepitope in forming the DNA binding site of the p70 (Ku) antigen. *J. Exp. Med.*, 175:1677–1684, 1992.
- [172] T. A. Kunkel and D. A. Erie. Dna mismatch repair. *Annu Rev Biochem*, 74:681–710, 2005.
- [173] B. Meissner, T. Bartram, C. Eckert, J. Trka, R. Panzer-Grümayer, I. Hermanova, E. Ellinghaus, A. Franke, A. Mörnicke, A. Schrauder, A. Teigler-Schlegel, P. Dörge, A. von Stackelberg, G. Basso, C. R. Bartram, R. Kirschner-Schwabe, B. Bornhäuser, J. P. Bourquin, G. Cazzaniga, J. Hauer, A. Attarbaschi, S. Izraeli, M. Zaliova, G. Cario, M. Zimmermann, S. Avigad, M. Sokalska-Duhme, M. Metzler, M. Schrappe, R. Koehler, G. Te Kronnie, and M. Stanulla. Frequent and sex-biased deletion of SLX4IP by illegitimate V(D)J-mediated

- recombination in childhood acute lymphoblastic leukemia. *Hum Mol Genet.*, 23:590–601, 2014.
- [174] A. Ciccia, N. McDonald, and S.C. West. Structural and functional relationships of the XPF/MUS81 family of proteins. *Annual review of biochemistry*, 77:259–287, 2008.
- [175] D. Klein Douwel, W. S. Hoogenboom, R. A. Boonen, and P. Knipscheer. Recruitment and positioning determine the specific role of the XPF-ERCC1 endonuclease in interstrand crosslink repair. *EMBO J.*, 36:2034–2046, 2017.
- [176] N.J. O’Neil, J.S. Martin, J.L. Youds, J.D. Ward, M.I.R. Petalcorin, A.M. Rose, and S.J. Boulton. Joint molecule resolution requires the redundant activities of MUS-81 and XPF-1 during *Caenorhabditis elegans* meiosis. *PLoS genetics*, 9:e1003582, 2013.
- [177] B. Wan, J. Yin, K. Horvath, J. Sarkar, Y. Chen, J. Wu, K. Wan, J. Lu, P. Gu, E. Y. Yu, N. F. Lue, S. Chang, Y. Liu, and M. Lei. SLX4 assembles a telomere maintenance toolkit by bridging multiple endonucleases with telomeres. *Cell Rep.*, 4:861–869, 2013.
- [178] J. S. Wilson, A. M. Tejera, D. Castor, R. Toth, M. A. Blasco, and J. Rouse. Localization-dependent and -independent roles of SLX4 in regulating telomeres. *Cell Rep*, 4:853–860, 2013.
- [179] M. R. Hodskinson, J. Silhan, G. P. Crossan, J. I. Garaycochea, S. Mukherjee, C. M. Johnson, O. D. Schärer, and K. J. Patel. Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair. *Mol Cell.*, 54:472–484, 2014.
- [180] A. F. Fagbemi, B. Orelli, and O. D. Schärer. Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair*, 10:722–729, 2011.
- [181] N. G. Jaspers, A. Raams, M. C. Silengo, N. Wijgers, L. J. Niedernhofer, A. R. Robinson, G. Giglia-Mari, D. Hoogstraten, W. J. Kleijer, J. H. Hoeijmakers, and W. Vermeulen. First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure. *Am J Hum Genet.*, 80:457–466, 2007.
- [182] L. J. Niedernhofer, G. A. Garinis, A. Raams, A. S. Lalai, A. R. Robinson, E. Appeldoorn, H. Odijk, R. Oostendorp, A. Ahmad, W. van Leeuwen, A. F. Theil, W. Vermeulen, G. T. van der Horst, P. Meinecke, W. J. Kleijer, J. Vijg, N. G. Jaspers, and J. H. Hoeijmakers. A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature*, 444:1038–1043, 2006.
- [183] A. M. Sijbers, W. L. de Laat, R. R. Ariza, M. Biggerstaff, Y. F. Wei, J. G. Moggs, K. C. Carter, B. K. Shell, E. Evans, M. C. de Jong, S. Rademakers, J. de Rooij, N. G. Jaspers, J. H. Hoeijmakers, and R. D. Wood. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell*, 86:811–822, 1996.
- [184] S. Flott and J. Rouse. Slx4 becomes phosphorylated after DNA damage in a Mec1/Tel1-dependent manner and is required for repair of DNA alkylation damage. *Biochem. J.*, 391:325–333, 2005.

- [185] S. Matsuoka, B. A. Ballif, A. Smogorzewska, E. R. 3rd McDonald, K. E. Hurov, J. Luo, C. E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S. P. Gygi, and S. J. Elledge. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science*, 316:1160–1166, 2007.
- [186] J. K. Holloway, S. Mohan, G. Balmus, X. Sun, A. Modzelewski, P. L. Borst, R. Freire, R. S. Weiss, and P. E. Cohen. Mammalian BTBD12 (SLX4) protects against genomic instability during mammalian spermatogenesis. *PLoS Genet.*, 7:e1002094, 2011.
- [187] Y. Zhou, J. H. Lee, W. Jiang, J. L. Crowe, S. Zha, and T. T. Paull. Regulation of the DNA Damage Response by DNA-PKcs Inhibitory Phosphorylation of ATM. *Mol Cell*, 65:91–104, 2017.
- [188] M. A. Rego, F. W. Kolling IV, and N. G. Howlett. The Fanconi anemia protein interaction network: Casting a wide net. *Mut. Res.*, 668:27–41, 2009.
- [189] K. Savitsky, A. Bar-Shira, S. Gilad, G. Rotman, Y. Ziv, L. Vanagaite, D. A. Tagle, S. Smith, T. Uziel, S. Sfez, M. Ashkenazi, I. Pecker, M. Frydman, R. Harnik, S. R. Patanjali, A. Simmons, G. A. Clines, A. Sartiel, R. A. Gatti, L. Chessa, O. Sanal, M. F. Lavin, N. G. Jaspers, A. M. Taylor, C. F. Arlett, T. Miki, S. M. Weissman, M. Lovett, F. S. Collins, and Y. Shiloh. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science*, 268:1749–1753, 1995.
- [190] K. Paeschke, K. R. McDonald, and V. A. Zakian. Telomeres: Structures in need of unwinding. *FEBS Lett.*, 584:3760–3772, 2010.
- [191] Vannier J. B., V. Pavicic-Kaltenbrunner, M. I. Petalcorin, H. Ding, and S. J. Boulton. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell*, 149:795–806, 2012.
- [192] B. Pfander and J. F. Diffley. Dpb11 coordinates Mec1 kinase activation with cell cycle-regulated Rad9 recruitment. *EMBO J.*, 30:4897–907, 2011.
- [193] P. Y. Ohouo, F. M. Bastos de Oliveira, Y. Liu, C. J. Ma, and M. B. Smolka. DNA-repair scaffolds dampen checkpoint signalling by counteracting the adaptor Rad9. *Nature*, 493:120–124, 2013.
- [194] D. Gritenaite, L. N. Princz, B. Szakal, S. C. Bantele, L. Wendeler, S. Schilbach, B. H. Habermann, J. Matos, M. Lisby, D. Branzei, and B. Pfander. A cell cycle-regulated Slx4-Dpb11 complex promotes the resolution of DNA repair intermediates linked to stalled replication. *Genes Dev.*, 28:1604–1619, 2014.
- [195] P. Y. Ohouo, F. M. Bastos de Oliveira, B. S. Almeida, and M. B. Smolka. DNA damage signaling recruits the Rtt107-Slx4 scaffolds via Dpb11 to mediate replication stress response. *Mol Cell*, 39:300–306, 2010.
- [196] R. T. Pedersen, T. Kruse, J. Nilsson, V. H. Oestergaard, and M. Lisby. TopBP1 is required at mitosis to reduce transmission of DNA damage to G1 daughter cells. *J. Cell Biol.*, 210:565–582, 2015.

- [197] S. Tanaka, T. Umemori, K. Hirai, S. Muramatsu, Y. Kamimura, and H. Araki. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature*, 445:328–332, 2007.
- [198] F. A. Barr, H. H. Sillje, and E. A. Nigg. Polo-like kinases and the orchestration of cell division. *Nat. Rev. Mol. Cell. Biol.*, 5:429–440, 2004.
- [199] J. Matos and West S. C. Holliday Junction resolution: regulation in space and time. *DNA repair*, 19:1–6, 2014.
- [200] H. Duda, M. Arter, Gloggnitzer. J., F. Teloni, P. Wild, M.G. Blanco, M. Altmeyer, and J.A Matos. Mechanism for Controlled Breakage of Under-replicated Chromosomes during Mitosis. *Dev. Cell*, 39:740–755, 2016.
- [201] N. Nair, D. Castor, T. Macartney, and J. Rouse. Identification and characterization of MUS81 point mutations that abolish interaction with the SLX4 scaffold protein. *DNA Repair*, 24:131–137, 2014.
- [202] D. Castor, N. Nair, A. C. Déclais, C. Lachaud, R. Toth, T. J. Macartney, D. M. Lilley, J. S. Arthur, and J. Rouse. Cooperative control of Holliday junction resolution and DNA repair by the SLX1 and MUS81-EME1 nucleases. *Mol. Cell*, 52:221–233, 2013.
- [203] H. Interthal and W. D. Heyer. MUS81 encodes a novel helix-hairpin-helix protein involved in the response to UV- and methylation-induced DNA damage in *Saccharomyces cerevisiae*. *Mol Gen Genet.*, 263:812–827, 2000.
- [204] V. Kaliraman and S. J. Brill. Role of SGS1 and SLX4 in maintaining rDNA structure in *Saccharomyces cerevisiae*. *Curr Genet.*, 41:389–400, 2002.
- [205] S. Dunin-Horkawicz, M. Feder, and J. M. Bujnicki. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, 7:98, 2006.
- [206] M. Sokolowska, H. Czapinska, and M. Bochtler. Hpy188I-DNA pre- and post-cleavage complexes - snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.*, 39:1554–1564., 2011.
- [207] F. Lian, F. Xiw, and C. Qian. Crystal structure and SUMO binding of Slx1-Slx4 complex. *Sci. Rep.*, 6:19331, 2016.
- [208] T. Putilina, P. Wong, and S. Gentleman. The DHHC domain: a new highly conserved cysteine-rich motif. *Mol Cell Biochem.*, 195:216–226, 1999.
- [209] Chen. X. B., R. Melchionna, C. M. Denis, P. H. Gaillard, A. Blasina, I. Van de Weyer, M. N. Boddy, P. Russell, J. Vialard, and C. H. McGowan. Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. *Mol Cell.*, 8:1117–1127, 2001.
- [210] T. Nishino, K. Komori, Y. Ishino, and K. Morikawa. X-Ray and Biochemical Anatomy of an Archaeal XPF/Rad1/Mus81 Family Nuclease: Similarity between Its Endonuclease Domain and Restriction Enzymes. *Structure*, 11:445–457, 2003.

- [211] T. Nishino, K. Komori, Y. Ishino, and K. Morikawa. Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. *Structure*, 13:1183–1192, 2005.
- [212] W. Yang. Nucleases: diversity of structure, function and mechanism. *Q Rev Biophys*, 44:1–93, 2011.
- [213] J.H. Chang, J.J. Kim, J.M. Choi, J.H. Lee, and Y. Cho. Crystal structure of the Mus81-Eme1 complex. *Genes & Dev*, 22:1093–1106, 2008.
- [214] G.H. Gwon, A. Jo, K. Baek, K.S. Jin, Y. Fu, J.-B. Lee, Y. Kim, and Y. Cho. Crystal structures of the structure-selective nuclease Mus81-Eme1 bound to flap DNA substrates. *The EMBO Journal*, 33:1061–1072, 2014.
- [215] R.C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Research*, 32:1792–1797, 2004.
- [216] A.M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, and G.J. Barton. Jalview version 2: A Multiple Sequence Alignment and Analysis Workbench. *Bioinformatics*, 25:1189–1191, 2009.
- [217] The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45 (D1):D158–D169, 2017.
- [218] A.J. Doherty, L.C. Serpell, and C.P. Ponting. The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Research*, 24(13):2488–2497, 1996.
- [219] A.J. Fadden, S. Schalbetter, M. Bowles, R. Harris, J. Lally, A.M. Carr, and N.Q. McDonald. A winged helix domain in human MUS81 binds DNA and modulates the endonuclease activity of MUS81 complexes. *Nucleic Acids Research*, 41:9741–9752, 2013.
- [220] K.S. Gajiwala and S.K. Burley. Winged helix proteins. *Current opinion in structural biology*, 10:110–116, 2000.
- [221] M. Nowotny and V. Gaur. Structure and mechanism of nucleases regulated by SLX4. *Curr Op Struct Biol*, 36:97–105, 2016.
- [222] J. H. Enzlin and O. D. Schärer. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J.*, 21:2045–2053, 2002.
- [223] W. Yang, J. Y. Lee, and M. Nowotny. Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell.*, 22:5–13, 2006.
- [224] M. Newman, J. Murray-Rust, J. Lally, J. Rudolf, A. Fadden, P. P. Knowles, M. F. White, and N. Q. McDonald. Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *EMBO J.*, 24:895–905, 2005.
- [225] O. V. Tsodikov, J. H. Enzlin, O. D. Schärer, and T. Ellenberger. Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc Natl Acad Sci U S A.*, 102:11236–11241, 2005.

- [226] K. Tripsianes, G. Folkers, E. Ab, D. Das, H. Odijk, N. G. Jaspers, J. H. Hoeijmakers, R. Kaptein, and R. Boelens. The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure.*, 13:1849–1858, 2005.
- [227] R. Coulthard, A. J. Deans, P. Swuec, M. Bowles, A. Costa, S.C. West, and N. Q. McDonald. Architecture and DNA recognition elements of the Fanconi anemia FANCM-FAAP24 complex. *Structure*, 21:1648–1658, 2013.
- [228] K. Hanada, M. Budzowska, M. Modesti, A. Maas, C. Wyman, J. Essers, and R. Kanaar. The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. *EMBO J.*, 25:4921–4932, 2006.
- [229] T. Shimura, M. J. Torres, M. M. Martin, V. A. Rao, Y. Pommier, M. Katsura, K. Miyagawa, and M. I. Aladjem. Bloom’s syndrome helicase and Mus81 are required to induce transient double-strand DNA breaks in response to DNA replication stress. *J Mol Biol.*, 375:1152–1164, 2008.
- [230] Y. K. Shin, T. Amangyeld, T. A. Nguyen, P. R. Munashingha, and Y. S. Seo. Human MUS81 complexes stimulate flap endonuclease 1. *FEBS J.*, 279:2412–2430, 2012.
- [231] J. Rouse. Control of genome stability by SLX protein complexes. *Biochem Soc Trans.*, 37:495–510, 2009.
- [232] K.T. Ehmsen and W.-D. Heyer. A junction branch point adjacent to a DNA backbone nick directs substrate cleavage by *Saccharomyces cerevisiae* Mus81-Mms4. *Nucleic Acids research*, 37:2026–2036, 2009.
- [233] F. Osman and M.C. Whitby. Exploring the roles of Mus81-Eme1/Mms4 at perturbed replication forks. *DNA repair*, 6:1004–1017, 2007.
- [234] S. C. West and Y. W. Chan. Genome Instability as a consequence of defects in the resolution of Recombination intermediates. *CSH Symp Quant Biol*, 82:pii: 034256, 2017.
- [235] S. H. Lee, L. N. Princz, M. F. Klügel, B. Habermann, B. Pfander, and C. Biertümpfel. Human Holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. *Elife*, 4:pii: e12256., 2015.
- [236] H.D. Wyatt and West S.C. SMX makes the cut in genome stability. *Oncotarget*, 8 (61):102765–102766, 2017.
- [237] Y. W. Chan and S. C. West. Spatial control of the GEN1 Holliday junction resolvase ensures genome stability. *Nat Commun.*, 5:4844, 2014.
- [238] B. Szakal and D. Branzei. Premature Cdk1/Cdc5/Mus81 pathway activation induces aberrant replication and deleterious crossover. *EMBO J.*, 32:1155–1167, 2013.
- [239] H. Lans and W. Vermeulen. Tissue specific response to DNA damage: *C. elegans* as role model. *DNA repair*, 32:141–148, 2015.

- [240] C. H. Lai, C. Y. Chou, L. Y. Ch'ang, C. S. Liu, and W. Lin. Identification of novel human genes evolutionarily conserved in *Caenorhabditis elegans* by comparative proteomics. *Genome Res.*, 10:703–713, 2000.
- [241] T.T. Saito, D.Y. Lui, H.-M. Kim, K. Meyer, and Colaicovo M.P. Interplay between Structure-Specific Endonucleases for Crossover Control during *Caenorhabditis elegans* Meiosis. *PLoS Genet.*, 9:e1003586, 2013.
- [242] A. Agostinho, B. Meier, R. Sonnevile, M. Jagut, A. Woglar, J. Blow, V. Jantsch, and A. Gartner. Combinatorial regulation of meiotic holliday junction resolution in *C. elegans* by HIM-6 (BLM) helicase, SLX-4, and the SLX-1, MUS-81 and XPF-1 nucleases. *PLoS genetics*, 9:e1003591, 2013.
- [243] S. Sarbajna, D. Davies, and S. C. West. Roles of SLX1-SLX4, MUS81-EME1, and GEN1 in avoiding genome instability and mitotic catastrophe. *Genes Dev.*, 28:1124–1136, 2014.
- [244] N. Dendouga, H. Gao, D. Moechars, M. Janicot, J. Vialard, and C. H. McGowan. Disruption of murine Mus81 increases genomic instability and DNA damage sensitivity but does not promote tumorigenesis. *Mol Cell Biol.*, 25:7569–7579, 2005.
- [245] J. P. McPherson, B. Lemmers, R. Chahwan, A. Pamidi, E. Migon, E. Matysiak-Zablocki, M. E. Moynahan, J. Essers, K. Hanada, A. Poonepalli, O. Sanchez-Sweatman, R. Khokha, R. Kanaar, M. Jasin, M. P. Hande, and R. Hakem. Involvement of mammalian Mus81 in genome integrity and tumor suppression. *Science*, 304:1822–1826, 2004.
- [246] S. S. Ho, W. Y. Zhang, N. Y. Tan, M. Khatoo, M. A. Suter, S. Tripathi, F. S. Cheung, W. K. Lim, P. H. Tan, J. Ngeow, and S. Gasser. The DNA Structure-Specific Endonuclease MUS81 Mediates DNA Sensor STING-Dependent Host Rejection of Prostate Cancer Cells. *Immunity*, 44:1177–1189, 2016.
- [247] J. Abraham, B. Lemmers, M. P. Hande, M. E. Moynahan, C. Chahwan, A. Ciccia, J. Essers, K. Hanada, R. Chahwan, A. K. Khaw, P. McPherson, A. Shehabeldin, R. Laister, C. Arrowsmith, R. Kanaar, S. C. West, M. Jasin, and R. Hakem. Eme1 is involved in DNA damage processing and maintenance of genomic stability in mammalian cells. *EMBO J.*, 22:6137–6174, 2003.
- [248] G. P. Crossan, L. van der Weyden, I. V. Rosado, F. Langevin, P. H. Gaillard, R. E. McIntyre, Sanger Mouse Genetics Project, F. Gallagher, M. I. Kettunen, D. Y. Lewis, K. Brindle, M. J. Arends, D. J. Adams, and K. J. Patel. Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. *Nat. Genet.*, 43:147–152, 2011.
- [249] C. Stoepker, K. Hain, B. Schuster, Y. Hilhorst-Hofstee, M. A. Rooimans, J. Steltenpool, A. B. Oostra, K. Eirich, E. T. Korthof, A. W. Nieuwint, N. G. Jaspers, T. Bettecken, H. Joenje, D. Schindler, J. Rouse, and de Winter J. P. SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. *Nat. Genet.*, 43:138–141, 2011.
- [250] A. D. Auerbach. Fanconi anemia and its diagnosis. *Mut. Res.*, 668:4–10, 2009.

- [251] V. Kaliraman, J. R. Mullen, W. M. Fricke, S. A. Bastin-Shanower, and S. J. Brill. Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. *Genes & Dev*, 15:2730–2740, 2001.
- [252] A. Shen, P.J. Lupardus, V.E. Albrow, A. Guzzetta, J.C. Powers, K.C. Garcia, and Bogyo M. Mechanistic and structural insights into the proteolytic activation of *Vibrio cholerae* MARTX toxin. *Nat Chem Biol*, 5:469–478, 2009.
- [253] A. Shen, P.J. Lupardus, M. Morell, E.L. Ponder, A.M. Sadaghiani, K.C. Garcia, and Bogyo M. Simplified, enhanced protein purification using an inducible, autoprocesing enzyme tag. *PLoS One*, 4:e8119, 2009.
- [254] J. G. Marblestone, S. C. Edavettal, Y. Lim, P. Lim, X. Zuo, and T. R. Butt. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci*, 15:182–189, 2006.
- [255] R. J. Peroutka, N. Elshourbagy, T. Piech, and T. R. Butt. Enhanced protein expression in mammalian cells using engineered SUMO fusions: Secreted phospholipase A2. *Protein Sci*, 17:1596–1595, 2008.
- [256] P. J. Lupardus, A. Shen, M. Bogyo, and K.C. . Garcia. Small molecule-induced allosteric activation of the *Vibrio cholerae* RTX cysteine protease domain. *Science*, 322:265–268, 2008.
- [257] D. Drew, S. Newstead, Y. Sonoda, H. Kim, G. von Heijne, and Iwata S. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nat. Protoc.*, 3:784–798, 2008.
- [258] U. Rothbauer, K. Zolghadr, S. Muyldermans, A. Schepers, M. C. Cardoso, and H. Leonhardt. A Versatile Nanotrap for Biochemical and Functional Studies with Fluorescent Fusion Proteins. *Mol Cel Proteomics*, 7:282–289, 2008.
- [259] N. Portolano, P. J. Watson, L. Fairall, C. J. Millard, C. P. Milano, Y. Song, S. M. Cowley, and J. W. R. Schwabe. Recombinant Protein Expression for Structural Biology in HEK 293F Suspension Cells: A Novel and Accessible Approach. *J. Vis. Exp.*, 92:51897, 2014.
- [260] I. Nigi, L. Fairall, and J. W. R. Schwabe. Expression and Purification of Protein Complexes Suitable for Structural Studies Using Mammalian HEK 293F Cells. *Curr. Protocols. Prot. Sci.*, 90:5.28.1–5.28.16, 2017.
- [261] J. R. Cussiol, D. Dibitto, A. Pellicoli, and M. B. Smolka. Slx4 scaffolding in homologous recombination and checkpoint control: lessons from yeast. *Chromosoma*, 126:45–58, 2017.
- [262] S. Zhuo, J. C. Clemens, D.J. Hakes, D. Barford, and J.E. Dixon. Expression, purification, crystallization, and biochemical characterization of a recombinant protein phosphatase. *J Biol Chem*, 268:17754–17761, 1993.
- [263] T. Barshevsky. unpublished results. <https://www.neb.com/products/protein-tools/protein-phosphatases/protein-phosphatases>, 2008.

- [264] N. Dephoure, K. L. Gould, S. P. Gygi, and D. R. Kellogg. Mapping and analysis of phosphorylation sites: a quick guide for cell biologists. *Mol. Biol. Cell*, 24:535–542, 2013.
- [265] Z. Liu, Y. Wang, H. Cheng, W. Deng, Z. Pan, S. Ullah, J. Ren, and Y. Xue. GPS 3.0: web servers for the prediction of protein post-translational modification sites. <http://gps.biocuckoo.org>, 2015.
- [266] P.V. Hornbeck, B. Zhang, B. Murray, J.M. Kornhauser, V. Latham, and E. Skrzypek. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research*, 43:512–520, 2015.
- [267] E. Petermann, M. L. Orta, N. Issaeva, N. Schultz, and T. Helleday. Hydroxyurea-Stalled Replication Forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair. *Mol Cell*, 37:492–502, 2010.
- [268] P. Sestili, C. Martinelli, and V. Stocchi. The fast halo assay: An improved method to quantify genomic DNA strand breakage at the single-cell level. *Mut. Res.*, 607:205–214, 2006.
- [269] Y. Pommier, E. Leo, H. Zhang, and C. Marchand. DNA Topoisomerases and Their Poisoning by Anticancer and Antibacterial Drugs. *Chem & Biol*, 17:421–433, 2010.
- [270] R. Linding, R. B. Russell, V. Neduva, and T. J. Gibson. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acid Res*, 31:3701–3708, 2003.
- [271] Y. Fu and W. Xiao. Functional domains required for the *Saccharomyces cerevisiae* Mus81-Mms4 endonuclease complex formation and nuclear localization. *DNA Repair*, 2:1435–1447, 2003.
- [272] T. W. Geders, K. Gustafson, and B. C. Finzel. Use of differential scanning fluorimetry to optimize the purification and crystallization of PLP-dependent enzymes. *Acta Crystallogr F*, 68:596–600, 2012.
- [273] F. Dupeux, M. Röwer, G. Seroul, D. Blot, and J. A. Márquez. A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr D*, 67:915–919, 2011.
- [274] M. Kipp, F. Göhring, T. Ostendorp, C.M. van Drunen, R. van Driel, M. Przybylski, and F.O. Fackelmayer. SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Molecular and cellular biology*, 20:7480–7489, 2000.
- [275] S. Okubo, F. Hara, Y. Tsuchida, S. Shimotakahara, S. Suzuki, H. Hatanaka, S. Yokoyama, H. Tanaka, H. Yasuda, and H. Shindo. NMR structure of the N-terminal domain of SUMO ligase PIAS1 and its interaction with tumor suppressor p53 and A/T-rich DNA oligomers. *J. Biol. Chem.*, 279:31455–31461, 2004.
- [276] L. Holm and L. M. Laakso. Dali server update. *Nucl. Acids Res.*, 44 (W1):W351–W355, 2016.

- [277] L. Holm, S. Kääriäinen, D Plewczynski, and C. Wilton. Using Dali for structural comparison of proteins. *Curr. Protocols of Bioinf.*, Chapter 5:Unit 5.5., 2006.
- [278] E. Skordalakes and J. M. Berger. Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell*, 114:135–146, 2003.
- [279] J.O.B. Jacobsen, M.D. Allen, S.M.V. Freund, and M. Bycroft. High-resolution NMR structures of the domains of *Saccharomyces cerevisiae* Tho1. *Acta Crystallographica. Section F, Structural biology communications*, 72:500–506, 2016.
- [280] H. W. Gabel and G. Ruvkun. The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. *Nat Struct Mol Biol.*, 15:531–533, 2008.
- [281] Q. Zhao, X. Xue, S. Longrich, P. Sung, and Y. Xiong. Structural insights into 5' flap DNA unwinding and incision by the human FAN1 dimer. *Nature communications*, 5:5726, 2014.
- [282] J.R. Walker, R.A. Corpina, and J. Goldberg. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*, 412:607–614, 2001.
- [283] D. Lu and J. L. Keck. Structural basis of *Escherichia coli* single-stranded DNA-binding protein stimulation of exonuclease I. *PNAS*, 105:9169–9174, 2008.
- [284] S. K. Korada, T. D. Johns, C. E. Smith, N. D. Jones, K. A. McCabe, and C. E. Bell. Crystal structures of *Escherichia coli* exonuclease I in complex with single-stranded DNA provide insights into the mechanism of processive digestion. *Nucleic Acids Res.*, 41:5887–5897, 2013.
- [285] J. Wang, X. Dong, and W.H. Reeves. A model for Ku heterodimer assembly and interaction with DNA. Implications for the function of Ku antigen. *J. Biol. Chem.*, 273:31068–31074, 1998.
- [286] Z. Zhang, L. Zhu, D. Lin, F. Chen, D. J. Chen, and Y. Chen. The Three-dimensional Structure of the C-terminal DNA-binding Domain of Human Ku70. *J. Biol Chem.*, 276:38231–38236, 2001.
- [287] M. E. Maguire and J. A. Cowan. Magnesium chemistry and biochemistry. *Biometals*, 15:203–210, 2002.
- [288] K. Huynh and C. L. Partch. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr. Protoc. Protein Sci.*, 79:28.9.1–28.9.14, 2016.
- [289] M. Vivoli, H. R. Novak, J. A. Littlechild, and N. J. Harmer. Determination of protein-ligand interactions using differential scanning fluorimetry. *J. Vis. Exp.*, 91:e51809, 2014.
- [290] J. A. Abbot, N. M. Livingston, S. B. Egri, E. Guth, and C. S. Francklyn. Characterization of aminoacyl-tRNA synthetase stability and substrate interaction by differential scanning fluorimetry. *Methods*, 113:64–71, 2017.
- [291] G. J. Crowther, P. He, P. P. Rodenbough, A. P. Thomas, K. V. Kovzun, D. J. Leibly, J. Bhandari, L. J. Castaneda, W. G. J. Hol, M. H. Gelb, A. J. Napuli, and W. C. Van Voorhis. Use of thermal melt curves to assess the quality of enzyme preparations. *Anal. Biochem.*, 399:268–275, 2010.

- [292] D. Radushev. Prism 7 .0c for MacOS. *GraphPad Prism Software*, La Jolla, CA, USA, 1994-2017.
- [293] A. Gijsbers, T. Nishigaki, and N. Sánchez-Puig. Fluorescence Anisotropy as a Tool to Study Protein-protein Interactions. *J Vis Exp.*, 116, 2016.
- [294] J. R. (eds) Lakowitz. Fluorescence Anisotropy. In Principles of Fluorescence Spectroscopy. *Springer, Boston, USA*, Third edition:353–382, 2006.
- [295] M. Mann, S.-W. Ong, M. Grønborg, H. Steen, O. N. Jensen, and A. Pandey. Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trend in Biotech.*, 20:261–268, 2002.
- [296] C. Lin, Z. Huang, W. Wen, A. Wu, C. Wang, and L. Niu. Enhancing protein expression in HEK-293 cells by lowering culture temperature. *Plos One*, 2015.
- [297] H. Kaufmann, X. Mazur, M. Fussenegger, and J. E. Bailey. Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotech Bioeng*, 63:573–582, 1999.
- [298] C. A. Lo, A. W. Greben, and B. E. Chen. Generating stable cell lines with quantifiable protein production using CRISPR/Cas9-mediated knock-in. *Biotechniques*, 62:165–174, 2017.
- [299] A. Vaisman, H. Ling, R. Woodgate, and W. Yang. Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis. *EMBO J.*, 24:2957–2967, 2005.
- [300] M. Nowotny and W. Yang. Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J.*, 25:1924–1933, 2006.
- [301] M. Nowotny, S. A. Sergei A. Gaidamakov, R. J. Crouch, and W. Yang. Crystal Structures of RNase H Bound to an RNA/DNA Hybrid: Substrate Specificity and Metal-Dependent Catalysis. *Cell*, 121:1005–1006, 2005.
- [302] Y. Tsunaka, K. Takano, H. Matsumura, Y. Yamagata, and S. Kanaya. Identification of Single Mn²⁺ Binding Sites Required for Activation of the Mutant Proteins of E. coli RNase HI at Glu48 and/or Asp134 by X-ray Crystallography. *J. Mol. Biol.*, 345:1171–1183, 2005.
- [303] Y. Tsunaka, M. Haruki, M. Morikawa, M. Oobatake, and S. Kanaya. Dispensibility of glutamic acid 48 and aspartic acid 134 for Mn²⁺-dependent activity of *Escherichia coli* ribonuclease HI. *Biochemistry*, 42:3366–3374, 2003.
- [304] I. Saugar, A. Jiménez-Martin, and J. A. Tercero. Subnuclear relocalization of structure-specific endonucleases in response to DNA damage. *Cell Rep.*, 20:1553–1562, 2017.
- [305] N. Nagaraj, J. R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Pääbo, and M. Mann. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol.*, 7:548, 2011.
- [306] K. Monier, J. C. Armas, S. Etteldorf, P. Ghazal, and K. F. Sullivan. Annexation of the interchromosomal space during viral infection. *Nat Cell Biol.*, 2:661–665, 2000.

- [307] H.D. Wyatt and West S.C. Holliday Junction Resolvases. *CSH Persp. Biol*, 6:a023192, 2014.
- [308] Y. Liu and S. C. West. Happy Holidays: 40th anniversary of the Holliday Junction. *Nat. rev Mol. Cell Biol.*, 5:937–946, 2004.
- [309] A. Hochschild and M Ptashne. Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell*, 44:681–687, 1986.
- [310] C.M. Mendel and D.B. Mendel. ‘Non-specific’ binding - The problem and a solution. *Biochem. J.*, 228:269–272, 1985.
- [311] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *PNAS*, 98:10037–10041, 2001.
- [312] G. M. Boratyn, C. Camacho, P. S. Cooper, G. Coulouris, A. Fong, N. Ma, T. L. Madden, W. T. Matten, S. D. McGinnis, Y. Merezuk, Y. Raytselis, E. W. Sayers, T. Tao, J. Ye, and I. Zaretskaya. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.*, 41:W29–W33, 2013.
- [313] P. Emsley and K. Cowtan. Coot: Model-building tools for molecular graphics. *Acta Crystallogr D*, 60:2126–2132, 2004.
- [314] J. Söding, A. Biegert, and A. N. Lupas. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, 33:W244–248, 2005.
- [315] The Inkscape Team. Inkscape. Draw freely. Version 0.48.2. <https://inkscape.org/en/>, 2010.
- [316] J. Cox and M. Mann. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, 26:1367–1372, 2008.
- [317] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and Baker N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucl. Ac. Res.*, 32:W665–W667, 2004.
- [318] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. Phaser crystallographic software. *J. Appl. Cryst*, 40:658–674, 2007.
- [319] P. D. Adams, P. V. Afonine, G. Bunkóczy, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, 66:213–221, 2010.
- [320] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols*, 10:845–858, 2015.
- [321] P. Artimo, M. Jonnalagedda, K. Arnold, D. Baratin, G. Csardi, E. de Castro, S. Duvaud, V. Flegel, A. Fortier, E. Gasteiger, A. Grosdidier, C. Hernandez, V. Ioannidis, D. Kuznetsov, R. Liechti, S. Moretti, K. Mostaguir, N. Redaschi, G. Rossier, I. Xenarios, and H. Stockinger.

- ExPASy: SIB bioinformatics resource portal. *ExPASy: SIB bioinformatics resource portal*, 40(W1):W597–W603, 2012.
- [322] W. L. Delano. The PyMOL molecular graphics system. 2002.
- [323] P. Schuck. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, 78:1606–1619, 2000.
- [324] Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D Biological Crystallography*, 66:133–144, 2010.
- [325] A. Drozdetskiy, C. Cole, J. Procter, and G. J. Barton. JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*, 43 (W1):W389 – W394, 2015.
- [326] C. Aslanidis and P. J. de Jong. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research*, 18:6069–6074, 1990.
- [327] U.K. Laemmli. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227 (5259):680–685, 1970.
- [328] The Regents of University of California. Protein Prospector, MS-Digest. 1995-2018.
- [329] A. Shevchenko, M. Wilm, O. Vorm, and M Mann. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analyt. Chemistry*, 68:850–858, 1996.
- [330] J. V. Olsen, L. M. F. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, and M. Mann. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. & Cel. Proteomics*, 4:20210–2021, 2005.
- [331] Z. Wu, X. Yang, G. Weber, and X. Liu. Plk1 Phosphorylation of TRF1 Is Essential for Its Binding to Telomeres. *J. Biol. Chem.*, 283:25503–25513, 2008.
- [332] D. C. Rodriguez Camargo, K. J. Korshavn, A. Jussupow, K. Raltchev, D. Goricanec, M. Fleisch, R. Sarkar, K. Xue, M. Aichler, G. Mettenleiter, A. K. Walch, C. Camilloni, F. Hagn, B. Reif, and A. Ramamoorthy. Stabilization and structural analysis of a membrane-associated hIAPP aggregation intermediate. *eLife*, 6:e31226, 2017.