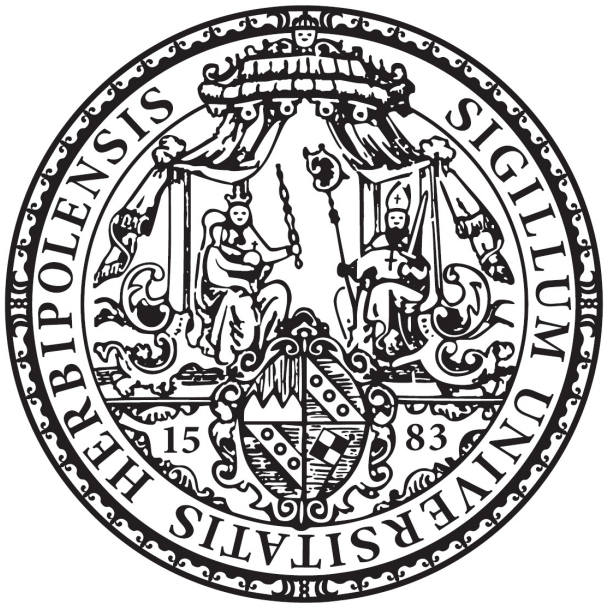


The Intricate Network of Replication-dependent Interstrand Crosslink DNA Repair

Das komplexe Netzwerk der replikationsabhängigen Reparatur von DNA-Quervernetzungen

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-University Würzburg,
Section Biomedicine

submitted by
Florian Rohleder
from Alheim-Hergershausen

Würzburg, 2014

Submitted on: .....
(Office stamp)

Members of the Supervisory Committee:

Chairperson:	Prof. Dr. Thomas Dandekar
Primary Supervisor:	Prof. Dr. Caroline Kisker
Second Supervisor:	Prof. Dr. Dr. h.c. mult. Roland Benz
Third Supervisor:	Prof. Dr. Detlev Schindler

Date of Public Defense: .....

Date of Receipt of Certificates: .....

Summary

The Fanconi anemia (FA) pathway is a replication-dependent DNA repair mechanism which is essential for the removal of interstrand crosslink (ICL) DNA damages in higher eukaryotes (Moldovan and D'Andrea, 2009). Malfunctions in this highly regulated repair network lead to genome instability (Deans and West, 2011). Pathological phenotypes of the disease FA which is caused by mutations in the eponymous pathway are very heterogeneous, involving congenital abnormalities, bone-marrow failure, cancer predisposition and infertility (Auerbach, 2009). The FA pathway comprises a complex interaction network and to date 16 FA complementation groups and associated factors have been identified (Kottemann and Smogorzewska, 2013). Additionally, components of nucleotide excision repair (NER), homologous recombination repair (HRR), and translesion synthesis (TLS) are involved and coordinated by the FA proteins (Niedzwiedz et al., 2004; Knipscheer et al., 2009). One of the FA proteins is the DEAH helicase FANCM. In complex with its binding partners FAAP24 and MHF1/2 it binds the stalled replication fork and activates the FA damage response (Wang et al., 2013). However, the exact steps towards removal of the ICL damage still remain elusive.

To decipher the underlying process of FA initiation by FANCM, this thesis mainly focuses on the archaeal FANCM homolog **h**elicase-associated **e**ndonuclease for **f**ork-structured DNA (Hef). Hef from the archaeal organism *Thermoplasma acidophilum* (taHef) differs from other archaeal Hef proteins and exclusively comprises an N-terminal helicase entity with two RecA and a thumb-like domain while others additionally contain a nuclease portion at the C-terminus. I solved the crystal structure of full-length taHef at a resolution of 2.43 Å. In contrast to the crystal structure of the helicase domain of Hef from *Pyrococcus furiosus* (pfHef), taHef exhibits an extremely open conformation (Nishino et al., 2005b) which implies that a domain movement of the RecA-like helicase motor domains of 61° is possible thus highlighting the flexibility of helicases which is required to translocate along the DNA. However, small-angle x-ray scattering (SAXS) measurements confirm an intermediate conformation of taHef in solution indicating that both crystal structures represent rather edge states. Most importantly, proliferating cell nuclear antigen (PCNA) was identified as an interaction partner of Hef. This interaction is mediated by a highly conserved canonical PCNA interacting peptide (PIP) motif. Intriguingly, the presence of PCNA does not alter the ATPase nor the helicase activity of taHef, thus suggesting that the interaction is entirely dedicated to recruit taHef to the replication fork to fulfill its function. Due to a high level of flexibility the taHef-taPCNA complex could not be crystallized and therefore SAXS was utilized to determine a low-resolution model of this quaternary structure.

This newly discovered PCNA interaction could also be validated for the eukaryotic FANCM homolog Mph1 from the thermophilic fungus *Chaetomium thermophilum* (ctMph1). As the first step towards the characterization of this interaction I solved the crystal structure of PCNA from *Chaetomium thermophilum* (ctPCNA).

Furthermore, it was possible to achieve preliminary results on the putative interaction between the human proteins FANCM and PCNA (hsFANCM, hsPCNA). In collaboration with Detlev Schindler (Human Genetics, Würzburg) and Weidong Wang (National Institute on Aging, Baltimore, USA) co-immunoprecipitation (CoIP) experiments were performed using hsFANCM and hsPCNA expressed in HEK293 cells. Although an interaction was reproducibly observed in hydroxyurea stimulated cells further experiments and optimization procedures are required and ongoing.

Zusammenfassung

Der Fanconi Anämie (FA) Signalweg ist ein replikationsabhängiger DNA-Reparaturmechanismus, der grundlegend zur Beseitigung von DNA-Schäden in Form von intermolekularen Quervernetzungen (ICL) beiträgt (Moldovan and D'Andrea, 2009). Fehlfunktionen in diesem stringent regulierten Reparaturnetzwerk führen somit zu Genominstabilität (Deans and West, 2011). Der pathologische Phänotyp der Krankheit FA, die durch Mutationen in dem gleichnamigen DNA-Reparatur Signalweg verursacht wird, ist sehr heterogen und umfasst angeborene Deformationen, Knochenmarksversagen, eine erhöhte Tumor Disposition sowie Infertilität (Auerbach, 2009). Der FA Mechanismus ist ein komplexes Netzwerk und bisher wurden 16 FA Komplementationsgruppen sowie weitere beteiligte Faktoren identifiziert (Kottemann and Smogorzewska, 2013). Zusätzlich sind Komponenten der Nukleotid-Exzisionsreparatur (NER), der homologen Rekombinationsreparatur (HRR) und Transläsionssynthese (TLS) involviert, die durch FA Proteine koordiniert werden (Niedzwiedz et al., 2004; Knipscheer et al., 2009). Eines der FA Proteine ist die DEAH Helikase FANCM. Im Komplex mit seinen Interaktionspartnern FAAP24 und MHF1/2 bindet FANCM an die durch den ICL Schaden zum Stillstand gekommene Replikationsgabel und aktiviert die FA Schadensantwort (Wang et al., 2013). Die weiteren Schritte, die zur Entfernung des ICL Schadens führen, sind jedoch weitestgehend ungeklärt.

Zur Aufklärung der Initiation des FA Mechanismus und der Rolle, die das FANCM dabei spielt, wurde in dieser Arbeit hauptsächlich das archaische FANCM Homolog *Helicase-associated Endonuclease for Fork-structured DNA* (Hef) analysiert. Hef aus dem archaischen Organismus *Thermoplasma acidophilum* (taHef) unterscheidet sich von anderen archaischen Hef Proteinen und besteht ausschließlich aus einem N-terminalen Helikase-Abschnitt mit zwei RecA und einer *thumb-like* Domäne, während andere Hef Proteine am C-Terminus zusätzlich eine Nuklease-Domäne besitzen. Ich habe die Kristallstruktur des taHef Proteins bei einer Auflösung von 2,43 Å gelöst. Im Gegensatz zur Kristallstruktur eines vergleichbaren Hef-Konstruktes aus *Pyrococcus furiosus* (pfHef) (Nishino et al., 2005b) liegt in taHef eine extrem offene Konformation der beiden RecA-Domänen vor, was impliziert, dass eine Bewegung der RecA-ähnlichen Helikase Motordomänen um 61° möglich ist und zudem die zur Translokation entlang der DNA notwendige Flexibilität von Helikasen verdeutlicht. Messungen mittels Kleinwinkelröntgenstreuung (SAXS) deuten hingegen auf eine intermediäre Konformation des taHef Proteins in Lösung hin, wodurch beide Kristallstrukturen als eher Randzustände angesehen werden können. Besonders hervorzuheben ist, dass das Protein *Proliferating Cell Nuclear Antigen* (PCNA) als Hef Interaktionspartner

identifiziert wurde. Diese Interaktion wird durch ein hoch-konserviertes kanonisches PCNA Interaktionspeptid-Motiv vermittelt. Interessanterweise beeinflusst PCNA aber weder die ATPase noch die Helikase Aktivität von taHef, was darauf hindeutet, dass diese Interaktion nur zur Rekrutierung des Hef Proteins zur Replikationsgabel dient. Wegen des hohen Maßes an Flexibilität konnte der taHef-taPCNA Komplex nicht kristallisiert werden, wohingegen SAXS Messungen erfolgreich waren und ein Model bei niedriger Auflösung konnte erhalten werden.

Diese nachgewiesene Interaktion zwischen Hef und PCNA konnte auch für das eukaryotische FANCM Homolog Mph1 aus dem thermophilen Pilz *Chaetomium thermophilum* (ctMph1) bestätigt werden. Als ersten Schritt zur Charakterisierung dieser Interaktion habe ich die Kristallstruktur von PCNA aus *Chaetomium thermophilum* (ctPCNA) gelöst.

Weiterhin war es möglich, vorläufige Resultate bezüglich der mutmaßlichen Interaktion zwischen den humanen Proteinen FANCM und PCNA (hsFANCM, hsPCNA) zu erhalten. In Kooperation mit Detlev Schindler (Humangenetik, Würzburg) und Weidong Wang (National Institute on Aging, Baltimore, USA) wurden Co-Immunopräzipitations-Experimente (CoIP) mit humanem FANCM und humanem PCNA aus HEK293-Zellen durchgeführt. Obwohl eine Interaktion in Hydroxyurea-stimulierten Zellen reproduzierbar nachgewiesen werden konnte, sind weitere Experimente notwendig, um diese Interaktion zu charakterisieren.

Contents

1 Introduction

1.1 DNA Maintenance

DNA as the carrier of the genetic information is a highly dynamic molecule and its preservation is fundamental for every living organism. The incorrect constitution of the DNA consequently provokes mutations, persistent damages, replication errors, and general genomic instability leading to severe diseases such as cancer and aging. The DNA is constantly attacked by endogenous and exogenous factors, e.g. spontaneous mutations, oxygen radicals, alkylating agents, UV-light, polycyclic aromatic hydrocarbons, crosslinking agents, and many others. Various commonly used anti-tumor agents also induce DNA lesions that constitute a severe thread to fast dividing cells. Therefore maintenance of genomic integrity at each phase of the cell cycle is a major task to the cell and distinct repair mechanisms have evolved which are adopted to a large variety of different DNA damages (Hoeijmakers, 2001).

Thus organisms of all kingdoms of life have developed various repair pathways with different modes of action and specificities. There are repair mechanisms that are specific for one particular type of damage such as DNA interstrand crosslink (ICL) repair (Chapter 1.2) (Dronkert and Kanaar, 2001). In contrast, nucleotide excision repair (NER) is able to remove a wide range of different lesions (Kuper and Kisker, 2012). NER is also a good example for the analogous development of repair mechanisms in prokaryotes and eukaryotes. While in prokaryotes damage detection and excision of a 12 nucleotide DNA patch containing the lesion is accomplished by only three proteins (UvrABC) (Kisker et al., 2013), in eukaryotes the entire process removing 24 - 32 nucleotides involves more than 30 different proteins (Schärer, 2013). Interestingly, archaeal and eukaryotic organisms share considerable similarities in their DNA processivity such as replication, transcription, and repair mechanisms. Thus archaeal DNA repair proteins are generally accepted as simplified models for their more complex eukaryotic homologs (White, 2003).

Importantly, *in vivo* many DNA repair pathways do not function independently. They are rather connected and build a complex and tightly regulated repair network. The central platform proliferating cell nuclear antigen (PCNA) is involved in diverse repair mechanisms and according to the nature of the damage it sets the course for the proper downstream reaction. The ubiquitination state of PCNA decides between polymerase switch in combination with translesion synthesis (TLS) or postreplication repair (PRR) (Lee and Myung, 2008). Another well characterized example for the crosstalk between distinct damage response pathways is the tumor suppressor protein p53 that controls

the cell fate in response to cellular stress (Elias et al., 2014). Activated by factors such as hypoxia, DNA damages, oncogenes, developmental stimuli, and others p53 targets various downstream genes that regulate transient cell-cycle arrest, differentiation, senescence, apoptosis, or that are connected to other pathways. Thereby p53 prevents proliferation and inhibits the generation of daughter cells with incorrect genetic content (Stiewe, 2007). Interestingly, PCNA also interacts with p53 and regulates its stability (Banks et al., 2006).

Although the entire DNA repair machinery is a scrupulously balanced and fine-tuned network to ensure genome stability, it is also important to keep the genetic content dynamic and guarantee evolutionary progress by the ability to adapt to environmental changes. A hypothetically perfect and absolutely error-free DNA repair machinery would cause stagnation and consequently result in competitive disadvantages to the organism. Hence, the restrained insertion of errors, e.g. as introduced by TLS polymerases, is crucial for life as well (Prakash et al., 2005).

1.2 Interstrand Crosslink DNA Damage

Interstrand crosslink (ICL) damages (Figure 1.1) constitute an extremely dangerous class of DNA lesions. The covalent and irreversible linkage between nucleotides of opposite DNA strands within the double-helix prevents strand separation during replication and transcription (Deans and West, 2011). Even a single unrepaired or not bypassed crosslink in yeast and about 40 ICLs in repair-deficient mammalian cells, respectively, can trigger apoptosis (Magana-Schwencke et al., 1982; Lawley and Phillips, 1996).

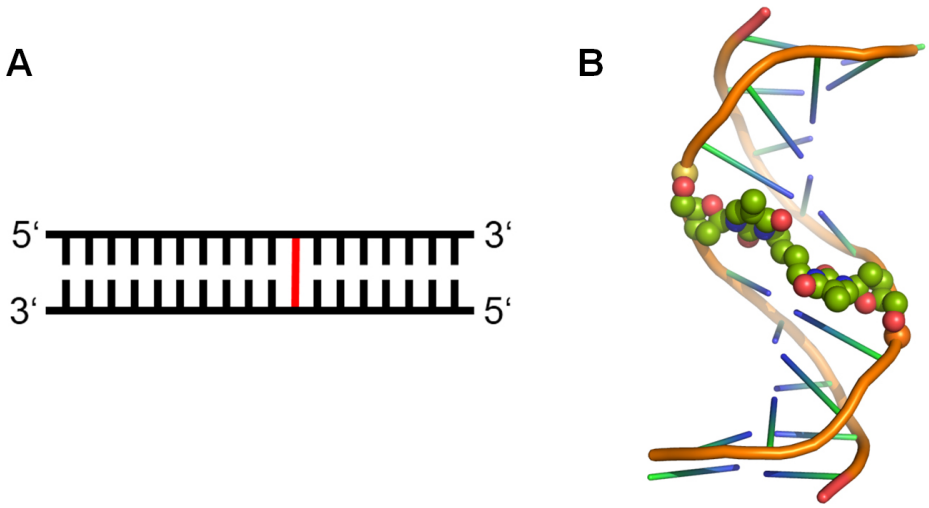


Figure 1.1: dsDNA containing an ICL lesion.
(A) Schematic illustration of an short dsDNA comprising an ICL (red) that covalently connects the two ssDNAs. **(B)** NMR solution structure of an N3T-butyl-N3T ICL (spheres) linked B-DNA (PDB code: 1XCI) (da Silva et al., 2005).

ICLs are caused by various endogenous and exogenous agents which contain two or more reactive groups and differ in size as well as in the extent of helix distortion. Endogenous sources are aldehydes (e.g. trans-4-hydroxynonenal, acetaldehyde, malondialdehyde, acrolein, formaldehyde, and crotonaldehyde) from lipid peroxidation, nitric oxide and nitrous acid occurring in cell signaling, 5'-(2-phosphoryl-1,4-dioxobutane) under hypoxic conditions, spontaneous hydrolysis of purines or base excision repair (BER) intermediates. Exogenous ICL-inducing agents such as nitrogen mustards (cyclophosphamide, melphalan, mechlorethamine, chlorambucil, ifosfamide, and bendamustine), platinum compounds (cisplatin, carboplatin, oxaliplatin, satraplatin, and picoplatin), mitomycin C, psoralens, and chloro-ethyl nitrosoureas are commonly used chemotherapeutics in clinical cancer treatment (Clauson et al., 2013). Distinct agents also vary in their cellular and nuclear uptake properties, crosslinking efficiency, and the resulting toxicity. Interestingly, ICL agents form a number of adducts with DNA and actual ICL damages only represent a small fraction. While only 1-5% of adducts formed by nitrogen mustards constitute ICLs, psoralens produce up to 40%. Furthermore, ICLs are not distributed randomly throughout the genome because the chromatin structure additionally influences adduct formation (Dronkert and Kanaar, 2001). For experimental analysis synthetic ICLs can be generated that mimic lesions caused by various natural sources (Guainazzi et al., 2010).

While in *E. coli* ICL damages are mainly repaired by NER and homologous recombination (HR) (Van Houten and McCullough, 1994; Cole, 1973), in higher eukaryotes specific ICL repair pathways have evolved that eliminate the lesion either during or outside of the S phase (Williams et al., 2013). Notably, an exceptional feature of ICL repair in comparison to other DNA repair pathways is the necessity to incise both DNA strands for damage removal.

1.2.1 Replication-Independent ICL Repair

ICL DNA damages do not constitute a threat only during S phase but also during regular cellular functions such as transcription which occurs in the G0 and G1 phases. For proper mRNA synthesis strand separation is essential and therefore the replication-independent ICL repair (RIR) exists next to the replication-dependent ICL repair. Although this mechanism is poorly understood it was shown that the classical repair pathways NER and TLS upon PCNA ubiquitination are involved. Interestingly, in the replication-dependent FA pathway (Chapter 1.2.2) TLS polymerases are coordinated by a different yet unknown mechanism (Ho and Schärer, 2010). However, further factors and the exact processes implicated in RIR still remain elusive. The incomplete RIR model proposed by J. Gautier is shown in Figure 1.2 (Williams et al., 2013).

On the way of understanding RIR various factors and correlations could be identified. Importantly, NER plays a major role. Cockaine syndrome proteins A and B (CSA and CSB) of transcription-coupled NER (TCR-NER) were shown to be necessary for

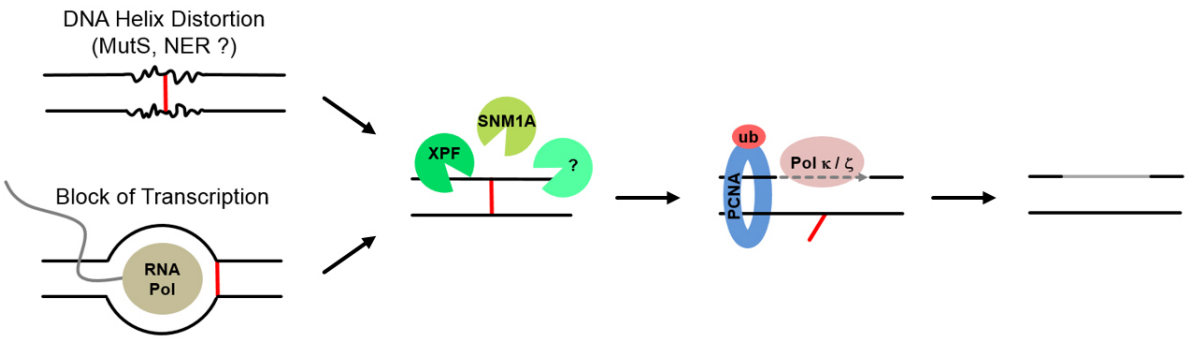


Figure 1.2: Simplified model of replication-independent ICL repair.
Initially, the ICL damage is detected by NER or mismatch repair (MMR) enzymes due to a helix distortion or the blocked replisome. Presumably, ICL collisions with proteins during other DNA transactions can also activate RIR. The nuclease XPF-ERCC1 in combination with SNM1A is involved in unhooking the lesion but the 3' acting nuclease is still unknown. TLS Pol κ and Pol ζ bypass the lesion and finally ICL repair is completed by additional factors (Williams et al., 2013).

mitomycin c (MMC) induced ICL repair during transcription (Zheng et al., 2003). Mutations in XPC also impair MMC induced ICL repair indicating that global genome NER (GG-NER) is important for RIR, as well. Interestingly, cisplatin induced ICL repair requires XPA, XPF, and XPG but not XPC (Enoiu et al., 2012). 5' incision to the lesion by the XPF-ERCC1 heterodimer in combination with the 5'-3' exonuclease SNM1A is crucial for unhooking the ICL damage while the 3' acting nuclease is not known (Wang et al., 2011). It was also suggested that SNM1A is recruited by Rad18 and ubiquitinated PCNA (Yang et al., 2010). Furthermore, Rad6/Rad18-dependent PCNA K164 monoubiquitination is important for the activation of TLS DNA polymerases kappa (Pol κ) and zeta (Pol ζ) in ICL repair (Williams et al., 2012; Sarkar et al., 2006). However, the data on the TLS polymerases are inconsistent. While the model above clearly states that in the replication-independent ICL repair exclusively Pols κ and ζ are involved another study assigns Pol ζ and Rev1 to this pathway (Enoiu et al., 2012). Since double-strand breaks (DSBs) do not occur in the RIR pathway the homologous recombination repair (HRR) protein Rad51 is not required (Williams et al., 2012).

Although the RIR pathway is so far poorly understood, it nicely demonstrates a closely intertwined network between different DNA repair pathways (Williams et al., 2013).

1.2.2 Fanconi Anemia Repair Pathway

The Fanconi anemia (FA) pathway promotes replication-dependent ICL repair in higher eukaryotes (Figure 1.3) (Knipscheer et al., 2009). This pathway is named after the Swiss pediatrician Guido Fanconi who lived from 1892 - 1979 and initially described this autosomal recessive and X-linked genetic disease (Fanconi, 1927). Patients with mutations in FA complementation group genes are characterized by their hypersensitivity towards DNA crosslinking agents and display chromosomal instability

including chromosomal alterations, double-strand breaks, hyper-recombination, and rearrangements, providing an explicit marker for the FA genotype. Pathological phenotypes of FA are very heterogeneous, involving congenital abnormalities, bone-marrow failure, cancer predisposition and infertility with a heterozygous carrier frequency of 1 in 300 (Auerbach, 2009). The median age of cancer onset is 14 years (Moldovan and D'Andrea, 2009). FA is categorized as a rare disorder with an incidence rate of 1 in 350,000 births (Shukla et al., 2012).

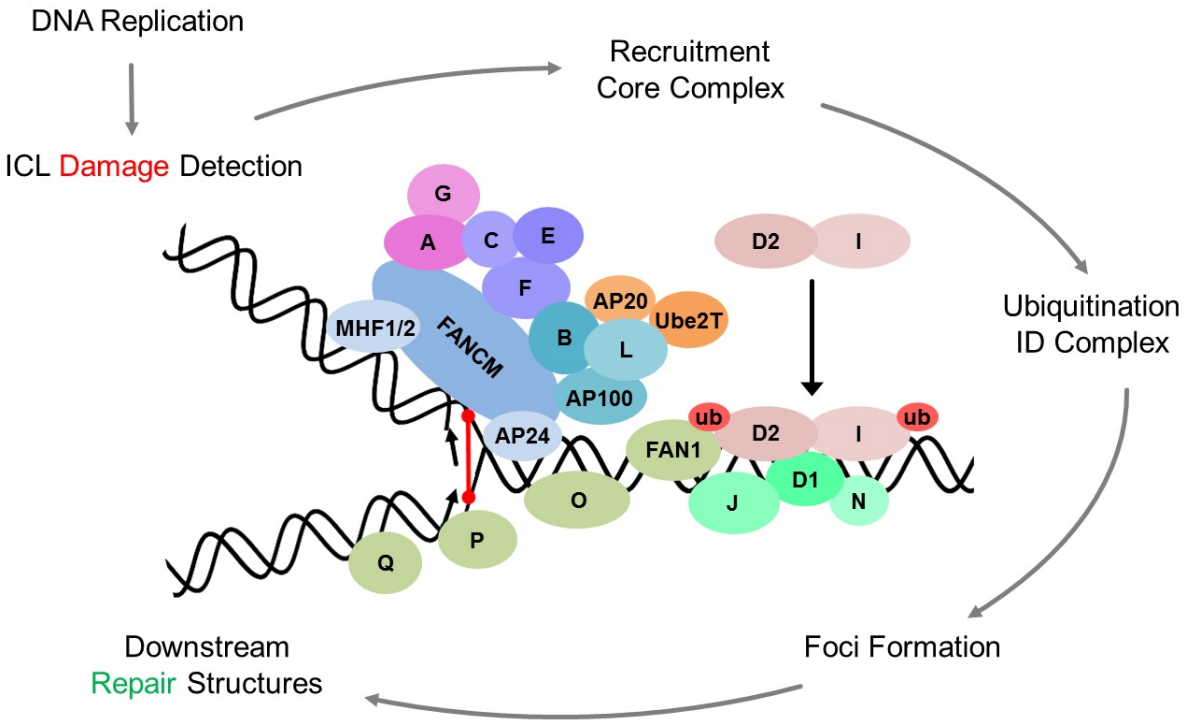


Figure 1.3: Schematic illustration of the Fanconi anemia DNA repair pathway. The DEAH helicase FANCM in complex with its binding partners FAAP24 and the heterotetrameric MHF1/2 complex recognizes the ICL stalled replication fork. Subsequently FANCM recruits the FA core complex (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG (XRCC9), and the associated proteins FAAP100 and FAAP20) comprising three subcomplexes to the lesion (Ciccia et al., 2007; Singh et al., 2010). The main function of the core complex is the ensuing mono-ubiquitination of FANCI and FANCD2 (ID-complex) by the integrated E3 ubiquitin ligase FANCL and the E2 ubiquitin conjugating enzyme Ube2T (Machida et al., 2006; Meetei et al., 2003). The ubiquitinated ID-complex then recruits FANCD1 (BRCA2), the iron-sulfur cluster containing helicase FANCJ (BRIP1), and FANCN (PALB2) to DNA foci triggering downstream repair structures that involve FANCO (Rad51C), FANCP (SLX4), FANCQ (XPF), and the ID-complex interacting nuclease FAN1 (Somyajit et al., 2012; Kim et al., 2011; Bogliolo et al., 2013; Liu et al., 2010). The exact downstream steps towards removal of the ICL damage still remain elusive.

ICL repair by the FA pathway is so far only poorly understood. To date 16 FA complementation groups and additional associated factors have been identified

(Kottemann and Smogorzewska, 2013).

Exposed ssDNA patches resulting from the ICL stalled replication fork can easily evoke double-strand breaks representing a serious threat to genome stability. Thus it is very important to quickly and efficiently process the stalled replication fork which is achieved by the FA ICL repair triggering a sophisticated repair cascade involving multiple protein complexes and several posttranslational modifications including the crucial ubiquitination of the ID-complex (Alpi and Patel, 2009). The FA pathway connects components of the classic repair systems NER, HRR, and TLS to remove ICL damages (Niedzwiedz et al., 2004; Knipscheer et al., 2009). XPF, which also functions in RIR, plays an important role in the downstream processing of the lesion during FA and is thus annotated as FANCQ (Bogliolo et al., 2013). In contrast to RIR, in FA the TLS polymerases Rev1 and Pol ζ are coordinated and the HRR protein Rad51C (also termed as FANCO) is involved (Williams et al., 2013; Somyajit et al., 2012). As mentioned before (Chapter 1.2.1) the assignment of the different TLS polymerases to the distinct RIR/FA pathways is unclear. Rev1 and Pol ζ are also found in RIR (Enoiu et al., 2012). Importantly, the actual removal of the ICL damage in downstream processes still remains elusive.

Notably, the FA repair pathway is not restricted to ICL damages. It is also activated by various types of DNA damages such as lesions caused by ultraviolet (UV) radiation, ionizing radiation (IR), hydroxyurea (HU), and even spontaneously occuring errors during replication (Moldovan and D'Andrea, 2009).

1.3 The Archaeal and Eukaryotic FANCM Protein Family

Helicases play a key role in DNA metabolism and thereby are also crucial for DNA repair processes. RecQ and iron-sulfur (Fe-S) helicase families have been shown to be essential for maintenance of the genomic integrity (Brosh, 2013). Besides their damage detection activity and the recruitment of further repair factors, additional functional domains often contribute to their substrate specificity and auxiliary properties. Rad5 and it's human ortholog HLTF facilitate replication fork regression and PCNA poly-ubiquitination via HIRAN and E3 RING domains (Unk et al., 2010).

The DNA remodeling enzyme Fanconi anemia complementation group M (FANCM) recognizes ICL-stalled replication forks and activates the FA pathway downstream repair cascade via multiple interactions (Thompson and Jones, 2010). Although the FA pathway exists only in higher eukaryotes, FANCM homologs (FANCM, Mph1, Fml1, and Hef) are present in all kingdoms of life acting on fork-structured DNA and recombination intermediates (Moldovan and D'Andrea, 2009). A common feature of this protein family is the N-terminal helicase entity consisting of the characteristic RecA-like motor domains Hel1 and Hel2 that comprise the seven conserved SF2

helicase motifs including a Walker A motif and a DEAH box which is a variant of the Walker B motif (Singleton et al., 2007). The thumb-like insertion domain (Ins) which is responsible for substrate specificity, is embedded within Hel2. In addition to an internal yet poorly characterized linker region of varying length, some family members also contain an enzymatically either active or inactive ERCC4-like nuclease portion and a tandem helix-hairpin-helix $(HhH)_2$ domain at the C-terminus that function as a homo- or heterodimerization interface (Whitby, 2010).

The domain architecture of FANCM and its homologs is illustrated in Figure 1.4.

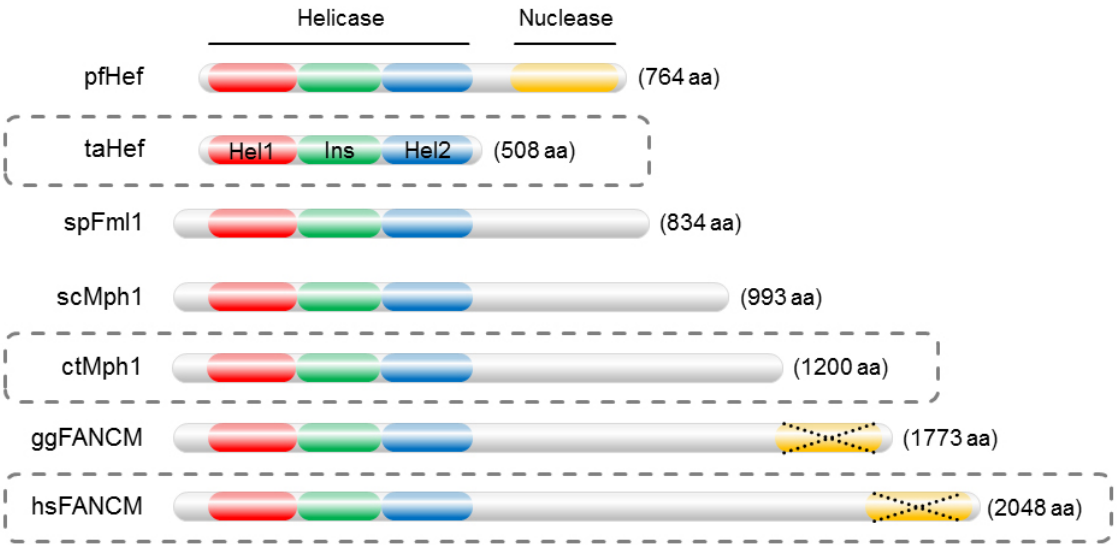


Figure 1.4: Domain architecture of FANCM protein family members.
FANCM and its homologs (Hef = helicase-associated endonuclease for fork-structured DNA; Fml1 = FANCM-like protein; Mph1 = mutator phenotype) from different archaeal and eukaryotic organisms (pf = *Pyrococcus furiosus*; ta = *Thermoplasma acidophilum*; sp = *Schizosaccharomyces pombe*; sc = *Saccharomyces cerevisiae*; ct = *Chaetomium thermophilum*; gg = *Gallus gallus*; hs = *Homo sapiens*). RecA-like helicase domains Hel1 and Hel2 are depicted in red and blue, respectively. The thumb-like insertion domain (Ins) is shown in green and uncharacterized regions (N-terminal elongations and the internal linker) are shown in grey. The enzymatically inactive nucleases within ggFANCM and hsFANCM are indicated by the black dotted crosses in the nuclease domain which is shown in yellow. Numbers in brackets reveal the lengths of the proteins in numbers of amino acids (aa). Dashed frames highlight the proteins that were analyzed in this thesis.

1.3.1 Helicase-Associated Endonuclease for Fork-Structured DNA

The helicase-associated endonuclease for fork-structured DNA (Hef) is an archaeal bi-functional DNA processing enzyme containing a DEAH helicase entity at the N-terminus and a structure-specific ERCC4-like endonuclease domain within the C-terminal third of the protein. The helicase part has been shown to comprise a

specific affinity to fork-structured DNA substrates and the nuclease cleaves nicked duplexes, 3'-flapped, 5'-flapped, and pseudo-X DNAs with the same efficiency incising at the 5'-side of the nick (Komori et al., 2002). Furthermore, Hef helicase and nuclease activities function cooperatively on stalled replication forks. DNA-dependent ATP hydrolysis by the N-terminal helicase part of Hef upon unwinding of forked-structured and Holliday junction (HJ) containing DNAs simultaneously stimulates the endonuclease activity of the protein, suggesting that Hef resolves stalled replication forks in a two-step process (Komori et al., 2004). The directionality of DNA unwinding has not been reported. It is assumed that the helicase function prepares the ideal conformation of the DNA for efficient cutting (Nishino et al., 2005a).

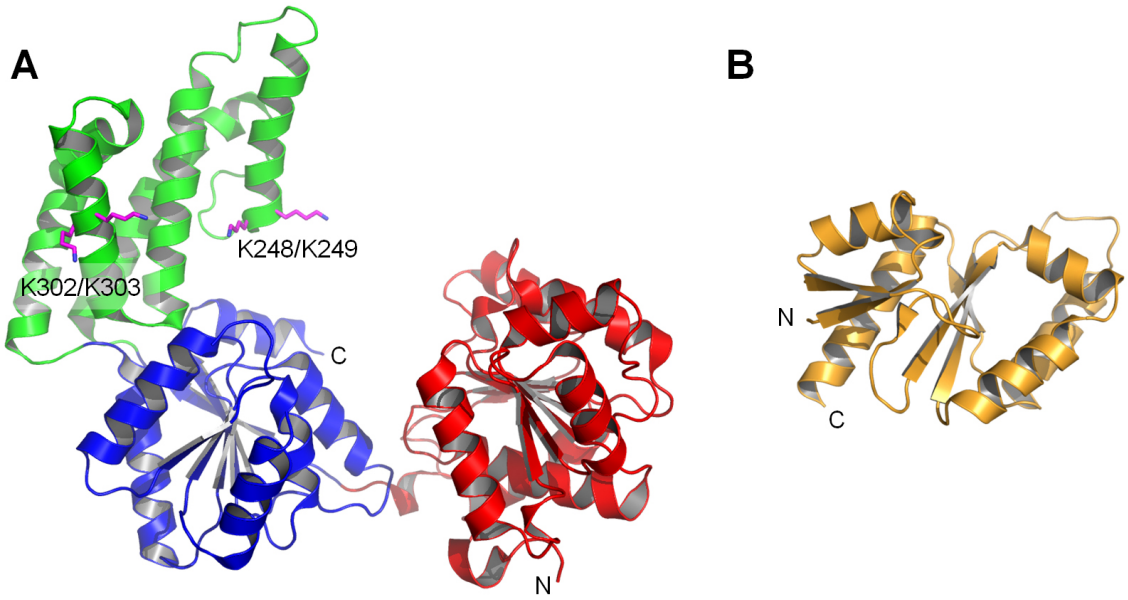


Figure 1.5: Crystal structure of Hef from *Pyrococcus furiosus* (pfHef).
(A) Cartoon representation of the helicase entity of pfHef (residues 1 - 494). SF2 helicase motordomains Hel1 (red) and Hel2 (blue) exhibit characteristic RecA-like folds. The thumb-like insertion domain (Ins, green) is embedded between β8 and α14 of Hel2. Residues K248/K249 and K302/303 important for fork DNA specificity are depicted in magenta (PDB code: 1WP9) (Nishino et al., 2005b). **(B)** Cartoon representation of the nuclease domain of pfHef (residues 547 - 759) comprising a type II restriction endonuclease fold (PDB code: 1J23) (Nishino et al., 2003).

Importantly, to date the Hef crystal structures by T. Nishino are the only available models for FANCM family proteins. It was not possible to crystallize full-length Hef and thus the individual helicase and nuclease entities of Hef from *Pyrococcus furiosus* (pfHef) were solved (Figure 1.5) (Nishino et al., 2003, 2005b). The helicase entity (Figure 1.5A) comprises the two characteristic SF2 helicase domains Hel1 and Hel2 (red and blue, respectively) that contain the conserved helicase motifs and are crucial for ATP binding and hydrolysis. The thumb-like insertion domain (Ins, green) is integrated within Hel2 and crucial for substrate specificity, indicating structural

and functional similarities to the thumb domain of *Taq* DNA polymerase (Kim et al., 1995). Interestingly, deletion of the entire Ins domain leads to the inability to unwind fork-structured DNAs but retains the capability to dissociate HJs with slightly decreased efficiency. Further mutational analysis revealed that residues K248/K249 and K302/K303 within the Ins domain play an important role for substrate specificity towards forked DNA (Nishino et al., 2005b). The C-terminal nuclease domain (Figure 1.5B) comprises the GDX_nERKX_3D signature motif of the Rad1/ERCC4 endonuclease superfamily which is responsible for the coordination of the divalent cation needed for catalysis followed by a DNA-binding HhH domain (Aravind et al., 1999). This nuclease portion is structurally related to type II restriction endonucleases and additionally serves as the interface for Hef homodimerization. Mutations leading to monomeric specimen also significantly reduce the endonuclease activity (Nishino et al., 2003). Upon dimerization through the nuclease domain Hef assumes a non-globular elongated conformation in solution (Lestini et al., 2013). These models of the individual domains already provide some insights into the structure-function correlation of Hef but they are insufficient for a comprehensive structural description including conformational dynamics needed for DNA processivity.

Hef was identified as an archaeal FANCM homolog and although the FA pathway is not conserved in archaea Hef is proposed to function in a DNA repair mechanism acting on fork-structured DNA (Meetei et al., 2005). Indeed, *in vivo* experiments in the euryarchaeon *Haloferax volcanii* showed Hef foci formation in response to replication arrest suggesting a stabilizing effect due to the direct interaction with collapsed replication forks (Lestini et al., 2013). This observation strongly supports the notion that Hef can be viewed as a simplified model for FANCM. So far there are no interaction partners of Hef known.

In the present study the recombinantly in *E. coli* expressed and purified archaeal proteins from the model organism *Thermoplasma acidophilum* were utilized. This thermoacidophilic archaeon can be found at self-heating cole refuse piles or solfatara fields and lives at temperatures between 45 and 62 °C and a pH range from 0.96 to 3.5. Optimal growth conditions comprise a temperature of 59 °C and a pH value of 2 (Darland et al., 1970). The 1.56 Mbp genome of *Thermoplasma acidophilum* consists of a single circular chromosome and sequencing revealed 1509 ORFs (Ruepp et al., 2000).

1.3.2 Fanconi Anemia Complementation Group M

The SF2 DEAH translocase Fanconi anemia complementation group M (FANCM) is the initiator of the FA DNA repair pathway (Xue et al., 2008). FANCM actually fails to unwind DNA *in vitro* but its translocase activity was demonstrated using a triple-helix displacement assay (Meetei et al., 2005). Additionally, previous studies showed that FANCM promotes branch point migration on HJs, replication fork reversal, and is also able to dissociate D-loops (Gari et al., 2008a,b). In contrast to archaeal Hef, FANCM has lost its ability to cleave DNA due to variations of key residues and electrostatic

properties at the active site region in the endonuclease domain (Yang et al., 2013).

So far it has only been possible to solve small fragments of FANCM in complex with binding partners (Figure 1.6).

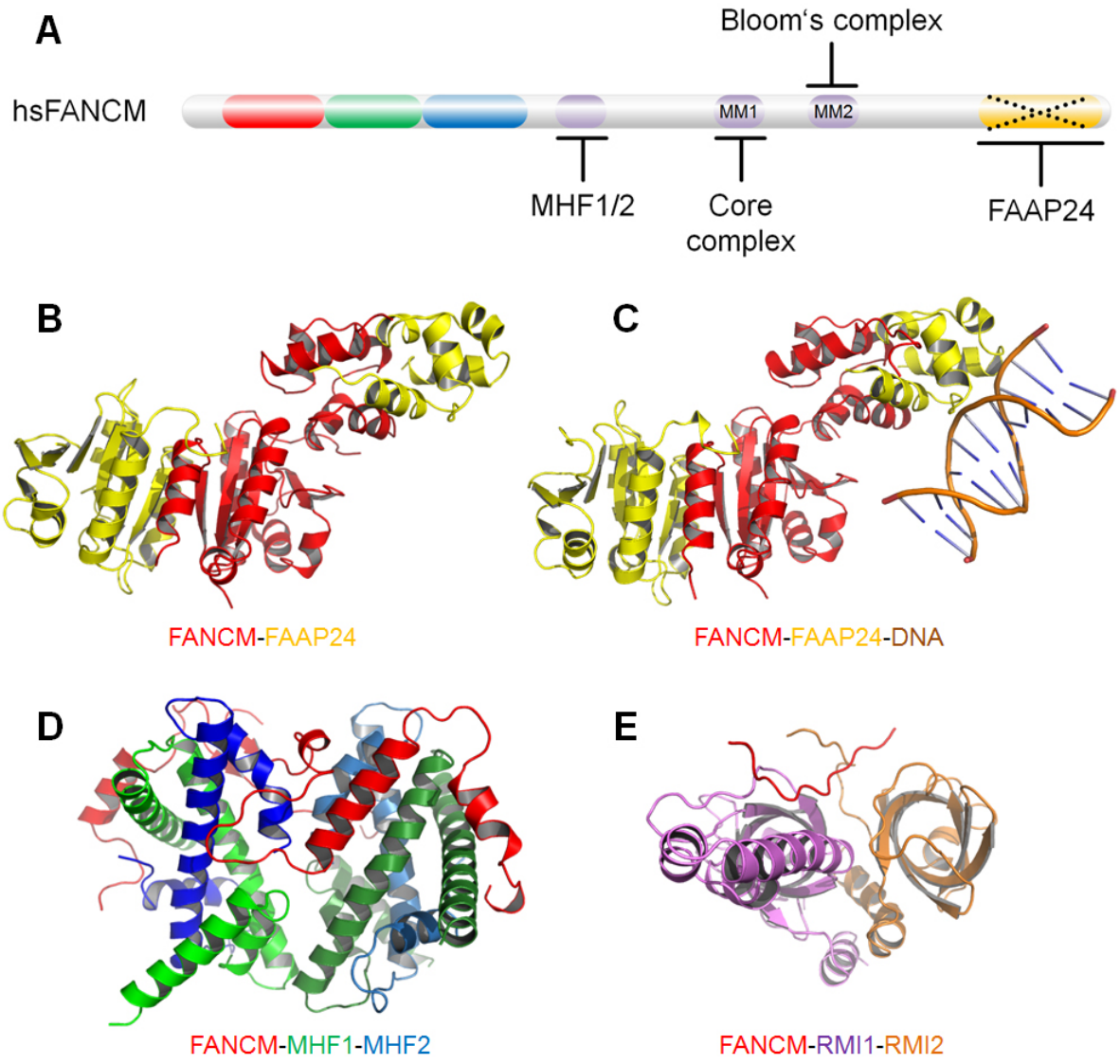


Figure 1.6: Crystal structures of hsFANCM fragments with interaction partners. **(A)** Schematic representation of hsFANCM with indicated locations of interaction interfaces. **(B)** FANCM$^{1813-2031}$ (red) and FAAP24 (yellow) (PDB code: 6M6W) (Yang et al., 2013). **(C)** FANCM$^{1818-2035}$ (red) FAAP24 (yellow) dsDNA (brown) (PDB code: 4XBO) (Coulthard et al., 2013). **(D)** FANCM$^{661-800}$ (red) MHF1$^{1-114}$ (green) MHF2 (blue) (PDB code: 4DRB) (Tao et al., 2012). **(E)** FANCM MM2 peptide in complex with the RecQ-mediated genome instability protein 1/2 dimer. FANCM$^{1218-1251}$ (red), RMI1$^{473-625}$ (magenta), and RMI2 (orange) (PDB code: 4DAY) (Hoadley et al., 2012).

FANCM interacts with FAAP24 via its C-terminal enzymatically inactive ERCC4-like nuclease domain. Interestingly, FAAP24 shares sequence homology with ERCC1 and contains a likewise inactive nuclease domain which is similar to the C-terminal domain of FANCM. This interaction is required for downstream FANCD2 monoubiquitination as well as ICL repair (Ciccia et al., 2007). Other direct interaction partners of FANCM are the histone-fold-containing FANCM-associated proteins MHF1 and MHF2. MHF1/2 forms a heterotetramer containing two units of each protein and binds FANCM C-terminally adjacent to the helicase entity (Tao et al., 2012). It was proposed that MHF1/2 mainly interacts with FANCM via MHF1 and MHF1 depletion also impairs FANCD2 ubiquitination (Singh et al., 2010). FANCM preferably binds branched DNA such as replication forks or HJs. FAAP24 specifically interacts with single-stranded DNA while MHF1/2 binds double-stranded DNA (Thompson and Jones, 2010). Thus the FANCM-FAAP24-MHF1/2 subcomplex is the only component of the core complex that directly binds to DNA. The FA core complex is recruited by binding of the highly conserved MM1 motif of FANCM to FANCF. Furthermore, the FANCM MM2 motif interacts with RM1 and topoisomerase IIIα which are both components of the genetic disorder Bloom's Syndrome (BS) complex. Interestingly, FA and BS show overlapping phenotypes such as an elevated frequency of sister chromatid exchanges. The observation that FANCM independently interacts with FA and BS proteins through motifs MM1 and MM2, respectively, indicates that FANCM functions as a bridge connecting both DNA repair pathways. Thereby FANCM also indirectly targets the genomic caretaker RecQ helicase BLM to the replication fork (Deans and West, 2009). Both motifs, MM1 and MM2, are located within the extensive linker region. It was also shown that FANCM is required for the activation of ATR by promoting TopBP1 loading on chromatin (Schwab et al., 2010). The interaction between FANCM-FAAP24 and ATR-ATRIP is mediated by HCLK2 which is important for DNA damage checkpoint signaling (Horejsi et al., 2009). Hence, FANCM seems to play a central role in genomic stability maintenance upon replication fork arrest.

FANCM posttranslational modifications are important for the regulation of the FA pathway. Phosphorylation of the phosphodegron ^{943}DSGYNS^{948} of hsFANCM by the kinase Plk1 triggers subsequent polyubiquitination by the β-TRCP component of the SCF E3 complex. This leads to the degradation of FANCM and simultaneously to the release of the core complex (Kee et al., 2009).

To date there are only three pathological mutations known within FANCM. The two nonsense mutations R658X and S724X as well as the deletion of exon 15 only affect the extensive linker region. The pathological effect of these mutations is still controversial because only a few FANCM defective individuals have been reported and these mutations have always been accompanied by mutations in other FA genes (Meetei et al., 2005; Harutyunyan et al., 2011).

1.3.3 Mutator Phenotype

The 3'-5' helicase mutator phenotype (Mph1) from *Saccharomyces cerevisiae* was the first member of the FANCM family that had been studied (Scheller et al., 2000). In contrast to Hef and FANCM, Mph1 does not comprise a nuclease domain at the C-terminus. It only consists of the characteristic helicase entity at the N-terminus and the adjacent long uncharacterized linker region (Whitby, 2010). There are no crystal structures of neither full-length nor fragments of Mph1 available so far.

Mph1 was shown to play an important role in yeast genome maintenance and various interaction partners could be identified. It is able to process a wide range of DNA substrates such as replication fork reversal, branch point migration of HJs, unwinding of 3' and 5' flaps, and the disruption of D-loop intermediates (Zheng et al., 2011b). Interestingly, MHF1/2 also appears to regulate Mph1 activity in replication fork repair and Mph1 dependent ICL repair is suggested as the yeast repair pathway homologous to FA (Yan et al., 2010; Singh et al., 2010). The Mph1 pathway is also epistatic with Rad5-mediated error-free DNA damage bypass but distinct from Rad18 and Pso2 ICL repair. Thus it is proposed that the PRR protein Rad5 polyubiquitinates PCNA and also recruits Mph1 to the ICL stalled replication fork to reverse and protect the fork during repair (Daee et al., 2012). Additional functions of Mph1 are the regulation of the Rad51-dependent, Rad59-independent homologous recombination pathway and also the suppression of chromosomal rearrangements through Rev3 (catalytic subunit of Pol ζ) (Schmidt et al., 2010). Hence, Mph1 seems to have a central function in the repair of stalled replication forks.

In the present study the recombinantly in *E. coli* expressed and purified proteins from the fungal model organism *Chaetomium thermophilum* were utilized (La Touche, 1948). This thermophilic eukaryote has evolved to a well established model organism and the derived thermostable proteins provide great advantages for biochemical and structural analyses (Amlacher et al., 2011).

1.4 Proliferating Cell Nuclear Antigen

Proliferating cell nuclear antigen (PCNA) is a co-factor for the replicative DNA polymerases delta and epsilon (Pols δ and ϵ) and additionally it serves as a central platform to coordinate multiple DNA metabolic processes such as replication and repair (Moldovan et al., 2007).

Ring-shaped DNA sliding clamps tether the polymerases to the DNA and significantly increase their processivity. PCNA and its functional homologs can be found in all branches of life and although there is barely any primary amino acids sequence similarity their 3D structure containing a pseudohexameric symmetry is well conserved (Indiani and O'Donnell, 2006). The sliding clamp of prokaryotes (β-clamp) consists of a homodimer whereas in eukaryotes (PCNA) and T4 bacteriophage (GP45) it

forms a homotrimer (Figure 1.7) (Oakley et al., 2003; Kontopidis et al., 2005; Moarefi et al., 2000). Interestingly, euryarchaeal PCNA also consists of a homotrimer while crenarchaeal PCNA arranges as a heterotrimer (Matsumiya et al., 2001; Dionne et al., 2003).

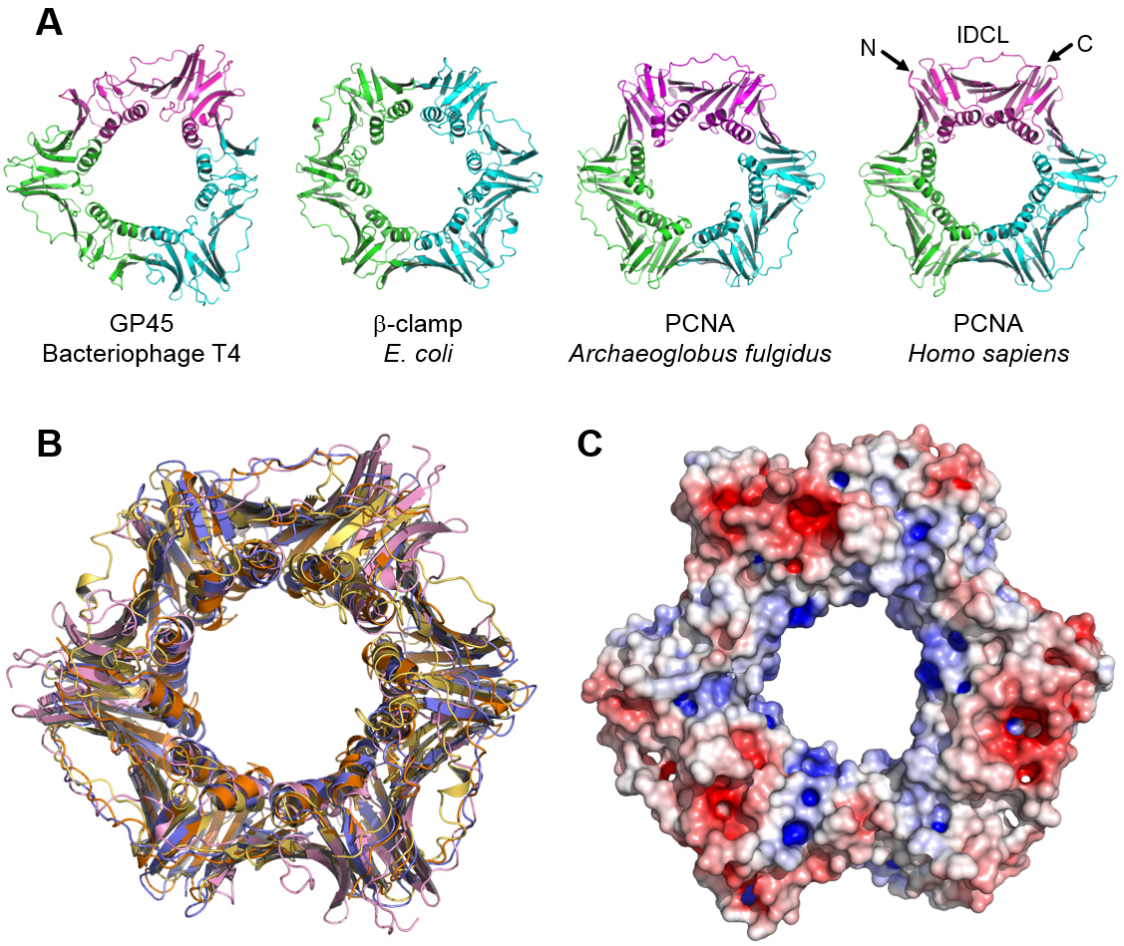


Figure 1.7: Crystal structures of DNA sliding clamps from different organisms. **(A)** Cartoon representation of DNA sliding clamps from bacteriophage T4 (GP45, PDB code: 1CZD) (Moarefi et al., 2000), *E. coli* (β-clamp, PDB code: 1MMI) (Oakley et al., 2003), *Archaeoglobus fulgidus* (PCNA, PDB code: 1RWZ) (Chapados et al., 2004), and *Homo sapiens* (PCNA, PDB code: 1VYM) (Kontopidis et al., 2005). GP45 and PCNA consist of a trimer while the prokaryotic β-clamp is a dimer. Monomers are depicted in cyan, green, and magenta. N- and C-termini as well as the interdomain connecting loop (IDCL) are indicated for one subunit of the hsPCNA homotrimer. **(B)** Superposition of the structures shown in (A) indicating the structural conservation of the DNA sliding clamps (GP45: pink; β-clamp: yellow; afPCNA: orange; hsPCNA: blue). **(C)** Electrostatic surface potential of hsPCNA (PDB code: 1VYM) (Kontopidis et al., 2005).

It was shown that DNA sliding clamps encircle the DNA and for this purpose the positively charged α-helices in the center of the protein play a crucial role

(Figure 1.7 C) (McNally et al., 2010). PCNA is loaded around the Pol α primed DNA by the chaperone-like AAA+ type ATPase replication factor C (RFC) clamp loader with the C-side (C-termini protruding from this side) facing towards the 3' end of the elongating DNA (Yao and O'Donnell, 2012). After initiation of the replication fork at the origins of replication Pols δ and ϵ bind to PCNA for leading and lagging strand synthesis. For the connection of the Okazaki fragments PCNA orchestrates the flap endonuclease 1 (FEN1) to remove the RNA primers and DNA ligase I to seal the resulting gap (Sakurai et al., 2005; Tom et al., 2001). When the replication fork is stalled by a damage that prevents progression PCNA coordinates the TLS and PRR pathways. Upon Rad6-Rad18 dependent PCNA monoubiquitination at K164 the replicative polymerases are exchanged by TLS polymerases that bypass the lesion in an error-prone pathway. Alternatively, further polyubiquitination at PCNA-ubiquitin K63 by Rad5-Ubc13-Mms2 triggers error-free PRR (Hoege et al., 2002).

Furthermore, PCNA participates in various additional genome maintenance pathways. It also plays an important role in the prevention of re-replication, DNA mismatch repair (MMR), BER, NER, chromatin assembly and remodeling, sister-chromatid cohesion and prevention of recombination, epigenetic DNA modification, and cell-cycle control. A broad range of DNA processing interaction partners such as polymerases, nucleases, ligases, helicases, and many others have been identified accentuating the importance of PCNA in DNA metabolism. Coordination and switching of binding partners are regulated by affinity-driven competition, proteolysis, phosphorylation, ubiquitination, and sumoylation (Moldovan et al., 2007).

The PCNA interacting peptide (PIP) is the most common binding motif for PCNA interaction but for some cases other interaction interfaces have also been described such as the BRCT domain of the TLS polymerase Rev1 and the APIM motif (Guo et al., 2006; Gilljam et al., 2009). The PIP-box amino acid consensus sequence is defined as Qxxhxxaa (Q = glutamine, x = not defined, h = hydrophobic, a = aromatic) (Warbrick, 1998). PIP-boxes are generally located within flexible and accessible areas of the protein. The glutamine sidechain sticks into the Q-pocket of PCNA and the hxxaa portion forms a 3_{10}-helix within the adjacent hydrophobic pocket which is located between the C-terminus and interdomain connecting loop (IDCL) (Figure 1.8) (Pascal et al., 2006). The PCNA binding interface can also be extended N- and C-terminally flanking the core PIP-box. C-terminally to the classical PIP-box FEN1 and p21 form additional intermoleculecular anti-parallel β-sheets with the IDCL (Sakurai et al., 2005; Gulbis et al., 1996). Interestingly, various PIP-boxes exhibit high variability within the consensus sequence. For example, in TLS polymerases Pols η, κ, and ι the Q is exchanged by M, R, or K (Hishiki et al., 2009). The human BER glycosylase MPG even comprises an inverted PIP-box (Xia et al., 2005). Consequently, these alterations also have an influence on the affinities towards PCNA. Notably, upon PCNA binding the enzymatic activity of some interaction partners such as FEN1 is stimulated (Frank et al., 2001).

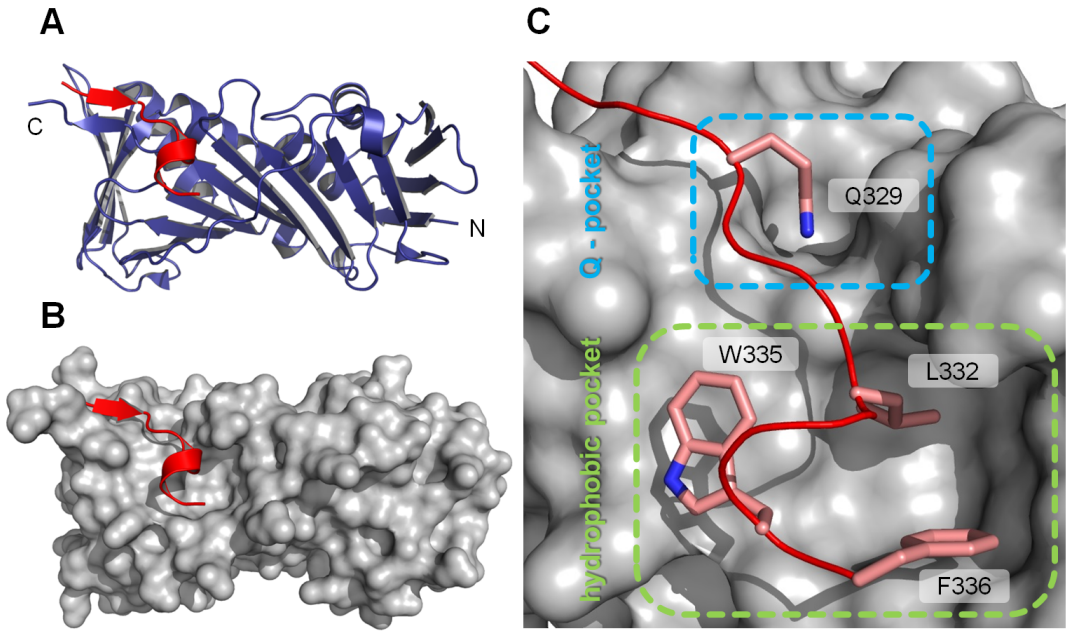


Figure 1.8: PIP_{afFEN1} bound to afPCNA.
Crystal structure of a PIP-box containing peptide derived from FEN1 (red) in complex with PCNA (blue & grey) from the archaeon *Archaeoglobus fulgidus* (aFEN1 and afPCNA, respectively). afPCNA is depicted as a monomer. A special feature of this PIP-box is the additional intermolecular β-sheet with the very C-terminus of PCNA. **(A)** Cartoon representation; N- and C-termini are indicated. **(B)** Surface of afPCNA. **(C)** Detailed view of the afPCNA binding pocket. The Q-pocket is highlighted in blue and the hydrophobic pocket that binds the 3_{10}-helix of the PIP-box is depicted in green. The amino acid sequence of this canonical afFEN1 PIP-box is ^{329}**Q**AT**L**ER**WF**^{336} (PDB code: 1RXZ) (Chapados et al., 2004).

Affinity studies of PIP-box peptides derived from p21 and FEN1 revealed very different K_D values of 80 nM and 60 µM, respectively. This variation in binding affinities is crucial for the coordinated recruitment of interaction partners (Bruning and Shamoo, 2004). Importantly, additional binding domains can contribute to the overall affinity towards PCNA. Y-family TLS polymerases contain two evolutionary conserved ubiquitin-binding domains (UBM and UBZ) that boost their affinity when PCNA is ubiquitinated to replace the replicative polymerase (Bienko et al., 2005).

1.5 Objectives

Although previous studies have already achieved great progress towards the understanding of FA, many pieces in the puzzle are still missing (Kottemann and Smogorzewska, 2013). The DEAH helicase FANCM is an important component of the FA pathway and plays a crucial role in the activation and initiation of this repair mechanism. So far, the only available structural information with respect to FANCM is derived from partial crystal structures of the archaeal homolog Hef from

Pyrococcus furiosus (pfHef) (Nishino et al., 2005a,b). Other crystal structures showing the FANCM interaction interfaces for FAAP24 or MHF1/2 binding only exhibit short patches of the entire protein and are not sufficient to derive a proper structure-function correlation (Yang et al., 2013; Coulthard et al., 2013; Tao et al., 2012; Hoadley et al., 2012). To clearly understand the function of FANCM and it's role in FA initiation more structural data are required. Due to the inability to recombinantly express and purify human FANCM at amounts that are needed for crystallization experiments homologous proteins from archaeal organisms present an elegant alternative.

Furthermore, it is known that the FA pathway removes ICL damages in a replication-dependent process (Knipscheer et al., 2009). Consequently, to gain insights into FA initiation by FANCM it is important to investigate the connection between FA and the DNA replication machinery. The DNA sliding clamp PCNA functions as a processivity factor for replicative DNA polymerases and serves as a central hub for various DNA metabolic processes (Moldovan et al., 2007). Since PCNA is involved in diverse DNA repair pathways such as NER or TLS that are also associated to FA, it could additionally play a key role in FA.

Thus the present thesis follows two major aims:

1. Structural characterization of the archaeal FANCM homolog Hef.
2. Examination of PCNA as a new Hef/FANCM/Mph1 interaction partner.

2 Materials

2.1 Chemicals

Chemicals and solutions were purchased from Carl Roth (Karlsruhe), Sigma-Aldrich (Seelze), Fluka (Neu-Ulm), Hampton Research (Laguna Hills, USA), or Applichem (Darmstadt) with analytical grade or better. The chemicals for protein crystallization were of the highest available purity.

2.2 DNA

DNA oligonucleotides were purchased from Biomers (Ulm) or Sigma-Aldrich (Hamburg).

Table 2.1: Primers for *Thermoplasma acidophilum* genes.

Name	*Sequence (5'-3')*
HEF_TA_InFu_5'	TTTCAGGGCGCCATGGTTATTCAGCCCAGGGAATATCAATTAAACG
HEF_TA_InFu_3'	GTGCGGCCGCAAGCTTGCAGAGATAGGCAGCAGGACAGATTAAAAGTCG
taHef_I505R/F508S_for	CAAGATTCGACTCTTAATCTGTCC
taHef_I505R/F508S_rev	GATTAAGAGTCGAATCTTGTTTTC
taHef_delC7_for	GAAGAGCGATGTTTAGAAAACAATATTCGAC
taHef_delC7_rev	GTCGAATATTGTTTTCTAAACATCGCTCTTC
taHef_delC33_for_L	AGCATAAGGAAATAATCTGTCCTGCTGCCTATCTCTGCAAGCTTG
taHef_delC33_rev_L	AGCAGGACAGATTATTTCCTTATGCTGCTGTAGTAGTATGCCATATC
taHef_K31A_for	ACGGGCCTGGGCGCAACCATAATTGCA
taHef_K31A_rev	TGCAATTATGGTTGCGCCCAGGCCCGT
taHef_D139F_for	CTATGCCTATGTTTTCATAGCCCAGGAAT
taHef_D139F_rev	ATTCCTGGGCTATGAAAACATAGGCATAG
taHef_D139K_for	CTATGCCTATGTTAAAATAGCCCAGGAAT
taHef_D139K_rev	ATTCCTGGGCTATTTTAACATAGGCATAG
PCNA_TA_InFu_NcoI_5'	TTTCAGGGCGCCATGGAGGCTGGATCTACAGATCCTCCG
PCNA_TA_InFu_HindIII_3'	GTGCGGCCGCAAGCTTTACTGCTCCATTCTCGGAGCCAG
PCNA_TA_5'korr._for	CTTTATTTTCAGGGCGCCATGATCAGACTGAACCTTTCGGTTAAGAATC
FEN_TA_For_NcoI_5'	AGGGATGATCCATGGGAACAGATATATCGGA
FEN_TA_Rev_HindIII	TATTAAGCTTATGTTAATCAAAACTATCGAG
XPD_TA_for_NcoI_5'	CGTCCCATGGACGAGAACAGGCAGTACCAG
XPD_TA_rev_HindIII	CCTCAAGCTTAGGCCCCATATTTTTCGCGTG
XPB_TA_For_PciI_5'	ATGGTCTCACATGTCCGATATCGTCTATTCTGGA
XPB_TA_Rev_HindIII	ACCGTCCTGAAGCTTACGTACCACCATCAGT
Bax1_TA_For_NcoI_5'	AGGAGAAACCATGGTCCCGGCAGAACTGAT
Bax1_TA_Rev_XhoI	ATGGACTCGAGTCAGGTGGAACGGACTATCT

Names and according nucleobase sequences (5'-3') of DNA oligonucleotides used as primers for cloning. (A = adenine, C = cytosine, G = guanine, T = thymine)

Table 2.2: Primers for *Homo sapiens* genes.

Name	*Sequence (5'-3')*
hFANCM_TEV_5'for	CTTTATTTTCAGGGCGCCATGAGCGGACGGCAAAGAACGCTTTTTCAG
hsFANCM_3'rev	GTGCGGCCGCAAGCTTGTCTTATATATCAGATTTCAGTCTATCTTGGTTAAG
hsFANCM_iso2_3'rev	GTGCGGCCGCAAGCTTGTCCTACTTTCCATCCCTATAGGAAAAGATAGATG
hFANCM_pIRES_5'for	GGATCGATATCTGCGGCCACCATGAGCGGACGGCAAAGAACGCTTTTTCAGAC GTGG
hFM-HA_pIRES_3'rev	GCACACTGGATCAGTTATTTAAGCGTAGTCTGGGACGTCGTATGGGTATATATC
hFANCM_killPIP-1_for	GCGTTTTCAGACGTCGGGCTCAAG
hFANCM_killPIP-1_rev	CCGACGTCTGAAAACGCGTTCTTTG
killPIP1_v2_5'for	GGATCGATATCTGCGGCCACCATGAGCGGACGGCAAAGAACGCGTTTTCAGAC GTCG
hFM_killPIP1_for_new	CGGCAAAGAACGCGTTTTCAGACGTCGGGCTCAAGTATCTCCCG
hFM_killPIP1_rev_new	CTTGAGCCCGACGTCTGAAAACGCGTTCTTTGCCGTCCGCTCATTAG
hFANCM_killPIP-2_for	GGCATAAGAGAGGGAGAGTCTGC
hFANCM_killPIP-2_rev	GCAGACTCTCCCTCTCTTATGCC
hFANCM_killPIP-3_for	CTGGAGAGAAGGAAACAGTCTCGTGAC
hFANCM_killPIP-3_rev	GAGACTGTTTCCTTCTCTCCAGTTGC
hFANCM_killPIP-4_for	AGTAGCCGAAAGAAAGATTCGTTC
hFANCM_killPIP-4_rev	GAACGAATCTTTCTTTCGGCTACT
hFM_killPIP4_for_new	GAGGCAAAGTAGCCGAAAGAAAGATTCGTTCTTATCAGAAGAAG
hFM_killPIP4_rev_new	GATAAGAACGAATCTTTCTTTCGGCTACTTTGCCTCATTCCATCCC
hFANCM_killPIP-5_for	CTCGCTCTGAATCGAGACTGTGGC
hFANCM_killPIP-5_rev	GCCACAGTCTCGATTCAGAGCGAG
hFM_killPIP5_for_new	CATCAACTCTCTCGCTCTGAATCGAGACTGTGGCAAGATCATCC
hFM_killPIP5_rev_new	GCCACAGTCTCGATTCAGAGCGAGAGAGTTGATGAATTCCAGTGG
hFANCM_killPIP-6_for	GGAAGATCGTACCTCAACATCTATTGC
hFANCM_killPIP-6_rev	GCAATAGATGTTGAGGTACGATCTTCC
hFM_killPIP6_for_new	CAAATGGAAGATCGTACCTCAACATCTATTGCTCCCAGGAATGAATC
hFM_killPIP6_rev_new	GGGAGCAATAGATGTTGAGGTACGATCTTCCATTTGTAAATAAG
hFANCM_killPIP-7_for	GATGAAAGTAGGCTGTTATCTGAAGATG
hFANCM_killPIP-7_rev	CATCTTCAGATAACAGCCTACTTTCATC
hFM_killPIP7_for_new	CAAGATGAAAGTAGGCTGTTATCTGAAGATGTTAATACAGAGTTCG
hFM_killPIP7_rev_new	GTATTAACATCTTCAGATAACAGCCTACTTTCATCTTGATCAGTGG
hFANCM_killPIP-8_for	GATTCCTCAAGACTTGACTCTTTAAATG
hFANCM_killPIP-8_rev	CATTTAAAGAGTCAAGTCTTGAGGAATC
hFM_killPIP8_for_new	GAACAAGATTCCTCAAGACTTGACTCTTTAAATGATGAAACTCAAC
hFM_killPIP8_rev_new	CATCATTTAAAGAGTCAAGTCTTGAGGAATCTTGTTCATTTTCTGAC
hsPCNA_TEV_5'for	CTTTATTTTCAGGGCGCCATGTTCGAGGCGCGCCTGGTCCAGGGCTCC
hsPCNA-nativ_5'for	CTAACAGGAGGAATTAACCATGTTCGAGGCGCGCCTGGTCCAGGGCTCC
hsPCNA_pBAD_3'rev	GTGCGGCCGCAAGCTTGTCCTAAGATCCTTCTTCATCCTCGATCTTGGG
hsPCNA->pIRES_3'rev	GCACACTGGATCAGTTATCTAAGATCCTTCTTCATCCTCGATCTTGGGAGCC
hsFEN1_TEV_5'for	CTTTATTTTCAGGGCGCCATGGGAATTCAAGGCCTGGCCAAACTAATTG
hsFEN1_HA_3'rev	TAATCTGGAACATCGTATGGGTATTTTCCCCTTTTAAACTTCCCTGCTGCCCC
hFEN-HA_pIRES_3'rev	GCACACTGGATCAGTTATCTAAGCGTAATCTGGAACATCGTATGGGTATTTTCC

Table 2.3: Primers for *Chaetomium thermophilum* genes.

Name	*Sequence (5'-3')*
ctFANCM_TEV_5'for	CTTTATTTTCAGGGCGCCATGGGCAAGGACTTCGATGAGTATGATGAC
ctFANCM_3'rev	GTGCGGCCGCAAGCTTGTCTCACTCGTCATCCTCACTATCCGTCAATATC
ctFANCM_korr_for	CTCAGGCTGTTGATGCCGCAGCCGAGGCCGCACTAACTGGGAATGAAACGAGGG GTAACTCC
ctFANCM_korr_rev	GAGGAGCTCCTCGTCTGAAAAGTCATCATACTCATCGAAGTCCTTGCCCATGGC GCCCTG
ctFM_korr_ins_for	GAGGAGCTCCTCGTGGCCCTTACTCAGGCTGTTGAT
ctFM_korr_ins_rev	ATCAACAGCCTGAGTAAGGGCCACGAGGAGCTCCTC
ctFM_N763_for	CGAAAACACCCAAGACAAGTCGTAACCCGAGCCGAAGAAGACCC
ctFM_N763_rev	GGGTCTTCTTCGGCTCGGGTTACGACTTGTCTTGGGTGTTTTCG
ctFM_N846_for	GCTCAACCGACGCTACAAGACGTAACCACGTGGCCAGAATGGCAAG
ctFM_N846_rev	CTTGCCATTCTGGCCACGTGGTTACGTCTTGTAGCGTCGGTTGAGC
ctFM_N902_for	CAAAGAACGCCGCGAGAAGAGCTAGGCCTATGAAGACCCTTGCCC
ctFM_N902_rev	GGGCAAGGGTCTTCATAGGCCTAGCTCTTCTCGCGGCGTTCTTTG
ctPCNA_TEV_5'for	CTTTATTTTCAGGGCGCCATGTTAGAAGCACGGTTGGAGCAGGCCAGC
ctPCNA_3'rev	GTGCGGCCGCAAGCTTGTCTTACTCATCATCACCAATCTTCGGAGCAAGGTAG

Primers with the name "ctFANCM" or "ctFM" were used for the gene encoding ctMph1 (FANCM homolog; Chapter 3.1).

Table 2.4: Vector primers.

Name	*Sequence (5'-3')*
TEV_5'rev_Vec-linear	GGCGCCCTGAAAATAAAGATTCTC
pET/pBAD_3'for_Vec-linear	GACAAGCTTGCGGCCGCACTCGAG
pBAD-nativ_5'rev_Vec-linear	GGTTAATTCCTCCTGTTAGCCC
pET-HA_for	CATACGATGTTCCAGATTACGCTATCCCCACTACTGAGAATCTTTATTTTCGG
pET-HA_rev	TAATCTGGAACATCGTATGGGTAGCTCATGGGGTGATGGTGATGGTGATGT TTC
pET-Flag_for	ACAAAGACGATGACGACAAGATCCCCACTACTGAGAATCTTTATTTTCAGG
pET-Flag_rev	TCGTCATCGTCTTTGTAGTCGCTCATGGGGTGATGGTGATGGTGATGTTTC
pET/BAD->pIRES_5'for	GGATCGATATCTGCGGCCACCATGAAACATCACCATCACCATCACCCCATG AGC
pIRES_Vec-lin_5'rev	GGCCGCAGATATCGATCCGAGCTCGGTAC
pIRES_long_5'rev	GCGCTAGCTAGGCCGCAGATATCGATCCGAGCTCGGTACCAAGC
pIRES_Vec-lin_3'for	TAGATAACTGATCCAGTGTGCTGGAAT
pIRES_long_3'for	GACTACGCTTAAATAACTGATCCAGTGTGCTGGAATTAATTCGC
pIRES_MCS_5'for	GGATCGATATCTGCGGCCTAGCTAGCGCTTAAGGC

Table 2.5: Sequencing and colony PCR primers.

Name	*Sequence (5'-3')*
pET upstream	GATGCGTCCGGCGTAGAG
T7 terbis	AACCCCTCAAGACCCG
T7prom	TAATACGACTCACTATAGGG
T7term	GCTAGTTATTGCTCAGCGG
pBAD for	TTATCGCAACTCTCTACTGT
pBAD rev	CTTCTGCGTTCTGATTTA
pBAD forward	CCATAGCATTTTTATCCATAAG
pBAD reverse	GATTTAATCTGTATCAGGCTG
pIRESf	CCAAGCTTGGTACCGAGCTCG
pIRESr	ACCCCAACAGCTGGCCCTCG
HEFmid	GACGAGTTTGCCACTCTCCG
FANMm1	CTACACAAGCTTCCACCAGG
FANMm2	GGAATTATGGATGGAACTAAAGGG
FANMm3	CAGAGTCAGTCCAACAAAAGAAG
FANMm4	CATCCTCAGTGATAGAATCTG
FANMm5	GTTGCCTTTATGATATACCTAATG
FANMm6	CTAGTGGAACTGTTATTATCCCATC
FANMm7	GCATGTAGCTAGGAAGTTTTTAG
FANMm8	GCAGAGCAAACAGACATCGCTG
FANMm9	GGTGAATAGTAATAAAAGTGAGGC
FANMio	GCCTTGGAGCTGCCAGGGCTC
ctFMmid1	CCAGGTCGAATCACAGCATCAC
ctFMmid2	CACGCAACATTGAGACGGTTACC
ctFMmid3	GCCACATCCATTGGCGAAGAGGG
ctFMmid4	CCTATGAAGACCCTTGCCCTGAC
XPD mid	GTCCGTGGCAGAGAAGTTTCTCAG

Table 2.6: DNA substrates.

Name	*Sequence (5'-3')*
Hef_top	AGCTACCATGCCTGCACGAATTAAGCAATTCGTAATCATGGTCATAGCT§
Hef_top_notag	AGCTACCATGCCTGCACGAATTAAGCAATTCGTAATCATGGTCATAGCT
Hef_top_3'Dabcyl	AGCTACCATGCCTGCACGAATTAAGCAATTCGTAATCATGGTCATAGCT#
Hef_top_15/20	ATGCCTGCACGAATTAAGCAATTCGTAATCATGGT
Hef_top_5'-15_3'Biotin	ATTCCAAGCAATTCGTAATCATGGTCATAGCT§
top_gegen18	CAATTCGTAATCATGGTC
Hef_bottom	AGCTATGACCATGATTACGAATTGCTTGGAATCCTGACGAACTGTAG
Hef_bottom_5'Cy3	*AGCTATGACCATGATTACGAATTGCTTGGAATCCTGACGAACTGTAG
Hef_bottom_15/20	ACCATGATTACGAATTGCTTGGAATCCTGACGAAC
Hef_bot_3'-15	AGCTATGACCATGATTACGAATTGCTTAATTCGT
Hef_bot_5'-15	TACGAATTGCTTAATTCGTGCAGGCATGGTAGCT
Hef_bot_comp	AGCTATGACCATGATTACGAATTGCTTAATTCGTGCAGGCATGGTAGCT
Hef_bot_12bub	AGCTATGACCATGATTACGCCCCATCCGGAACGTGCAGGCATGGTAGCT
bot_gegen18	GACCATGATTACGAATTG
Hef_short	AATTCGTGCAGGCATGGTAGCT
Hef_short_15	AATTCGTGCAGGCAT
Hef_fm	CTACAGTTCGTCAGGATTCCAAGCAATT
Hef_sm	AATTGCTTAATTCGTGCAGGCATGGTAGCT
Hef_fg4	CTACAGTTCGTCAGGATTCC
Hef_fg4_15	GTTCGTCAGGATTCC
BGAA-F	ATTGCGGATTTTTTGGAGCCGAAGGCTCCATCF%CTACCGCAAT
Jiang14	CGACGCTAGCGTCG

(§ = Biotin, # = Dabcyl, * = Cy3, % = Fluorescein)

Table 2.7: Composition of DNA substrates.

Annealed DNA substrate	*Single strand composition*			
Fork I (F-I)	top	bottom	fm	short
Fork II (F-II)	top	bottom	fg4	short
Fork III (F-III)	top	bottom	fg4	sm
Half Fork I (HF-I)	top	bottom	fm	-
Half Fork II (HF-II)	top	bottom	g4	-
Half Fork III (HF-III)	top	bottom	-	short
Half Fork IV (HF-IV)	top	bottom	-	sm
Open Fork (OF)	top	bottom	-	-
5' Overhang (5'OH)	top	bot_3'-15	-	-
3' Overhang (3'OH)	top	bot_5'-15	-	-
ssDNA (ss49)	top	-	-	-
dsDNA (ds49)	top	bot_comp	-	-
12mer Bubble (12Bub)	top	bot_12bub	-	-
Fork 15/20 (F15/20)	top_15/20	bottom_15/20	fg4_15	short_15
Open Fork 15/20 (OF15/20)	top_15/20	bottom_15/20	-	-
Open Fork Cy3* (OF_Cy3)	top	bottom_5'Cy3	-	-

(% = Fluorescein, * = Cy3)

Table 2.8: Plasmids.

Name	Properties	*Source*
pBADM-11	AraBAD promoter, pUC origin, Amp resistance, N-His_6-tag (TEV cleavable), 4,100 bp	EMBL
pETM-11	T7-lac promoter, pBR332 origin, Kan resistance, N-His_6-tag (TEV cleavable), C-His_6 (optional), 6,029 bp	EMBL
pCDF-11	T7-lac promoter, CloDF13 origin, Strep resistance, N-His_6-tag (TEV cleavable), C-His_6 (optional), 4,276 bp	EMBL
pIRESneo3	CMV promoter, ColE1 origin, Amp and G418 (IRES) resistances, SV40 poly A^+, 5,260 bp	Clontech

Vectors used for recombinant protein expression in *E. coli* and human cell culture. Sequences have been modified by the implementation of HA-/Flag-tags or deletion of fusion components to create native expression constructs.

2.3 Cell Culture

Table 2.9: Expression strains.

Name	*Characteristics*	*Source*
DH5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 endA1 thi-1 hsdR17	NEB
BL21-CodonPlus(DE3)-RIL	*E.coli* B F^- ompT hsdS($r_B{}^-$ $m_B{}^-$) dcm^+ Tet^r gal λ(DE3) endA Hte [argU ileY leuW Cam^r]	Agilent
HEK293	human embryonic kidney, epithelial, adherent	Biocenter
HeLa	human cervix, epithelial, adherent	RVZ

Cells used for recombinant protein production. *E.coli* strains (top) were acquired commercially while human derived cells (bottom) were kindly provided by Detlev Schindler (Human Genetics, Biocenter, Würzburg) and Mike Friedrich (Bio Imaging Center, RVZ, Würzburg).

Table 2.10: Media.

Medium	*Composition*
LB	10 g/l Tryptone, 10 g/l Yeast Extract, 5 g/l NaCl
LB-Agar	10 g/l Tryptone, 10 g/l Yeast Extract, 5 g/l NaCl, 15 g/l Agar
Cell Culture	DMEM (high glucose, pyruvate), 10 % FCS, 1 % Pen/Strep

Sterile media are supplemented with required antibiotics or stimulating agents (Table 2.11).

Table 2.11: Medium additives.

Name	*Abbreviation*	*Final concentration*	*Supplier*
Arabinose	Ara	0.02 %	Roth
Isopropyl-β-D-thiogalactopyranoside	IPTG	0.5 mM	Roth
Ampicillin	Amp	100 µg/ml	Roth
Kanamycin	Kan	34 µg/ml	Roth
Chloramphenicol	Cam	50 µg/ml	Roth
Streptomycin	Strep	50 µg/ml	Roth
Geneticin	G418	600 µg/ml	Roth
Mitomycin C	MMC	40-80 ng/ml	Roth

2.4 Proteins and Peptides

Table 2.12: Properties of recombinantly expressed and purified proteins.

Protein	*MW*	ϵ	*PI*	*AA*
taHef_WT	60959 (57961)	38280 (35300)	6.94 (7.10)	534 (510)
taHef_ΔC7	60079 (57081)	38280 (35300)	6.94 (7.10)	527 (503)
taHef_ΔC33	57100 (54101)	38280 (35300)	8.47 (9.04)	501 (477)
taHef_I505R/F508S	60942 (57944)	38280 (35300)	7.15 (7.78)	534 (510)
taHef_K31A	60902 (57904)	38280 (35300)	6.76 (6.72)	534 (510)
taHef_D139K	60972 (57974)	38280 (35300)	7.45 (8.46)	534 (510)
taHef_D139F	60991 (57993)	38280 (35300)	7.15 (7.78)	534 (510)
taPCNA	30708 (27710)	13410 (10430)	5.23 (4.76)	272 (248)
taPCNA_addN25	33539 (30541)	19035 (16055)	5.02 (4.67)	297 (273)
taFEN1	41898 (38900)	29925 (26945)	6.20 (5.72)	362 (338)
taXPD	74509 (71511)	68120 (65140)	6.47 (6.24)	646 (622)
taXPB	54756 (51757)	61770 (58790)	7.92 (8.67)	479 (455)
taBax1	49771 (46773)	52955 (49975)	6.21 (5.76)	433 (409)
ctMph1	137408 (134410)	89770 (86790)	7.89 (8.16)	1226 (1202)
ctPCNA	31715 (28717)	13785 (10805)	4.49 (4.26)	285 (261)
hsPCNA	31895 (28897)	19285 (16305)	4.55 (4.31)	287 (263)
hsFEN1_C-HA	46804 (43805)	30745 (27765)	8.14 (8.53)	415 (391)

Molecular weight (MW; g/mol), extinction coefficient (ϵ; l/mol/cm), isoelectric point (PI; pH), and length counted in number of amino acids (AA) of recombinantly expressed and purified proteins. All designations refer to the proteins as they are expressed in *E. coli*, i.e. including His_6-tag, TEV site, and HA-tag. Values for the TEV cleaved proteins are given in parentheses.

Table 2.13: Peptides for ITC experiments.

Name	*Sequence*	*MW*	ϵ
$hsFANCM^{1-17}$	MSGR**Q**RT**L**FQT**W**GSSIS	1942	5500
$hsFANCM^{1-17}$ L5R/W12S	MSGR**Q**RT<u>R</u>FQT<u>S</u>GSSIS	1886	0

Peptides for ITC experiments comprise the very N-terminal 17 amino acids of human FANCM (hsFANCM). Residues marked in bold constitute the putative PIP-box #1 and underlined residues highlight the corresponding "killPIP" mutations (Chapter 4.13.1). Molecular weight (MW; g/mol), extinction coefficient (ϵ; l/mol/cm).

Table 2.14: Antibodies.

Antigen	*Properties*	*Supplier*
α-PCNA	mouse, monoclonal (PC10), rat PCNA, ab29	Abcam
α-PCNA	rabbit, monoclonal (EPR3821), human PCNA, NBP1-40733	Novus
α-FANCM	mouse, polyclonal, human FANCM isoform 2, H00057697-B01P	Novus
α-HA	mouse, monoclonal (16B12), CYPYDVPDYASL, MMS-101P	Covance
α-Mouse	goat, polyclonal, conjugated HRP, IgG (H+L), 115-035-003	Dianova
α-Mouse	goat, polyclonal, conjugated Alexa647, IgG (H+L) (highly cross-absorbed), A-21236	Invitrogen
α-Rabbit	goat, polyclonal, conjugated Alexa488, IgG (H+L) (highly cross-absorbed), A-11034	Invitrogen

Properties describe host, clonality (clone ID), conjugation, immunogen, and specific order number for the antibodies.

2.5 Buffers

All buffers, solutions, and screens were prepared using ultra-pure ddH_2O (Millipore) and respective pH values were adjusted by adding either 32 % NaOH or 37 % HCl.

Table 2.15: Initial crystallization screens.

Name	*Supplier*
Index Screen HT	Hampton Research
Crystal Screen I & II	Hampton Research
Additive Screen	Hampton Research
Wizard Screen I & II	Emerald Biosciences
Protein Complex Suite	Qiagen
Nucleix Suite	Qiagen
Nextal PEG Suite	Qiagen
pH Clear Suite	Qiagen
Optimix 1-5	Topaz

Commercially available crystallization screens for initial trials with the sitting drop method were prepared using the Lissy liquid handling system (Zinsser Analytic). Compositions of follow-up fine screens are described in the corresponding sections within Chapter 4.

Table 2.16: Buffers and solutions.

Name	*Components*	*pH*
Standard (Std)	20 mM Tris/HCl, 200 mM NaCl, 5 mM $MgCl_2$	7.5
His-A	50 mM Tris/HCl, 300 mM NaCl, 20 mM Imidazole	8.0
His-B	50 mM Tris/HCl, 300 mM NaCl, 500 mM Imidazole	8.0
Lysis	His-A, 1 µl/50 ml DNaseI, 1 Tablet Protease Inhibitor EDTA-free (order number: 11873580001, Roche)	8.0
Sdx	20 mM Tris/HCl, 200 mM NaCl	7.5
Sdx-HS	20 mM Tris/HCl, 500 mM NaCl	7.5
Sdx-DTT	20 mM Tris/HCl, 200 mM NaCl, 1 mM DTT	7.5
TAE	40 mM Tris/HCl, 0.1 % Acetic acid, 1 mM EDTA	8.5
TBE	89 mM Tris, 89 mM Boric acid, 2 mM EDTA	8.3
TE	10 mM Tris/HCl, 1 mM EDTA	7.5
DNA-Annealing	0.1x TE, 100 mM KCl	7.5
DNA-Binding	20 mM Tris/HCl, 100 mM NaCl, 5 mM $MgCl_2$	7.5
SDS Running	25 mM Tris, 192 mM Glycine, 0.1 % SDS	8.6
Staining	50 % Ethanol, 10 % Acetic acid, 0.1 % Coomassie blue	ND
Destaining	10 % Ethanol, 5 % Acetic acid	ND
Loading-SDS	50 mM Tris/HCl, 100 mM DTT, 2 % SDS, 10 % Glycerol, 0.1 % Bromphenol blue	6.8
Loading-Native	10 mM Tris/HCl, 5 mM Na-Acetate, 2 mM EDTA, 10 % Glycerol, 0.06 % Orange G	7.5
Blotting	25 mM Tris, 192 mM Glycine, 20 % Methanol, 0.02 % SDS	8.3
DPBS	1.47 mM KH_2PO_4, 8.06 mM Na_2HPO_4-$7H_2O$, 138 mM NaCl, 2.67 mM KCl (order number: 14190, Gibco)	7.2
DPBS-T	DPBS, 0.05 % Tween-20	7.2
CD	20 mM KH_2PO_4/K_2HPO_4	8.0
Fixer	50 % Methanol, 10 % Acetic acid	ND
Sensitizer	0.02 % $Na_2S_2O_3$, 0.2 % $AgNO_3$, 0.075 % Formaldehyde	ND
Developer	6 % Na_2CO_3, 0.05 % Formaldehyde	ND
ATPase	50 mM Tris/HCl, 50 mM KCl, 4 mM $MgCl_2$, 1 mM DTT, 10 u/ml PK, 10 u/ml LDH, 2 mM PEP, 300 µM NADH	7.5
Helicase	20 mM Tris/HCl, 50 mM NaCl, 10 mM KCl, 10 mM $MgCl_2$, 1 mM EDTA	7.5
Helicase-Stop	20 mM Tris/HCl, 10 mM KCl, 10 mM EDTA, 0.5 % SDS, 3 mg/ml Proteinase K	8.0
CoIP-LP	20 mM Tris/HCl, 150 mM NaCl, 1 mM $MgCl_2$, 1 % Triton-X	7.5
CoIP-LS/HS	10mM Hepes, 150/500mM NaCl, 1mM $MgCl_2$, 5% Glycerol	7.5
CoIP-RIPA	50 mM Tris/HCl, 150 mM NaCl, 1 % NP40, 0.1 % SDS, 0.5 % Sodium Deoxycholate	8.0
AB-Staining	DPBS, 4 % PFA, 1 % Triton-X, 0.5 % NP40	7.2

ND = pH value was not adjusted.

Table 2.17: Electrophoresis gels.

Name	*Components*	*pH*
1% Agarose-Gel	TAE, 1 % Agarose	8.5
15% SDS-Gel (resolving)	15 % AA:BisAA (37.5:1), 375 mM Tris/HCl, 0.1 % SDS, 0.25 % APS, 0.05 % TEMED	8.8
5% SDS-Gel (stacking)	5 % AA:BisAA (37.5:1), 125 mM TrisHCl, 0.1 % SDS, 0.25 % APS, 0.15 % TEMED	6.8
10% Native-Gel	10 % AA:BisAA (37.5:1), 0.5 x TBE, 10 mM $MgCl_2$, 0.2 % APS, 0.1 % TEMED	8.3

Gels for native and denaturing electrophoretic separation of proteins and DNA.

Table 2.18: Thermofluor buffer screen.

	1	2	3	4	5	6
A	Citrate 4.5	Bis-Tris 7.0	Imidazole 6.5	Hepes 8.0	Tris 8.5	*
B	Acetate 4.6	Ada 6.5	Imidazole 8.0	Hepes 8.5	Tris 9.0	*
C	Mes 5.5	Ada 7.0	Na/K PO_4 6.8	Bicine 8.0	Taps 8.0	*
D	Mes 6.5	Bis-Tris Propane 6.0	Na/K PO_4 7.55	Bicine 8.5	Taps 9.0	*
E	Cacodylate 6.0	Bis-Tris Propane 7.0	Mops 7.0	Bicine 9.0	Glycylglycine 8.5	*
F	Cacodylate 6.5	Pipes 6.5	Mops 7.5	Tris 7.0	Ches 9.0	#
G	Bis-Tris 5.5	Pipes 7.0	Hepes 7.0	Tris 7.5	Ches 9.5	#
H	Bis-Tris 6.5	Pipes 7.5	Hepes 7.5	Tris 8.0	Caps 9.8	#

All buffers are present at a concentration of 100 mM. Numbers indicate the adjusted pH value. Column 6 represents the reference measurements (* = reference buffer; # = reference buffer w/o protein).

2.6 Enzymes and Kits

Table 2.19: Enzymes.

Category	*Enzymes (Supplier)*
Polymerase	Herculase (Agilent), Phusion (Finnzymes/NEB), Pfu (Fermentas), T4 DNA Pol, Taq (NEB)
Nuclease	DNaseI (Invitrogen), DpnI, HindIII, NcoI, PciI, SalI, ScaI, SmaI, StuI, XhoI (NEB)
Ligase	Quick Ligase (NEB)
Protease	TEV (EMBL), Proteinase K (Invitrogen)

Table 2.20: Kits.

Name	*Application*	*Supplier*
Quick Ligation	DNA ligation	New England Biolabs
InFusion	Cloning	Clontech
QuikChange	Site-directed mutagenesis	Stratagene
Colony PCR	PCR-Master-Mix S	Peqlab
NucleoSpin Plasmid	Plasmid miniprep	Macherey-Nagel
NucleoSpin PCR Clean-up	PCR DNA purification	Macherey-Nagel
ECL Western Blotting	HRP detection	Thermo Scientific
NE-PER	Nuclear extraction	Thermo Scientific

Table 2.21: Ladders.

Name	*Application*	*Range*	*Supplier*
GeneRuler 100 bp	DNA	100 - 1,000 bp	Fermentas
GeneRuler 1 kb	DNA	250 - 10,000 bp	Fermentas
PageRuler Prestained	Protein	10 - 170 kDa	Fermentas

2.7 Equipment

Table 2.22: Software and databases.

Software/Database	*Application*	*Link (www)*
ATSAS Suite	SAXS Analysis	embl-hamburg.de
CCP4 Suite	Data processing and refinement	ccp4.ac.uk
ClustalW2	Multiple sequence alignment	ebi.ac.uk
Coot	Model building	lmb.bioch.ox.ac.uk/coot
EndNote	Literature management	endnote.com
ESPript	PostScript alignment generator	espript.ibcp.fr
ExPASy	Bio-Informatic tools collection	expasy.org
LaTeX	Typesetting	latex-project.org
Microsoft Office	Productivity	office.microsoft.com
MolProbity	Structure validation	molprobity.biochem.duke.edu
Mosflm	Indexing	mrc-lmb.cam.ac.uk
MxPro	Thermofluor	agilent.com
NCBI	Biomedical database	ncbi.nlm.nih.gov
Origin Pro	Graphing and data analysis	originlab.com
PDB	Protein structure databank	rcsb.org
Phaser	Phasing	phaser.cimr.cam.ac.uk
Phenix	Refinement	phenix-online.org
Phyre	Protein structure prediction	sbg.bio.ic.ac.uk
PsiPred	Secondary structure prediction	bioinf.cs.ucl.ac.uk
PyMol	3D viewer	pymol.org
Quantity One	Gel documentation	bio-rad.com
Refmac5	Refinement	ysbl.york.ac.uk
Scala	Scaling	mrc-lmb.cam.ac.uk
Spectra Manager	CD spectroscopy	jascoinc.com
UCLA MBI	Diffraction Anisotropy Server	services.mbi.ucla.edu
Unicorn	Chromatography	gehealthcare.com
UniProt	Protein database	uniprot.org

Bio-Informatic tools, programs, online applications, and databases.

Table 2.23: Instruments.

Application	*Device*	*Company*
Biolayer Interferometry	Octet RED	Forté Bio
CD Spectropolarimeter	J-810	Jasco
Cell Disruptor	E615	Constant Systems
	M-110P	Microfluidics
Confocal Microscope	TCS STED	Leica
Crystal Dehydration	HC1	EMBL
Crystallization Robot	HoneyBee 963	Zinsser Analytic
DLS	DynaPro Titan	Wyatt Technology
DNA Gel Documentation	Gel Doc XR+	BioRad
Fluorescence Spectrometer	FluoroMax 4	Horiba Jobin Yvon
FPLC Systems	Äkta purifier 10	GE Healthcare
	Äkta avant	GE Healthcare
ITC	VP-ITC	MicroCal
	ITC200	GE Healthcare
Liquid Handling System	Lissy	Zinsser Analytic
Molecular Imager	Pharos FX	BioRad
PCR Thermocycler	Mastercycler EPgradient S	Eppendorf
Plate Reader	FluoStar Optima	BMG Labtech
Spectrophotometer	NanoDrop ND 1000	Peqlab
Thermofluor (qPCR)	Mx3005P	Agilent
X-ray Detector	R-AXIS HTC	Rigaku
X-ray Generator	MicroMax-007HF	Rigaku

Relevant instruments; standard lab equipment is not listed.

3 Methods

3.1 Cloning and Mutagenesis

For cloning, mutagenesis, and other modifications of the plasmids conventional restriction/ligation based cloning, sequence and ligation-independent cloning (SLIC), and QuikChange protocol as well as combined approaches were applied. The exact strategies and details for the particular construct preparations are provided with the respective cloning experiments in Chapter 4.2.

Detailed experimental protocols have been described previously (Sambrook et al., 1989; Li and Elledge, 2007; Liu and Naismith, 2008).

Genes encoding Hef (NCBI: CAC12619), PCNA (NCBI: Q9HJQ0), FEN-1 (NCBI: NP_394495), XPD (NCBI: Q9HM14), XPB (NCBI: NP_394476) and Bax1 (NCBI: NP_394477) from *Thermoplasma acidophilum* (taHef, taPCNA, taFEN-1, taXPD, taXPB, taBax1) were cloned from genomic DNA (ATCC 25905). Genes encoding Mph1 (NCBI: EGS17949) and PCNA (NCBI: EGS18086 w/o exon 4) from *Chaetomium thermophilum* (ctMph1, ctPCNA) were cloned from mRNA (Amlacher et al., 2011) derived cDNA kindly provided by Agnes Elias. The expression vector pIRESneo3 (Clontech) containing FANCM (NCBI: NP_065988) cDNA from *Homo sapiens* (hsFANCM) was kindly provided by Weidong Wang (Xue et al., 2008) and the gene encoding human PCNA (hsPCNA, NCBI: NP_002583) was extracted from cDNA (Agilent MegaMan Human Transcriptome Library).

For bacterial expression in *E. coli* the vectors pETM-11, pBADM-11, pCDF-11 (EMBL) were used. For expression in stably transfected HEK293 cells the vector pIRESneo3. was used. Furthermore, the vectors pETM-11 and pIRESneo3 were modified by addition of an N- or C-terminal HA-/Flag-tag. In pBADM-11 the N-terminal His_6-tag/TEV-site was deleted to create a native hsPCNA expression construct.

For verification of the correct nucleobase sequences all constructs were sequenced by Seqlab (Göttingen) or Eurofins MWG (Ebersberg).

3.2 Recombinant Protein Production

3.2.1 Cell Culture

E. coli were cultivated in LB medium under permanent selection with the appropriate antibiotics (Table 2.11) at 37 °C. For general purposes liquid medium was used while for cloning purposes and to obtain single clones cells were grown on solid LB agar plates. For miniprep DNA isolation and small-scale expression tests 5 ml cultures were used. Large-scale protein expression (up to a total volume of 30 l) was conducted in multiple 2.5 l flasks at 200 rpm and expression was induced at OD_{600} = 0.5 - 0.8 by adding 0.02 % arabinose (pBADM-11) or 0.5 mM IPTG (pETM-11). All proteins were expressed using *E.coli* RIL cells at 15 °C over night in LB medium.

For long-term storage 20 % glycerol was added to the LB medium and then 1 ml of *E. coli* cells was shock frozen in liquid N_2 and stored at -80 °C.

HEK293/HeLa were cultivated following standard cell culture protocols as provided by the European Collection of Cell Cultures (ECACC Handbook - Fundamental Techniques and Protocols, Sigma-Aldrich). Stably transfected HEK293 cell lines were constantly grown in the presence of G418. In 2-3 day intervals cells were split according to the respective population density using DPBS washing and trypsine digestion procedures. Due to constitutive protein expression from the pIRESneo3 vector (CMV promoter) a particular induction for expression initiation was not needed.

For long-term storage of HEK293/HeLa cells 20 % DMSO and 40 % FCS were added to the medium and cells (1 ml) were frozen to -80 °C over night using a cryo-tool filled with isopropanol. The cells were then stored in liquid N_2.

3.2.2 Transformation

Plasmids were transformed into chemo-competent *E. coli* cells (NEB, Agilent) via a heat shock transformation protocol. 1 µl of plasmid DNA (miniprep scale; ca. 100 ng/µl / SLIC: 5 µl of annealing reaction) was added to a 50-100 µl aliquot of competent *E. coli* cells and incubated on ice for 30 min. Then a heat shock was applied for 1 min at 42 °C in a water bath and subsequently the cells were transferred to ice for 5 min. 900 µl LB-Medium was added and cells were grown for 60 min at 37 °C and 600 rpm. Afterwards 150 µl of the cells were plated on an LB-Agar plate containing the appropriate antibiotics and incubated overnight at 37 °C.

Transfection and generation of stable HEK293 cell lines via G418 selection (pIRESneo3) was kindly performed by Detlev Schindler (Human Genetics, Biocenter, University of Würzburg).

3.2.3 Cell Lysis

To gain access to the recombinantly expressed target protein cells were lysed mechanically and then separated into soluble and insoluble fractions. Cells either obtained from a fresh culture or from a frozen cell pellet were resuspended in cold Lysis buffer (1g/10ml). The suspension was subjected to the cell disrupter system (E615 Constant Cell Disruption Systems or M-110P Microfluidics) twice at a pressure of 1.8 kbar and constant cooling. Subsequently the cell lysate was centrifuged at 34,000 g and 4 °C for 1 h and the resulting supernatant was used for further protein purification procedures.

HEK293 cells were lysed osmotically by incubation in ice-cold CoIP-LP buffer for 5 min immediately followed by centrifugation at 10,000 g for 10 min at 4 °C to separate soluble and insoluble fractions.

3.2.4 Affinity Chromatography

In general affinity chromatography is a method for protein isolation based on the specific and reversible binding of the target protein to an immobilized ligand. Thereby the target protein is separated from the cell lysate and after defined washing steps, to remove unspecific contaminants, it is eluted from the column. The final grade of protein purity thus depends on the affinity properties and the amount of target protein within the initially applied solution.

Detailed information regarding this purification method is provided by (Porath et al., 1975) and the Affinity Chromatography Handbook (GE Healthcare).

All recombinant proteins purified in this thesis contained an N-terminal hexahistidine (His_6)-tag and hence immobilized metal ion affinity chromatography (IMAC) using Ni^{2+} cations was applied. After equilibration with His-A buffer, the soluble cell lysate fraction was subjected to a 1 ml NiMAC (Novagen) or a 5 ml HisTrap FF crude (GE Healthcare) column at 4 °C. The column was extensively washed with approx. 20 CV His-A buffer containing 20 mM imidazole until a stable baseline of the UV absorption was achieved. Then proteins were eluted gradually, step-wise, or using His-B buffer containing 500 mM imidazole. Chromatographic purification was performed using Äkta FPLC (purifier or avant, GE Healthcare) systems. Alternatively, the column loading step was conducted manually using a peristaltic pump and subsequently the column was connected to the Äkta system for the washing and elution procedures. Finally, samples of the different fractions were analyzed via SDS-PAGE and fractions containing the desired protein were pooled and concentrated for subsequent size exclusion chromatography.

3.2.5 Size Exclusion Chromatography

Size exclusion chromatography (SEC), also known as gel filtration, is a chromatographic method to separate proteins by size and shape under native conditions. SEC experiments

were performed using a HiLoad Superdex 200 26/60 pg column (GE Healthcare) with a resolution of 10 to 600 kDa in combination with an Äkta FPLC system (purifier or avant, GE Healthcare) at 4 °C. The Superdex 200 material consists of dextran which is covalently linked to a highly cross-linked agarose matrix. Under a constant flow, small proteins enter the porous Superdex material thereby covering a longer distance than larger proteins that are unable to enter. Thus proteins of high molecular weight elute earlier than small proteins. During the experiment parameters such as absorption at 280 and 260 nm, buffer concentrations, pH value, conductivity, column and system pressure, and temperature were monitored.

Protein samples with a volume of 5 or 10 ml were injected automatically from a sample loop and eluted isocratically with a flow-rate of 1 ml/min over a length of 1.3 column volumes (CV). Peak fractions were collected for subsequent protein content determination via SDS-PAGE (Chapter 3.3.1). Afterwards chosen fractions were pooled, concentrated, flash-frozen in liquid N_2 and stored at -80 °C.

3.3 Protein Characterization

3.3.1 Gel Electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a method to separate proteins according to their molecular weight under denaturing conditions (Laemmli, 1970). SDS is a negatively charged detergent that binds to the protein backbone and provides the protein with negative charges that are proportional to it's size. Additionally, a reducing agent such as β-mercaptoethanol is added to disrupt disulfide bonds and thus supporting protein denaturation. In an electric field, consequently, all denatured SDS-coated proteins move towards the anode. The SDS coating ensures that the migration speed is only determined by the protein size, independently of the natural charge or conformation, and the meshwork of the polyacrylamide fibers that construct the gel.

SDS-PAGE was used to determine the content and the grade of purity of the protein samples. The samples were mixed with Loading-SDS buffer and incubated for 5 min at 95 °C prior to loading on the gel. To estimate the molecular weight of the separated proteins, a marker containing proteins of known sizes (Fermentas PageRuler Prestained) was also applied as a reference on every gel. After electrophoretic separation the protein content was visualized by different staining techniques such as Coomassie-staining, silver staining, or by Western blot. Gels were digitized using a Scanjet G2710 (HP) scanner and finally dried.

Proteins were also analyzed under native conditions and separated according to their charge via poly-acrylamide and agarose gel electrophoresis. This method is described in Chapter 3.4.1.

3.3.2 UV Spectrophotometry

Protein and DNA concentrations were determined via absorption measurements using a NanoDrop ND 1000 (Peqlap) UV/Vis spectrophotometer. Proteins are measured at a wavelength of 280 nm and DNA at 260 nm, respectively, while the ratio of A_{280}:A_{260} was used as an indicator for sample purity. Concentrations (c) were calculated utilizing the Lambert-Beer law (Equation 3.1) from the experimentally obtained absorption (A) as well as the known values for thickness of the cuvette (d) and the sample specific extinction coefficient (ϵ) (Pfeiffer and Liebhafsky, 1951).

$$c = \frac{A}{\epsilon \cdot d} \tag{3.1}$$

Protein specific extinction coefficients at 280 nm (ϵ_{280}) can be calculated from the quantity (n) of its tryptophan, tyrosine, and cysteine residues (Equation 3.2).

$$\epsilon_{280} = n_{Trp} \cdot 5690 + n_{Tyr} \cdot 1280 + n_{Cys} \cdot 120 \tag{3.2}$$

Additionally, it is also possible to specify protein concentrations via the peptide bond absorbance at 205 nm.

Molecular weights and calculated extinction coefficients of all recombinantly expressed and from *E. coli* purified proteins are specified in Table 2.12.

3.3.3 Thermofluor

Thermofluor, also known as differential scanning fluorimetry, is an indirect experimental approach for the determination of protein melting temperatures (T_M) reflecting a measure for protein stability. The fluorophore Sypro Orange (Sigma-Aldrich) becomes activated through its interaction with hydrophobic amino acid side chains. Upon heat-dependent protein unfolding, buried hydrophobic residues within the soluble protein will be exposed to the aqueous phase and trigger Sypro Orange activation, thereby increasing the fluorescence signal. Sypro Orange fluorescence is detected at wavelengths for excitation = 492 nm and emission = 610 nm.

Thermofluor is typically used for protein specific buffer optimization to improve the stability of the protein in solution. Furthermore, this method can also be used for protein-DNA, protein-ligand, and protein-protein interaction analyses. Upon interaction the protein is stabilized and thereby the T_M is increased.

The experimental setup constitutes multiple measurements at the same time in a 96-well plate with a volume of 25 µl per well, 0.5 mg/ml protein, and 5x concentrated

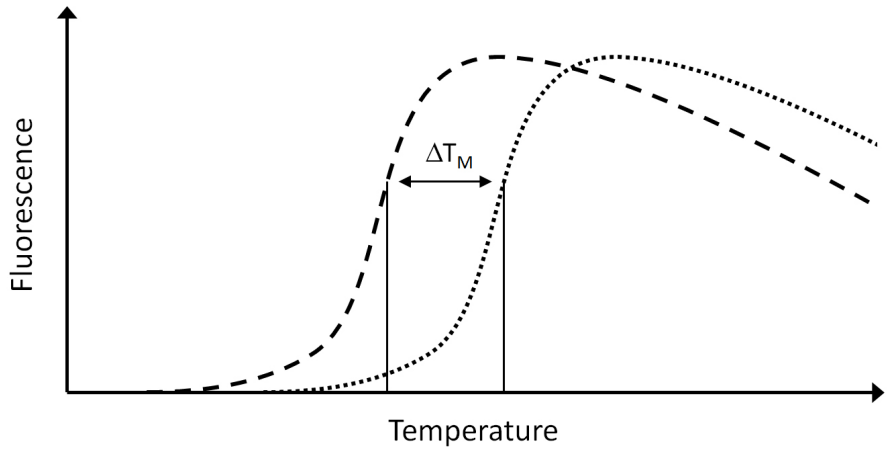


Figure 3.1: Thermofluor scheme.
Stabilization of the protein by buffer optimization or DNA binding results in a shift of the T_M to a higher temperature (ΔT_M). T_M is calculated from the turning point (f" = 0) of the melting curve.

Sypro Orange in the test buffer. While monitoring the fluorescence signal the mixtures are heated up from 25 °C to 95 °C with a speed of 1 °C/min using a qPCR instrument (Agilent Mx3005P). In addition to the standard buffer screen (Table 2.18) customized screens were developed to analyze stabilizing effects of additional components such as salt and DNA. The resulting data were evaluated using the data analysis file (Excel) provided by the Structural Genomics Consortium (SGC) in Oxford.

A detailed description of this method is provided by F. Niesen (Niesen et al., 2007).

3.3.4 Dynamic Light Scattering

Dynamic light scattering (DLS), also known as photon correlation spectroscopy, is a spectroscopic method to determine the size distribution of particles in solution. Monochromatic laser light is scattered by macromolecules and other dissolved particles revealing the polydispersity of the solution. Based on the difference between the input and the scattered light, the diffusion constant and size of the particles are calculated describing the hydrodynamic radius. Assuming a globular and spherical shape of the protein, the Stokes-Einstein relation can be applied to deduce the apparent hydrodynamic radius R_h(app) (Equation 3.3).

$$R_h(app) = \frac{kT}{6\pi\eta D(c)} \tag{3.3}$$

The concentration-dependent translational diffusion coefficient (D(c)), the Boltzmann constant (k), the temperature (T), and the viscosity of the solvent (η) are known from

the experimental setup.

Protein samples (0.5 mg/ml) in standard buffer were extensively centrifuged for 30 min at 16,000 g and 20 °C. Subsequently the samples were analyzed in a quartz cuvette (DynaPro, Wyatt Technology) in multiple measurements at various temperatures and then the data of the same temperature were averaged. For crystallization purposes the polydispersity of the protein solution should not exceed 20 %. The data were automatically evaluated by the Dynamics V6 software supplied with the equipment and presented as a distribution of molecular radii and molecular weights.

Further information about this method and particularly its application in protein crystallization is supplied by W. William Wilson (William Wilson, 2003).

3.3.5 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy is a method to analyze the secondary structure content of a given protein and to verify the correct folding of the protein or its variants of interest. Circular polarized light is absorbed by chiral substances such as amino acids while the molecular environment of the amino acids influences their absorption properties. The resulting ellipticity is plotted against the wavelength and characteristic spectra for α-helices, β-sheets, and disordered protein regions are obtained (Figure 3.2). Thus CD spectroscopy was especially used to confirm that mutations within a target protein do not affect its overall conformation thereby impairing its function.

A detailed explanation of this method with the focus on protein characterization is provided by S.M. Kelly (Kelly et al., 2005).

CD measurements were conducted at 20 °C in a 0.1 cm quartz cuvette utilizing a J-810 CD spectropolarimeter (Jasco). The protein was used at a concentration of 0.25 mg/ml in a volume of 200 µl and monitored in CD buffer (20 mM K-phosphate, pH 8.0) to exclude buffer artifacts such as caused by the chiral Tris molecule. Spectra were recorded from 260 to 190nm and accumulated eight times to optimize the signal-to-noise ratio with a scan speed of 20 nm/min. Spectra of different measurements (e.g. wild type and variants) were normalized.

3.3.6 ATPase Activity Assay

Helicases are DNA dependent ATPases and vice versa ATP hydrolysis provides the energy required for DNA translocation and strand separation. To analyze the ATP consumption properties and the stimulating effects of various DNA substrates, an enzyme coupled ATPase assay was performed. The principle of this assay is depicted in Figure 3.3.

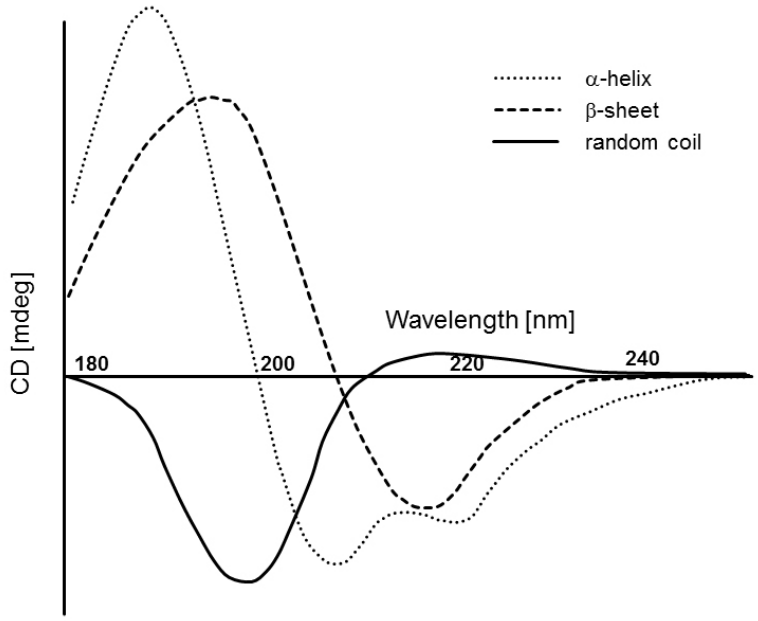


Figure 3.2: Schematic illustration of secondary structure specific CD spectra.
α-helices, β-sheets, and disordered protein entities exhibit different characteristic spectra. Proteins comprising α-helices as well as β-sheets also display combined spectra.

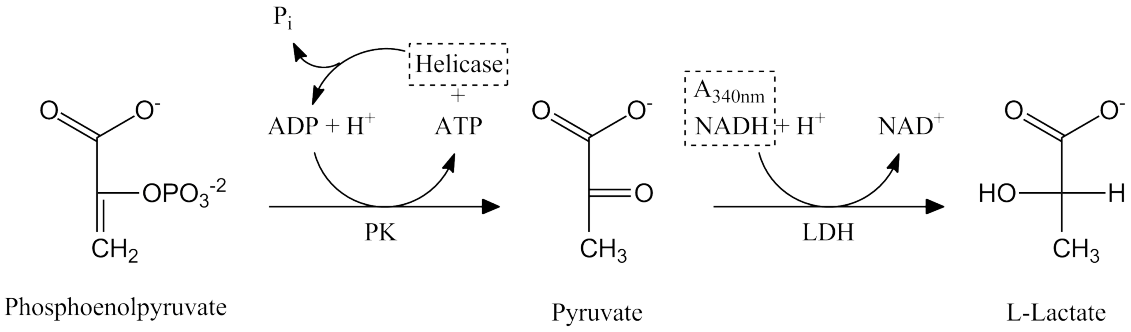


Figure 3.3: ATPase Assay.
Reaction scheme of the enzyme coupled ATPase assay with the participating enzymes pyruvate kinase (PK) and lactate dehydrogenase (LDH). In the presence of an active ATPase (i.e. helicase) ADP is generated which permits the dephosphorylation of phosphoenolpyruvate by PK to yield pyruvate. The oxidation of NADH is monitored spectrophotometrically by measurement of the absorption at 340 nm.

DNA dependent conversion of ATP to ADP was monitored using a coupled enzyme assay system consisting of pyruvate kinase and lactate dehydrogenase to link the hydrolysis of ATP to the oxidation of NADH (ϵ_{340nm} = 6220 M^{-1} cm^{-1}). Multiple measurements were conducted simultaneously utilizing a 96 well plate reader (FluoStar Optima, BMG Labtech). The reaction mixture contained the ATPase-buffer and a protein concentration gradient in the presence of 0.5 µM DNA substrate. The reaction mixture (0.1 ml) was equilibrated at 37 °C for 5 min followed by the addition of 1 mM ATP to initiate the reaction. The rate of hydrolysis was calculated from the linear change in absorbance at 340 nm after ATP injection. The concentration of conversed ATP equals the concentration of oxidized NADH (300 µM) in a 1:1 stoichiometry. The rates were determined multiple times and blank corrected for the oxidation of NADH in the absence of protein. The data are reported as the mean rate (M ATP / M protein * sec) ± the standard error of the mean.

3.3.7 Helicase Activity Assay

DNA unwinding activity was analyzed in a native gel based assay using Cy3-labelled DNA substrates. Various protein concentrations in the presence of labeled open fork DNA were incubated ± 5 mM ATP for 30 min at 37 °C. The helicase reaction buffer contained 2 µM capture oligonucleotide to prevent re-annealing of the separated substrate. The reaction was interrupted by the addition of Helicase-Stop buffer containing 1 µM capture oligonucleotide for 10 min at 37 °C.

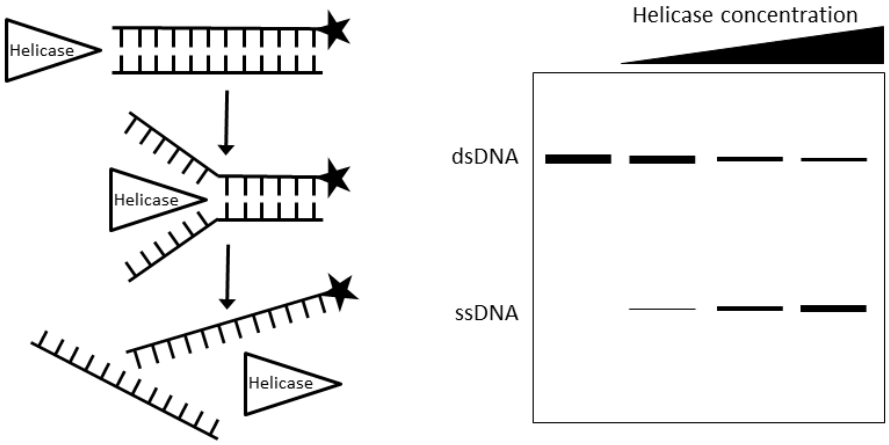


Figure 3.4: Schematic illustration of the helicase assay principle.
In an ATP-dependent process helicases separate DNA double-strands (dsDNA) resulting in single-stranded DNA (ssDNA). One strand is labeled with the fluorophore Cy3 which is indicated by the asterix and can be detected via native gel electrophoresis in combination with fluorescence imaging (PharosFX, BioRad). Coherent with an increasing helicase concentration the DNA unwinding efficiency is improved.

Subsequently, the products were resolved by non-denaturing polyacrylamide gel electrophoresis followed by fluorescence imaging detection of the Cy3 signal at an

excitation wavelength of 532 nm and an emission wavelength of 605 nm. The principle of this assay is shown in Figure 3.4.

Exact concentrations, different DNA substrates, and deviations from the standard protocol are listed within the section describing the particular experiment.

3.4 Interaction Analysis

3.4.1 Native Gel Electrophoresis

Protein-protein as well as protein-DNA interactions can be detected and analyzed via native gel electrophoresis. Due to the absence of any denaturing agents such as SDS, β-mercaptoethanol, DTT, heat or other factors, macromolecules are separated by their charge and conformation in the electric field. Polyacrylamide and agarose gels were used with this method. The pH values of the gel and the buffer have a major influence on the protein charge and thereby control the protein's running behavior in the gel.

Upon interaction the charge and overall conformation of the molecule are altered resulting in different gel migration. Markers are not used for native polyacrylamide or agarose gels. Complex formation is indicated by a new appearing band and concurrently the disappearance of the individual protein bands. Proteins were visualized by Coomassie or silver staining and the DNA was detected by silver staining or Midori green. Combinations were used as well.

The exact experimental procedures are provided with the respective experiments in Chapter 4.

3.4.2 Analytical Size Exclusion Chromatography

Protein complexes or protein-nucleic acid complexes can also be characterized by analytical size exclusion chromatography (ASEC). The principle of size exclusion chromatography (SEC) is described in Chapter 3.2.5. The complex constitutes a higher molecular weight than the individual components. Thus complex formation is indicated by a peak shift towards a lower elution volume and the reduction of the single component peaks. If proteins do not interact, the mixture only displays an overlay of the single component elution profiles.

For analytical size exclusion measurements a Superdex 200 10/300 GL (GE Healthcare) column was used. Proteins with a concentration of 50 µM in a volume of 500 µl were eluted isocratically with standard buffer for 1.3 CV at 4 °C. Samples were collected in 0.5 ml fractions and analyzed by SDS-PAGE for their protein content.

3.4.3 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a biophysical method for the determination of thermodynamic and kinetic properties of biomolecular binding interactions. Through the direct measurement of the change in enthalpy (ΔH) during a binding event under isothermal conditions (T), further parameters such as the dissociation constant (K_D), reaction stoichiometry (N), and the contribution of the entropy (ΔS) can be calculated. The definition of the Gibb's free energy (ΔG) is given in Equation 3.4.

$$\Delta G = \Delta H - T\Delta S \tag{3.4}$$

To obtain evaluable sigmoidal heat release profiles the ITC experiments should be conducted at a recommended c-value (Equation 3.5) between 5 and 1,000, while 40 is optimal ([M] = concentration of the macromolecule in the cell) (Broecker et al., 2011).

$$c = \frac{[M]N}{K_D} \tag{3.5}$$

A detailed description of this method is provided by Pierce and Freyer (Pierce et al., 1999; Freyer and Lewis, 2008).

Prior to the actual ITC experiment all protein samples were extensively dialyzed against standard buffer over night at 4 °C and subsequently centrifuged for 20 min at 25,000 g to remove aggregates, followed by filtration and degassing (only for VP-ITC). The buffer used for dialysis was also the reference buffer for the ITC measurements. All experiments were performed using VP-ITC and ITC200 instruments (MicroCal, GE Healthcare) at 25 °C and buffer-to-buffer titrations were carried out for background subtraction prior to curve fitting. The final molar ratio of the two components (syringe : cell) was $>$ 2:1. For oligomeric proteins the indicated concentrations always refer to the monomeric specimen. The signal of the first injection was discarded and data were fitted to a single binding site model using the MicroCal ITC data evaluation software (Origin) with no fixed parameters.

Exact values and experiment specific details such as used components, concentrations, volumes, number of injections, reference power and spacing time are indicated with the respective measurement in Chapter 4.

3.4.4 Co-Immunoprecipitation

Co-immunoprecipitation (CoIP) is an *ex vivo* method for the detection of protein-protein interactions. One protein component of a particular multi-protein complex is extracted from whole cell lysate (bait, see Figure 3.5) using a specific antibody

which is immobilized to protein A/G agarose beads. Protein A/G specifically binds immunoglobulins. Interacting proteins (prey, see Figure 3.5) are thereby also pulled out of the lysate and can be detected by Western blot or mass spectrometry. Because CoIP does not distinguish between a direct or an indirect interaction which can be mediated by additional factors, a mutated variant of one binding partner is needed that specifically disrupts complex formation. The basic principle of this method is depicted in Figure 3.5.

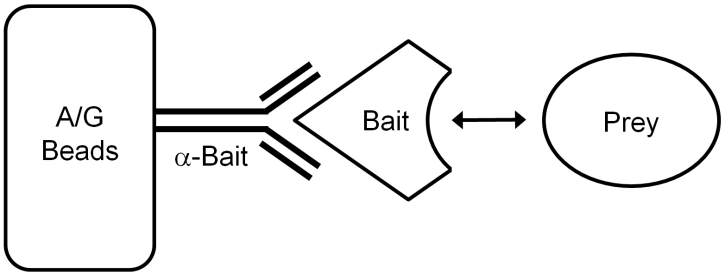


Figure 3.5: Principle of co-immunoprecipitation.
A specific antibody (α-bait) is immobilized on protein A/G agarose beads and binds the respective target protein (bait) from the whole cell lysate. Interaction partners (prey) are also extracted (co-immunoprecipitated) through interaction with the bait protein. After a washing procedure to remove unspecifically bound proteins, the proteins contained in the complex are detected via Western blot analysis using bait and prey specific antibodies.

A detailed description and standard protocols are provided by the Protein Methods Library (Pierce / Thermo Scientific).

Preliminary CoIP experiments were established using stably transfected HEK293 cells constitutively expressing HA-tagged proteins from the pIRESneo3 vector. Thus for immunoprecipitation and also subsequent Western blot (WB) analysis of the bait proteins (hsFANCM and hsFEN1) an α-HA antibody (Covance) was used. For WB detection of the prey protein (hsPCNA) a highly specific α-PCNA (Abcam) was applied. A 12 ml high density HEK293 cell culture was used for each CoIP experiment. First, the growth medium was discarded and the adherent cells were washed with 5 ml DPBS buffer. Then the cells were lysed osmotically and with the utilization of a scratching tool in 1 ml ice-cold CoIP buffer. The lysate was incubated for 20 min rotating at 10 rpm, followed by centrifugation at 10,000 g for 10 min. 5 µl α-HA antibody was added to the supernatant and incubated for 2 h rotating at 10 rpm. Meanwhile A/G-beads were washed 3x with buffer and blocked with 3 % BSA for 1 h. After washing the beads 3x, 20 µl were added to the antibody containing supernatant. At this point a "beads only" control was prepared, i.e. beads added to supernatant without antibody to test unspecific binding to the A/G beads. The reaction mixture was again incubated for 2 h rotating at 10 rpm. Finally, the beads were washed 3x, heated up to 95 °C for 5 min in SDS-loading buffer and objected to SDS-PAGE and Western blot analysis. The entire protocol was performed at 4 °C or on ice.

Further CoIP experiments were performed by the project associated collaboration partners Yutong Xue and Weidong Wang (NIA/NIH, Baltimore, USA).

3.4.5 Immunofluorescence

Immunofluorescence is a method to specifically stain proteins within a cell, utilizing fluorescently labeled antibodies (AB). Either fluorophore conjugated primary (direct) or secondary (indirect) ABs can be applied. For proper discrimination the attached fluorophores comprise different excitation and emission wavelengths resulting in unequal colors that ideally do not overlap. Upon interaction or being present in the same complex, the different colors merge and result in the complementary color. Imaging is performed by standard fluorescence or confocal microscopy.

Detailed information with respect to this method and its applications is provided by the reference book Immunohistochemical Staining Methods (Dako).

200 µl of low density HEK293 or HeLa cultures per well were grown for 1 d under standard conditions in a 12-well Chamber (Ibidi). Then a stimulating agent (MMC for ICL induction) was added at various concentrations and cells were grown for another day. The growth medium was removed and the adherent HEK293/HeLa cells were washed twice with 200 µl DPBS, respectively. Subsequently, cells were fixed and permeabilized simultaneously with AB-staining buffer for 30 min at RT followed by blocking with 5 % BSA in DPBS for 30 min. For antibody staining of proteins inside the nucleus the addition of Triton-X in the buffer was essential. The cells were washed 3x with 200 µl DPBS-T and the first AB (1:100 in 50 µl DPBS) was applied for 1.5 h. Again the cells were washed 3x with 200 µl DPBS-T and the second AB (1:300 in 50 µl DPBS) was applied for 1 h. After final washing (3x with 200 µl DPBS-T) and removal of the liquid, the samples were mounted using Fluoroshield with DAPI (Sigma) mounting medium for additional nucleus staining with DAPI. Alexa488, Alexa647, and DAPI were measured utilizing a TCS STED (Leica) microscope.

3.5 Structure Determination

3.5.1 X-Ray Crystallography

X-ray crystallography is a method to determine the 3D structure of micro- and macromolecules at atomic resolution. To solve the crystal structure of a specific target protein, multiple steps have to be accomplished and although the crystallization behavior of each individual protein is different and unpredictable some general rules can be applied. Exceptions are possible. First, a very pure ($>$ 90 %) and highly concentrated ($>$ 1 mg/ml) protein solution is needed. Well-diffracting crystals are obtained by screening and optimization procedures. X-ray diffraction datasets are recorded and processed while the phase information is acquired in additional experiments or extracted from a homologous pre-existing crystal structure. Next, the

model is built either manually or automatically and has to be refined. Eventually, the structural model of the target molecule is interpreted while complementary biochemical experiments cannot be overrated.

Actual experimental details are provided with the respective experiment in Chapter 4.

A detailed description of this complex method including extensive background information is provided by B. Rupp and C. Giacovazzo (Rupp, 2010; Giacovazzo et al., 2011). Significant parts of the following explanations have been published previously by H.M. Maric (Maric, 2012).

Protein Crystallization

To determine the 3D structure of a protein by means of x-ray diffraction, a well-ordered crystal is required which is characterized by a consistent homogeneous crystal lattice. Thus proteins were crystallized utilizing the vapor diffusion method. Proteins were mixed with the precipitant solutions (crystallization drop) and equilibrated against a significantly larger reservoir of the precipitant solution. Due to the concentration gradient of the precipitant between the drop and the reservoir, H_2O diffuses towards the reservoir via the gas phase, thereby slowly increasing the protein concentration within the drop. When the nucleation barrier is reached, crystal nuclei can form. Ideally, crystal growth at the nuclei and solvent diffusion together allow the protein concentration to stabilize in a condition where protein crystallization can occur. This technique was performed either by placing the drop on a shelf that is surrounded by the reservoir solution (sitting drop) or by hanging the drop from a cover slide towards the reservoir solution (hanging drop).

The sitting drop method was used for initial screening as well as subsequent fine screens using the HoneyBee 963 crystallization robot (Zinsser Analytic). The drop contained 0.3 µl protein solution and 0.3 µl crystallization solution and was equilibrated against 40 µl reservoir solution in a 96-well crystallization plate (Greiner Bio-One). Customized fine screens were generated using the Lissy liquid handling robot (Zinsser Analytic). The hanging drop method was used for follow-up screens to enhance crystal size where drops (1 µl protein solution plus 1 µl crystallization solution) were equilibrated against 1 ml reservoir solution. Crystal growth time varied from hours to weeks.

Data Collection

X-rays are scattered by electrons that are distributed periodically in either direction within the highly ordered crystal lattice. When x-rays encounter the lattice plane under the glancing angle θ and Bragg's law (Equation 3.6) is satisfied, constructive interference occurs and reflections can be detected. Intensities and positions of the reflections are measured and the Miller indices (h,k,l) are determined from the positions.

$$n\lambda = 2d_{hkl}sin\theta \tag{3.6}$$

Prior to the actual x-ray diffraction experiments, single crystals were transferred into a cryo-protectant solution using a nylon loop (Hampton Research) and shock frozen in liquid N_2. Initial data collection was performed using a MicroMax-007HF x-ray generator at a wavelength of 1.5418 Å and an R-AXIS HTC imaging plate detector (both Rigaku). Complete data sets were recorded using synchrotron radiation at the European Synchrotron Radiation Facility (ESRF, Grenoble) and the Helmholtz Zentrum Berlin (BESSY). All diffraction experiments were conducted at a temperature of 100 K.

Data Processing and Molecular Replacement

Diffraction images were indexed and integrated using Mosflm (Leslie and Powell, 2007). Scaling and all other subsequent steps were performed using the CCP4 software package and Phenix (Winn et al., 2011). Data processing was accomplished in five steps:

- Indexing: Selection of a preliminary space group and determination of unit cell parameters that describe the crystal lattice.
- Integration: Correction for background and determination of spot intensities.
- Merging: Correction for relative strengths of the spots in different images as well as calculation of a single averaged intensity for each structure factor.
- Evaluation: R_{sym} is calculated from the agreement between the separate observations of a reflection and its symmetry mates.
- Calculation of the structure factor amplitudes from the scaled and merged intensities.

The electron density map is related to the diffracted waves (Equation 3.7).

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} F_{hkl}e^{-2\pi i(hx+ky+lz)} \tag{3.7}$$

The measured intensity I(hkl) of a reflection is proportional to the square of the structure factor amplitude |F(hkl)| (Equation 3.8).

$$I(hkl) = |F(hkl)|^2 \tag{3.8}$$

To calculate the structure factor F(hkl) from the intensity, the phase ϕ of the wave is needed (Equation 3.9).

$$F(hkl) = |F(hkl)|e^{i \cdot \phi(hkl)} \tag{3.9}$$

The phase information is lost during data collection which is commonly known as the phase problem of crystallography. To solve the phase problem molecular replacement (MR) was applied. MR uses the phase information of a pre-existing and homologous structure as a search model. Correct positioning in the unit cell of the unknown structure is achieved by calculating the phase independent Patterson function of the measured data and the search model (Equation 3.10).

$$\rho(x, y, z) = \sum_{hkl} |F_{hkl}|^2 e^{-2\pi i(hx+ky+lz)} \tag{3.10}$$

The number of molecules in the asymmetric unit can be estimated with the help of the Matthews coefficient (V_M, Equaition 3.11).

$$V_M = \frac{V_{EZ}}{M_r \cdot z \cdot n} \tag{3.11}$$

V_M is calculated employing the molecular weight of the protein (M_r), the volume of the unit cell (V_{EZ}), the number of symmetry operators (z) in the space group, and number of molecules (n) in the asymmetric unit. The solvent content of a protein crystal is typically between 30 % and 70 %.

Model Building and Refinement

Structural models were built manually using Coot allowing real space fitting (Emsley et al., 2010) and optimized models were refined with Refmac or Phenix (Murshudov et al., 1997; Adams et al., 2010). Upon refinement the phases become more accurate and an improved electron density is calculated. Alternating rounds of model building and automated refinement were carried out to successively improve the atomic model.

The R-factor is an indicator for the quality of the model in comparison to the measured data. A decreasing R-factor thereby implies an improvement from the previous model (Equation 3.12).

$$R = \frac{\Sigma_{hkl}||F_o(hkl)| - |F_c(hkl)||}{\Sigma_{hkl}|F_o(hkl)|} \tag{3.12}$$

Because the R-factor is derived from data used during refinement it is biased towards the resulting model. Therefore the free R-factor (R_{free}) is calculated as an additional

control for the R-factor. The R_{free} typically comprises randomly chosen 5 % of the data that are not used in the refinement process thus reflecting the un-biased accuracy of the model.

For the final quality control the model was analyzed using MolProbity (Chen et al., 2010).

3.5.2 Small Angle X-Ray Scattering

Small angle x-ray scattering (SAXS) is a method for the structural characterization of macromolecules in solution at low resolution. The major advantage of SAXS is its ability to analyze a) proteins under native conditions, i.e. prevent crystallization artifacts, and b) flexible proteins or quaternary complexes. Particularly in combination with x-ray crystallography SAXS is a powerful technique.

A detailed description of this method is given by (Putnam et al., 2007) and a comprehensive summary can be found at (Mertens and Svergun, 2010). SAXS data analysis was performed by the project associated collaboration partner Adam Round (EMBL, Grenoble) who also wrote significant parts of this Section 3.5.2.

Prior to the actual SAXS measurements the protein samples were re-purified by additional SEC runs in Standard buffer and subsequently concentrated to ca. 10 mg/ml. Polydispersity below 10 % was validated by DLS (Chapter 3.3.4) and 5 mM DTT was added to prevent radiation damage of the samples. A dilution row ranging from approx. 10 to 0.5 mg/ml of the protein was monitored. All SAXS experiments were conducted at the ESRF BioSAXS beamline ID14-eh3 with additional measurements at the upgraded endstation BM29 (Pernot et al., 2010, 2013).

Standard experimental protocols were used with sample handling and scripting via the robotic sample changer and automated data collection available at the beamline. 10 individual frames were collected for every exposure, each 2 seconds in duration using the Pilatus 1M detector (Dectris). Individual frames were processed automatically and independently within the EDNA framework, yielding individual radially averaged curves of normalized intensity versus scattering angle (Equation 3.13).

$$s = \frac{4\pi SIN\Phi}{\lambda} \tag{3.13}$$

Additional data reduction within EDNA utilizes the automatic data processing tools of the EMBL-Hamburg ATSAS package, to combine timeframes, excluding any data points affected by aggregation induced by radiation damage, yielding the average scattering curve for each exposure series (Petoukhov et al., 2007). Matched buffer measurements taken before and after every sample were averaged and used

for background subtraction. Merging of separate concentrations and further analysis steps were performed manually using the tools of the ATSAS package. The forward scattering I(0) and the radius of gyration (Rg) were calculated from the Guinier approximation, the hydrated particle volume was computed using the Porod invariant and the maximum particle size (Dmax) was determined from the pair distribution function computed by GNOM using PRIMUS (Guinier, 1938; Porod, 1982; Svergun, 1992). 40 ab initio models were calculated for each construct using DAMMIF, and then averaged, aligned and compared using DAMAVER (Franke and Svergun, 2009; Volkov and Svergun, 2003). The most representative model for each construct was selected by DAMAVER. The theoretical scattering of the known structures was calculated with CRYSOL and OLIGOMER was used for fitting of a combination of multiple structures simultaneously (Svergun et al., 1995). Rigid body modeling using the high resolution structures of the known domains was performed with CORAL (Petoukhov et al., 2012).

4 Results

The Fanconi anemia (FA) DNA repair pathway is crucial for the replication-dependent removal of interstrand crosslink (ICL) DNA damages in higher eukaryotes (Knipscheer et al., 2009). Importantly, the SF2 DEAH translocase Fanconi anemia complementation group M (FANCM) protein plays an important role in the initiation and coordination of this process by recruiting downstream repair factors to the lesion (Xue et al., 2008). The following chapter presents insights into the structure and function of the FANCM protein family and demonstrates PCNA as a new interaction partner. For these studies the archaeal FANCM homolog Hef was primarily used as a model for the eukaryotic system. The Hef crystal structure as well as a SAXS in solution model elucidate the extreme conformational flexibility of this helicase. Strikingly, the interaction between Hef and PCNA via a conserved PCNA-interacting peptide (PIP) motif could be identified and was biochemically, structurally, and functionally characterized. Furthermore, complex formation with PCNA could also be validated for the fungal Mph1 protein which likewise belongs to the FANCM protein family. Finally, preliminary data on a putative interaction between human FANCM and PCNA are very promising but still require definite proof. Together these results provide new insights into the function of the FANCM protein family and indicate a role for PCNA in replication-dependent ICL repair.

4.1 Sequence Analysis

In previous studies the archaeal protein Hef was identified as a functional FANCM homolog (Mosedale et al., 2005). Interestingly, Hef from the archaeal model organism *Thermoplasma acidophilum* (taHef) holds a unique position. It is the only Hef protein family member of genome-sequenced archaea that comprises only the N-terminal helicase entity. All other Hefs additionally contain a C-terminal ERCC4-like nuclease portion that cleaves fork-structured DNA substrates (Nishino et al., 2003) and is also necessary for homo-dimerization forming a non-globular elongated constitution (Lestini et al., 2013). The Hef helicase entity is also conserved in retinoic acid inducible gene I (RIG-I) which senses viral RNA in the cytoplasm and plays an important role in the innate antiviral immune response (Schmidt et al., 2012). In contrast to taHef all other homologs are extended in length in either direction and contain further, functionally distinct domains (Figure 4.1).

A common feature of taHef, pfHef, hsFANCM, and hsRIG-I is the presence of the two RecA-like motor domains (Hel1 & Hel2) characteristic for SF2 helicases and

the thumb-like insertion domain decisive for substrate specificity (Hall and Matson, 1999). Together these three domains constitute the helicase entity. Interestingly, the hsFANCM C-terminal nuclease domain is enzymatically inactive and lacks residues essential for this activity (Enzlin and Schärer, 2002). Instead, the degenerated ERCC4-like domain conjoined with adjacent tandem helix-hairpin-helix (HhH) motifs of hsFANCM serves as an interface for FAAP24 binding which likewise contains an inactive ERCC4-like domain and HhH motifs (Ciccia et al., 2007). In contrast, RIG-I is functionally distinct from the FANCM protein family. In addition to the conserved helicase core hsRIG-I contains CARD1/CARD2 domains at the N-terminus and the Zn^{2+}-containing C-terminal regulatory domain (CTD) both of which are not present in the FANCM protein family (Kowalinski et al., 2011).

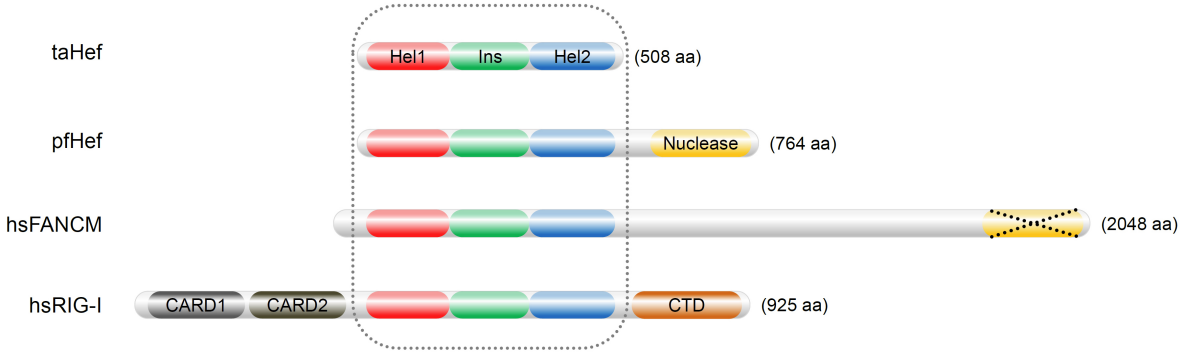


Figure 4.1: Domain architecture of Hef and the homologs FANCM and RIG-I. RecA-like motor domains Hel1 (red) and Hel2 (blue) comprise the seven SF2 helicase motifs and enclose the thumb-like insertion domain (green). Hef from *Pyrococcus furiosus* (pfHef) contains an ERCC4-like nuclease domain (yellow) and the corresponding enzymatically inactive domain of FANCM from *Homo sapiens* (hsFANCM) is illustrated by the dotted cross. The helicase core of RIG-I from *Homo sapiens* (hsRIG-I) is flanked at the N terminus by the two caspase activation and recruitment domains 1 and 2 (CARD1 in grey; CARD2 in black). At the C-terminus of hsRIG-I the C-terminal regulatory domain (CTD) is located (brown). The dotted frame marks the region displayed in detail in Figure 4.2.

Notably, this inimitable domain architecture of taHef is not based on an annotation error of the *T. acidophilum* genome sequencing results (Ruepp et al., 2000). The directly adjacent and inverted 3' downstream sequence to the taHef (gene Ta1501) TAG stop codon encodes the replication factor C small subunit (taRFCS; gene Ta1500) which is part of the PCNA clamp loader (Miyata et al., 2004). This genetic proximity of taHef and taRFCS might also indicate the functional connection between taHef and taPCNA (Chapter 4.9). On the 5' upstream side of the taHef gene Ta1501 the tRNA-Ala gene is encoded. There is thus no contact to a nuclease encoding sequence, strengthening the special character of taHef among the archaeal Hef protein family. However, the lack of DNA cleavage activity could be complemented by a directly or indirectly interacting, so far unknown and evolutionary related protein that resembles the pfHef nuclease properties, but it is also conceivable that taHef is functionally

distinct from other Hef proteins.

Generally, Hel1 and Hel2 show a higher conservation among these different enzymes than the thumb-like insertion domain (Figure 4.2). Full-length taHef and the helicase core of hsFANCM (residues 81-679) share a sequence identity of 25.4% (similarity 46.1%) while full-length taHef and the helicase core of hsRIG-I (residues 233-832) are 24.3% identical (similarity 39.8%). Nevertheless, the helicase motifs of hsRIG-I exhibit more differences than hsFANCM to the two Hef proteins (pfHef and taHef). Most strikingly, at the very C-terminus of taHef (residues 502-508) a canonical PCNA-interacting peptide (PIP) box motif (orange) is present which can also be found in pfHef (yellow) emphasizing their functional importance (Chapter 1.4). Overall, the C-terminal PIP box, adjacent to the helicase entity, is conserved among archaeal Hef proteins (Chapter 4.9). Notably, a PIP-box at the very N-terminus is also conserved among FANCM proteins which is most likely accessible for protein-protein interactions (Chapter 4.13.1). taHef and hsFANCM both contain only one conserved aromatic residue within the PIP-box. However, this deviation from the PIP-box consensus sequence is a common alteration present in multiple other PCNA interaction partners (Matsumiya et al., 2002). In contrast, the crystal structure of RIG-I demonstrates that sequences resembling putative PIP-boxes are all located in secondary structure elements and thereby not accessible for interaction with other proteins (Kowalinski et al., 2011).

To date taHef has been incorrectly annotated as “RNA helicase (RIG-I) related protein” (NCBI: CAC12619; UniProt: Q9HI46). On the basis of sequence identity, a precise assignment of taHef to FANCM or RIG-I is not possible. However, due to (a) similar functional domain composition, (b) higher conservation within the SF2 helicase motifs, and most importantly (c) the presence of the PIP-box, taHef should be considered as a simplified model of FANCM.

Importantly, to date taHef (Chapter 4.4) and pfHef proteins constitute the only available crystal structures of FANCM family members (Nishino et al., 2005b). Merely crystal structures of small fragments of hsFANCM in complex with binding partners could be solved (Figure 1.6) while the full-length constitution of eukaryotic FANCM or its homologs still remains elusive. To obtain further insights into these intriguing proteins I thus pursued secondary structure predictions of ctMph1 and hsFANCM using PsiPred which may reveal the feasibility to pursue structural studies of at least parts of these proteins. Since the *Chaetomium thermophilum* genome has not been fully annotated yet, the ATP-dependent DNA helicase-like protein (UniProt accession code: G0SET3) was identified as ctMph1 by a protein BLAST search (Figure 1.4). Secondary structure predictions revealed several features of ctMph1 (Figure 4.3). In accordance with the similarity to taHef with respect to the primary amino acid sequence level, the helicase core of the protein also displays a conserved structural constitution. Additionally, ctMph1 comprises long unstructured extensions at its N- and C-termini of about 200 and 440 amino acids (aa), respectively, and two classical

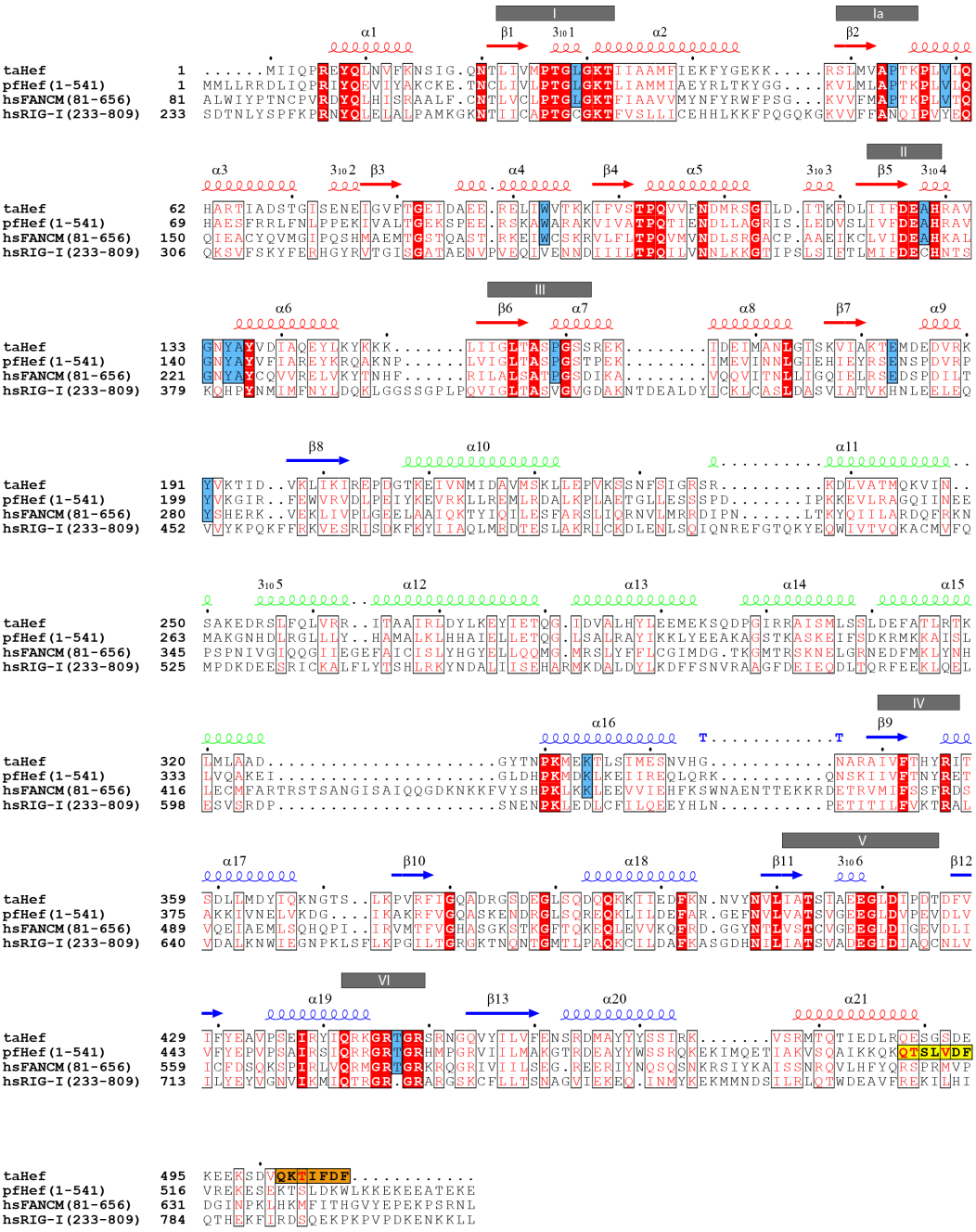


Figure 4.2: Alignment of taHef and corresponding helicase domains from homologs. From top to bottom: full-length taHef and homologous helicase domains of pfHef, hsFANCM and hsRIG-I. Secondary structure elements derived from taHef are indicated above the sequence and are color coded according to their domains; arrows indicate β-strands and coils $\alpha/3_{10}$-helices. Helicase motifs are shown as grey boxes. Conserved residues are shown by red letters and strictly conserved residues in white with a red background. Residues strictly conserved only among Hef and hsFANCM are marked with a blue background. The PIP-box at the very C-terminus of taHef is highlighted in orange whereas the putative PIP-box within pfHef is depicted by a yellow background. The alignment was generated with ClustalW2 (Larkin et al., 2007).

PIP-box sequences. The first putative motif comprises residues 108-114 and the second one is located at residues 918-925. Notably, both putative PIP-boxes are not conserved and e.g. Mph1 from *Saccharomyces cerevisiae* (scMph1) does not contain either of both (UniProt accession code: P40562).

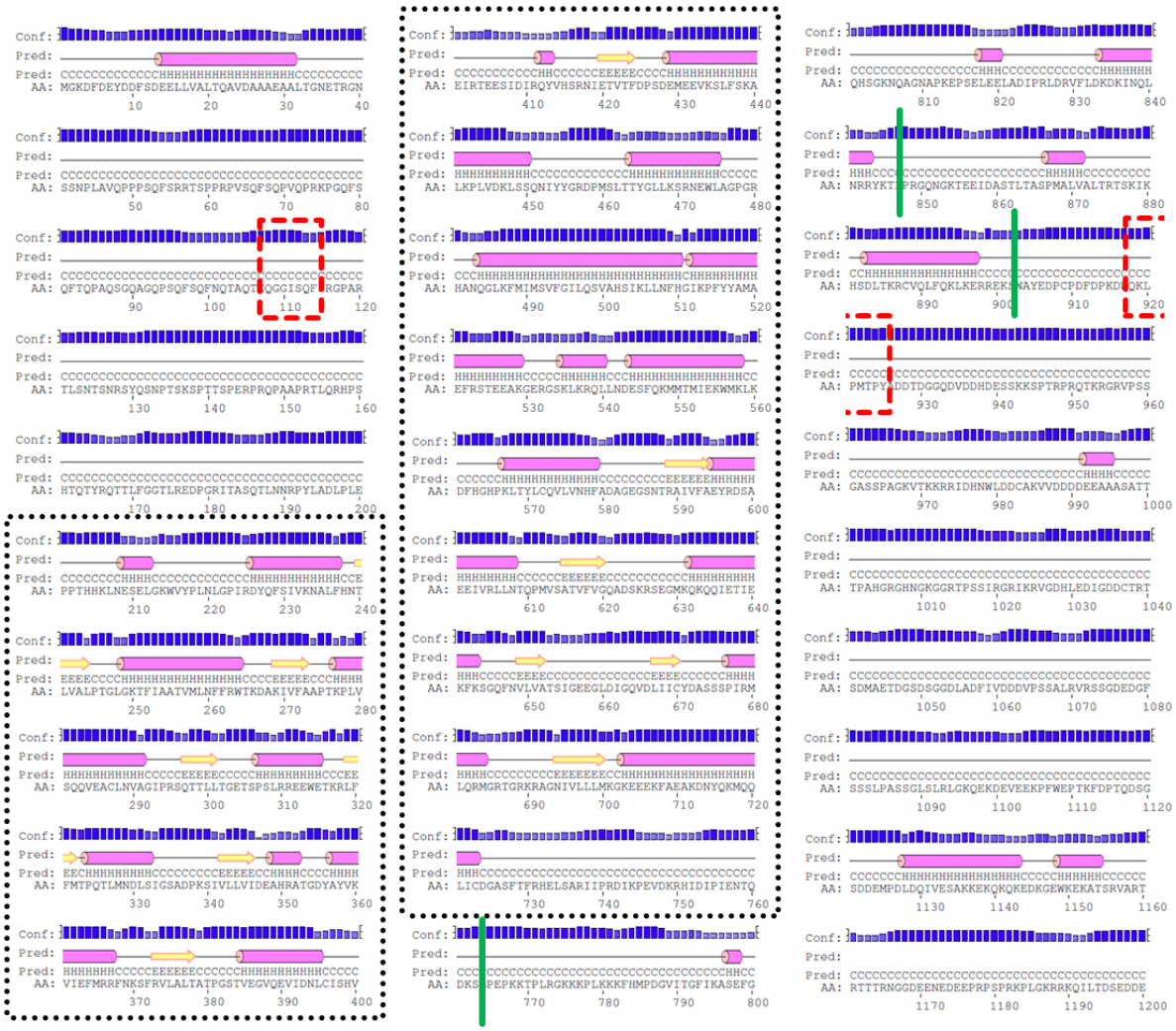


Figure 4.3: Secondary structure prediction for ctMph1.
ctMph1 is predicted to contain long unstructured extensions N- and C-terminally to the helicase core which is equivalent to taHef (black dotted frame). Magenta cylinders symbolize helices (H), yellow arrow β-strands (E), and the black lines in between represent random coils (C). The blue bars indicate the confidence of prediction. AA shows the amino acids sequence. The green lines indicate C-terminal degradation products and the red dotted frames depict putative PIP-box sequences (Chapter 4.12).

The secondary structure prediction of hsFANCM confirms that the helicase entity is N-terminally flanked by a 40 aa unstructured extension and that between the helicase and the C-terminal well-structured nuclease domain there is also a long unstructured

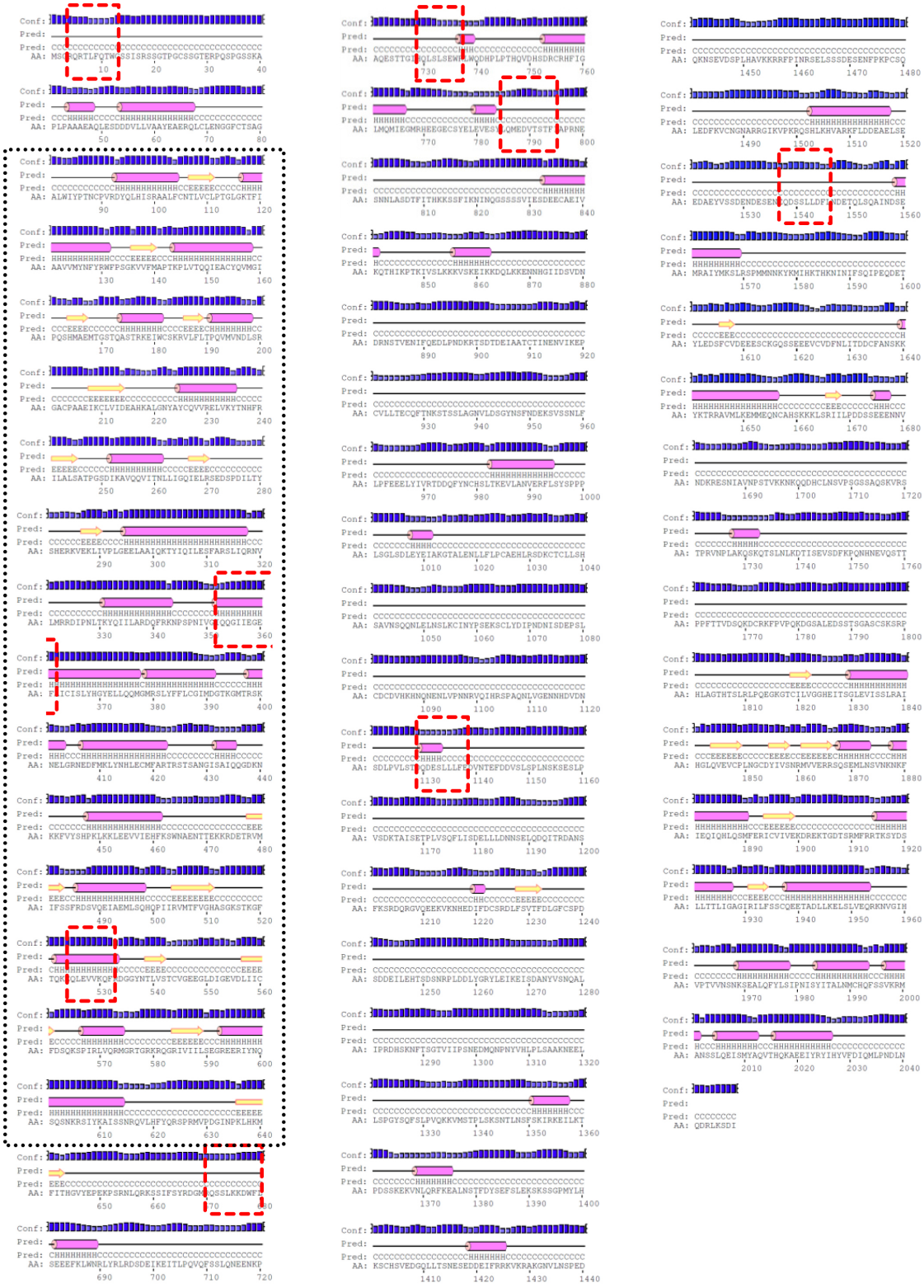


Figure 4.4: Secondary structure prediction for hsFANCM.
hsFANCM is predicted to contain an unstructured N-terminus as well as an unstructured linker region between the helicase entity (black dotted frame) and the C-terminal nuclease domain. The red dotted frames depict putative PIP-box sequences (Chapter 4.13).

linker region of about 1150 aa. Furthermore, there are eight putative PIP-boxes distributed over the entire protein. All motifs except the two within the helicase are located in unstructured areas which is an important property of functional PIP-boxes. Putative PIP-box number 7 (residues 1130-1137) is predicted to be partially located in an α-helix but the confidence for this secondary structure element is very low and thus uncertain. This region is therefore likely to be unstructured as well and thus PIP-box 7 still remains a possible candidate for a potential interaction with PCNA.

4.2 Cloning and Mutagenesis

For all *in vitro* as well as *ex vivo* experiments the genes encoding the corresponding target proteins first had to be recombinantly inserted into vectors and then transferred into expression hosts for subsequent protein production. Thus genes from *Thermoplasma acidopilum* (ta), *Chaetomium thermophilum* (ct), and *Homo sapiens* (hs) were cloned into various vectors (Table 4.1). For bacterial protein expression mainly pBADM-11/pCDF-11 and pETM-11/30 vectors were used because the different antibiotic resistances, regulatory promotor elements, purification tags, and origins of replication allow co-transfection coherent with co-expression of two proteins. For constitutive protein expression in HEK293 cells the vector pIRESneo3 was used. During the cloning procedures all PCR products, linearized vectors, test digestions and colony-PCRs of assembled plasmids were analyzed by 1 % agarose gel electrophoresis and ethidium bromide, Midori Green, or GelRed staining. Finally, to verify the correctness of the constructs, the entire inserts and 5'/3'-flanking regions were sequenced and compared to the theoretical sequences.

Table 4.1: Cloned expression constructs.

Insert	*Vector*	*Modification*
taHef	pBADM-11	—
taHef_K31A	pBADM-11	K31A exchange
taHef_D139F	pBADM-11	D139F exchange
taHef_D139K	pBADM-11	D139K exchange
taHef_I505R/F508S	pBADM-11	I505R/F508S double exchange
taHef_ΔC7	pBADM-11	Deletion of C-terminal 7 aa
taHef_ΔC33	pBADM-11	Deletion of C-terminal 33 aa
taPCNA	pBADM-11	—
taPCNA	pETM-11	—
taPCNA_addN25	pBADM-11	N-terminal extension of 25 aa
taPCNA_addN25	pETM-11	N-terminal extension of 25 aa
taFEN1	pBADM-11	—
taFEN1	pETM-11	—
taXPD	pBADM-11	—
taXPD	pETM-11	—

(Continued)

Insert	*Vector*	*Modification*
taXPB	pBADM-11	—
taXPB	pETM-11	—
taBax1	pBADM-11	—
taBax1	pETM-11	—
ctMph1	pETM-11	—
ctMph1_1-763	pETM-11	Stop codon at position 764
ctPCNA	pETM-11	—
ctPCNA	pCDF-11	—
hsFANCM	pETM-11	—
hsFANCM	pETM-30	—
hsFANCM_iso2	pETM-11	Isoform 2
hsFANCM_iso2	pETM-30	Isoform 2
hsFANCM_iso2	pFastBacHTa	N-terminal GST-tag from pETM-30 added
hsFANCM	pIRESneo3	—
hsFANCM_killPIP-1	pIRESneo3	L8R/W12S double exchange
hsFANCM_killPIP-2	pIRESneo3	I357R/F361S double exchange
hsFANCM_killPIP-3	pIRESneo3	V528R/V529R/F532S triple exchange
hsFANCM_killPIP-4	pIRESneo3	L674R/W678S double exchange
hsFANCM_killPIP-5	pIRESneo3	L733R/W736S double exchange
hsFANCM_killPIP-6	pIRESneo3	V790R/F794S double exchange
hsFANCM_killPIP-7	pIRESneo3	L1134R/F1137S double exchange
hsFANCM_killPIP-8	pIRESneo3	L1542R/F1545S double exchange
hsPCNA	pBADM-11	—
hsPCNA_native	pBADM-11	Native / no tag or other additional aa
hsPCNA-HA	pBADM-11	HA-tag inserted between His_6 and TEV site
hsPCNA-HA	pIRESneo3	Expression cassette from pBADM-11
hsPCNA-Flag	pIRESneo3	HA-tag substituted by Flag-tag
hsFEN1-HA	pETM-11	C-terminal HA-tag added
hsFEN1-HA	pIRESneo3	Expression cassette from pETM-11

Several cloning methods including classical restriction/ligation, LIC, InFusion (Clontech), and SLIC were applied. Notably, SLIC appeared to be the easiest, fastest, and thus most efficient strategy (Li and Elledge, 2007). The great advantage of SLIC is that there is no need for restriction enzymes and thereby any given sequences can be combined while the fragments are produced by PCR. For the success of this cloning method the length of the overlapping sequences is very important. The overlaps between insert and target plasmid should contain at minimum 18 bases. The following example shows the primer combinations for cloning of the hsPCNA gene into pBADM-11 both tagged and native:

hsPCNA_TEV_5'for
CTTTATTTTCAGGGCGCC ATGTTCGAGGCGCGCCTGGTCCAGGGCTCC

hsPCNA-native_5'for
CTAACAGGAGGAATTAACC ATGTTCGAGGCGCGCCTGGTCCAGGGCTCC

hsPCNA_pBAD_3'rev
GTGCGGCCGCAAGCTTGTC CTA AGATCCTTCTTCATCCTCGATCTTGGG

TEV_5'rev_Vec-linear
GGCGCCCTGAAAATAAAGATTCTC

pBAD-nativ_5'rev_Vec-linear
GGTTAATTCCTCCTGTTAGCCC

pET/pBAD_3'for_Vec-linear
GACAAGCTTGCGGCCGCACTCGAG

The PCR product for the gene (cyan) which is amplified by the primers hsPCNA_TEV_5'for / hsPCNA_pBAD_3'rev (tagged) or hsPCNA-native_5'for / hsPCNA_pBD_3'rev (native) is flanked by sequences that correspond to the vector (plum, orange, yellow; reverse-complementary). Notably, these specific vector based overlaps do not bind to the template during the first PCR cycle. Only from the third cycle on the primers are 100% complementary to the amplified target gene which is then flanked on both 5' and 3' sides by the short vector sequences. To prevent unspecific side products during gene amplification, the primer region that binds to the target gene sequence (cyan) needs to be considerably longer than the vector specific overlapping region attached to the 5' and 3' sides (plum, orange, yellow). Notably, the size of the overlapping region is only defined by the attached sequence of the hsPCNA primers. The shown vector primer sequences produced very clean products and in combination with the corresponding flanking sequences cloning succeeded with high efficiency. Furthermore, exactly the same vector linearization primers or rather the resulting PCR product, the linearized vectors, could also be used for the cloning of other genes into these respective vectors with the same flanking regions attached. Practically, this means for primer design concerning new constructs that only the gene/insert sequence is substituted starting from the start codon in the forward primer and from the stop codon (reverse complementary) in the reverse primers while the flanking sequences do not change. This leads to another great advantage concerning primer design. There is no need for the addition, deletion, or exchange of bases to retain the correct frame or enzyme recognition site. For primer design only the overlapping region sequence (plum, orange, yellow) is added 5' to the start codon or 3' to the stop codon (grey). If the C-terminal His_6-tag of the pETM-11 vector would be needed, the stop codon (grey) must be removed with the correct frame ensured.

4.3 Recombinant Protein Expression and Purification

Before large-scale protein expression and purification can proceed in most cases small-scale expression and His-trap enrichment tests were performed. If not explicitly mentioned otherwise, for all following procedures a standardized protocol was utilized consisting of *E. coli* RIL cells as the expression host at 15 °C induced with 0.5 mM IPTG (for pETM-11/30) or 0.02 % arabinose (for pBADM-11) for overnight expression followed by mechanical cell lysis and a two-step His-trap and size exclusion chromatography purification. Eventually, the purified proteins were concentrated, flash-frozen in liquid nitrogen, and stored at -80 °C for further usage. Proteins used in this study were expressed and purified several times. For the purpose of simplification in the following section only one exemplary procedure for each protein is illustrated.

taHef

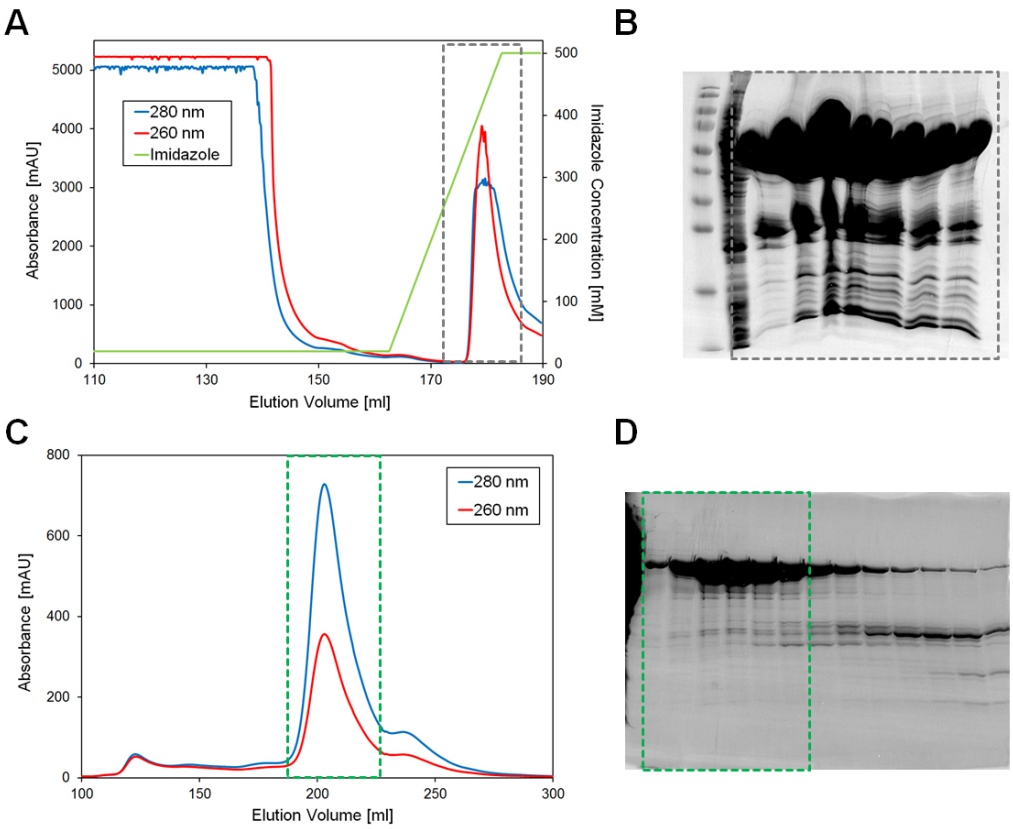


Figure 4.5: Purification of taHef.
(A) Chromatogram of washing and elution of taHef from the NiMAC 1 ml column (Novagen). **(B)** SDS-PAGE analysis of the protein content after the His-trap purification. **(C)** Chromatogram of the SEC using a Superdex 200 26/60 column (GE Healthcare). **(D)** SDS-PAGE analysis of the protein content after the SEC purification. Dashed boxes indicate the fractions in the corresponding gels. Final fractions are highlighted by the green dashed box.

taHef expression in a 14 l culture in LB medium yielded a 27 g cell pellet and after purification (Figure 4.5) a total amount of about 30 mg of pure protein could be obtained which was concentrated to 43 mg/ml (740 µM).

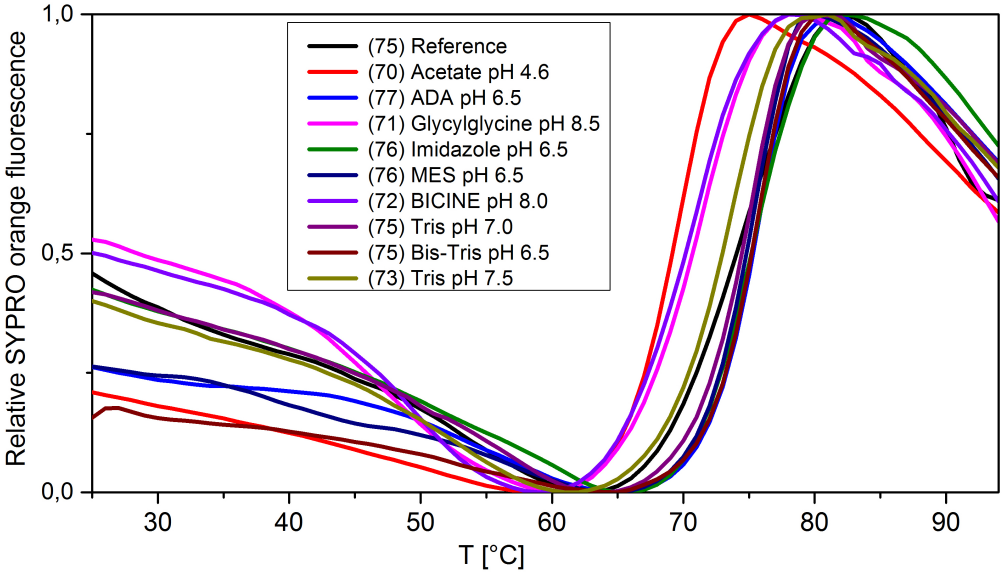


Figure 4.6: Thermofluor buffer screen of taHef.
Normalized protein melting curves of taHef in different buffers (colored) compared to the reference buffer (20 mM Tris, 500 mM NaCl, pH 7.5; black). The fluorescence signal refers to the signal emitted by SYPRO orange. Numbers in brackets in front of the buffer names in the legend indicate the according T_M values (°C).

A thermofluor buffer screen of apo taHef was performed and a selection of various representative buffers is depicted in Figure 4.6. Each buffer was used at a concentration of 100 mM and does not contain salt or other additives. Although the slope of the curve for the reference buffer (20 mM Tris, 500 mM NaCl, pH 7.5) is not optimal, the corresponding T_M of 75 °C is in the same range as the best buffers (MES, Imidazole and Bis-Tris, each at pH 6.5) with a T_M of about 76 °C. Apparently taHef seems to be most stable at pH 6.5. This analysis also showed that imidazole has no negative effect on the stability of taHef and should not harm the protein during IMAC purification. A comparison of the reference buffer with Tris/HCl pH 7.5 and pH 7.0 illustrates that a lower pH value and also the addition of salt contribute to taHef stability. The ideal buffer condition for taHef is a pH value of 6.5 in combination with up to 500 mM NaCl. However, since taHef crystallized well in the reference buffer (Chapter 4.4.1) this buffer condition with a reduced salt concentration (200 mM NaCl) was retained for all subsequent purifications.

For the purpose of a consistent system and to prevent drastic buffer changes for interaction experiments, all other proteins were purified utilizing the same buffer conditions with slight variations such as the addition of 5 mM $MgCl_2$ or 1 mM DTT.

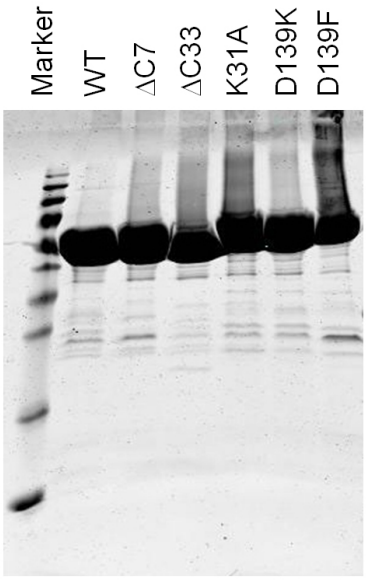


Figure 4.7: Purified taHef variants.
Equal concentrations (5 µM) of each purified and concentrated taHef protein variant purified as described for the wild type protein and separated by standard SDS-PAGE and subsequent Coomassie staining.

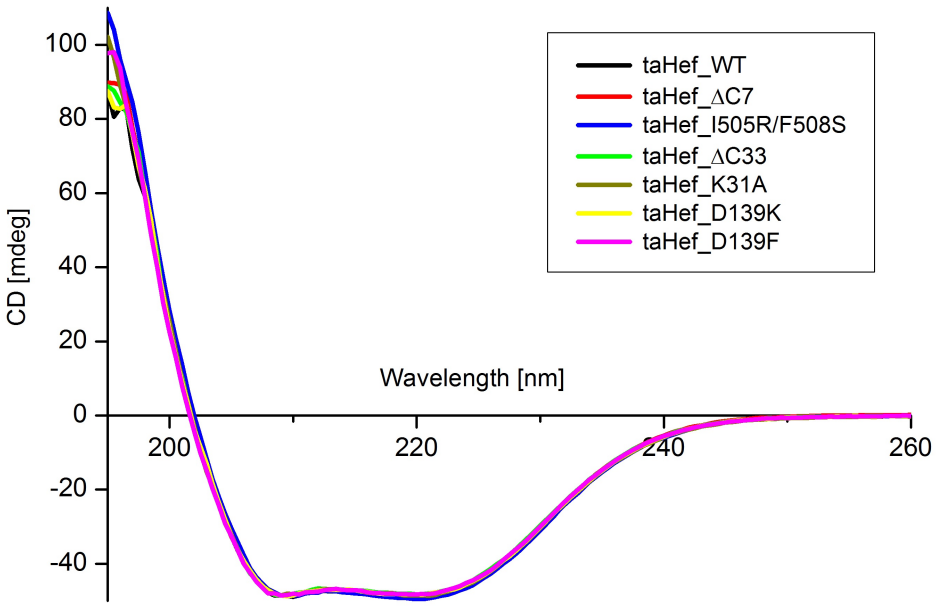


Figure 4.8: CD Spectra of taHef WT & Variants.
Circular dichroism (CD) spectra of taHef WT and variants display a very similar profile indicating correct folding of all analyzed proteins and a predominantly α-helical secondary structure composition.

All taHef protein variants were purified to homogeneity by the same protocol as the wild type protein and chromatograms as well as SDS-PAGE analyses were comparable (Figure 4.7). Furthermore, the correct folding of the taHef wild type protein as well as the variant proteins was confirmed by the according circular dichroism (CD) spectra that also indicated a relatively predominant α-helical constitution (Figure 4.8).

taPCNA

taPCNA expression in a 4 l culture in LB medium yielded a 27 g cell pellet and after purification (Figure 4.9) a total amount of 30 mg pure protein could be obtained which was concentrated to 10 mg/ml (320 µM).

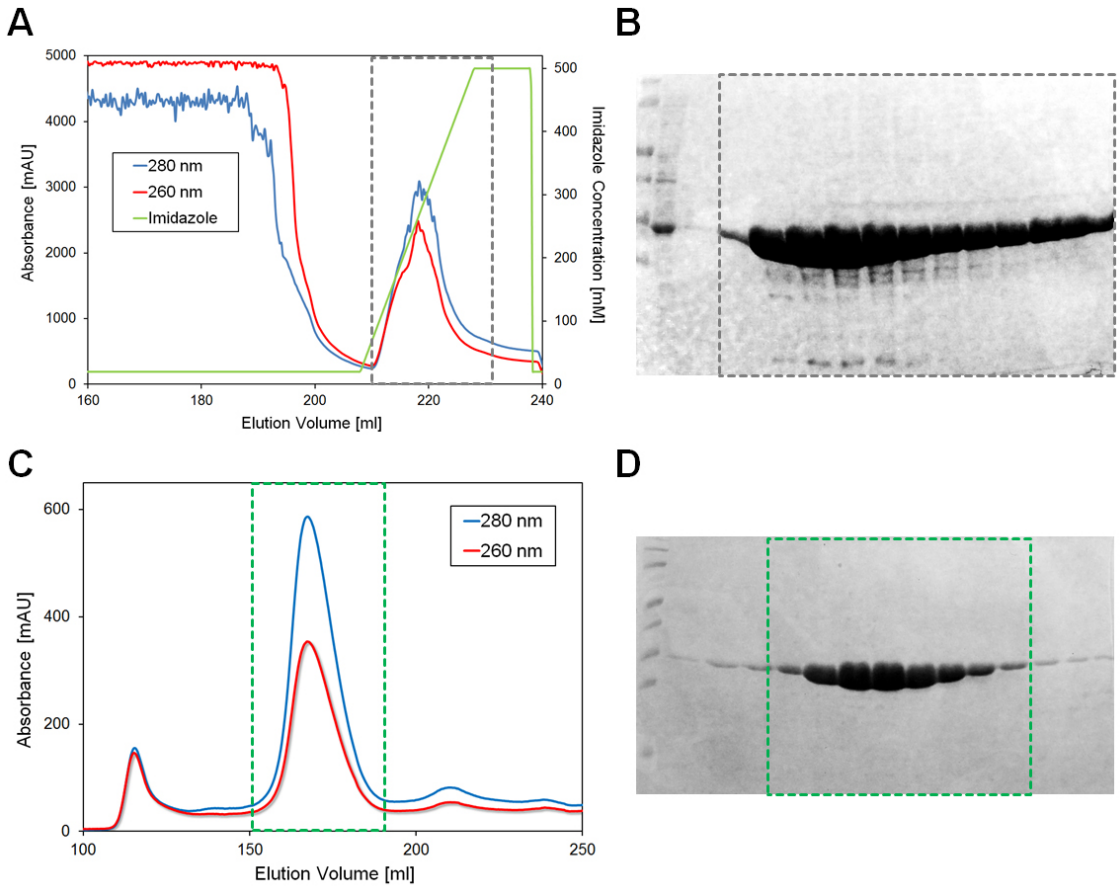


Figure 4.9: Purification of taPCNA.
(A) Chromatogram of washing and elution of taPCNA from the NiMAC 1 ml column (Novagen). **(B)** SDS-PAGE analysis of the protein content after the His-trap purification. **(C)** Chromatogram of the SEC using a Superdex 200 26/60 column (GE Healthcare). **(D)** SDS-PAGE analysis of the protein content after the SEC purification. Dashed boxes indicate the fractions in the corresponding gels. Final fractions are highlighted by the green dashed box.

taFEN1

taFEN1 expression in a 3 l culture in LB medium yielded a 20 g cell pellet and after purification (Figure 4.10) a total amount of about 13 mg of pure protein could be obtained which was concentrated to 43 mg/ml (1 mM). Notably, in addition to the His-trap and SEC an anion exchange chromatography step was inserted between the two chromatographic steps. The additional bands after the final purification step (Figure 4.10 F) were confirmed as degradation products of taFEN1 by mass spectrometry (MS).

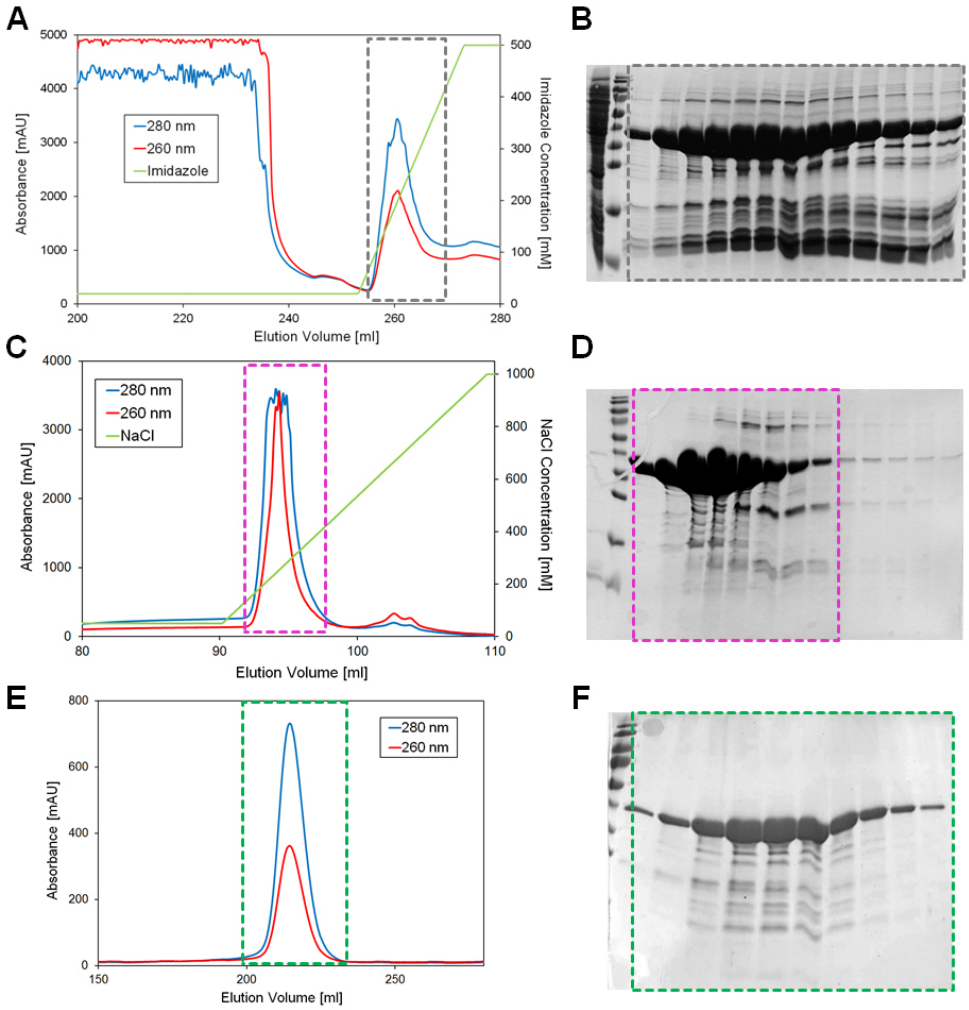


Figure 4.10: Purification of taFEN1.
(A) Chromatogram of washing and elution of taFEN1 from the NiMAC 1 ml column (Novagen). **(B)** SDS-PAGE analysis of the protein content after (A). **(C)** Chromatogram of the anion exchange chromatography using a HiTrap Q FF 1ml column (GE Healthcare). **(D)** SDS-PAGE analysis of the protein content after (C). **(E)** Chromatogram of the SEC using a Superdex 200 26/60 column (GE Healthcare). **(F)** SDS-PAGE analysis of the protein content after (E). Dashed boxes indicate the fractions in the corresponding gels. Final fractions are highlighted by the green dashed box.

To determine the oligomeric state as well as the polydispersity of taHef, taPCNA, and taFEN1 dynamic light scattering measurements were conducted (Figure 4.11). The proteins showed low polydispersity (9-14 %) and taHef as well as taFEN1 was present as a monomer while taPCNA constituted a homotrimer.

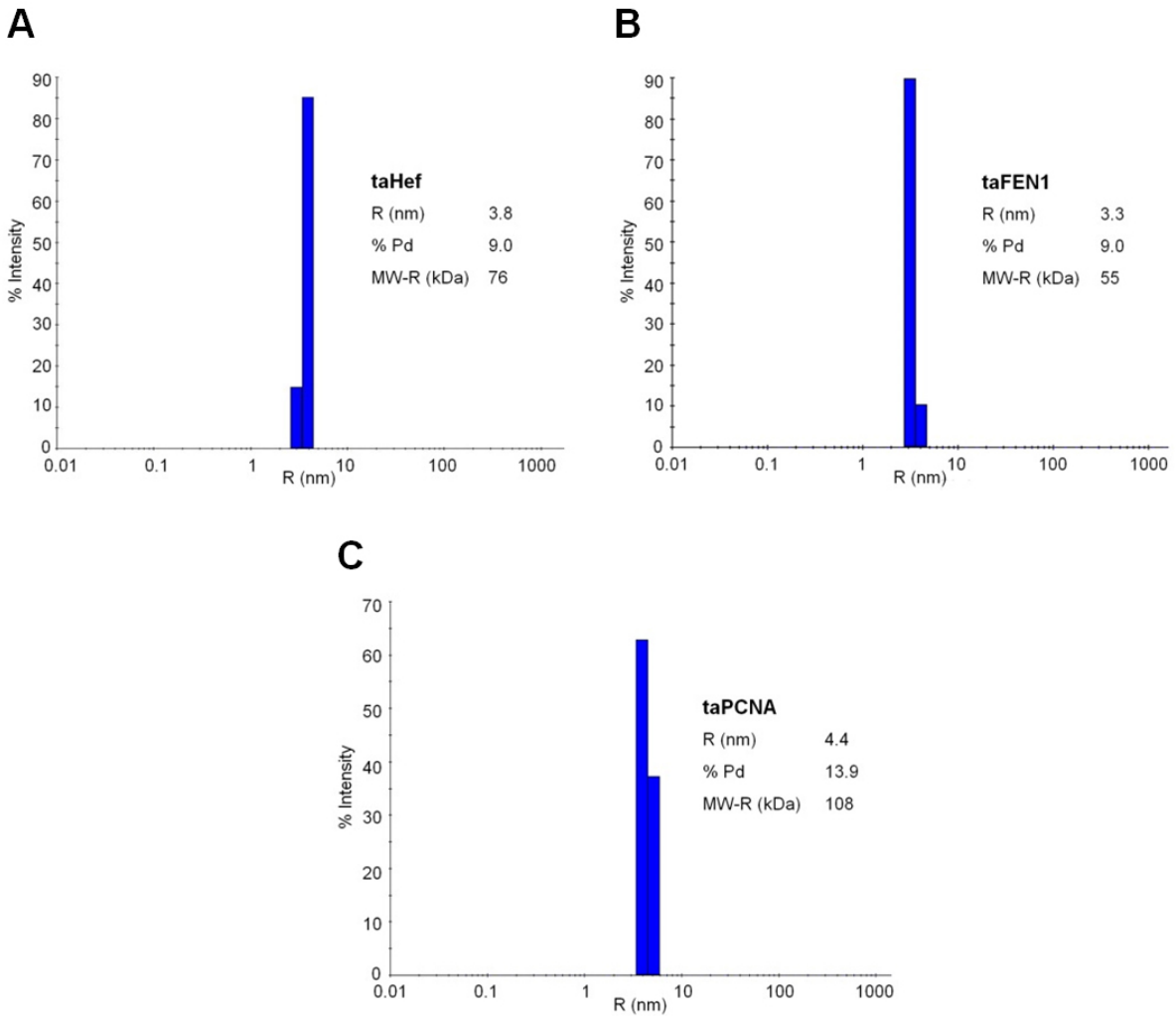


Figure 4.11: Dynamic light scattering analysis of taHef, taFEN1, and taPCNA. Dynamic light scattering (DLS) measurements for recombinantly expressed and purified *Thermoplasma acidophilum* proteins indicate mono-disperse solutions. **(A)** taHef is present as a monomer as indicated by the calculated MW of 76 kDa (actual MW = 61 kDa) and a polydispersity of 9 % with an apparent hydrodynamic radius (R) of 3.8 nm. **(B)** taFEN1 is present as a monomer as indicated by the calculated MW of 55 kDa (actual MW = 42 kDa) and a polydispersity of 9 % with an apparent hydrodynamic radius (R) of 3.3 nm. **(C)** taPCNA is present as a homotrimer as indicated by the calculated MW of 108 kDa (actual MW = 93 kDa) and a polydispersity of 14 % with an apparent hydrodynamic radius (R) of 4.4 nm.

ctMph1

Initial expression and His-trap enrichment tests of ctMph1 revealed that soluble protein could only be expressed in *E. coli* RIL cells compared to the Rosetta and Arctic Express cell strains. Strikingly, while the small-scale test expression and purification predominantly resulted in the soluble full-length (FL) ctMph1 with the correct size of 137 kDa including the N-terminal His_6-tag, ctMph1 displayed extensive degradation during large-scale purification attempts (Figure 4.12). Due to the rough but also very fast protocol of the test purification utilizing a simplified His-trap enrichment in addition to full-length protein only the typical impurities were obtained. However, when switching to large-scale expression and purification ctMph1 displayed massive degradation (Deg) already during His-trap enrichment and after the final size exclusion chromatography step no full-length protein was present at levels which could be detected by Coomassie staining. Instead the ctMph1 protein degraded to stable fragments of sizes between 80 and 110 kDa. These results indicate that initially ctMph1 is correctly expressed by *E. coli* in its full-length form but during the time-consuming large-scale His-trap and SEC purification steps the protein is degraded to multiple stable truncated fragments.

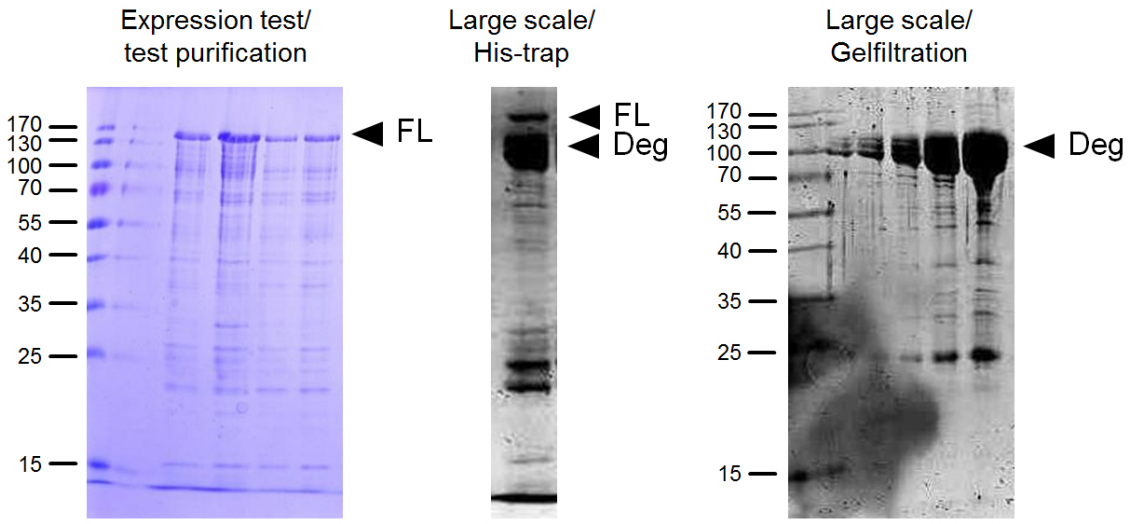


Figure 4.12: Degradation of purified ctMph1.
15 % SDS-PAGE (Coomassie-stained) analysis of purified ctMph1 which was recombinantly expressed in *E. coli* RIL cells. A fast 5 ml test expression with subsequent His-trap enrichment of ctMph1 showed predominantly full-length (FL) protein coherent with impurities that are typical for this rough procedure. In contrast, during large-scale expression and the two days purification procedure ctMph1 massively degraded (Deg) to a stable product.

However, a 15 l expression culture yielded a 27 g cell pellet and about 1 mg of degraded ctMph1 could finally be obtained. Further analyses of the ctMph1 degradation product via mass spectrometry (MS) and the resulting generation of a new expression construct that resembles the truncated variant is depicted in Chapter 4.12.

ctPCNA

ctPCNA expression in a 15 l culture in LB medium yielded a 24 g cell pellet and after purification (Figure 4.13) a total amount of about 350 mg of pure protein could be obtained which was concentrated to 70 mg/ml (2.2 mM). Notably, the SEC elution profile (green box in Figure 4.13 C) does not exhibit a single peak but rather an overlay of multiple peaks indicating that the ctPCNA homotrimer engages an equilibrium of multiple oligomeric species. However, these various species could not be separated and therefore the entire peak was pooled.

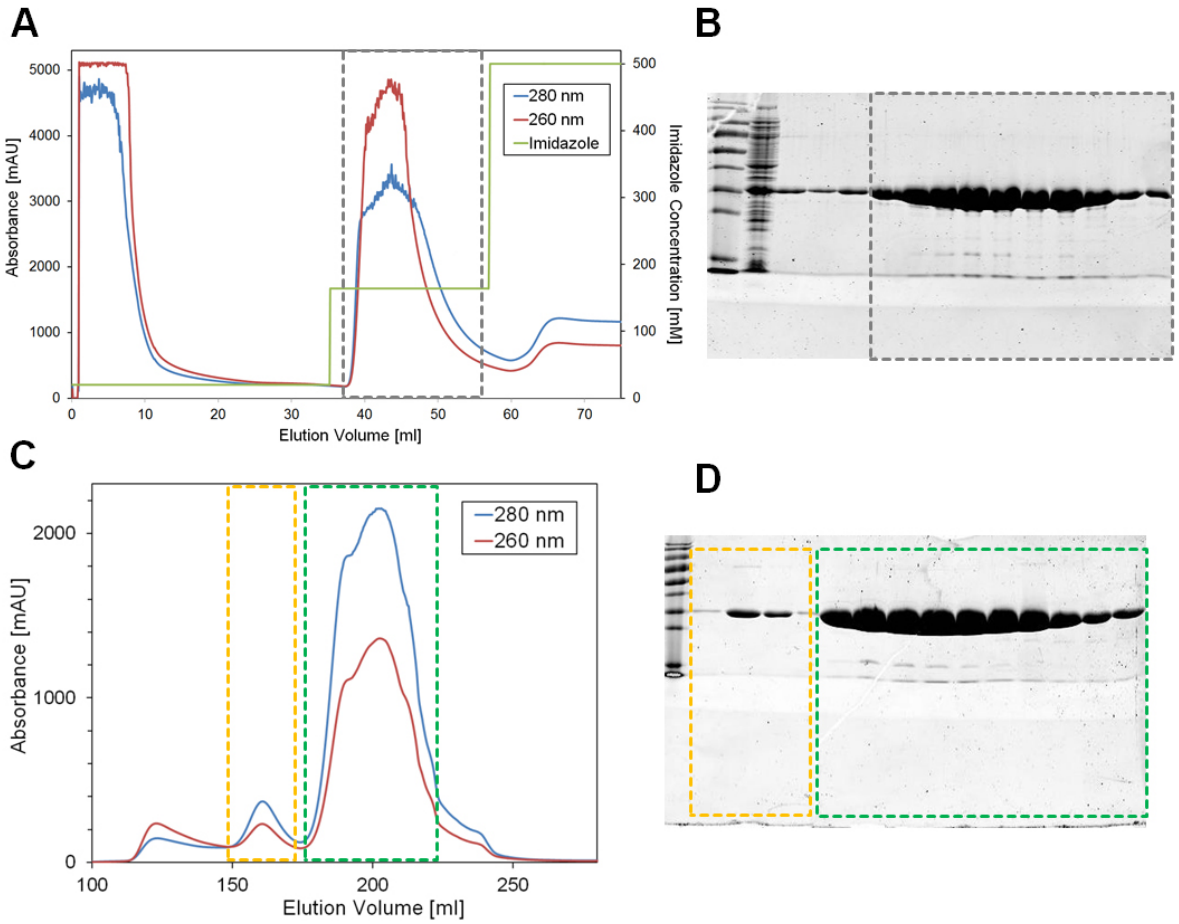


Figure 4.13: Purification of ctPCNA.
(A) Chromatogram of washing and elution of ctPCNA from the HisTrap FF 5 ml column (GE Healthcare). **(B)** SDS-PAGE analysis of the protein content after the His-trap purification. **(C)** Chromatogram of the SEC using a Superdex 200 26/60 column (GE Healthcare). **(D)** SDS-PAGE analysis of the protein content after the SEC purification. Dashed boxes indicate the fractions in the corresponding gels. Final fractions are highlighted by the green dashed box.

hsFANCM isoform 2

Full-length hsFANCM consists of 2048 amino acids with a corresponding MW of 232 kDa. Unfortunately, it is not possible to recombinantly express intact full-length hsFANCM in *E. coli* and thus it was tried to express and purify the hsFANCM isoform 2 (hsFANCMiso2) which has a MW of 76 kDa. hsFANCMiso2 comprises the helicase core of the protein which is equivalent to taHef plus the N-terminal extension and a stop codon at position 670. Expressed from the pETM-11 vector hsFANCMiso2 including the N-terminal His_6-tag would comprise a MW of 79 kDa. A purification approach yielded hsFANCMiso2 mainly in the insoluble form and small amounts in an impure higher oligomeric fraction (red arrows in Figures 4.14 B & D).

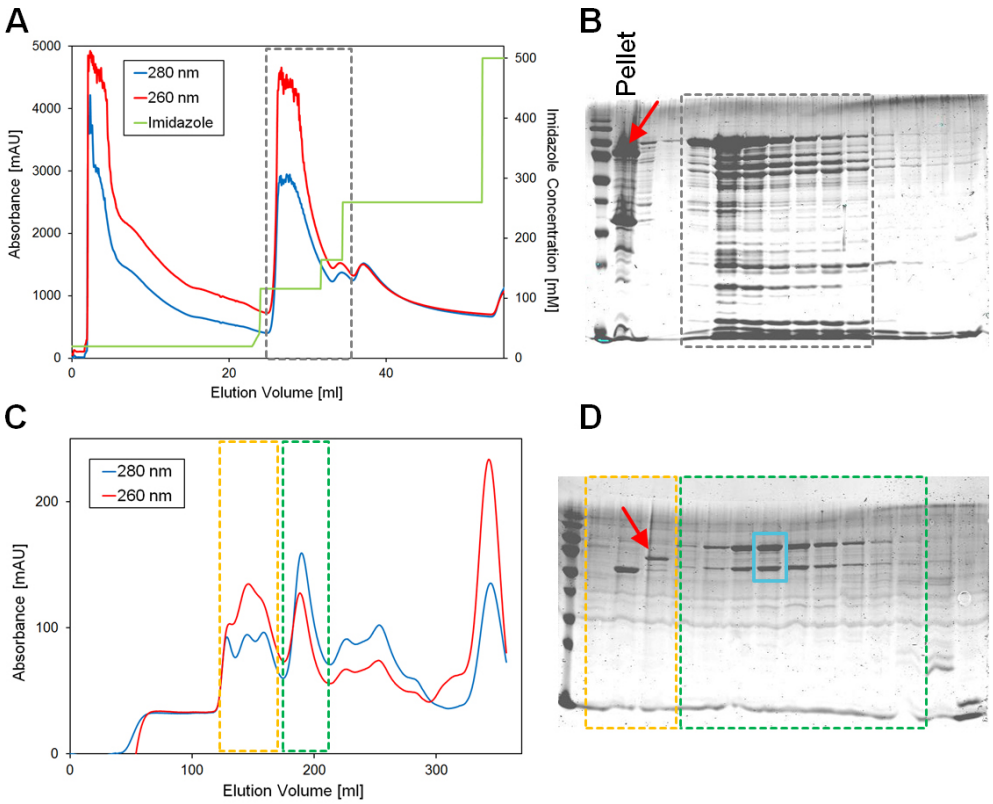


Figure 4.14: Purification approach of hsFANCMiso2 from pETM-11.
(A) Chromatogram of washing and elution of hsFANCMiso2 from the NiMAC 1ml column (Novagen). **(B)** SDS-PAGE analysis of the protein content after the His-trap purification. **(C)** Chromatogram of the SEC using a Superdex 200 26/60 column (GE Healthcare). **(D)** SDS-PAGE analysis of the protein content after the SEC purification. Red arrows indicate putative hsFANCMiso2 bands. Dashed boxes indicate the fractions in the corresponding gels. Final fractions are highlighted by the green dashed box. The blue box marks protein bands that were analyzed by MS.

MS analysis of the final purification revealed endogenous *E. coli* proteins (blue box in Figure 4.14 D (pyruvate dehydrogenase E1 component, 100 kDa, and

glucosamine–fructose-6-phosphate aminotransferase, 67 kDa) rather than the expected hsFANCMiso2 protein.

hsPCNA

hsPCNA expression in a 15 l culture in LB medium yielded a 54 g cell pellet and after purification (Figure 4.15) a total amount of about 100 mg of pure protein could be obtained which was concentrated to 10 mg/ml (300 mM). Remarkably, after HisTrap purification there was still a huge amount of soluble hsPCNA in the flow-through but a second purification procedure was not possible due to aggregation. This aggregation is also visible in the SEC profile (yellow box in Figure 4.15 C) and could be explained by an intolerance towards imidazole because the purified hsPCNA after SEC for which no imidazole is present in the buffer no aggregation was observed.

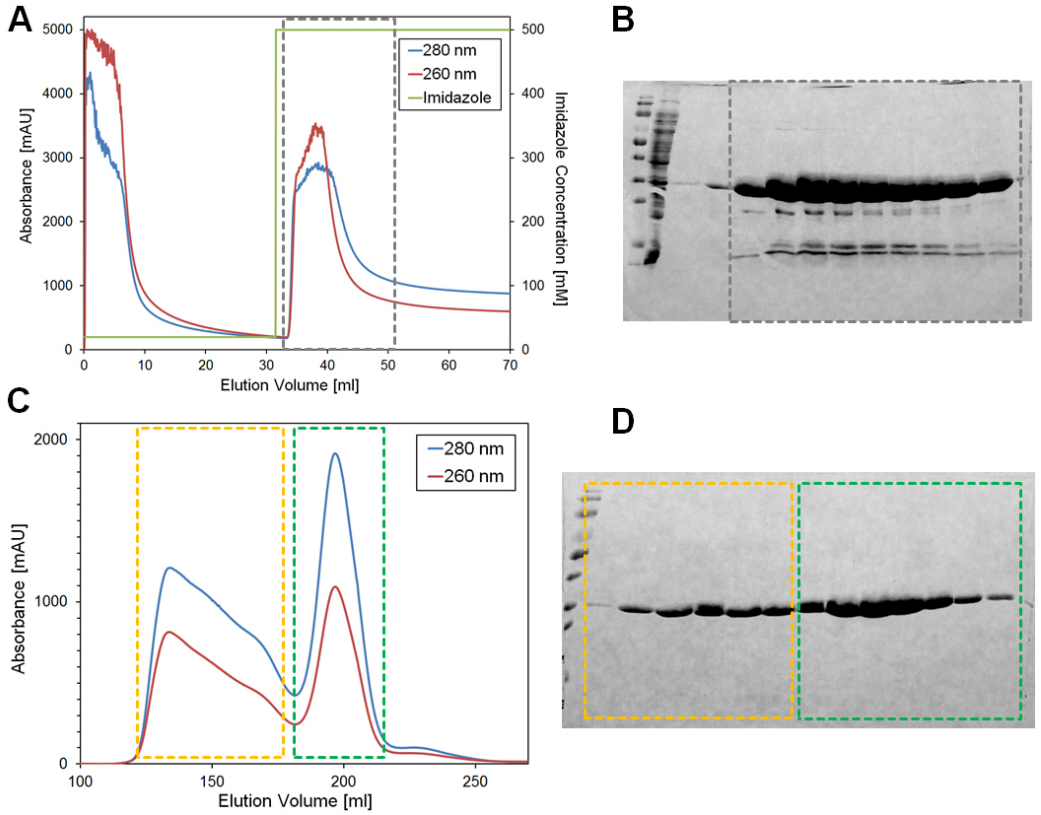


Figure 4.15: Purification of hsPCNA.
(A) Chromatogram of washing and elution of hsPCNA from the NiMAC 1ml column (Novagen). **(B)** SDS-PAGE analysis of the protein content after the His-trap purification. **(C)** Chromatogram of the SEC using a Superdex 200 26/60 column (GE Healthcare). **(D)** SDS-PAGE analysis after the protein content of the SEC purification. Dashed boxes indicate the fractions in the corresponding gels. Final fractions are highlighted by the green dashed box.

hsFEN1

hsFEN1 expression in a 10 l culture in LB medium yielded 24 g cell pellet. In a first two-step His-trap and SEC purification procedure hsFEN1 was heavily degraded and mostly lost and thus the protocol was changed to a one-step approach utilizing only His-trap enrichment (Figure 4.16).

Due to the high imidazole concentration in the His-trap elution buffer, it was not possible to determine the precise protein concentration utilizing the absorption at 280 nm. However, the fraction with a roughly estimated protein concentration of ~ 5 µM in a volume of 5 ml (red arrow in Figure 4.16 B) was used for ASEC experiments (Chapter 4.14).

The double-band in the SDS-PAGE still indicates degradation of hsFEN1 and there were also additional impurities present but this preparation was sufficient for its implementation into interaction studies with hsPCNA.

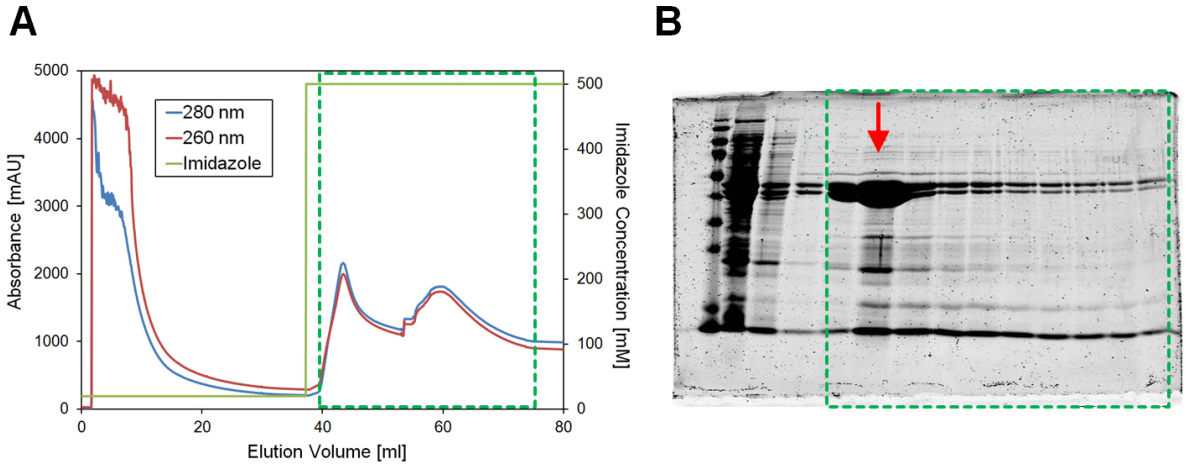


Figure 4.16: Purification of hsFEN1.
(A) Chromatogram of washing and elution of hsFEN1 from the HisTrap FF 5 ml column (GE Healthcare). **(B)** SDS-PAGE analysis of the protein content after the His-trap purification. Dashed boxes indicate the fractions in the corresponding gels that were pooled and stored. The fraction marked by the red arrow was used for ASEC experiments with hsPCNA (Figures 4.74 D & 4.75).

Circular dichroism (CD) spectra of taPCNA, ctPCNA, and hsPCNA indicate correct folding as well as the relatively similar content of α-helices and β-sheets in the three proteins (Figures 4.17).

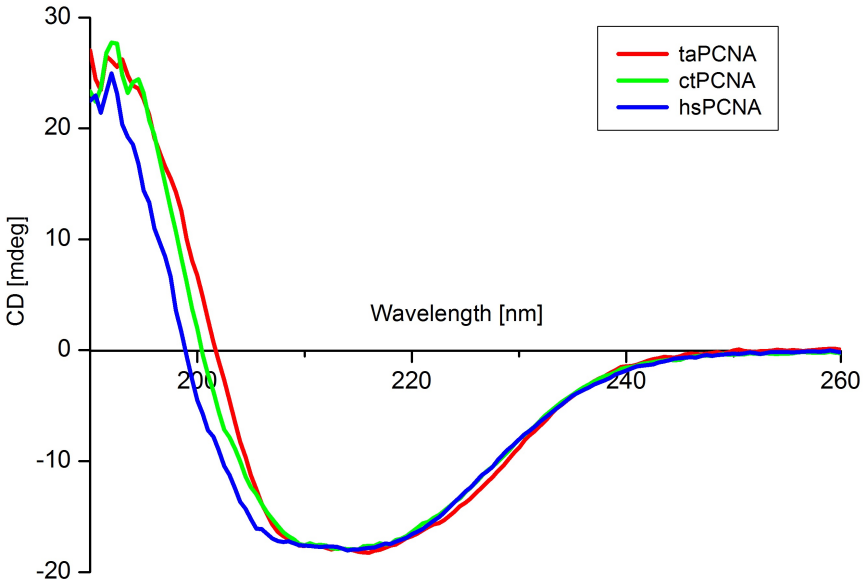


Figure 4.17: CD Spectra of PCNA from Different Organisms.
Circular dichroism (CD) spectra of PCNA from *Thermoplasma acidophilum*, *Chaetomium thermophilum*, and *Homo sapiens* indicate correct protein folding as well as a similar relative α-helix and β-sheet content.

4.4 Crystal Structure of taHef

4.4.1 Crystallization, Data Collection, and Refinement

Crystals of recombinantly expressed and purified taHef were grown by vapor diffusion in sitting drops containing equal volumes of 0.3 µl protein and reservoir solution equilibrated against 40 µl reservoir solution. taHef was used at a concentration of 43 mg/ml (740 µM) and was dissolved in 20 mM Tris/HCl, 500 mM NaCl, pH 7.5. Setups were performed automatically using the Zinsser Honeybee liquid handling system. Initially, the standard crystallization screens (Table 2.15) were tested and crystals grew under various PEG-containing conditions within one week at 20 °C. Notably, all protein crystals revealed a similar needle-like and clustered shape with the same space group and unit cell dimensions (Figure 4.18 A). To separate single crystals the clusters were carefully broken apart and individual needles extracted. Prior to x-ray diffraction testing and data collection, the crystals were transferred into a cryo-protective solution containing the reservoir solution supplemented with 20% PEG 400. The crystals were flash cooled in liquid nitrogen and data collection was performed at 100 K. Glycerol and MPD were also tested as cryo-protectants but PEG 400 led to deliver the best diffraction results. taHef crystals of a maximum size of 200 x 10 x 10 μm^3 were obtained from the crystallization solution consisting of 0.1 M sodium acetate, 0.2 M ammonium acetate, 20 % PEG 4000, pH 5.0 and diffracted to 2.43 Å (Figure 4.18 B).

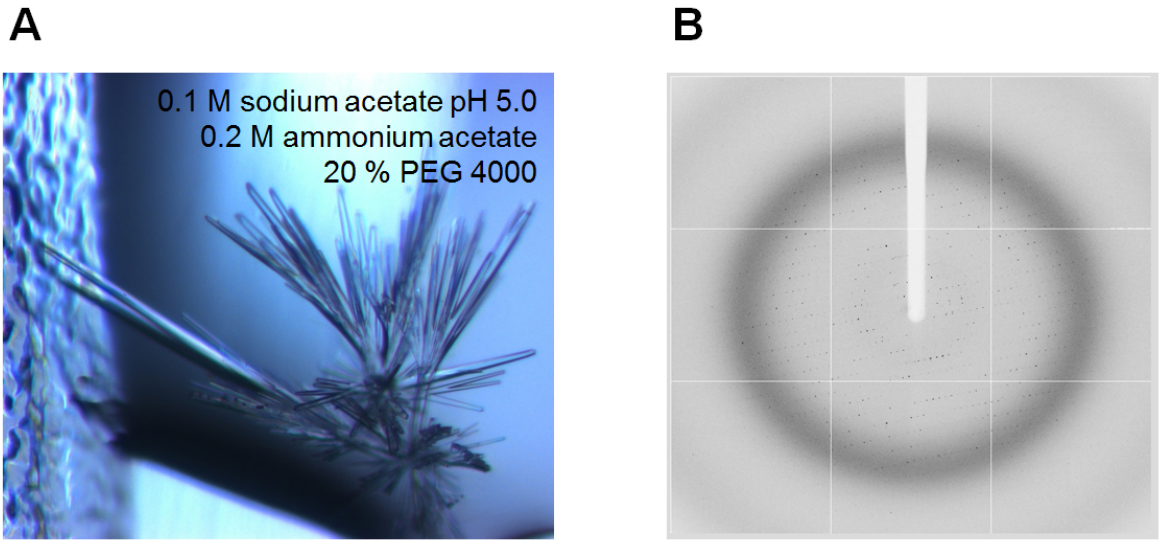


Figure 4.18: taHef crystals and exemplary diffraction image.
(A) Crystals of taHef (43 mg/ml; 740 µM) that were grown by the sitting drop vapor diffusion method. 0.3 µl protein solution combined with 0.3 µl reservoir solution were equilibrated against 40 µl reservoir solution and revealed crystals of 200 x 10 x 10 μm^3 after one week. **(B)** One exemplary diffraction image of taHef that was taken at beamline ID23.1 (ESRF) indicating a resolution of 2.43 Å.

To obtain high resolution data of apo taHef, the crystals were objected to synchrotron radiation and diffraction data were collected at beamline ID23.1 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at a wavelength of 0.916 Å. The data were indexed and processed using the programs Mosflm and Scala, respectively (Leslie and Powell, 2007; Evans, 2006). The crystals belonged to the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 40.7 Å, b = 90.5 Å, c = 177.8 Å, $\alpha = \beta = \gamma = 90°$ and contained one molecule per asymmetric unit. Remarkably, $P2_12_12_1$ is the most frequent space group in protein crsytals and occurs in about one-third of known protein crystal structures (Wukovitz and Yeates, 1995). Structure solution was achieved by molecular replacement (MR) using single domains of Hef from *Pyrococcus furiosus* (pfHef; PDB code: 1WP9) as search models and the programs Molrep and Phaser (Nishino et al., 2005b; Vagin and Teplyakov, 1997; McCoy et al., 2007). The model was built by alternate rounds of manual model building in Coot and automated refinement using Refmac5 and phenix.refine (Emsley et al., 2010; Adams et al., 2010; Murshudov et al., 1997). Refinement was carried out against the highest resolution data set of up to 2.43 Å yielding a final R factor of 18.5 % (R_{free} = 23.3 %). Structure validation was performed using MolProbity (Chen et al., 2010). The data and refinement statistics indicate reliable and good-quality data (Table 4.3).

Table 4.2: Data collection and refinement statistics taHef.

Data set	
Space group	$P2_12_12_1$
Wavelength (Å)	0.916
Cell dimensions	
a, b, c (Å)	40.7, 90.5, 177.8
α, β, γ (deg)	90, 90, 90
Resolution (Å)	49.5-2.43 (2.56-2.43)
Completeness (%)	99.5 (97.2)
R_{sym}[a] / R_{pim}[b]	0.142 (0.594) / 0.057 (0.260)
Mean $I/\sigma I$[c]	8.7 (3.1)
Multiplicity	6.8 (5.7)
No. of unique reflections	25702
Refinement	
Resolution (Å)	49.5-2.43
R_{cryst} / R_{free}[d]	0.185 / 0.233
No. of working / test reflections	24339 / 1305
No. of protein / solvent atoms	3833 / 186
Coordinate error	0.33
Wilson B-factor (Å^2)	34.27
Overall average B-factor (Å^2)	47.76
Rms[e] deviations from ideal values	
Bond length (Å)	0.002
Bond angles (deg)	0.616
Dihedral angles (deg)	12.321
Planar groups (Å)	0.002
Ramachandran statistics[f] (%)	98.7 / 1.1 / 0.2

Values in parentheses refer to the highest resolution shell.

[a] $R_{sym} = \Sigma_{hkl}\Sigma_i|I_i - I|/\Sigma_{hkl}\Sigma_i I$ where I_i is the ith measurement and I is the weighted mean of all measurements of I.

[b] $R_{pim} = \Sigma_{hkl}(1/N-1)^{\frac{1}{2}}\Sigma_i|I_i - I|/\Sigma_{hkl}\Sigma_i I$ where I_i is the ith measurement, I is the weighted mean of all measurements of I and N is the redundancy of the data.

[c] $I/\sigma I$ indicates the average of the intensity divided by its standard deviation.

[d] $R_{cryst} = \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is the same as R_{cryst} for 5% of the data randomly omitted from the refinement.

[e] Root mean square.

[f] Ramachandran statistics indicate the fraction of residues in the favoured, allowed and outlier regions of the Ramachandran diagram as defined by MolProbity (Chen et al., 2010).

In contrast to the taHef structure with a resolution of 2.43 Å (R factor: 18.5 %; R_{free}: 23.3 %), the resolution of the previously known pfHef$^{1-494}$ structure is only 2.9 Å with an R factor of 25.7 % (R_{free} = 28.6 %) (Nishino et al., 2005b). Another difference between the two crystal structures lies in the Ramachandran statistics that function as an additional quality indicator. While the taHef structure was refined to 98.7 % favored and only 1.1 % allowed (0.2 % outliers) residues, pfHef$^{1-494}$ exhibits 90.8 % favored and 9.2 % allowed residues. Thus comparing the data and refinement statistics of the previously known pfHef$^{1-494}$ and taHef crystal structures, it can be concluded that due to the higher resolution and better refinement statistics the taHef structure exhibits a model with an improved quality as compared to the previously available data.

4.4.2 Structure Analysis

The taHef crystal structure shows that the protein can be divided into three distinct domains: two characteristical RecA-like helicase motordomains (Hel1 and Hel2) that contain the seven SF2-typical helicase motifs (Caruthers and McKay, 2002) and the thumb-like insertion domain (Ins). The Ins domain is integrated in Hel2 and the C-terminal helix α21 folds back onto the N-terminal domain Hel1. Furthermore, the very C-terminal 17 amino acids (residues 492-508) which also contain the PIP-box motif are flexible and thus not visible in the crystal structure (Figure 4.19).

In Hel1 (residues 1-193; red) seven central parallel β-strands are flanked by nine α-helices and four 3_{10}-helices. This domain contains the SF2 helicase motifs I-III. Interestingly, the very C-terminal helix α21 (residues 478-490; orange) is part of Hel1 and is integrated between α7 and β7 to complete this domain.

The second helicase motor domain Hel2 (residues 194-203, 329-475; blue) consists of six parallel β-strands surrounded by five α-helices and one 3_{10}-helix, thereby assuming the second typical RecA-like fold including helicase motifs IV-VI with the thumb-like domain inserted between residues 203 and 329. A loop comprising 10 residues between β10 and α18 (residues 382-391) is disordered and may engage a defined conformation upon DNA binding. Remarkably, this loop is also disordered in the pfHef structure (Nishino et al., 2005b). Furthermore, a short loop of three amino acids (residues 448-450) at the downstream end of helicase motif VI is unstructured. Hel2 also contains two short stretches that resemble putative PIP motif sequences (1: residues 395-402, 2: residues 454-460). However, both are located within secondary structure elements (1: α18, 2: β13) and are not solvent accessible. Functional PIP boxes are commonly located in unstructured areas of the interacting protein and only upon binding to a hydrophobic pocket in PCNA a 3_{10}-helix is formed (Sakurai et al., 2005). A canonical PIP box containing the sequence ^{502}QKTIFDF^{508} is present at the disordered C-terminus of taHef.

An illustration of the 2D topology of taHef displays its secondary structure composition and also portrays the protein's segmentation into the three distinct domains

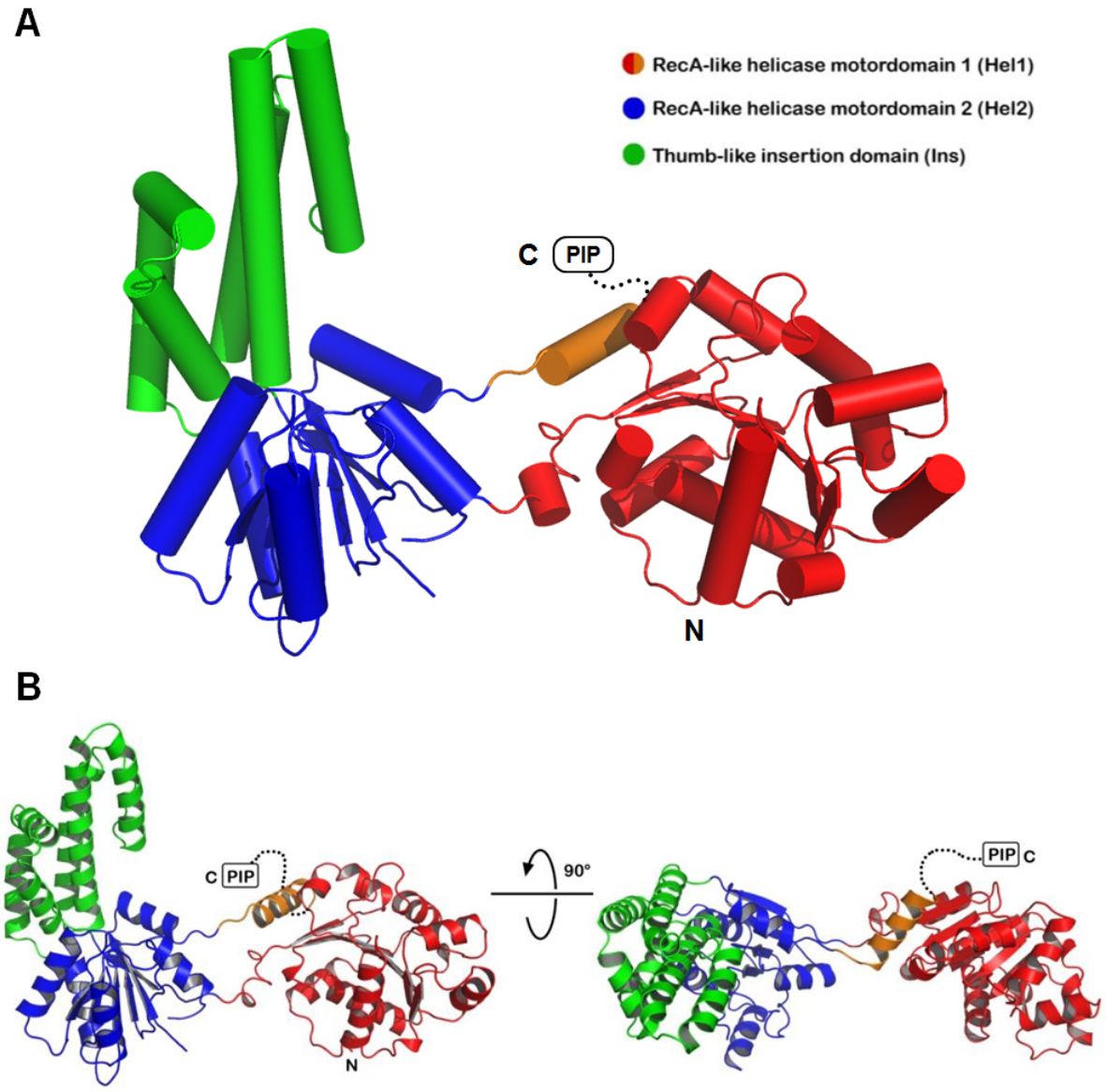


Figure 4.19: Crystal structure of taHef.
(A) Hel1 (red / orange) and Hel2 (blue) form the RecA-like motordomains characteristic for SF2 helicases. The thumb-like insertion domain (Ins, green) is integrated within Hel2 between β8 and α16. Helix α21 (orange) folds back onto Hel1 and is integrated between α7 and β7. The flexible C-terminus (residues 492-508), not visible in the crystal structure, is represented by the dotted line and comprises the PIP-box (residues 502-508, labeled with "PIP"). N- and C-termini are indicated. **(B)** Rotation of the protein by 90° reveals its elongated conformation.

(Figure 4.20). Another interesting feature in addition to the C-terminal helix α21 which folds back to domain Hel1 is the bridging β-strand β8. β8 is the only secondary structure element in Hel2 upstream to the integrated Ins domain. Together with the downstream helix α16 it positions the Ins domain on the top of Hel2. Since the Ins domain is inserted between the two Hel domains the two-way hinge between Hel1 and Hel2 is thus constructed by α9/β8 and α20/α21.

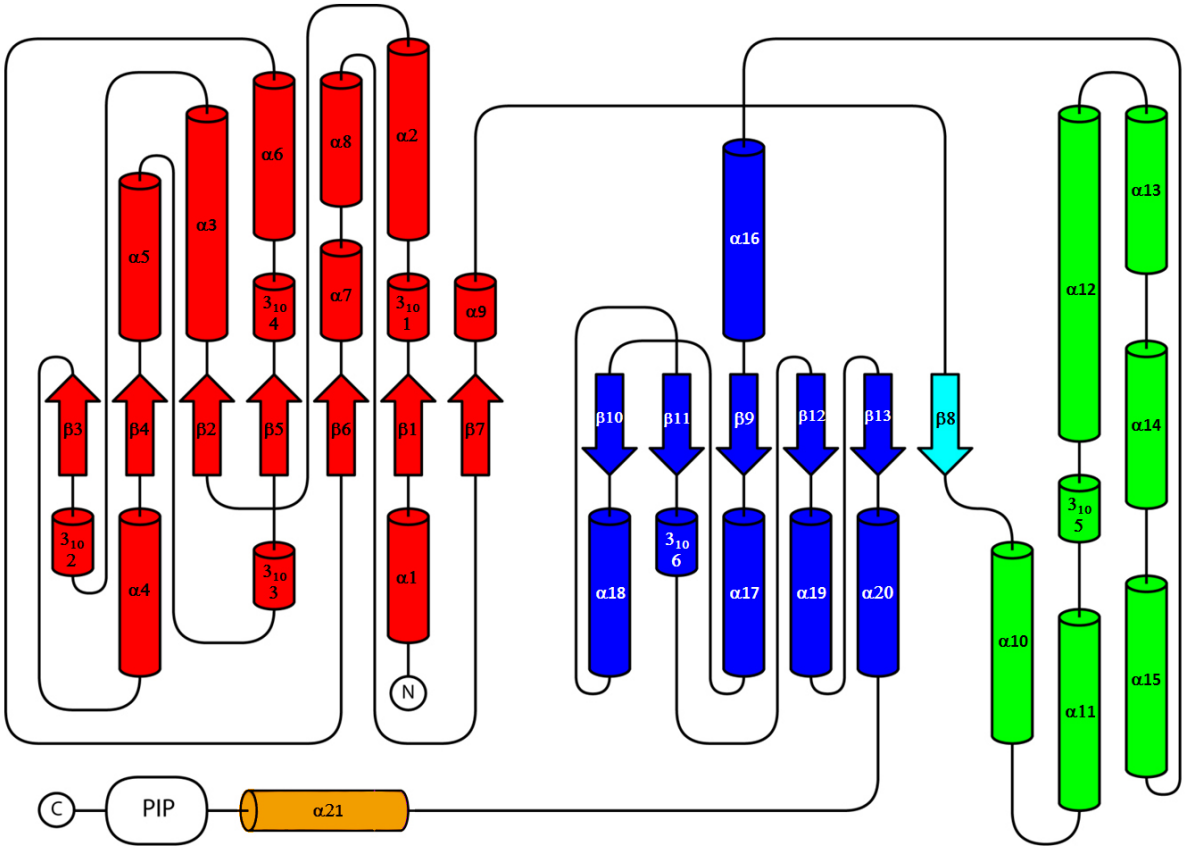


Figure 4.20: 2D topology of taHef.
Schematic representation of the taHef 2D topology. Tubes symbolize $\alpha/3_{10}$-helices and arrows represent β-strands. Secondary structure elements are labeled and N-/C-termini are indicated. The PIP-box within the unstructured C-terminus is also depicted. The color code (Hel1: red; Hel2: blue; thumb-like insertion domain (Ins): green; helix α21: orange; bridging β-strand 8: cyan) is based on Figure 4.19.

Furthermore, a 360° rotation in 90° steps illustrates the overall shape of taHef that resembles a left hand in the "front view" (Figure 4.21). The thumb-like domain (Ins) literally stands for the thumb while the Hel2 symbolizes palm and the aligned four residual fingers are represented by Hel1. Interestingly, taHef exhibits an elongated conformation which resembles an open conformation of this helicase. Presumably, the two-way hinge between Hel1 and Hel2 provides some flexibility to permit the formation of a closed conformation as symbolized by a closed hand to grab the DNA (Chapter 4.6).

This conformational mobility is assumed to be crucial for the ATP-dependent worm-like movement of helicases/translocases along the DNA (Lohman et al., 2008). Moreover, the open conformation of taHef stands in stark contrast to the closed conformation of pfHef (Chapter 4.6).

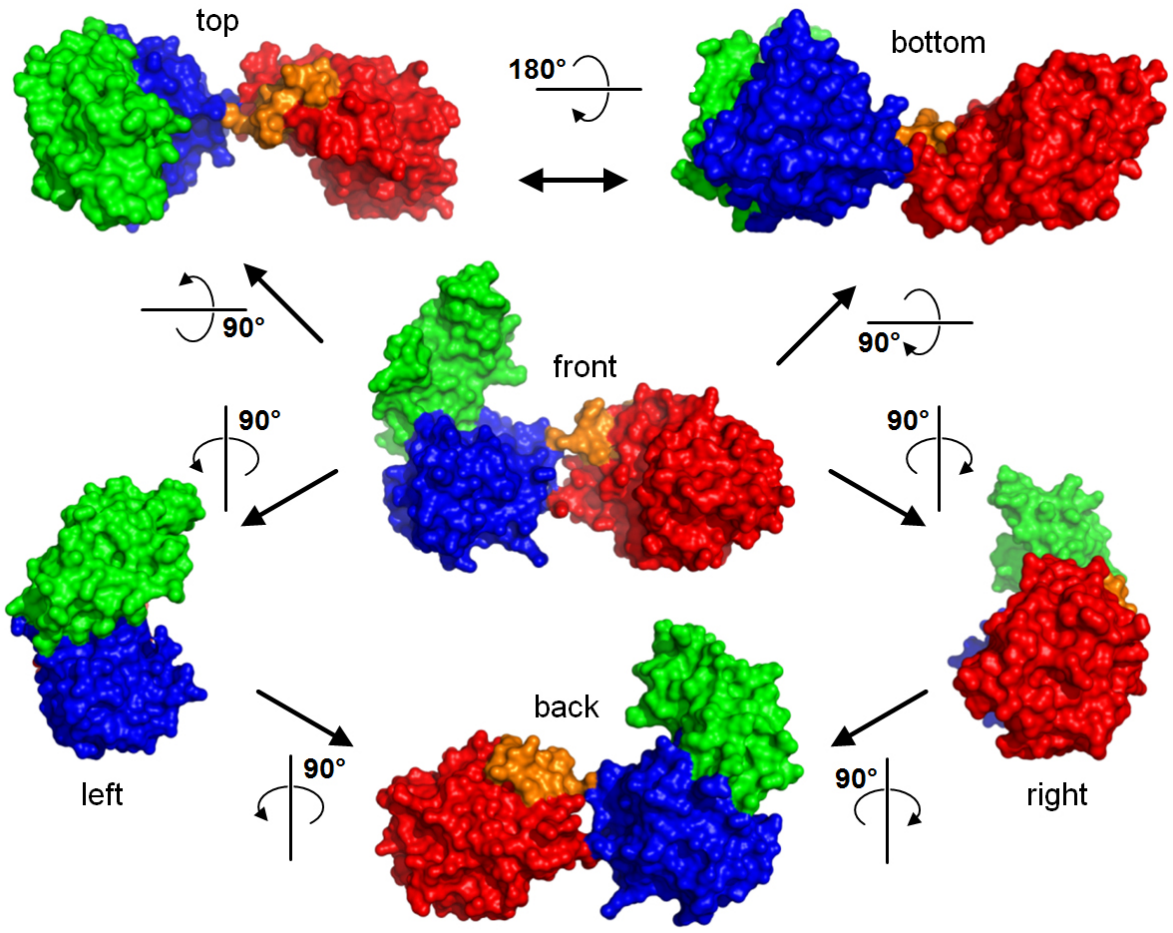


Figure 4.21: taHef in various orientations.
Surface representation of taHef rotated by 90° in all possible directions. Rotation operators are indicated by arrows and the domains Hel1 (red), Hel2 (blue), Ins (green) as well as the C-terminal helix α21 (orange) are colored according to Figure 4.19. Designations of the respective orientations are indicated.

Visualization of the electrostatic surface potential of taHef reveals extensive positively charged areas at two positions of the protein: (1) in the interspace between Hel1 and Hel2 and (2) on the insertion domain (Figure 4.22). These sections of taHef might thus be involved in promoting DNA interaction. Interestingly, an inchworm-like mechanism for the movement of helicases along the DNA was proposed. The translocation process thereby involves conformational changes of the two RecA-like helicase motor domains embracing the DNA coherent with ATP hydrolysis (Patel and Donmez, 2006). Importantly, this theory is impressively supported by structural analyses of

the two helicases Rep and PcrA both in complex with DNA (PDB accession codes: 1UAA and 3PJR, respectively) (Korolev et al., 1997; Velankar et al., 1999). In both examples single-stranded DNA is bound between the two helicase domains triggering an open/closed conformational switch. Remarkably, taHef also contains a positively charged patch between Hel1 and Hel2. Furthermore, DNA binding was confirmed (Chapters 4.6 & 4.7) and it can be assumed that the flexibility of these domains permits the above described movements. Hence, it is valid to speculate that taHef displays a similar behavior in sliding along the DNA.

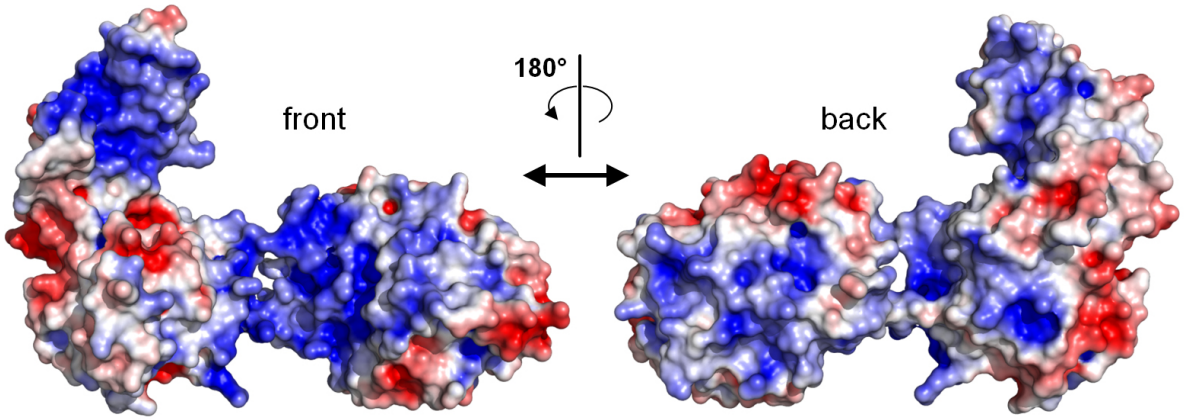


Figure 4.22: Electrostatic surface potential of taHef.
The surface potential has been calculated with PyMOL/APBS at an ionic strength of 150 mM and is contoured at ± 4 kBT. Positively charged areas are indicated in blue and negatively charged areas in red. The interspace between Hel1 and Hel2 as well as the surface of the insertion domain exhibit extensive positively charged surface potentials indicating their feasible role in DNA binding.

Insertion domains determine the function of helicases and can be integrated at various positions within the two motor domains (Caruthers and McKay, 2002). The α-helical thumb-like insertion domain of taHef (green) is arranged between β8 and α16 of Hel2 and is responsible for Hef substrate specificity. Thumb-like domains are conserved among various DNA processing enzymes such as polymerases or HIV-1 reverse transcriptase (Steitz, 1999).

Similar to the thumb domain of DNA polymerases, the structurally conserved insertion domain of Hef promotes specificity for branched DNA structures (Nishino et al., 2005b; Li et al., 1998). Thus the 6 α-helices (α10-α15) of the thumb-like insertion domain of taHef expose 9 arginine as well as 9 lysine residues to the surface that are likely to be involved in DNA interaction and may contribute to the specificity for fork-structured DNA substrates (Figure 4.23). This arrangement of arginines and lysines creates two channels that merge at the interspace between Hel1 and Hel2. One channel is located on the front side and the other on the back side of the Ins domain. So far it remains unclear which part exactly and how it interacts with the DNA substrates.

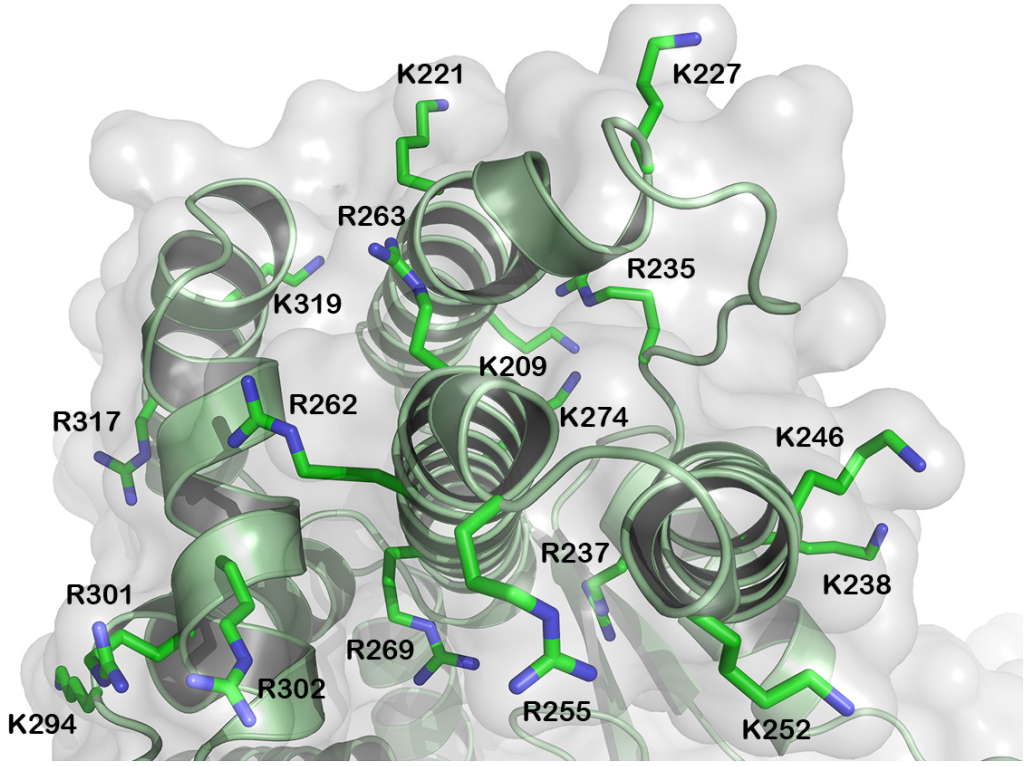


Figure 4.23: Top view of the thumb-like insertion domain (Ins) of taHef.
Surface-exposed positively charged arginine and lysine side chains of the Ins domain are likely to mediate DNA interactions and substrate specificity.

Although the authors of the pfHef crystal structure conclude from their biochemical data that residues K248/K249 and K302/K303 which are located in the Ins domain in helices α11 and α13, respectively, play important roles in substrate recognition, the effect of the mutations at these two sites varies (Nishino et al., 2005b). While the K248E/K249E double reversion results in a strongly reduced Holliday junction (HJ) unwinding and a complete inhibition of fork unwinding, the data on the K302E/K303E variant are more complex. This variant has no effect on HJ unwinding compared to the WT protein but it influences pfHef's capability on fork unwinding. Interestingly, the open fork intermediate is not visible during fork dissociation which is very prominent in the WT and positive control (Q9R mutation) experiments. However, the K302E/K303E variant also disassembles the fork substrate to its basic ssDNA components similarly to WT and the positive control. Thus the actual role of pfHef residues K302/K303 in DNA processivity still remains elusive. Notably, the deletion of the entire Ins domain of pfHef results in a total loss of fork unwinding activity but retains its ability to dissolve HJ DNA comparably to the WT protein. Hence, residues K248/K249 are very interesting because their charge reversal has a bigger impact on protein functionality than deletion of the entire domain.

These particular residues, i.e. pfHef K248/K249 and K302/K303, were therefore compared to their taHef counterparts (Figure 4.24). Only the positively charged residues K248/K249 are conserved in taHef (R237/K238) while K302/K303 are surprisingly reversed to the negatively charged E290/E291 in taHef. This observation in combination with previous data (Nishino et al., 2005b) leads to the conclusion that residues K302/K303 (pfHef) and E290/E291 (taHef) play only a minor role in DNA interaction. On the other hand the K302E/K303E double-reversion in pfHef displays a phenotype and thus further experiments are necessary to elucidate the exact function of these residues. The conserved residues K248/K249 (pfHef) and accordingly R237/K238 (taHef) protrude from the Ins domain and point into the interspace between Hel1 and Hel1. Interestingly, in hsFANCM there is only one lysine (K331) conserved at this position (Figure 4.2) and a mutational analysis of this particular amino acid regarding DNA binding and unwinding would be very interesting. At the position of K302/K303 (pfHef) there are cysteine and glycine residues (C387/G388) present in hsFANCM supporting a rather neglectable character of these side chains regarding DNA binding.

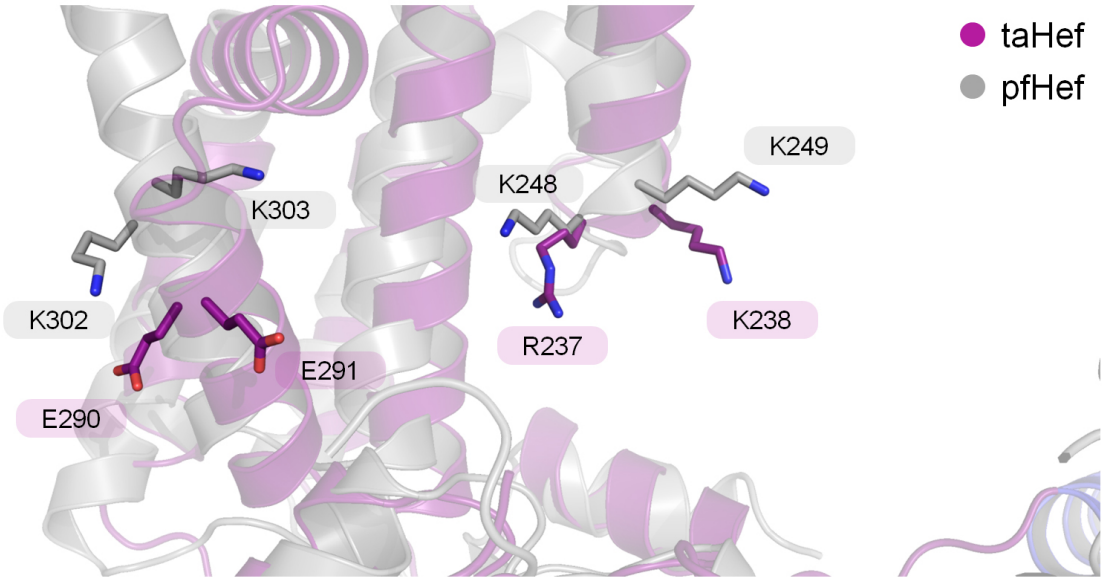


Figure 4.24: Superposition of taHef and pfHef Ins domains.
Comparison of residues within the Ins domain that were identified to be involved in fork and HJ unwinding by pfHef (Nishino et al., 2005b). taHef (magenta) residues R237/K238 are located at the same position in helix α11 as pfHef (grey) K248/K249 and thus may resemble a similar function. In contrast to the positively charged residues K302/K303 in pfHef, taHef comprises the negatively charged residues E290/E291 at this position in helix α13 and there are no other basic amino acids present in close proximity.

Besides positively charged amino acid side chains that primarily interact with the negatively charged sugar-phosphate backbone of the DNA often aromatic side chains also play an important role in mediating protein-DNA interactions. Thereby the aromates intercalate and form π-stacking interactions with the nucleobases of the DNA

(Anjana et al., 2012). For this purpose the taHef Ins domain was analyzed for its aromatic amino acids content (Figure 4.25).

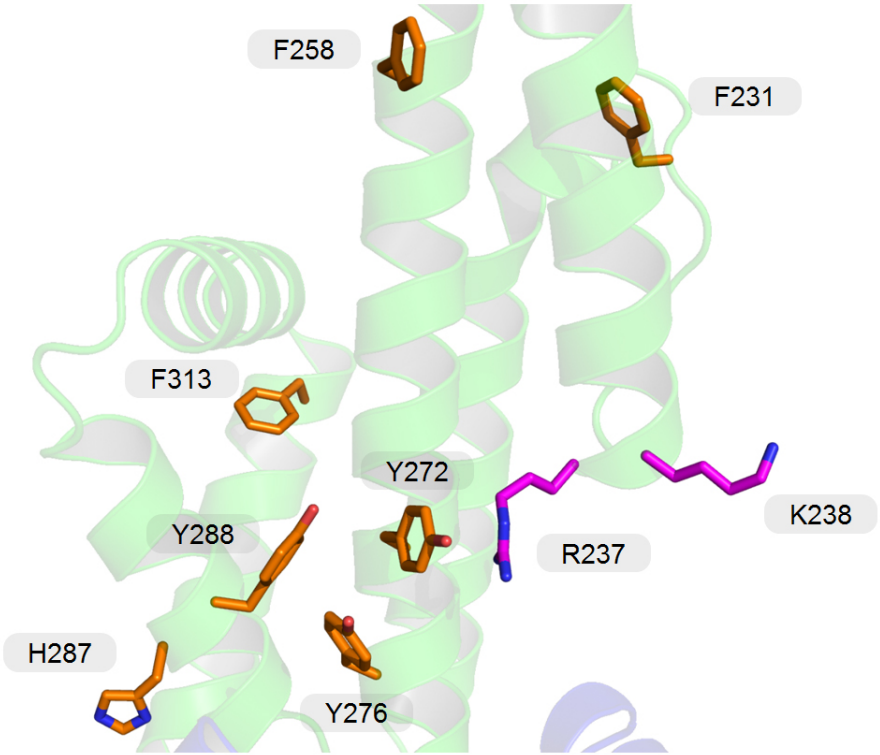


Figure 4.25: Aromatic side chains within the taHef Ins domain.
Detailed view of the taHef Ins domain (green) with all aromatic residues highlighted (orange). Residues R237/K238 are depicted as an orientation reference point (magenta). Aromatoc residues might be involved in promoting DNA interaction and thus could contribute to ICL damage detection.

The taHef Ins domain comprises seven aromatic amino acids but only two positions are conserved among its homologs. Firstly, the aromatic character of Y272 is preserved in pfHef and hsFANCM as histidines. In contrast, hsRIG-I contains an asparagine at this position instead. Interestingly, one residue upstream in taHef, where taHef contains an aspartate (D271), pfHef, hsFANCM, and hsRIG-I comprise a histidine and tyrosines that contribute to a generally aromatic-rich area. Secondly and even more strikingly, Y288 represents the only fully conserved aromate. In hsFANCM a phenylalanine is located at this position instead which is N-terminally flanked by another phenylalanine and an additional tyrosine. In contrast to other aromatic residues such as phenylalanine or tryptophan, tyrosine binds strongly to DNA and is found in many protein-DNA interfaces (Anjana et al., 2012). Thus it can be concluded that the aromatic residues at the position of Y288 might play an important role in DNA interaction.

4.5 Hinge Region and Helix α21

An important difference between the crystal structures of taHef and pfHef is the diverging conformational state of the two proteins. While the previously known structure of pfHef exhibits the compact and closed state, taHef represents the elongated and open conformation. A superposition of both structures with the aligned Hel2 and Ins domains in the top view clearly shows the relative movement of Hel1 (Figure 4.26). This movement leads to a shift of Hel1 by 61° as represented by the angle between G113 (taHef) and G120 (pfHef) to K193 (taHef) (Figure 4.31). Based on the pfHef structure *in silico* calculations were performed to predict conformational changes as well as flexibility and the results showed only a minor movement towards an even more closed conformation (Flechsig et al., 2011).

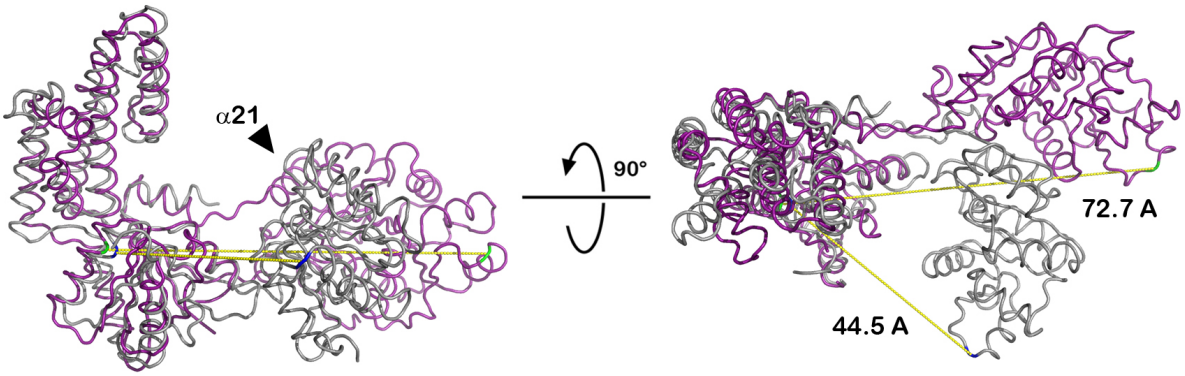


Figure 4.26: Experimentally determined and predicted conformational change. The distance between residues D85 and S359 of taHef is 72.7 Å and between corresponding residues S92 and A375 of pfHef 44.5 Å highlighting the conformational differences. taHef comprises the open and pfHef the closed state. The C-terminal helix α21 of taHef is indicated by an arrow.

The authors explained this theoretical behavior by the ATP binding and hydrolysis site between Hel1 and Hel2. As an indicator for this conformational change the distance between residues S92 (Hel1) and A375 (Hel2) was used. The widest predicted distances reached from 37.9 Å to 44.5 Å with the pfHef crystal structure representing the presumably open state. Interestingly, the taHef crystal structure clearly contradicts these *in silico* predictions. In taHef the distance between the corresponding residues D85 (Hel1) and S359 (Hel2) is increased to 72.7 Å (Figure 4.26). In contrast to the predicted movement of 6.6 Å towards the more compact state, taHef thus further extends towards the opposite and elongated direction by 28.2 Å.

A very interesting feature of the taHef structure is the presence of helix α21 that complements Hel1 and thereby builds the second component of the two-way hinge between Hel1 and Hel2. This α-helix is not present in the pfHef structure. For the pfHef crystallization and x-ray diffraction experiments different constructs were tested

(Nishino et al., 2005b). A pfHef variant that comprises the N-terminal 546 amino acids and thereby containing helix α21 yielded protein crystals that unfortunately diffracted only to 6 Å. Thus pfHef was further truncated to the N-terminal 494 amino acids which resulted in the final 2.9 Å structure lacking helix α21. Remarkably, the authors state that pfHef$^{1-494}$ exhibits ten-fold decreased helicase and ATPase activities compared to pfHef$^{1-546}$ while the helicase motifs and the substrate specificity were not affected.

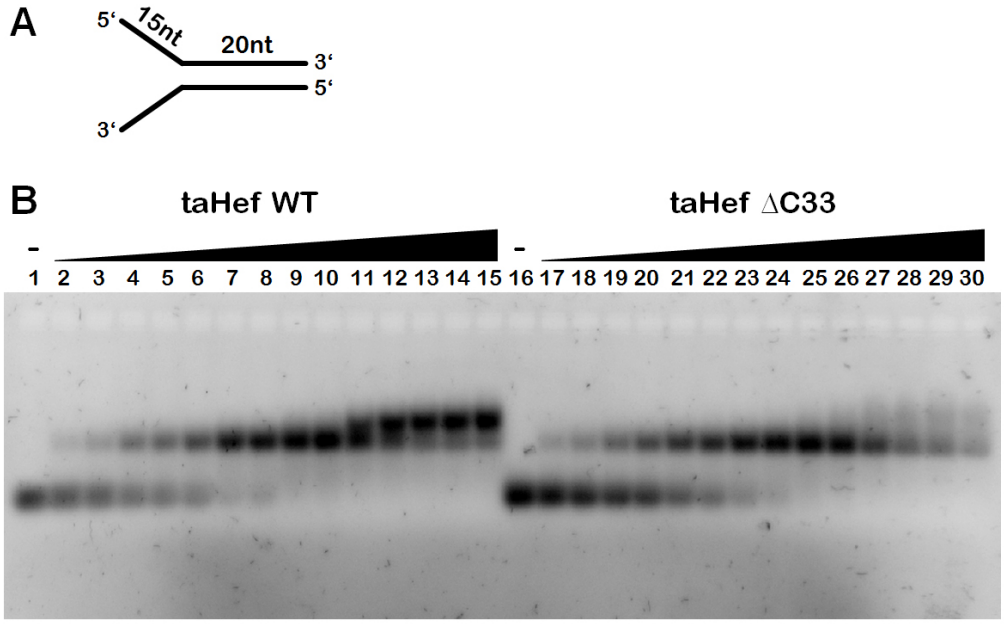


Figure 4.27: Influence of helix α21 on DNA binding.
(A) The open fork 15/20 (OF15/20) DNA substrate is assembled by two 35 nucleotides (nt) long DNA single-strands. 20 nt are complementary and constitute the double-stranded stem while 15 nt cannot align and produce single-stranded Y-arms. **(B)** Native gel electrophoresis on a 1% agarose gel with Midori-Green staining to visualize DNA while the protein is not visualized by this technique. 5 µM DNA were incubated with varying protein concentrations ranging from 1 to 10 µM (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 8, 9, 10 µM). The first lanes (1 and 16) represent the negative controls showing the DNA in the absence of the taHef protein. Binding reactions were conducted in DNA-binding buffer at RT and a volume of 20 µl for 30 min prior to running the samples on the gel in TAE-buffer. taHef WT produces a new band occurring at protein concentrations from 6-10 µM (lanes 11-15). At the same concentrations of the taHef ΔC33 variant (lanes 26-30) the DNA disappears presumably due to a nuclease contamination.

The taHef ΔC33 variant lacks the C-terminal 33 amino acids and thus resembles the pfHef$^{1-494}$ construct with K475 being the last amino acid of the protein. This variant was initially generated for co-crystallization experiments of taHef in complex with DNA (Chapter 4.7.2) because I assumed that helix α21 was responsible for the conformational difference between the taHef and pfHef crystal structures. I thus speculated that the deletion of this particular helix could force taHef in a likewise

closed conformation to facilitate DNA binding. The structural characterization of the taHef ΔC33 variant is shown in Chapter 4.6.

A gel-shift experiment displays the DNA binding behavior of taHef WT and the ΔC33 variant (Figure 4.27). As a DNA substrate the open fork 15/20 was used which comprises a 20 nucleotides double-stranded stem and 15 nucleotides single-stranded Y-arms. Interestingly, both proteins are capable to bind the OF15/20 substrate but at higher concentrations or rather a protein : DNA ratio > 1 the WT protein produces an additional band on the gel which is not generated by ΔC33. In this experiment there was no ATP present and thus the new band cannot be explained by classical ATP-dependent DNA unwinding. Strikingly, the ΔC33 variant binds to the substrate with a comparable affinity. At a first glance taHef WT and the ΔC33 variant seem to differ in their DNA binding behavior at high protein concentrations (6-10 µM, lanes 11-15 and 26-30 in Figure 4.27) but remarkably the same samples display a strong reduction of the DNA signal. This indicates a nuclease contamination of the taHef ΔC33 protein leading to the disappearance of the DNA. This leads to the conclusion that helix α21 is not involved in the initial DNA binding event. It may rather play a role in the downstream processing of the DNA. The exact function and mechanistic details of helix α21 need to be addressed in future experiments.

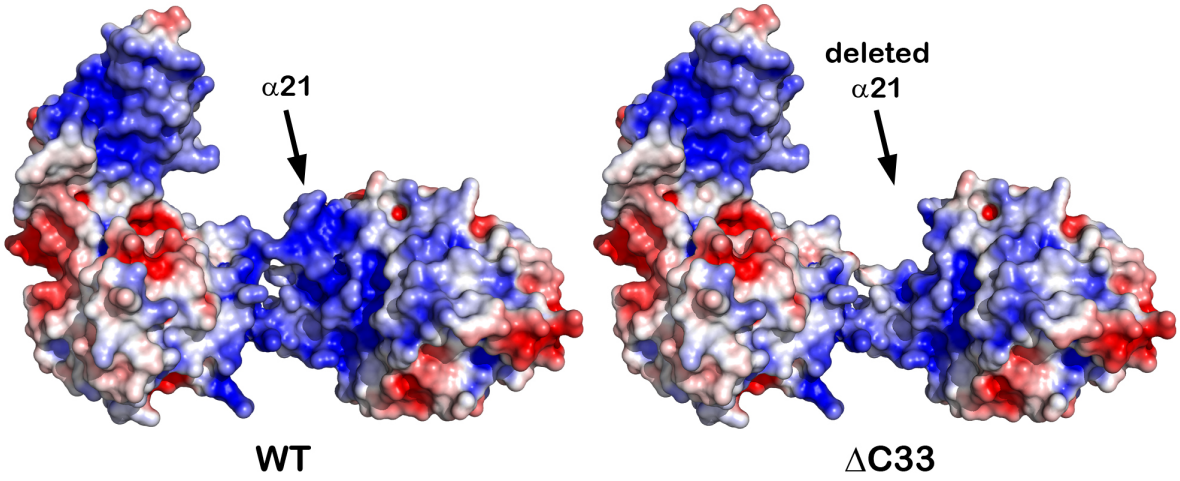


Figure 4.28: Contribution of helix α21 to the electrostatic surface potential. The surface potential has been calculated as in Figure 4.22. The arrows indicate the position of helix α21 in taHef WT (left) and the deleted area in the modeled ΔC33 variant (right), respectively. The model of the ΔC33 variant was generated with the crystal structure of the WT protein and the deletion of the C-terminal 33 residues.

Remarkably, helix α21 significantly contributes to the positively charged surface potential of the interspace between Hel1 and Hel2 which is a potential DNA binding site (Figure 4.28). This observation supports the assumption that helix α21 is involved in DNA interaction and that its loss influences Hef's ability to correctly process the substrate.

4.6 SAXS Analysis of taHef

The crystal structures of full-length taHef and the helicase domain of pfHef (PDB code: 1WP9) (Nishino et al., 2005b) reveal two extreme conformations of the Hef protein. Both the open state of taHef and the closed state of pfHef might not represent the physiological state and possibly exhibit crystallization artifacts. However, crystal structures only show snapshots of proteins which are highly dynamic and flexible molecules. Thus to gain insight into the native apo conformation of taHef and to analyze the influence of helix α21 small-angle x-ray scattering (SAXS) was utilized (Tsutakawa et al., 2007). This method allows the determination of low-resolution (1-2 nm) protein *ab initio* models in solution (Mertens and Svergun, 2010). Obviously, this resolution is insufficient to specify molecular details such as the exact orientations of amino acid side chains but it presents a very effective tool to determine overall shapes and conformations of proteins under in solution conditions.

Monodisperse solutions of taHef WT and the ΔC33 variant were objected to small angle x-ray scattering (SAXS) measurements to analyze their in solution conformations. Prior to data collection all samples were re-purified by SEC and a polydispersity below 10% was validated by DLS. Experiments were conducted at the ESRF BioSAXS beamlines ID14-eh3 and BM29 (Pernot et al., 2010, 2013) and data analysis was performed by our collaboration partner Adam Round (EMBL, Grenoble) who also wrote significant parts of the corresponding result sections (Chapters 4.6, 4.8.2, 4.9.1).

Comparison of the WT scattering (WT exp.data) to the theoretical scattering of the known high resolution structures for taHef (taHef X-tal fit) and pfHef (pfHef X-tal fit) resulted in poor fits (chi of 6.97 and 3.5, respectively) but suggested that the pfHef structure was closer to the shape under physiological conditions in solution than the taHef structure (Figure 4.29).

However, even the pfHef structure shows systematic deviations indicating the solution conformation is different to that observed in the crystal structure. The program OLIGOMER was used to probe for the presence of a mixture of both crystal forms in solution, but did not yield an improvement to the fit, suggesting that neither of the two extreme conformations is predominant in solution. As the data could not be described by either known structure or as a mixture of the two known forms, we assume that the domains are free to move and adopt a different conformation than observed in either crystal form.

For recovery of the relative domain positions in solution, a combination of *ab initio* and rigid body modeling was used utilizing the programs Dammif and CORAL, respectively. To allow movement of Hel1 within taHef, additionally to the unstructured parts that are not visible in the crystal structure (N-terminal His_6-tag: yellow spheres; C-terminal 17 amino acids including the PIP-box residues 492-508 and the loop with residues 382-391: blue spheres), we defined a flexible hinge between Hel1 and Hel2

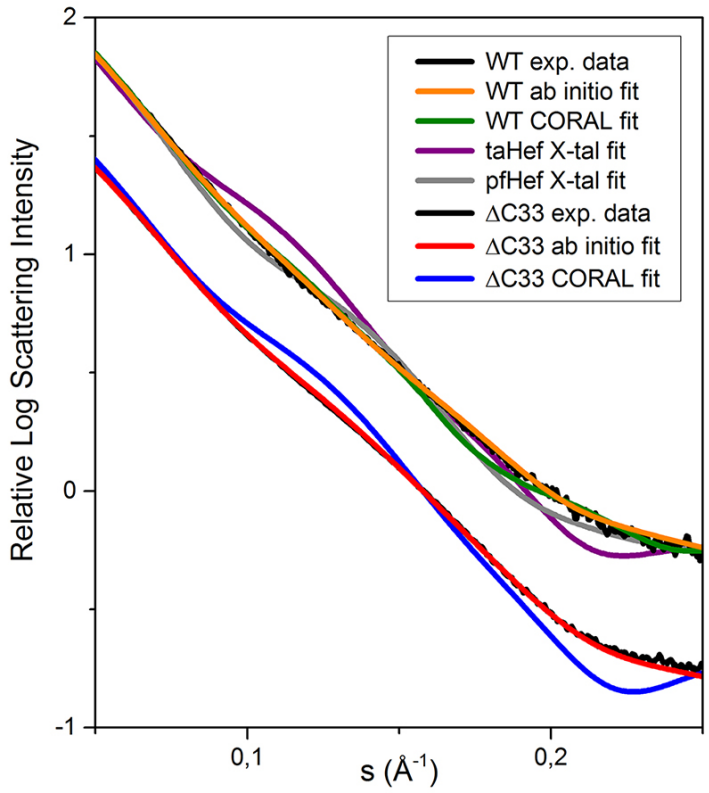


Figure 4.29: 1D SAXS intensity fits of taHef.
Top: structure derived from theoretical SAXS intensity curves for taHef (magenta) and the pfHef helicase domain (grey) as well as fits for the *ab initio* model (brown) and the corresponding adjusted taHef model by CORAL (green) overlaid with the experimental data (black). Bottom: ΔC33 taHef experimental data (black dots) with overlaid theoretical scattering of the WT rigid body CORAL model for which the C-terminal 33 residues were removed (blue) and the optimized ΔC33 *ab initio* model (red).

(residues 192-195 and 471-476: red spheres) (Figure 4.30). The resulting models yielded reasonable fits to the data with chi values of 1.8 and 2.07 for the *ab initio* (WT ab initio fit) and rigid body (WT CORAL fit) models, respectively. Furthermore, a good correspondence between the resulting models was observed (Figure 4.29).

The resulting solution *ab initio* and rigid body models of taHef WT (Figure 4.30) reveal an intermediate conformation between the open and closed crystal structures (Figure 4.31) and suggest that, although the helicase domains are able to reach the conformations observed in the crystals, an intermediate position seems to be more likely in solution. Even though the data suggest some flexibility of the domain positions as shown by the slightly larger *ab initio* envelope, the close correspondence suggests that the average positions are reasonably well described by the proposed models.

As discussed in Chapter 4.5 a major difference between the taHef and the pfHef structure is the presence of the C-terminal helix α21 (blue in Figure 4.31) in

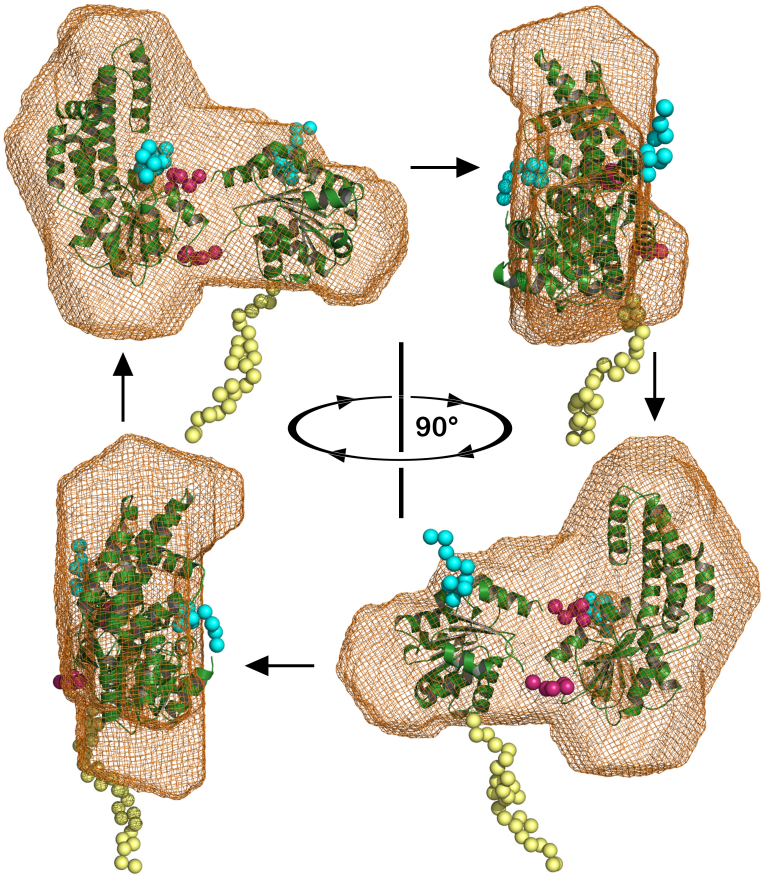


Figure 4.30: taHef SAXS in solution model.
The two taHef WT SAXS models (*ab initio* model as orange mesh; rigid body CORAL model as green cartoon) are rotated horizontally in 90° steps. The defined flexible hinge between Hel1 and Hel2 (residues 192-195 and 471-476) is depicted as red spheres; the disordered loop (residues 382-391) and the likewise invisible C-terminal 17 amino acids including the PIP-box linked to α21 (residues 492-508) are indicated as blue spheres. The disordered N-terminal His_6-tag is highlighted by yellow spheres.

taHef, which is not present in the pfHef structure. It is therefore tempting to speculate that this helix may be responsible for the open conformation in taHef which differs from pfHef by a 61° shift of Hel1. Broad motional dynamics are an important property of helicases and crucial for translocation along the DNA (Lohman et al., 2008).

Comparison of taHef WT with the ΔC33 variant, which lacks the C-terminal helix α21, reveals an overall similar shaped scattering profile but ΔC33 displays a smaller size with a radius of gyration (Rg) of 31.6 Å compared to 32.4 Å for the WT protein. Comparison of the theoretical scattering of the WT rigid body CORAL model (with the corresponding 33 residues removed) and the ΔC33 experimental data did not yield a satisfactory fit (chi 9.9). Following optimization with CORAL, using the same strategy as for the WT protein, the fit improved (chi 2.33) and the Rg of the optimized model was 30.8 Å (Figure 4.29). However, this observed difference in Rg is subtle and upon

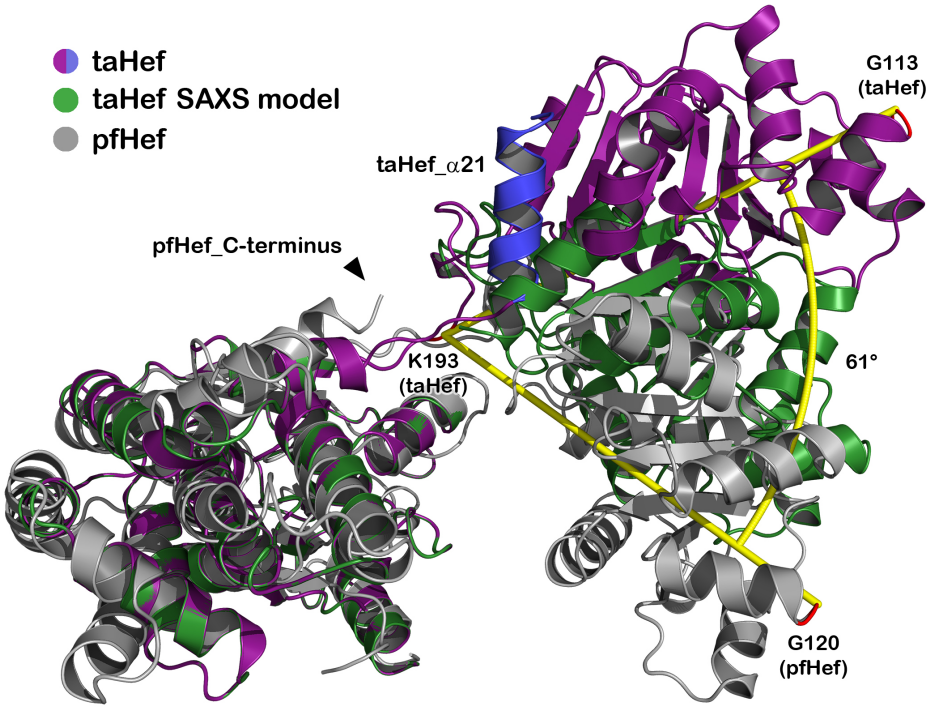


Figure 4.31: Superposition of the Hef crystal structures and the rigid body CORAL SAXS model. Superposition of full-length taHef (magenta/blue) and the helicase domain of pfHef (grey, PDB: 1WP9) (Nishino et al., 2005b) as well as the taHef rigid body SAXS model (green). The C-terminus of the pfHef helicase domain is indicated by an arrow. The C-terminal helix α21 of taHef is depicted in blue. The conformational change of Hel1 relative to the Hel2/thumb-like domain is illustrated by the angle (yellow) between taHef_G113 and pfHef_G120 to taHef_K193 (red) of 61°. The SAXS model exhibits an intermediate conformation between the open taHef and the closed pfHef crystal structures.

overlaying multiple CORAL runs of the WT protein and the ΔC33 variant, we observed some variation in the domain positions while the Rg is compensated for by the position of the linkers. Therefore it was not possible to accurately determine quantitative differences in the domain positions between WT and ΔC33 and the relative domain positions appear similar even though the ΔC33 variant seems to rotate more as might be expected with only one linker connecting the domains being present in this variant.

These results indicate that helix α21 has no significant influence on the overall conformation of taHef. Therefore the different orientations of the helicase domains as visible in the taHef and pfHef crystal structures are not induced by the presence or absence of helix α21.

Notably, taHef WT was also analyzed in the presence of a 20 nucleotides long ssDNA (fg4) and an open fork (OF15/20) DNA substrate in combination with the non-hydrolyzable ATP analogon ATPγS. While it was not possible to obtain good quality and meaningful data on the taHef-DNA complexes due to aggregation and too

low monodispersity of the samples, the addition of of ATPγS in the absence of DNA did not modify the structural properties of taHef. The ATP binding and hydrolysis site is constructed by helicase motifs 1 and 2 (Walker A and B) which coordinate the nucleotide between Hel1 and Hel2 and couple ATP hydrolysis to DNA translocation. The observation from the SAXS experiment is in agreement with an isothermal titration calorimetry (ITC) experiment that did not indicate any ATP binding by taHef in the absence of DNA (Figure 4.32).

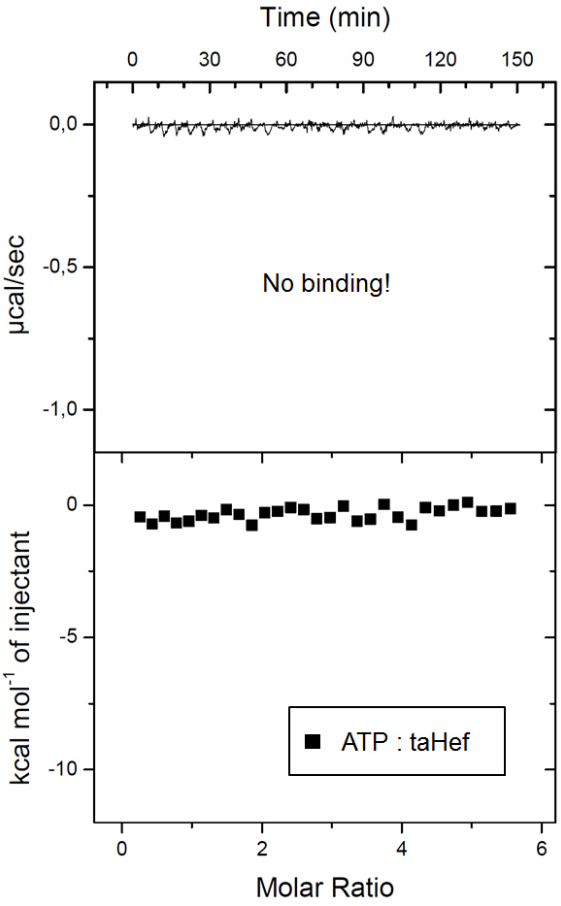


Figure 4.32: No ATP binding by taHef in the absence of DNA.
Isothermal titration calorimetry (ITC) measurements of taHef and ATP in standard buffer (20 mM Tris/HCl, 200 mM NaCl, 5 mM $MgCl_2$, pH 7.5) using a VP-ITC (MicroCal). 300 µl ATP at a concentration of 500 µM were injected in 30 x 10 µl steps into a 1.4 ml taHef solution at a concentration of 20 µM resulting in a final ATP:taHef ratio of 5.4. No heat release and thus no binding could be detected.

In contrast, DNA binding by taHef does not require ATP (Chapter 4.7) indicating that the DNA is bound prior to ATP thus preventing a wasteful ATP consumption by the enzyme. Moreover, these experiments confirm that for the structural characterization of the taHef-DNA interaction the addition of ATP is not necessary. Instead a non-hydrolyzable ATP analogon such as ATPγS or ADP-BeF_3 could be used to tightly trap the complex and hinder DNA dissociation.

4.7 Interaction with DNA

To elucidate the function of taHef as a helicase and its potential role during replication, various experiments were performed to analyze its interaction with DNA. As it was known from previous studies that Hef specifically processes fork-structured DNA (Komori et al., 2004; Nishino et al., 2005b) my analysis particularly focused on several distinct substrates that resemble fork-like secondary structures as well as double-stranded and single-stranded DNA (Figure 4.33).

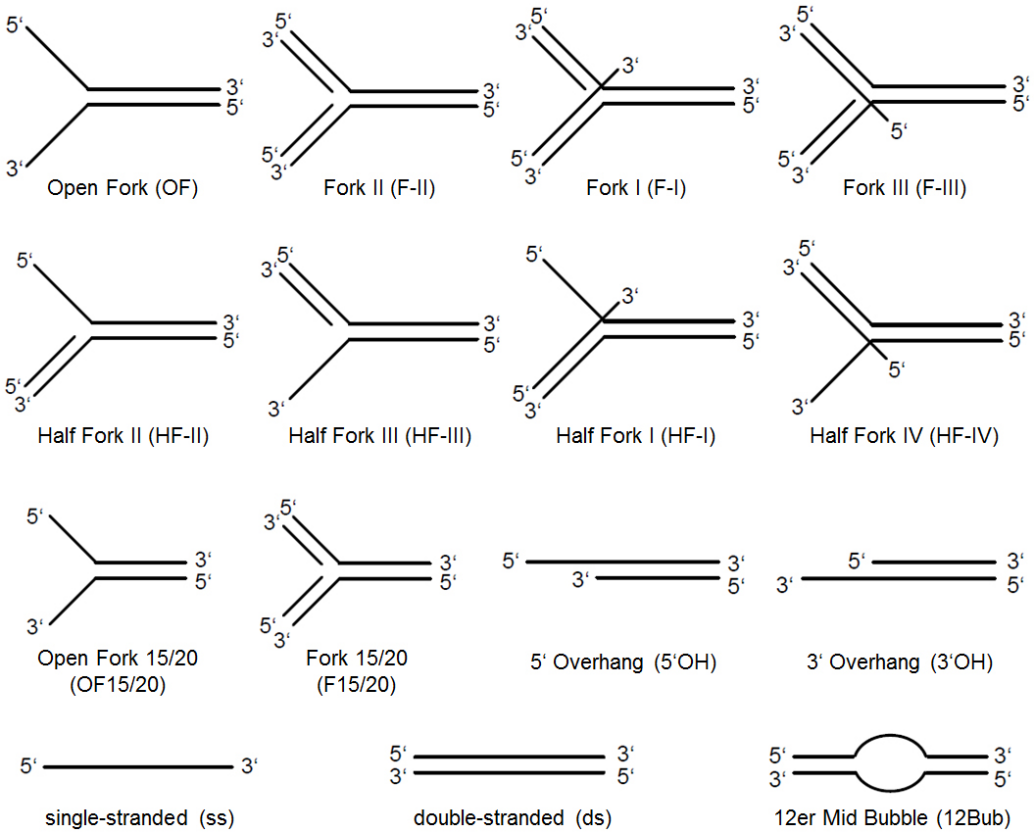


Figure 4.33: DNA substrates.
Schematic representation of the secondary structures of the annealed DNA substrates. Substrates Open Fork to Half Fork IV comprise equal lengths (double-stranded stem: 27 nucleotides; 5' Y-arm: 22 nucleotides; 3' Y-arm: 20 nucleotides; flap-extensions at branch in F.I, F-III, HF-I, HF-IV: 8 nucleotides). OF15/20 and F15/20 comprise the same sequences as OF and F-II, respectively, but were truncated at each end to generate a double-stranded stem of 20 nucleotides and equally long Y-arms of 15 nucleotides. 5'OH and 3'OH comprise a total size of 49 nucleotides while the according overhangs in both contain 15 nucleotides. If not mentioned differently, the ssDNA had a length of 49 nucleotides but also cloning primers of various lengths were used as ssDNA substrates. ds and 12Bub are 49 nucleotides long and blunt-ended and 12Bub contains a central unpaired stretch of 12 nucleotides. Sequences and compositions are provided in Tables 2.6 & 2.7, respectively. Sequences of the used single-stranded DNAs were adopted from Nishino et al. (2005b) and modified to generate additional substrates.

To generate these DNA substrates the single-strands were mixed equimolar (10 µM) in annealing buffer consisting of 0.1 x TE and 100 mM KCl. The mixture was heated up to 95 °C for 10 min and slowly cooled down to RT over night. The DNA was then stored on ice or at 4 °C for further usage. Importantly, after each DNA annealing reaction the respective batch was checked via native PAGE followed by silver staining to assure the presence of the correct products.

4.7.1 DNA Binding

Firstly, binding of taHef to different DNA substrates with various secondary structures was tested in a native gel electrophoresis experiment (Figure 4.34). The complex DNA substrates consisting of multiple DNA single-strands showed some minor annealing by-products (F-I, F-II, HF-I, HF-II, HF-III in Figure 4.34 A) but they were nevertheless used for further analysis. Interestingly, taHef binds all tested substrates which is indicated by a clear shift in the native PAGE coherent with a strong reduction of the apo DNA bands when incubated in the presence of taHef (Figure 4.34 B). The single-stranded substrate ss49 produces a more defined complex band while all others in addition to the complex display a smear on the gel.

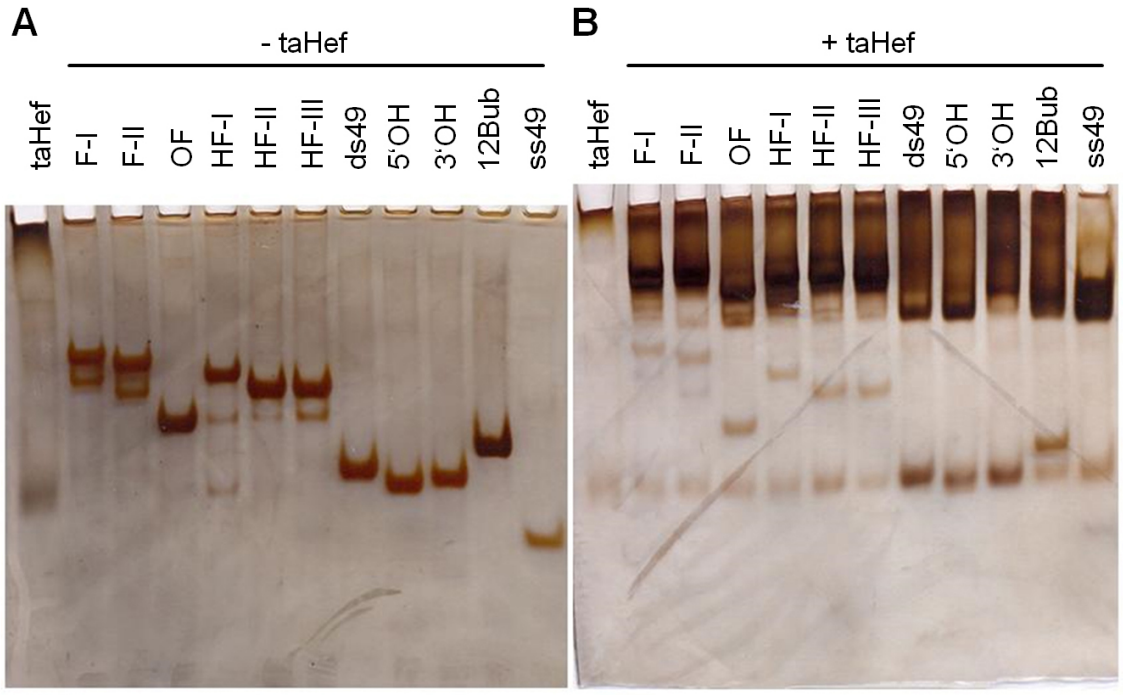


Figure 4.34: taHef binds to different DNA substrates.
Silver stained native 10 % PAGE. **(A)** Annealed DNA substrates that constitute various secondary structures as indicated in Figure 4.33 in the absence of taHef. **(B)** Incubation of the DNA substrates in the presence of taHef results in a shift of the bands indicating complex formation while signals for the DNA are strongly reduced. In the assay the following components were used: 5 µM taHef, 10 µM DNA, 10 mM ATPγS, 10 mM $MgCl_2$, annealing buffer.

Notably, ITC and biolayer interferometry measurements for an in-depth characterization of the thermodynamic parameters of the interaction between taHef (WT and variants) and various DNA substrates were inconclusive. Biolayer interferometry experiments were performed by Stefanie Wolski and a clear binding of taHef to a biotinylated ss50mer, F-I, and F-II DNAs were observed. But unfortunately, it was not possible to fit the binding curves to determine k_{on} and k_{off} for the calculation of the corresponding K_D values. Furthermore, preliminary ITC analyses showed binding of taHef WT, ΔC7, K31A, and D139K to the OF15/20 DNA substrate in 20 mM Tris/HCl, 200 mM NaCl, 5 mM $MgCl_2$, pH 7.5 with a 1:1 stoichiometry and a K_D of about 50 nM while the ΔC33 variant exhibited under these conditions a strongly reduced affinity. Unfortunately, these results were not reproducible and the heat-release signals interfered with a strong background noise that might be due to partial opening and re-annealing events of the double-stranded DNA substrate. Nevertheless, since neither ATP nor a non-hydrolyzable analogon were added in these experiments it could be concluded that ATP is not required for exclusive DNA binding.

Additionally, binding to single-stranded (ss) DNA was tested in thermofluor experiments. Due to the high melting temperature (T_M) of apo taHef of 75 °C (Figure 4.6) in the high-salt gel filtration buffer (Sdx-HS) DNA substrates composed of multiple strands could not be analyzed by this technique. It was assumed that the interaction with DNA stabilizes taHef resulting in an increased T_M of the protein.

At first the influence of the salt concentration of the buffer on apo taHef compared to taHef in the presence of a ss 47mer DNA was analyzed (Figure 4.35 A). Interestingly, salt had an opposite effect on taHef protein stability in the presence of DNA. While a boost from 20 mM to 200 mM NaCl raises the T_M of apo taHef from 73 to 76 °C, the same NaCl concentrations led to a reduction of the T_M from 79 to 77 °C in the presence of the ss47 DNA. Nevertheless, high salt in combination with DNA was favored compared to high salt without DNA. The difference between taHef in the presence of DNA (79 °C) and in the absence of DNA (73 °C) in low salt buffer is the most striking. This observation can be explained by a prevention of the protein-DNA interaction at high salt concentrations through ions that occupy the positively charged taHef surface which is presumably responsible for DNA binding (Figure 4.22) and also anions that shield the negatively charged DNA, respectively.

After identification of the best buffer conditions various lengths of ssDNA (range of 10-47 nucleotides) were tested (Figure 4.35 B). Remarkably, a ss 10mer DNA does not alter taHef's stability compared to taHef in the absence of DNA. In contrast, a ss 15mer slightly increases the T_M and the biggest step is achieved by the addition of a ss 20mer (orange in Figure 4.35 B) with an enhancement (ΔT_M) of 4 °C with respect to taHef in the absence of DNA. Longer DNAs do not significantly further increase the T_M and thus it can be concluded that a single-stranded DNA substrate with a length of about 20 nucleotides in combination with a low salt buffer constitute ideal conditions for the taHef-DNA interaction. Nevertheless, a buffer containing 100 mM salt was chosen as a

compromise for protein stability and DNA binding capability.

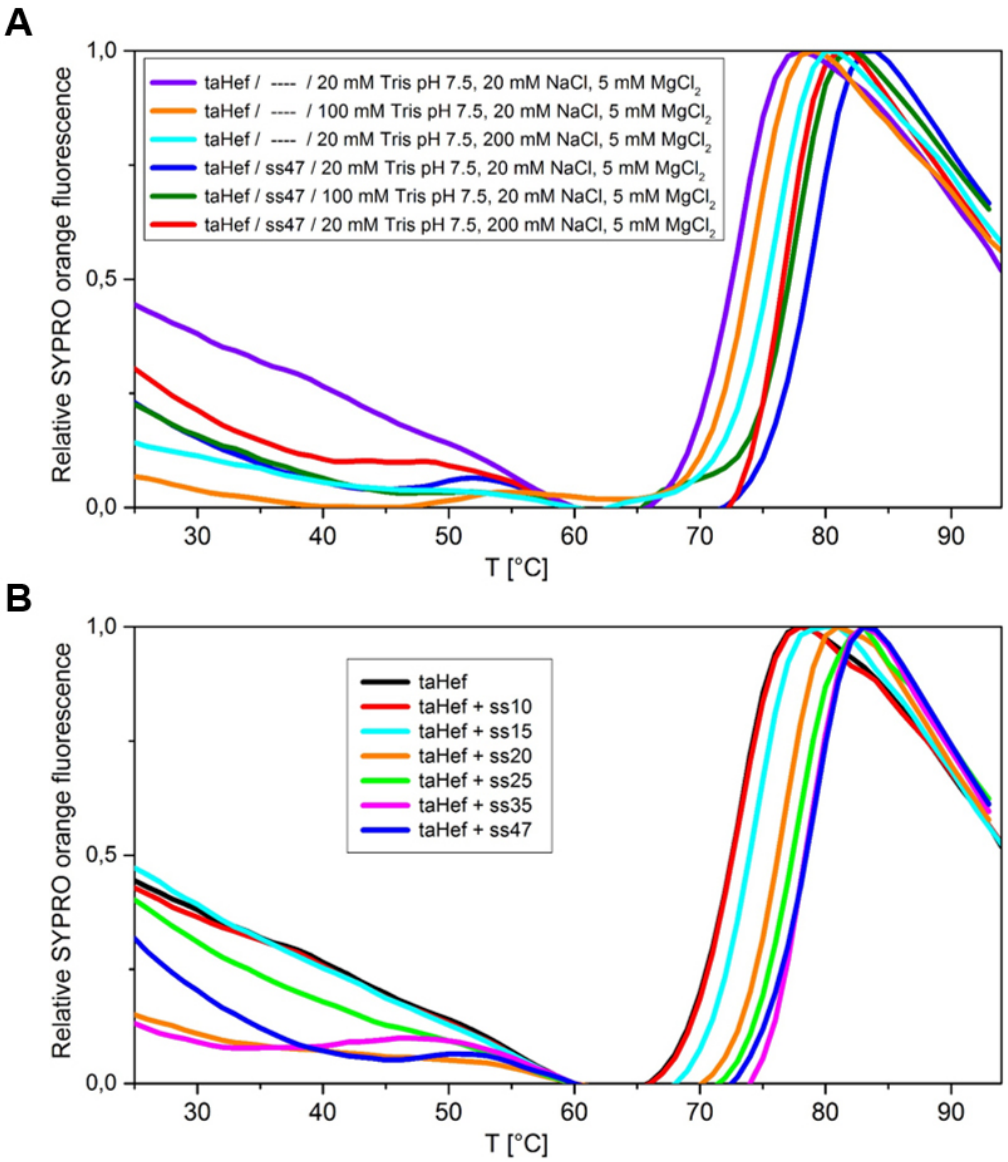


Figure 4.35: Optimization of taHef-ssDNA complex formation.
Protein melting curves determined by thermofluor analysis indicate the stability of taHef in different buffers and in combination with single-stranded (ss) DNA of various lengths. **(A)** The T_M of taHef was measured in three different buffers (low to high salt) in the absence (magenta, orange, cyan) and presence (blue, green, red) of ssDNA with a length of 47 nucleotides. **(B)** The stabilizing effect of taHef by ssDNA of various lengths (red: 10mer; cyan: 15mer; orange: 20mer; green: 25mer; pink: 35mer; blue: 47mer) compared to apo taHef (black) was analyzed in a low salt buffer (20 mM Tris pH 7.5, 20 mM NaCl, 5 mM $MgCl_2$). Protein and DNA were used at equimolar concentrations of 10 µM and no ATP was included.

The DNA-dependent ATPase activity of translocases and helicases is crucial for their function to move along the DNA (Singleton et al., 2007). Thus the effect

of single-stranded DNA with various lengths was tested towards the possibility to stimulate taHefs specific ATPase activity. This experiment shows that a ssDNA substrate with a length between 28 and 35 nucleotides is required to trigger the ATPase activity of taHef in this indirect enzyme-coupled assay (Figure 4.36).

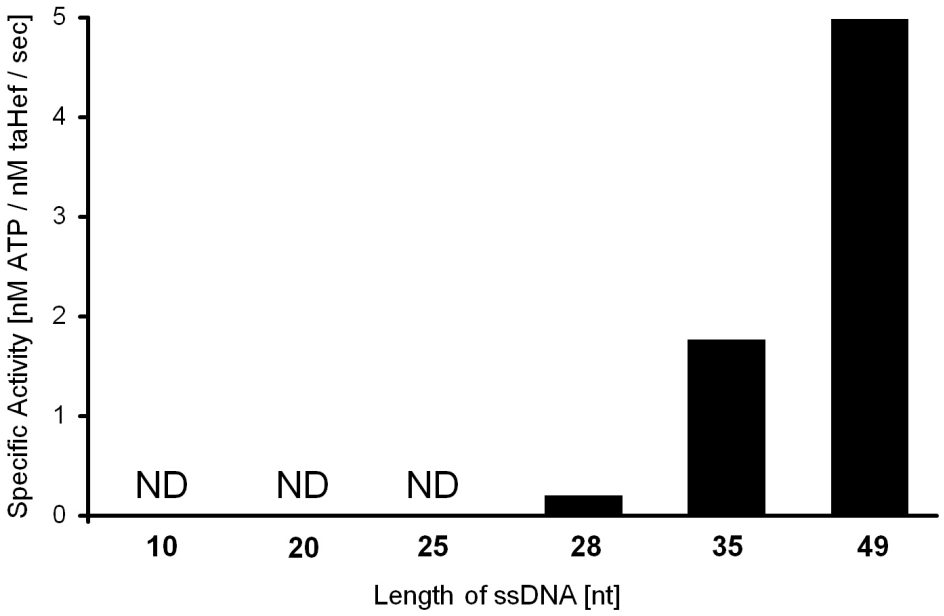


Figure 4.36: ATPase activity of taHef triggered by ssDNA of various lengths. Specific ATPase activity of taHef stimulated by single-stranded DNA substrates of various lengths. ATP hydrolysis by taHef was not induced in the presence of 10, 20, and 25 nucleotides long ssDNAs. Instead, a weak activity (0.2 nM ATP / nM protein / sec) could be observed in the presence of a 28 nucleotides long ssDNA while a 35mer elevated the activity to 1.8 nM ATP / nM protein / sec, respectively. A 49 nucleotides long ssDNA had the strongest effect and triggers 5.0 nM ATP / nM protein / sec. This experiment was performed only once and thus the results are preliminary.

Interestingly, ssDNA with lengths of 10, 20, and 25 nucleotides (black, red, green in Figure 4.36) did not stimulate ATP hydrolysis while, in contrast, substrates with 35 and 49 nucleotides (blue and orange) clearly displayed ATPase activity. A length of 28 nucleotides (magenta) induced an insignificant activity indicating the necessity of about 35 nucleotides to properly stimulate the ATPase actvity of taHef. Notably, the reaction buffer (50 mM Tris pH 7.5, 50 mM KCl, 4 mM $MgCl_2$, 1 mM DTT, 300 µM NADH, 2 mM phosphoenolpyruvate, 10 u/ml lactate dehydrogenase and pyruvate kinase) contained a similar salt concentration compared to the thermofluor experiments.

These results were of particular importance for the following co-crystallization approaches because large and flexible molecules can interfere with the crystallization process and therefore the preferably shortest ssDNA that still binds to taHef should be used.

4.7.2 Attempts for Co-Crystallization of taHef and DNA

To better understand the functionality of taHef as a helicase or its possible role in ICL repair and subsequently to draw conclusions for the human homolog FANCM, a crystal structure of a taHef-DNA complex was pursued and therefore multiple different strategies were applied. Based on the preceding binding analyses different DNA substrates were used such as single-stranded and fork-structured DNA. The protein-DNA complexes were formed in DNA binding buffer consisting of 20 mM Tris/HCl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$ both with and without the addition of AMP-PNP, ATPγS or ADP-BeF_3 in a 1:1.2 (protein:DNA) ratio. Crystallization trials were set up at 4 as well as 20 °C with varying concentrations. Notably, soaking experiments were not performed because the crystal packing would not allow access of the DNA towards the interspace between domains Hel1 and Hel2 (Figure 4.38). Unfortunately, taHef-DNA co-crystallization attempts only yielded crystals in the same space group and with the same cell dimensions as apo taHef which was crystallized in the absence of DNA ($P2_12_12_1$; 40.7 Å, 90.5 Å, 177.8 Å) and DNA could never be observed. Crystals with a different shape consisted either of salt or after a long growth time of degraded taHef without DNA (Figure 4.37).

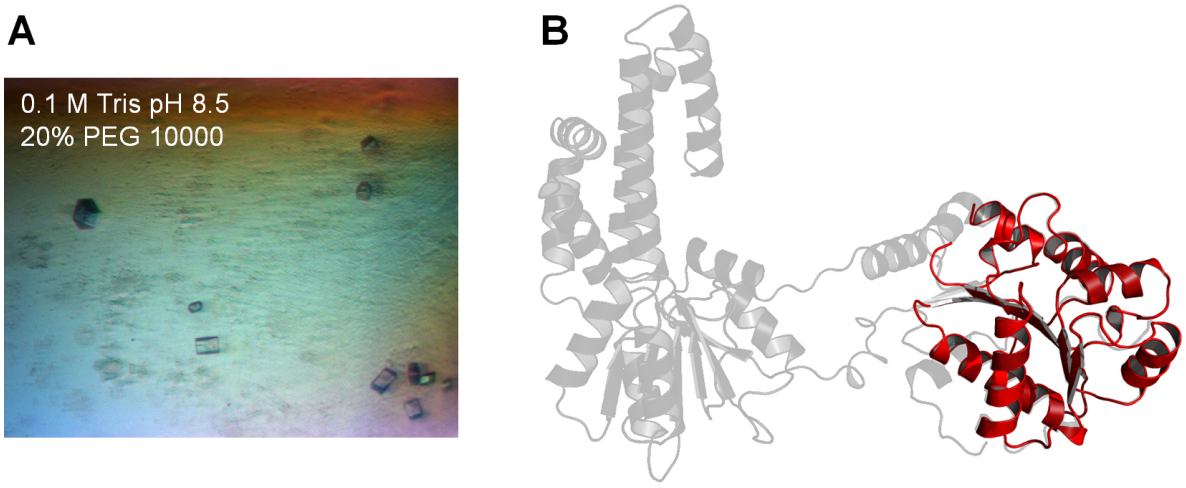


Figure 4.37: Crystallization of taHef in the presence of DNA and crystal structure of degraded taHef. (A) Co-crystallization attempts of taHef together with a single-stranded 20 nucleotides long DNA (fg4) in 0.1 M Tris/HCl pH 8.5, 20 % PEG 10000 yielded crystal growth after 3 months. X-ray diffraction experiments using these crystals revealed that the protein was degraded and only the taHef domain Hel1 (red) was partially present. Crystals belonged to space group $P2_1$ and diffracted to 2 Å. **(B)** A superposition of the obtained Hel1 structure (red) with full-length taHef (grey) is shown for comparison.

A detailed analysis of the crystal lattice of apo taHef (Figure 4.38) revealed an extensive contact area in which the outer surface of Hel1 from one taHef sticks into the cavity between Hel1 and Hel2 of another symmetry-related taHef molecule. These contacts within the crystal may interrupt DNA binding or could lead to the dissociation of the DNA during crystal formation. .

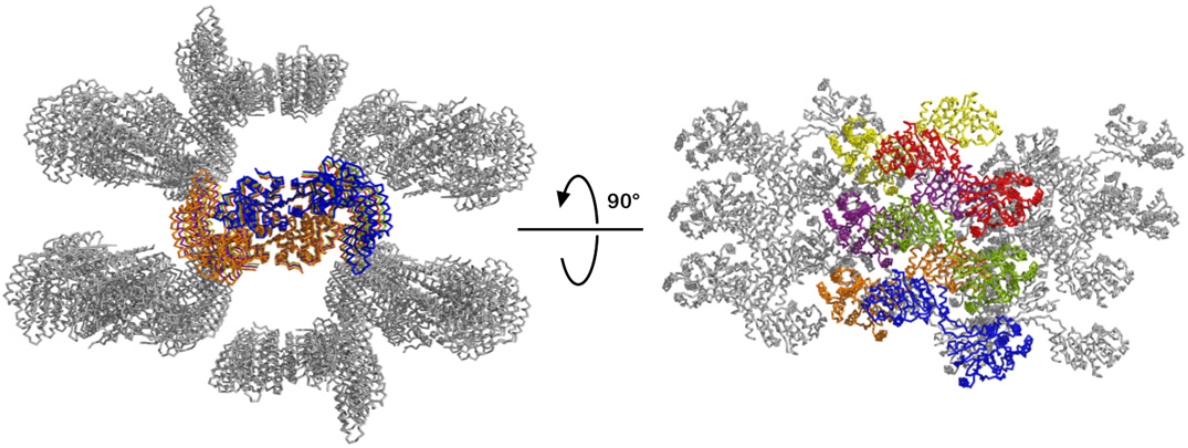


Figure 4.38: Crystal lattice of apo taHef.
The periodical 3D crystal lattice of apo taHef in space group $P2_12_12_1$. taHef monomers within one row are depicted in different colors, respectively, and surrounding molecules are shown in grey.

Although the thumb-like insertion domain also contributes to contacts in the crystal lattice, the contact between Hel1 and the Hel1-Hel2 interspace represents the major interface which occupies 1669.6 Å^2 (7.2 %) of one taHef protein and 1635.8 Å^2 (7.1 %) of the other with ΔG of -4.9 kcal/mol. This corresponds to 45 residues (9.4 %) and 52 residues (10.9 %) of the two neighboring protein molecules, respectively (Table 4.3).

Table 4.3: Interface between two taHef symmetry molecules in the crystal lattice.

	taHef A		*taHef B*	
Number of atoms				
Interface	176	(4.6 %)	185	(4.8 %)
Surface	2068	(54.0 %)	2068	(54.0 %)
Total	3833	(100.0 %)	3833	(100.0 %)
Number of residues				
Interface	45	(9.4 %)	52	(10.9 %)
Surface	421	(87.9 %)	421	(87.9 %)
Total	479	(100.0 %)	479	(100.0 %)
Solvent-accessible area (Å^2)				
Interface	1669.6	(7.2 %)	1635.8	(7.1 %)
Total	23165.5	(100.0 %)	23165.5	(100.0 %)
ΔG	-4.9 kcal/mol		P-value = 0.358	

Major crystal contact between two taHef molecules (orange and blue in Figures 4.38 & 4.39 A).

Residue D139 was identified to play an important role in this non-physiological interface and taHef variants D139K and D139F were generated to disrupt this particular hydrogen-bond network and should thereby prevent further crystallization in this space group (Figure 4.39).

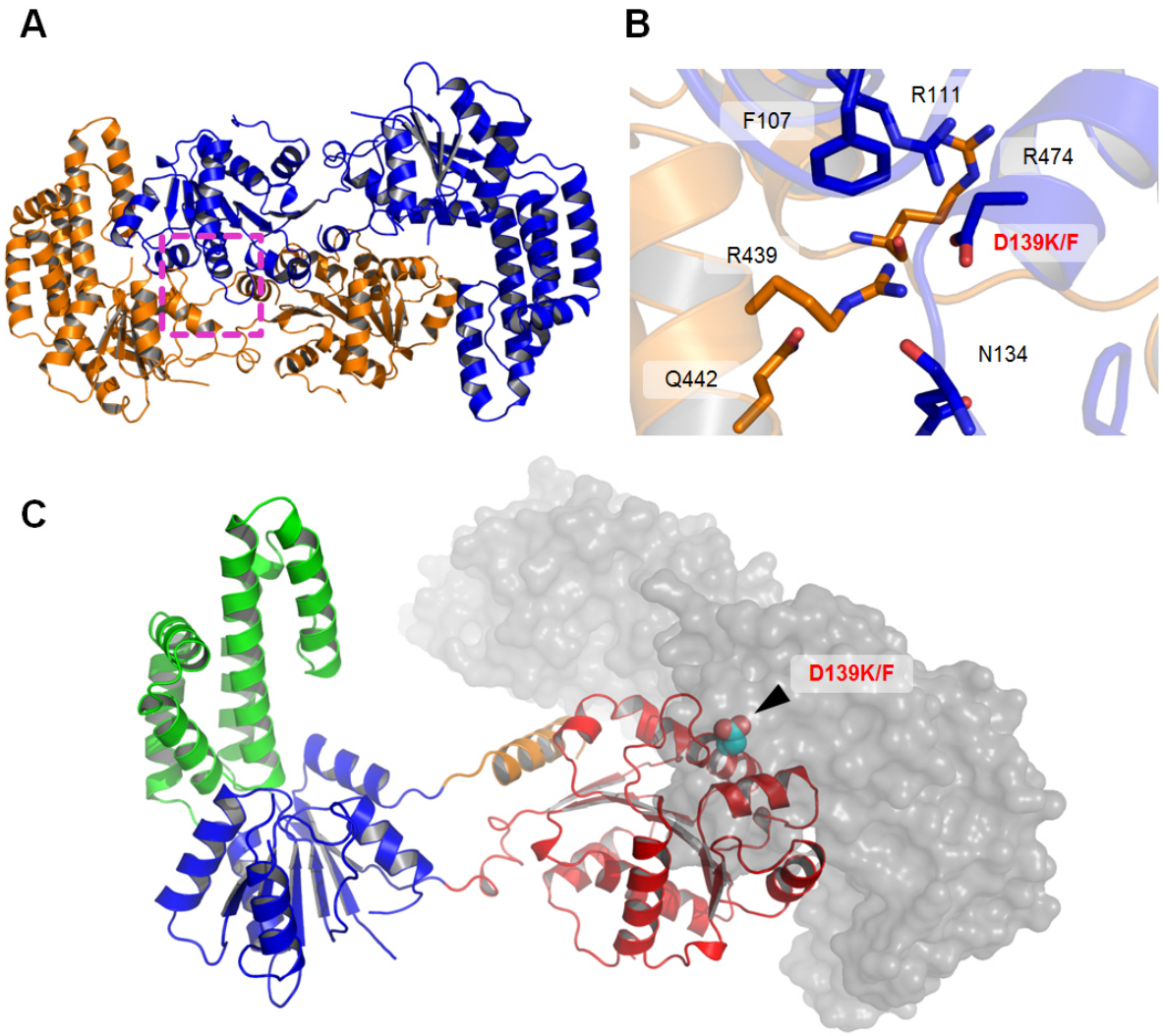


Figure 4.39: Major crystal contact in taHef.
(A) Two symmetry-related taHef molecules (orange and blue) of the $P2_12_12_1$ crystal lattice as shown in Figure 4.38. The main contact area is highlighted by the pink dashed square which is depicted in detail in (B). **(B)** Residues of the two taHef molecules that primarily contribute to the contact interface. R439 and D139 are both strongly involved in this interaction. Due to the fact that R439 is located in the positively charged area between Hel1 and Hel2 that is presumably responsible for DNA binding, this residue was not altered but D139K and D139F modifications were chosen instead. **(C)** Residue D139 is located on the back of Hel1 and thus not expected to be involved in DNA binding. The symmetry-related taHef molecule is shown in surface representation.

Strikingly, taHef D139K and D139F did not crystallize under the same conditions as taHef WT although correct protein folding was confirmed by CD spectroscopy analysis (Figure 4.8). The D139K and D139F variants thus successfully prevented the alignment of the taHef molecules into a crystal lattice as observed for the WT protein.

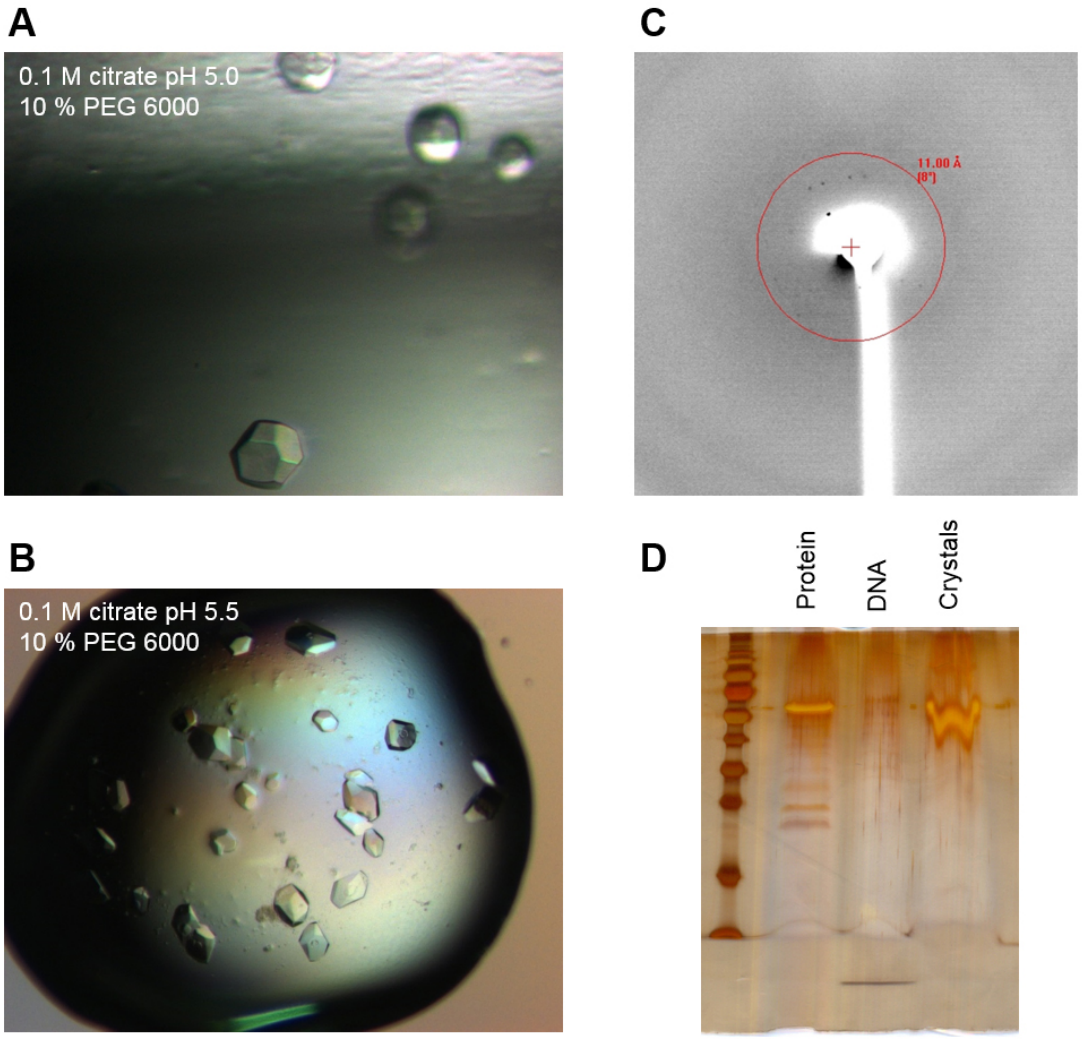


Figure 4.40: taHef D139K Jiang14 ssDNA complex crystallization.
Co-crystallization of taHef D139K (330 µM / 20 mg/ml) and the Jiang14 DNA substrate (400 µM double-stranded) in the presence of 5 mM ADP-BeF_3. **(A)** Initial taHef-Jiang14 crystals were grown in sitting drop vapor diffusion experiments containing 0.3 µl of the protein-DNA solution mixed with 0.3 µl reservoir solution and equilibrated against 40 µl reservoir solution at 20 °C for one week. Unfortunately, the crystals did not diffract. **(B)** Optimized crystals grown in hanging drop vapor diffusion experiments using 1 µl of the protein-DNA solution mixed with 1 µl reservoir solution and equilibrated against 1 ml reservoir solution at 20 °C for one week. The crystals displayed sharp edges and an increased size. **(C)** X-ray diffraction tests with an exposure time of 15 min and 0.5° oscillation revealed that the crystals were poorly diffracting. **(D)** Extracted and washed crystals on a silver stained standard SDS-PAGE exhibited full-length taHef D139K but no DNA could be observed.

Unfortunately, apo taHef D139K/F did not yield reasonable protein crystals at all. taHef in a different space group could possibly display another conformational state and thus further help to understand the movement of domains Hel1 and Hel2. However, the main purpose of these variants was not to obtain the apo structure but rather their application in co-crystallization approaches in the presence of DNA. Thus taHef D139K and D139F were co-crystallized with different fork-structured and single-stranded DNA substrates utilizing various crystallization conditions. Initially, the protein was mixed with DNA (1:1.2) and then dialyzed against the DNA binding buffer (20 mM Tris/HCl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$) over night. In a next step 5 mM ADP-BeF_3 was added and after a short incubation time the crystallization trials were set up.

In addition to the standard DNAs, also the Jiang14 substrate was utilized which was derived from the RIG-I-DNA complex crystal structure (Jiang et al., 2011). It comprises the palindromic sequence 5'-CGACGCTAGCGTCG-3' that is also able to assume a blunt-end dsDNA (single-stranded: 4.3 kDa; annealed double-strand: 8.6 kDa) Interestingly, only in the presence of the Jiang14 DNA substrate taHef D139K could be crystallized and the crystal form clearly differed from taHef WT crystals (Figure 4.18). In contrast to the long needles bunches, taHef D139K ssJiang14 ADP-BeF_3 produced unclustered and compact crystals while the general crystallization conditions were similar. Apo taHef WT also crystallized at pH 5.0 and in the presence of the long-chained PEG 3350 used as the precipitant ingredient. Importantly, the diffraction pattern showed that the tested crystal consisted of protein but due to the low resolution ($\sim$ 11 Å) it was not possible to determine further parameters. SDS-PAGE and silver staining analysis of the crystals revealed the presence of full-length taHef D139K but the DNA could not be detected.

It can thus be concluded that the mutation of taHef residue D139 alters the crystallization behavior of the protein as indicated by the different crystal form. This taHef variant is therefore recommended for further co-crystallization experiments.

4.8 Structural Analysis of taPCNA

The sequence analysis of taHef revealed a PCNA interacting peptide motif (PIP-box) at the very C-terminus of the protein (Figure 4.2) and thus the DNA sliding clamp PCNA from *Thermoplasma acidophilum* (taPCNA) was cloned, expressed, and purified from *E. coli* to analyze a putative interaction with taHef (Chapter 4.9). When analyzing protein complexes it is essential to also explore the individual components as well and therefore in addition to the taHef apo crystal structure and SAXS model (Chapters 4.4 & 4.6) taPCNA was also structurally characterized.

4.8.1 Crystallization of taPCNA

Recombinantly expressed and purified taPCNA in Sdx-buffer (20 mM Tris/HCl, 200 mM NaCl, pH 7.5) crystallized at a concentration of 9 mg/ml (300 µM) in a sitting drop vapor diffusion crystallization setup with 0.3 µl protein mixed with 0.3 µl reservoir solution (100 mM HEPES, 30 % PEG 300, pH 7.5), respectively, and equilibrated against 40 µl reservoir solution at 20 °C. taPCNA crystallized in the highly symmetrical face-centered cubic space group F23 and crystals diffracted up to ~ 4.5 Å resolution (Figure 4.41). Several cryo-protective solutions were tested but the condition without any additional cryo-protectant yielded the best diffraction properties. Crystals were flash cooled in liquid nitrogen and x-ray diffraction was then tested at 100 K.

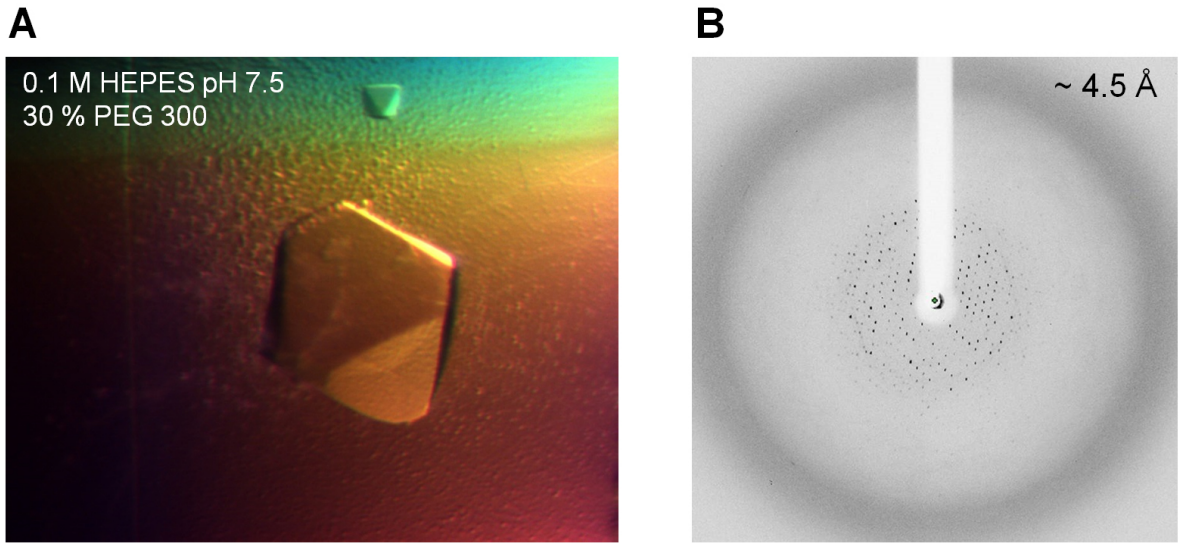


Figure 4.41: taPCNA protein crystals and an exemplary diffraction image.
(A) Protein crystals of taPCNA (9 mg/ml; 300 µM) that were grown by the sitting drop vapor diffusion method. 0.3 µl protein solution combined with 0.3 µl reservoir solution were equilibrated against 40 µl reservoir solution and led to crystals of 200 x 200 x 200 μm^3 after two days. **(B)** One exemplary diffraction image of the best taPCNA protein crystal shows a maximum resolution of ~ 4.5 Å.

taPCNA crystals could be reproduced and also the size could be improved in further fine screens with larger volumes (1 µl protein + 1 µl reservoir solution; equilibrated against 1 ml reservoir) in hanging drop setups. Unfortunately, the diffraction behavior could not be improved.

Controlled dehydration of a protein crystal utilizing the humidity control device HC1 can improve the diffraction properties (Sanchez-Weatherby et al., 2009). Thus taPCNA diffraction experiments were pursued at RT and decreasing hydration states were tested (Figure 4.42). In contrast to the fully hydrated crystals, x-ray diffraction significantly declined with coherent reduction in hydration. While at 99% hydration a distinct

diffraction pattern was still visible it was almost depleted at 95%. It could thus be concluded that neither the switch to data collection at RT nor a reduction of the hydration state of the taPCNA protein crystals improved diffraction.

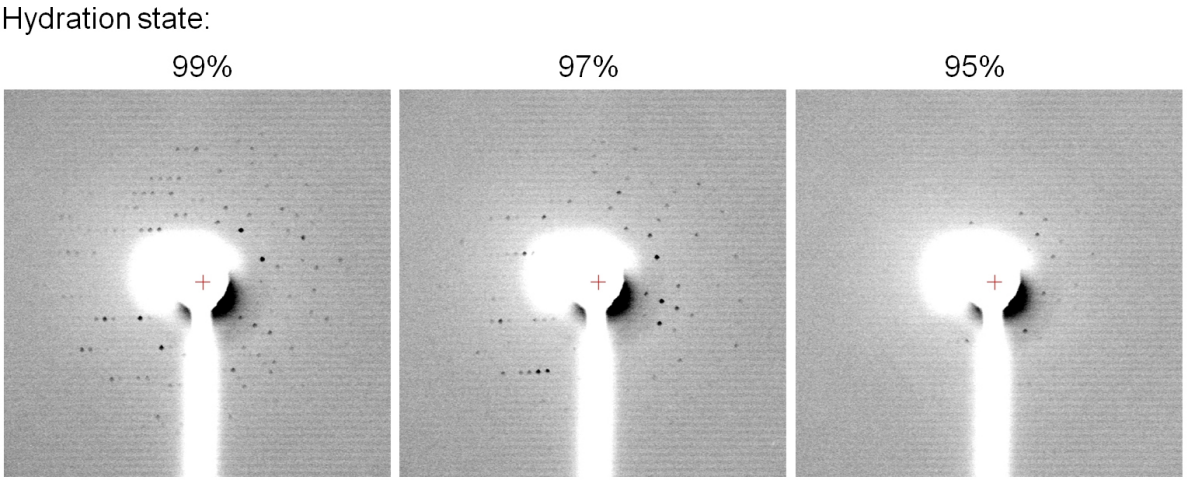


Figure 4.42: taPCNA x-ray diffraction pattern at different hydration states. Diffraction pattern of taPCNA protein crystals at room temperature (RT) and in decreasing hydration states provided by the humidity control device HC1.

Since there are already several crystal structures of PCNA from various source organisms (archaea and eukaryotes) both apo and in complex with interaction partners and in the presence of DNA available, the optimization of the taPCNA crystals was discontinued. However, as shown in Chapter 4.14, a high-resolution crystal structure of taPCNA still remains highly interesting. Especially a detailed view of the PIP-box binding pocket (see Chapter 4.14) might provide insight into the differences to the eukaryotic system. An important option for crystal optimization could be the removal of the N-terminal His_6-tag. This step could drastically change the crystallization and x-ray diffraction properties of taPCNA.

4.8.2 SAXS Analysis of taPCNA

As a control and to facilitate the interpretation of the taHef-taPCNA SAXS model (Chapter 4.9.1) it was important to verify that the taPCNA ring can be approximated as a single rigid body. Since there is no crystal structure of taPCNA available, the structure of human PCNA (hsPCNA; PDB entry 1VYM) (Kontopidis et al., 2005) was utilized and matched the experimental data reasonably well (chi 4.9), although systematic deviations consistent with a small amount of flexibility were observed (Figure 4.43 A).

This flexibility was visualized in the superposition of the *ab initio* model generated with hsPCNA, where the *ab initio* model appeared to be slightly bent and more

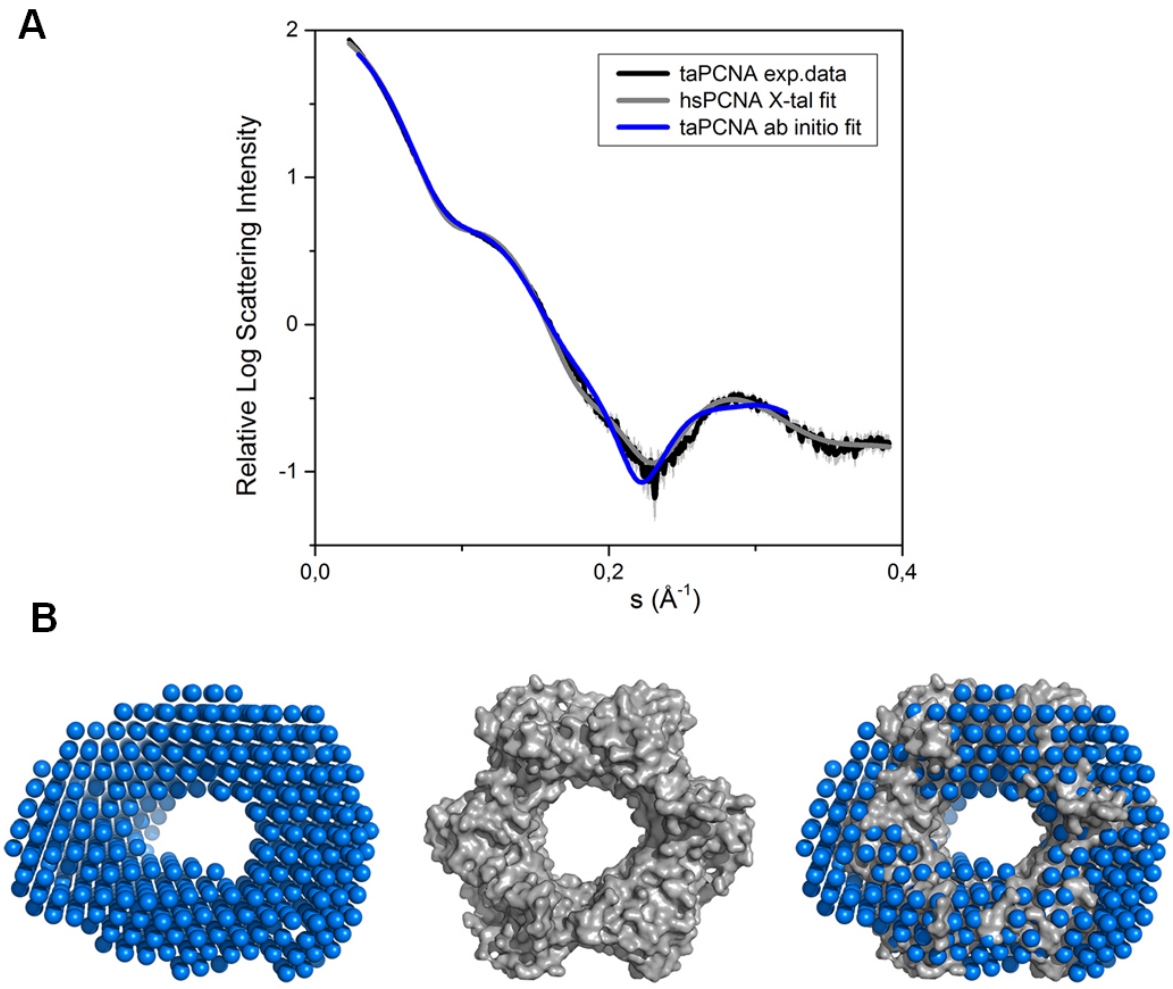


Figure 4.43: SAXS analysis of taPCNA.
(A) 1D SAXS data of taPCNA (black) overlaid with the theoretical scattering of the human PCNA structure (grey) yielding a chi value of 4.0 and the taPCNA *ab initio* Dammif model (blue) with a corresponding chi value of 3.7 indicating that the shape of the structure is compatible with the solution data but suggesting some flexibility. **(B)** Superposition of the taPCNA *ab initio* Dammif model (blue) and the hsPCNA crystal structure (grey, PDB code: 1VYM) (Kontopidis et al., 2005). The taPCNA model resembles the overall structure of ring-shaped homotrimeric hsPCNA. The PCNA ring is likely to bend slightly and expand and contract leading to a more oval average structure.

elliptical, indicating the possibility of twisting, expansion, and contraction of the ring (Figure 4.43 B). Although taPCNA shows some flexibility, it was assumed that the hsPCNA structure is a reasonable approximation of the taHef average position in solution. Thus the SAXS model nicely confirmed the donut-shaped constitution of taPCNA.

4.9 Interaction between taHef and taPCNA

Based on multiple sequence analyses (Chapter 4.1) a canonical PIP-box in taHef (residues 502-508) was identified which is located in a flexible area of the protein (Figure 4.44). The PIP-box is highly conserved among Hef proteins of various archaeal organisms, thus suggesting a possible interaction interface for PCNA binding. The PIP-box is located C-terminally adjacent to the helicase entity within an elsewise undefined and unstructured region which is a major requirement for functional PIP-boxes. Since taHef is the only member of the Hef protein family that solely comprises the N-terminal helicase domain the PIP-box is located at the very C-terminus of this particular protein. Other Hef proteins are however C-terminally extended by a linker as well as the nuclease domain and thus in these proteins the PIP-box is positioned centrally.

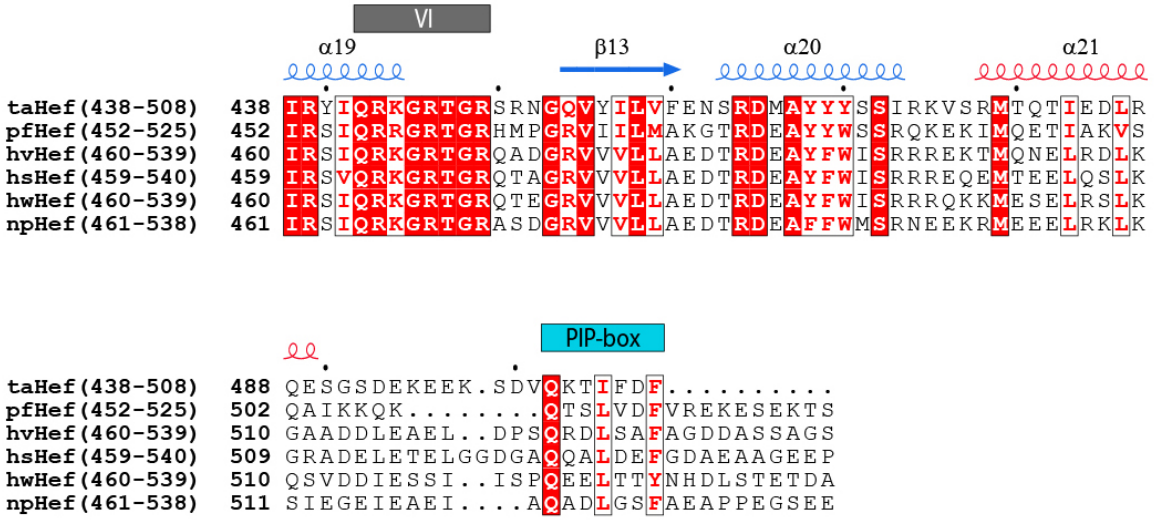


Figure 4.44: Conserved PIP-box in archaeal Hef proteins.
Sequence alignment of the C-terminus of the helicase portion and the beginning of the adjacent linker towards the nuclease domain of Hef from different archaeal organisms (ta = *Thermoplasma acidophilum*; pf = *Pyrococcus furiosus*; hv = *Haloferax volcanii*; hs = *Halobacterium salinarum*; hw = *Haloquadratum walsbyi*; np = *Natronomonas pharaonis*). taHef is the only member that does not comprise the C-terminal nuclease domain. Secondary structure elements were derived from the taHef crystal structure and labeled according to Figure 4.2. The conserved canonical PIP-box within an undefined and unstructured region is highlighted by the cyan box.

In order to verify whether the conserved PIP-box is functional interaction studies between taHef and taPCNA were performed. Analytical size exclusion chromatography (ASEC) analysis of a 1:1 mixture of taHef and taPCNA confirmed our hypothesis and showed a direct interaction between both proteins (Figures 4.45 A & C). taHef WT (red) elutes from the calibrated SEC column with a single peak at 14.3 ml (calculated MW: 59 kDa) and taPCNA (grey) at 13.6 ml (calculated MW: 86 kDa). These values

correspond to a taHef monomer (actual MW: 61 kDa) and a taPCNA homotrimer (actual MW: 93 kDa; monomer: 31 kDa). The single peak including both proteins (black) is shifted to 11.1 ml, indicating stable complex formation in solution. The calculated molecular weight of the complex peak eluting at 11.1 ml (~280 kDa) suggests a 1:1 stoichiometry. Despite an equimolar input of both proteins (500 µM each) the taPCNA peak appears lower because of a lower extinction coefficient ($\epsilon_{\text{taPCNA}}/\epsilon_{\text{taHef}} = 0.35$).

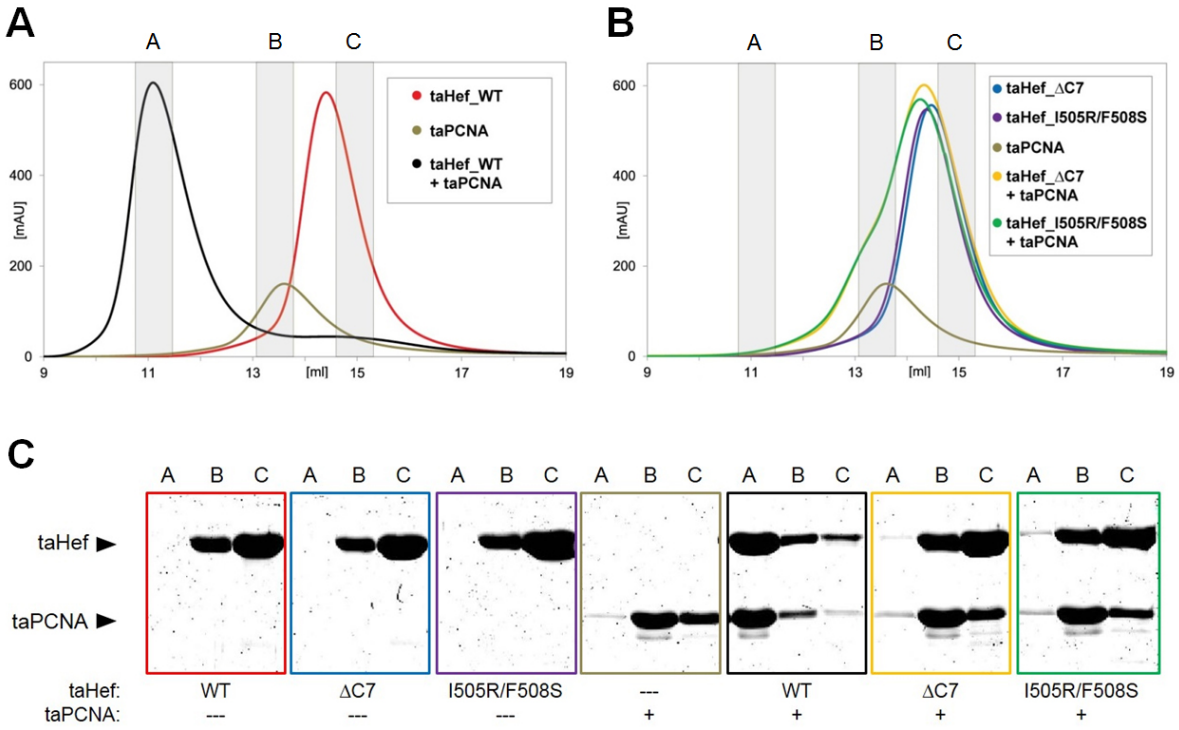


Figure 4.45: taHef-taPCNA interaction.
(A) Analytical size exclusion chromatography (ASEC) of taHef WT (red), taPCNA (grey), and an equimolar mixture of both (black; $\epsilon_{taPCNA}/\epsilon_{taHef} = 0.35$). The peak at a lower elution volume (11 ml) representing a higher molecular weight indicates stable complex formation. **(B)** ASEC profiles of taHef ΔC7 (blue) and taHef I505R/F508S (purple) show a similar elution behavior as the wild type protein (red in (A)). Combinations of taHef ΔC7 and taPCNA (yellow) as well as taHef I505R/F508S and taPCNA (green) do not show any complex formation. **(C)** Peak fractions A, B, and C (labels above the gels) from (A) and (B) were analyzed via SDS-PAGE. Corresponding elution chromatograms are indicated by respective frame colors and the particular content is additionally specified below the gels.

To verify the C-terminal PIP-box $_{502}$QKTIFDF$_{508}$ (Figure 4.44) as the PCNA interaction site of taHef, we deleted this patch (taHef ΔC7, blue in Figures 4.45 B & C) which clearly led to a disruption in complex formation (yellow in Figures 4.45 B & C). It was previously shown that PIP boxes specifically bind to a hydrophobic pocket in PCNA by constituting a 3_{10}-helix. Hydrophobic and aromatic residues within the PIP box consensus sequence are crucial for this particular 3_{10}-helix formation (Chapados et al., 2004). To impair taPCNA binding and simultaneously prove the classical PIP

box binding mechanism of taHef, we generated a double taHef I505R/F508S variant and analyzed it in the presence of PCNA by SEC (magenta in Figures 4.45 B & C). Similar to the ΔC7 deletion, this double mutation also abolishes taPCNA binding (green in Figures 4.45 B & C) indicating the inhibition of 3_{10}-helix formation.

To further validate the interaction between taHef and taPCNA and also to gain insight into the binding kinetics of the complex, isothermal titration calorimetry (ITC) was pursued which confirmed our previous observation and showed that taHef WT binds to taPCNA with a dissociation constant (K_D) of 1 µM and a 1:1 stoichiometry (Figure 4.46). Since taPCNA forms a homotrimer, the complex comprises three taHef monomers resulting in a heterohexamer. The Gibb's free energy (ΔG) of this complex formation constitutes -7.6 kcal mol^{-1} and shows a spontaneous reaction. The free enthalpy (ΔH) contribution of -15 kcal mol^{-1} indicates an exothermic reaction and the reduction of entropy (ΔS) of -24.8 cal mol^{-1} K^{-1} can be partially explained by the transition of the PIP-box from an unstructured state to a well-defined 3_{10} helix (Sakurai et al., 2005). The entropy term (TΔS) accounts for -7.4 kcal mol^{-1} and counteracts the overall enthalpy driven reaction. In accordance to our analytical size exclusion chromatography experiments both PIP-box variants (taHef ΔC7 & taHef I505R/F508S) also abolish taPCNA binding in ITC (Figure 4.46).

These results further confirm the C-terminal canonical PIP-box of taHef (residues 502-508) as the major interaction interface for taPCNA binding. ITC experiments with other PIP-box mediated PCNA interaction partners show high variability in binding affinities. Human p21 exhibits strong binding to PCNA with a K_D of 80 nM while it is 60 µM for the FEN1 interaction and thus three orders of magnitude lower (Bruning and Shamoo, 2004). On the basis of these values, taHef can be classified as a moderate PCNA binding partner.

Interestingly, the PIP-box sequence of taHef (^{502}**Q**K**T**I**F**D**F**^{508}) is very similar to the one of replication factor C large subunit from *Pyrococcus furiosus* (pfRFCL; UniProt accession code: Q9UWR2) (^{470}**Q**A**T**L**F**D**F**^{476}). Replication factor C is a clamp loader protein which is responsible for the loading of PCNA onto the DNA (Bowman et al., 2004). Although residues $K503_{taHef}$/$A471_{pfRFCL}$ and $I505_{taHef}$/$L473_{pfRFCL}$ (underlined) differ within the two PIP-boxes, PIP_{pfRFCL} constitutes the best available structural model for PIP_{taHef}. PIP_{pfRFCL} is also located at the C-terminus of the protein and it additionally contains the three C-terminal residues ^{477}LKK^{479}. Figure 4.47 shows the crystal structure of the pfPCNA-PIP_{pfRFCL} complex (PDB code: 1ISQ) (Matsumiya et al., 2002). Unfortunately there are no kinetic data such as ITC experiments available for this interaction (Matsumiya et al., 2002).

Interestingly, the T at position 3 within the PIP motif is conserved among FEN1 from Archaeglobus fulgidus (afFEN1), pfRFCL, and taHef and might also play a role in the interaction with PCNA additionally to the conserved PIP-box sequence. This residue will be discussed in Chapter 4.14.

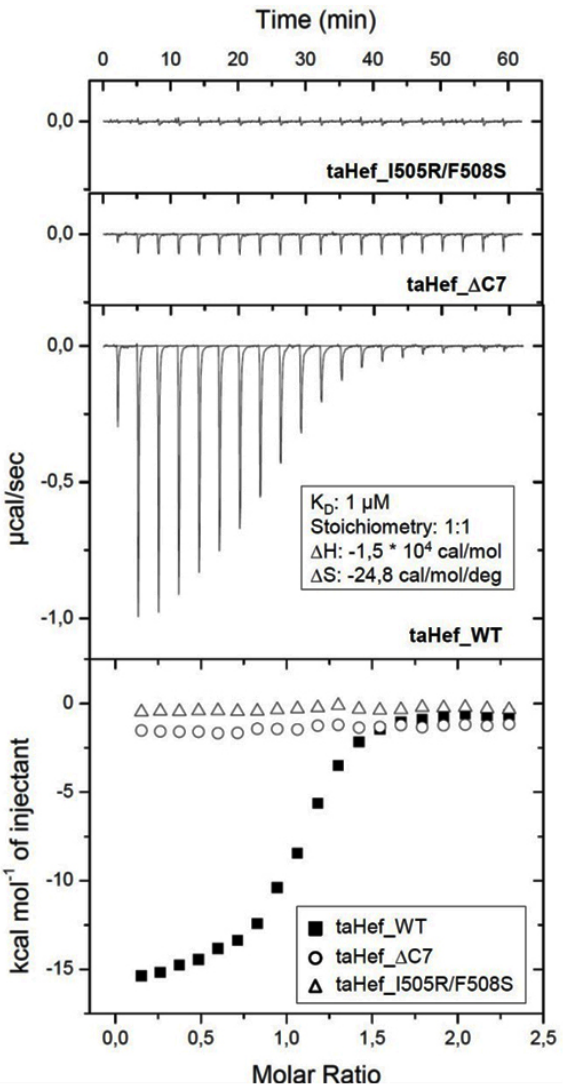


Figure 4.46: Binding kinetics of the taHef-taPCNA interaction.
Isothermal titration calorimetry (ITC) measurements of the interaction between taHef and taPCNA. taHef_WT-taPCNA *in vitro* binding kinetics feature a dissociation constant (K_D) of 1 µM and a 1:1 stoichiometry. Thermodynamic parameters of the binding reaction: ΔH = -15 kcal mol^{-1}, ΔS = -24.8 cal mol^{-1} K^{-1} and calculated ΔG = -7.6 kcal mol^{-1}. For the two taHef PIP-box mutants ΔC7 and I505R/F508S no heat release and thus no binding can be detected.

It was extensively tried to crystallize the taHef-taPCNA complex utilizing the removal of the N-terminal His_6-tags, limited proteolysis, broad range of protein concentrations, re-purification of the complex prior to crystallization, different temperatures, addition of ATPγS and various DNA substrates. Unfortunately, the only protein crystals that could be obtained did not diffract and solely comprised taPCNA instead of the complex (Figure 4.48).

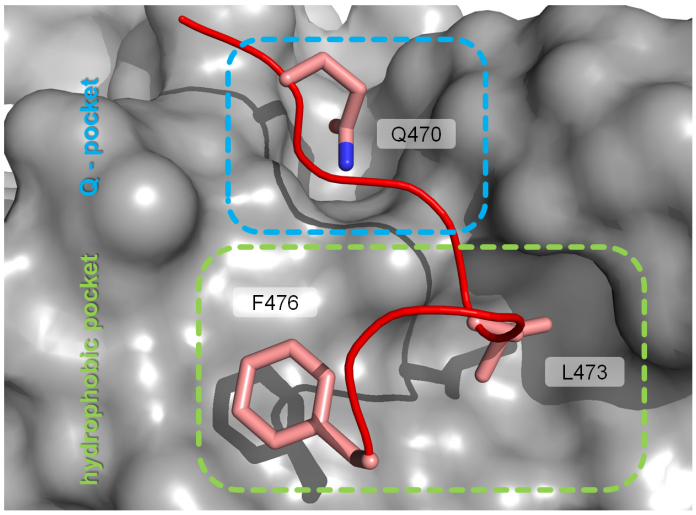


Figure 4.47: PIP_{pfRFCL} bound to pfPCNA.
Crystal structure of a PIP-box containing peptide derived from RFCL (red) in complex with PCNA (grey) from the archaeon *Pyrococcus furiosus* (pfRFCL and pfPCNA, respectively) (PDB code: 1ISQ) (Matsumiya et al., 2002). The Q-pocket is highlighted in blue and the hydrophobic pocket that binds the 3_{10}-helix of the PIP-box is depicted in green. The amino acid sequence of the pfRFCL PIP-box (^{470}QATLFDF^{476}) is very similar to the taHef PIP-box (^{502}QKTIFDF^{508}) and thus the present structure can be used as a model for PIP_{taHef}-taPCNA.

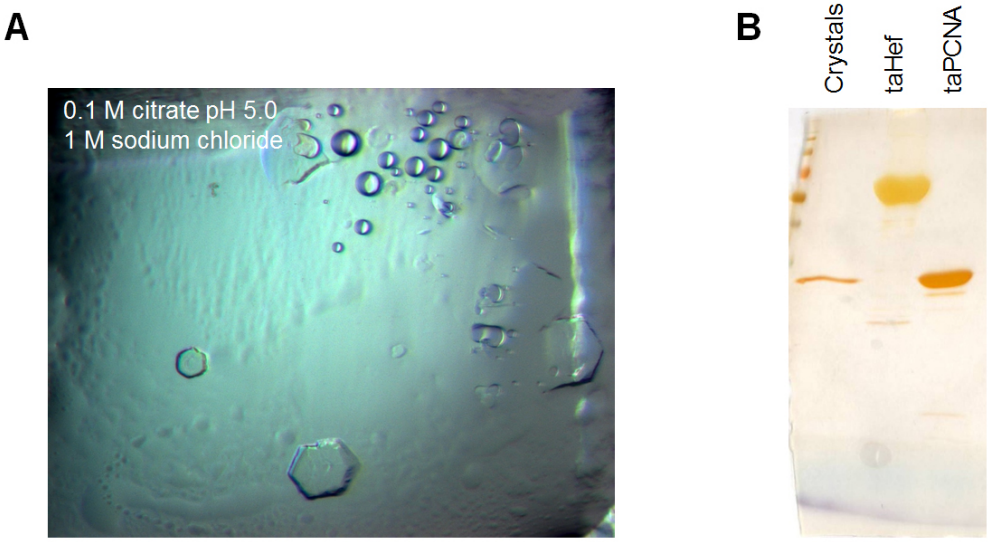


Figure 4.48: Crystallization of the taHef-taPCNA complex.
(A) Crystallization trials of the freshly purified taHef-taPCNA complex in a hanging drop vapor diffusion setup yielded crystals that did not diffract. **(B)** Visualization of the crystals via standard SDS-PAGE and silver staining revealed that only taPCNA instead of the complex was present in the crystals.

4.9.1 SAXS Analysis of the taHef-taPCNA Complex

Presumably, due to flexibility it was not possible to crystallize the taHef-taPCNA complex. Interestingly, the crystal structure of the human FEN1-PCNA complex (PDB accessory code: 1UL1) shows three hsFEN1 molecules bound to the hsPCNA homotrimer while each hsFEN1 adopts a different relative orientation (Sakurai et al., 2005). This structure also provides a good example for the anchoring of interaction partners on PCNA by the PIP-box (and adjacent areas) without affecting their free movement and functionality. For the determination of the structure of the taHef-taPCNA complex in solution SAXS was utilized.

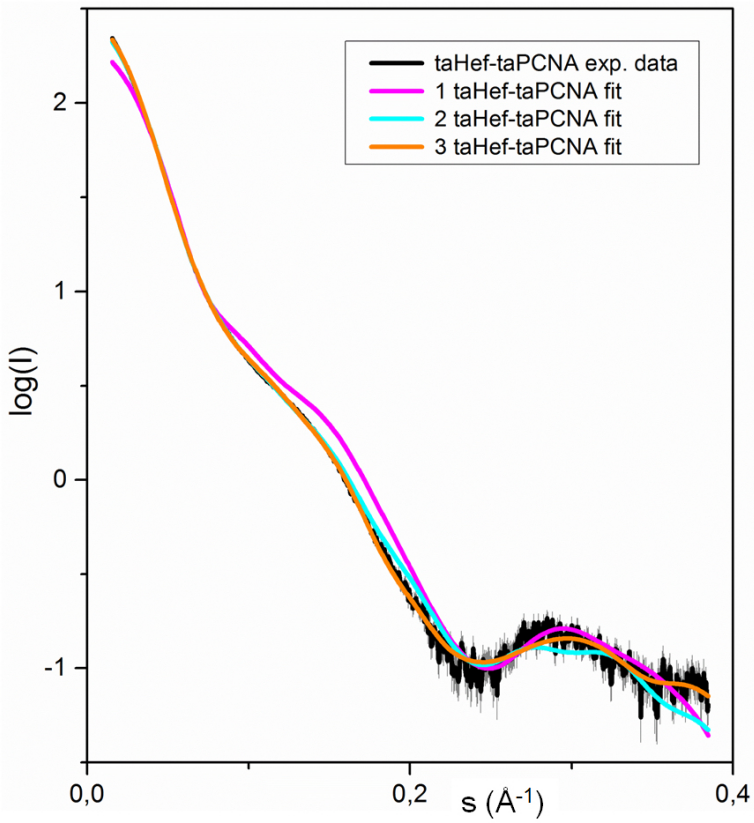


Figure 4.49: 1D SAXS intensity fits of the taHef-taPCNA complex.
Experimental SAXS data for the taHef-taPCNA complex (black) with indicated error (grey) and fits for one (pink), two (green) and three (orange) taHef proteins bound to the taPCNA trimer.

The apo taHef and apo taPCNA SAXS models (Chapters 4.6 & 4.8.2) served as controls to facilitate modeling and interpretation of the taHef-taPCNA complex (Figure 4.50). Rigid body modeling is facilitated using the minimum number of bodies and since taPCNA constitutes a stable donut-shape it was treated as a rigid body. Moreover, any effects caused by the flexibility of taPCNA will be minor compared to domain movements in taHef and the possible movement of taHef relative to taPCNA (flexible linker with 26 residue). The hsPCNA structure (PDB code: 1VYM) (Kontopidis et al.,

2005) was used for the modeling of taPCNA within the taHef-taPCNA complex.

Using CORAL we analyzed the possibility of one, two, or three taHef monomers bound to taPCNA. As there is a limitation with respect to the number of individual domains that can be described in CORAL the 3 taHef model had to be constrained in P3 symmetry (grey lines in Figure 4.50 A), whereas the others were not. Still, the best fit was obtained with 3 taHefs bound, (even with this additional constraint) followed closely by 2 taHef molecules, whereas 1 taHef provided a very poor fit (chi values of 1.98, 2.5 and 18.8, respectively) (Figure 4.49). Thus the SAXS data show that the 3 taHef bound model was the preferred stoichiometry in solution which is in agreement with the ITC and analytical size exclusion experiments (Chapter 4.9).

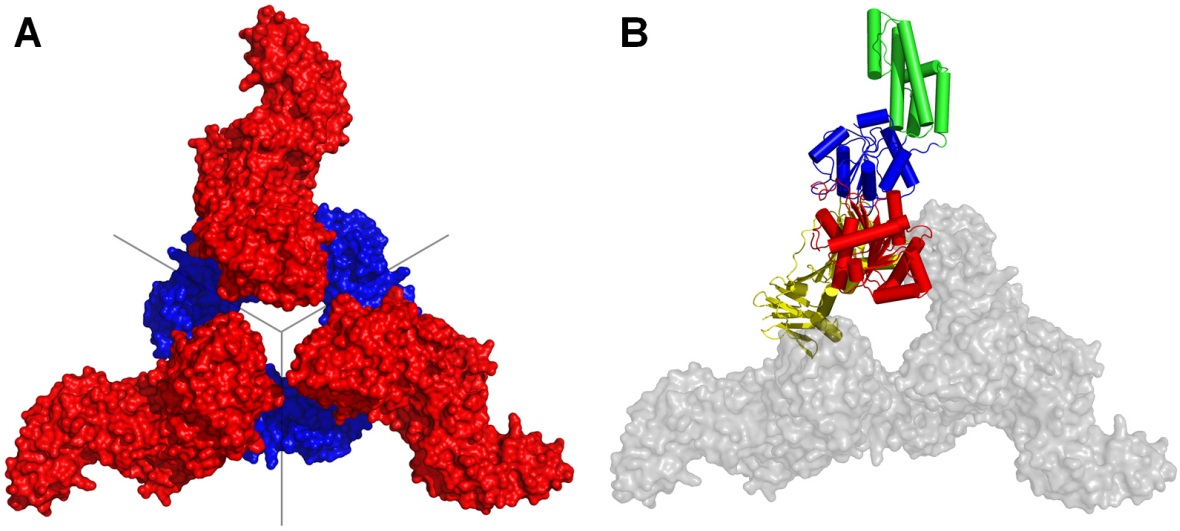


Figure 4.50: SAXS rigid body model of the taHef-taPCNA complex.
(A) Rigid body CORAL model of the complex with three taHef monomers (red) bound to the PCNA trimer (blue). For PCNA modeling the human crystal structure (PDB entry 1VYM) (Kontopidis et al., 2005) was used. The applied P3 symmetry is indicated by the grey lines. **(B)** One monomeric subunit of taPCNA (yellow) together with one taHef molecule (Hel1: red; Hel2: blue; Ins: green) is depicted in cartoon representation.

Although it could be shown that the taPCNA ring structure is present in solution and that it has 3 taHefs bound in the taHef-taPCNA complex, the positions of the taHef molecules cannot be definitively determined as the taHef-taPCNA complex exhibits significant flexibility which can be demonstrated by the variation in the domain positions in multiple runs by CORAL each providing reasonable fits to the data. However, none of them leads to an ideal fit and therefore it is highly likely that the assumption of P3 symmetry is only partially valid due to flexibility (Figure 4.43). The PIP-box binding site may still be close to P3 symmetry but the known flexibility of the taHef molecule may disrupt the symmetry of the entire complex and hence we observed the systematic deviations of the model fit to the data (Chapter 4.6).

4.9.2 SAXS Analysis of the taFEN1-taPCNA Complex

The interaction between flap-endonuclease 1 (FEN1) and PCNA as well as its physiological function has been described for the archaeal and eukaryotic proteins (Chapados et al., 2004; Zheng et al., 2011a). Besides extensive biochemical analysis the FEN1-PCNA complex was also structurally characterized and thus FEN1 served as a positive control for experiments with taHef-taPCNA and hsFANCM-hsPCNA in terms of interaction with PCNA (Chapters 4.9 & 4.13). Therefore the proteins taFEN1, taPCNA (Chapter 4.8.2) and also the complex of both were measured via SAXS for method validation.

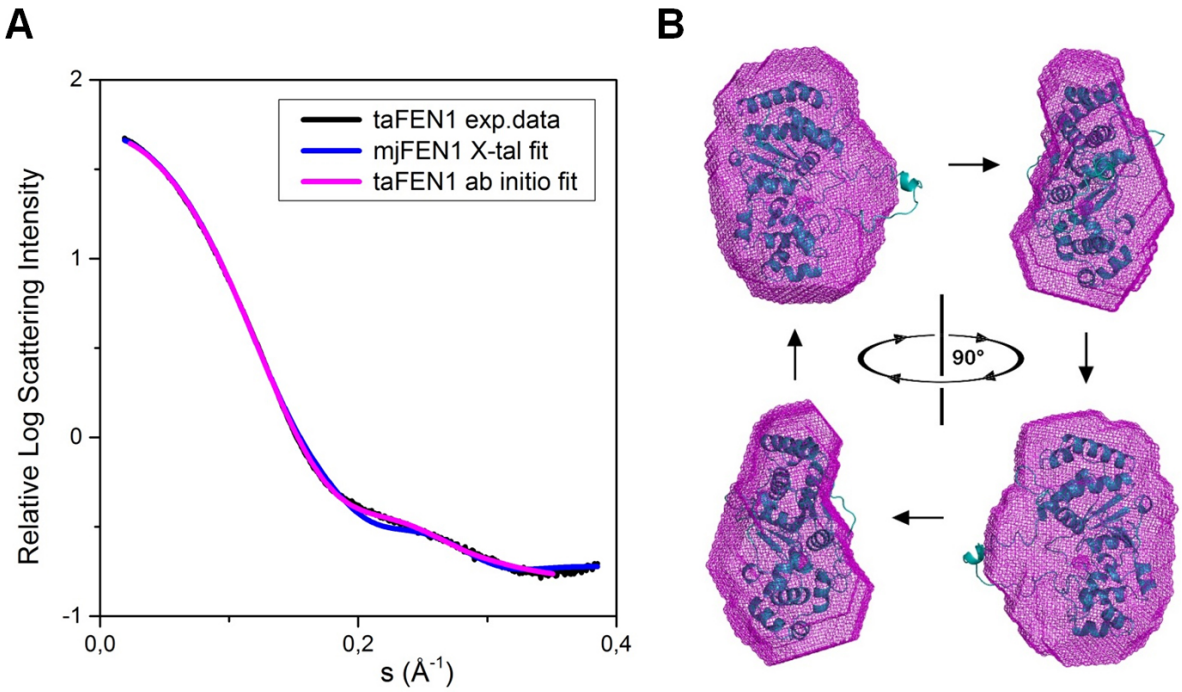


Figure 4.51: SAXS analysis of taFEN1.
(A) 1D SAXS data of taFEN1 (black) overlaid with the theoretical scattering of the mjFEN1 structure (blue) yielding a chi value of 3.3 and the taFEN1 *ab initio* Dammif model (magenta) with a corresponding chi value of 2.3 indicating that the shape of the structure is compatible with the solution data but suggesting some flexibility. **(B)** Superposition of the taFEN1 *ab initio* Dammif model (magenta) and the mjFEN1 crystal structure (blue, PDB code: 1A77) (Hwang et al., 1998). The taFEN1 model resembles the overall shape of mjFEN1 indicating a similar structural conformation.

Comparison of the taFEN1 scattering (taFEN1 exp.data) to the theoretical scattering of the known high resolution structure of FEN1 from the archeaon *Methanococcus jannaschii* (mjFEN1) (PDB code: 1A77) (Hwang et al., 1998) resulted in a reasonable fit (chi of 3.3) suggesting that taFEN1 constitutes a similar constitution in solution (blue in Figure 4.51). The comparable overall shape of in solution taFEN1 and the mjFEN1 crystal structure is also indicated by similar Rg values (24.3 Å and 23.5 Å,

respectively). The *ab initio* Dammif model of taFEN1 (magenta in Figure 4.51; chi of 2.3) supports this assumption which is also illustrated by the superposition of the taFEN1 *ab initio* Dammif model and the mjFEN1 crystal structure (Figure 4.51 B).

The stoichiometry of the taFEN1-taPCNA complex was analyzed by modeling one, two, or three taFEN1 monomers bound to taPCNA (Figure 4.52). The best fit was obtained with 1 taFEN1 bound, followed by 2 taFEN1 molecules, whereas 3 taFEN1 provided a very poor fit (chi values of 9.2, 14.4, and 27.9, respectively). Comparison of the experimental data with the fit for apo PCNA without any FEN1 bound also exhibits a bad fit (chi of 25.0). Thus the SAXS data clearly show that the 1 taFEN1 bound model was the preferred stoichiometry in solution. Notably, the chi values are relatively high because the PCNA structure was not minimized, resulting in variations as also observed for apo taPCNA.

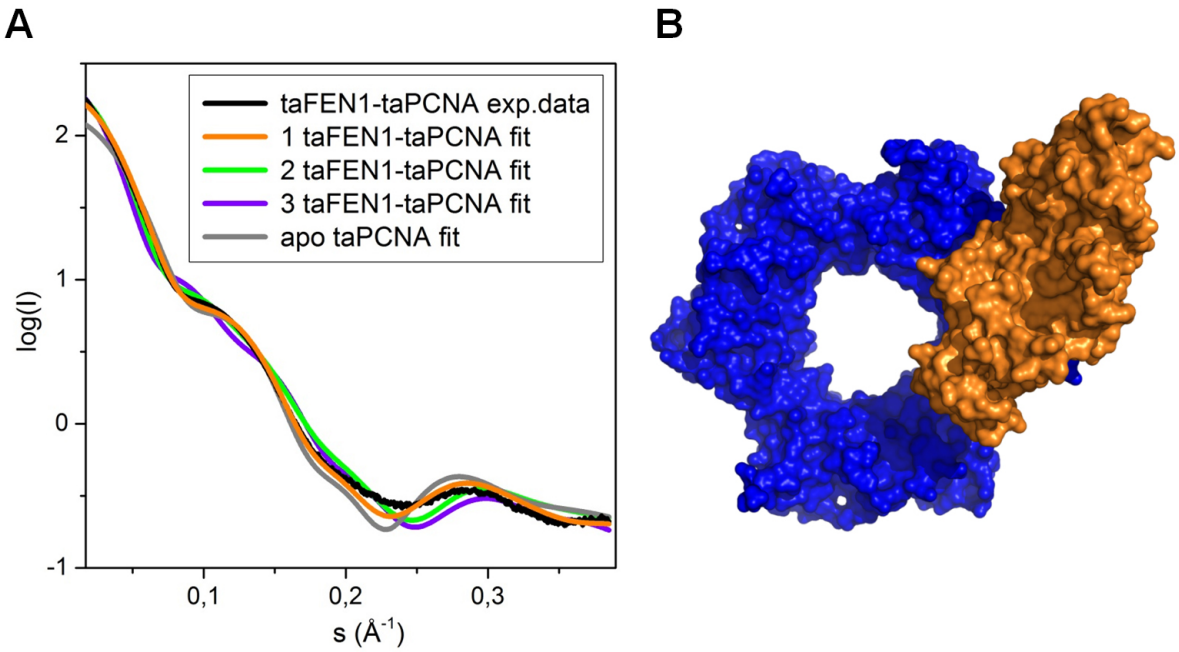


Figure 4.52: SAXS model of the taFEN1-taPCNA complex.
(A) Experimental 1D SAXS intensity data for the taFEN1-taPCNA complex (black) and fits for one (orange; chi of 9.2)), two (green; chi of 14.4) and three (magenta; chi of 27.9) taFEN1 proteins bound to the taPCNA trimer. A fit for apo taPCNA without any taFEN1 molecules bound is shown in grey (chi of 25.0). **(B)** Rigid body CORAL model of the complex with one taFEN1 monomers (orange) bound to the PCNA trimer (blue). For PCNA modeling the crystal structure of hsPCNA (PDB entry 1VYM) (Kontopidis et al., 2005) and for FEN1 the crystal structure of mjFEN1 (PDB code: 1A77) (Hwang et al., 1998) were used, respectively.

This observation is in agreement with ASEC interaction studies on taFEN1 and taPCNA (Figure 4.53). The chromatogram shows single peaks for the taFEN1 monomer (16 ml) and the taPCNA homotrimer (13.5 ml) while the mixture of both

indicates stable complex formation by the new appearing peak at a lower elution volume (12.5 ml). Remarkably, the peak representing apo taFEN1 (16 ml) is only partially reduced upon complex formation indicating that not all taFEN1 molecules bind to taPCNA.

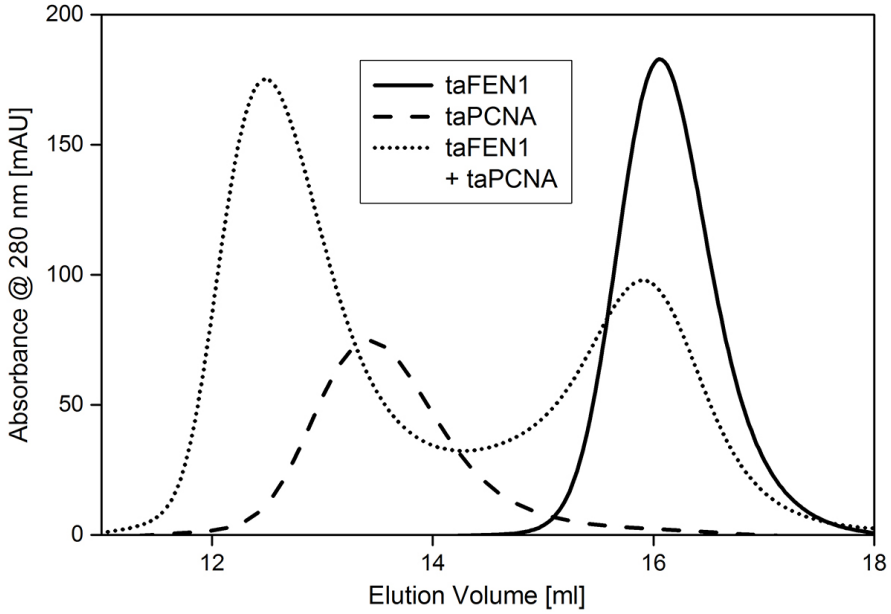


Figure 4.53: taFEN1-taPCNA interaction.
Analytical size exclusion chromatography (ASEC) profiles of taFEN1 (solid), taPCNA (dashed), and an equimolar mixture of both (dotted). The peak at a lower elution volume (12.5 ml) representing a higher molecular weight indicates stable complex formation. Only partial reduction of the apo taFEN1 peak upon complex formation indicates incomplete binding of taFEN1 to taPCNA. Protein concentrations of 20 µM were applied.

Since FEN1 binds PCNA with a K_D of 60 µM (Bruning and Shamoo, 2004), the here observed 1:3 ratio of taFEN1 to taPCNA can be explained by the used protein concentrations. SAXS experiments were conducted at a range of 0.5-10 mg/ml which, assuming a 1:1 mixture of taFEN and taPCNA, corresponds to 2-40 µM. For ASEC analysis a concentration of 20 µM of each protein was applied. Steric hindrances resulting from the flexibility of the first bound taFEN1 molecule to the taPCNA trimer might disfavor binding of a second taFEN1 unless all other taPCNA trimers are also occupied at least once.

4.10 DNA Processing Activity

Due to the difference in domain architecture and the unique position among the Hef protein family, taHef was also analyzed with respect to its DNA processing activity in helicase and ATPase assays. It was shown previously that the helicase and nuclease entities of pfHef exhibit a cooperative function on stalled replication forks (Komori

et al., 2004). However, the pfHef helicase part alone is also able to dissociate fork-structured DNA (Nishino et al., 2005b). Accordingly, two different DNA substrates were used presenting secondary structures that resemble replication fork intermediates in combination with a wide taHef concentration range (50 nM, 500 nM, and 5 µM): (1) A fork with double-stranded Y-arms and (2) an open fork with single-stranded Y-arms.

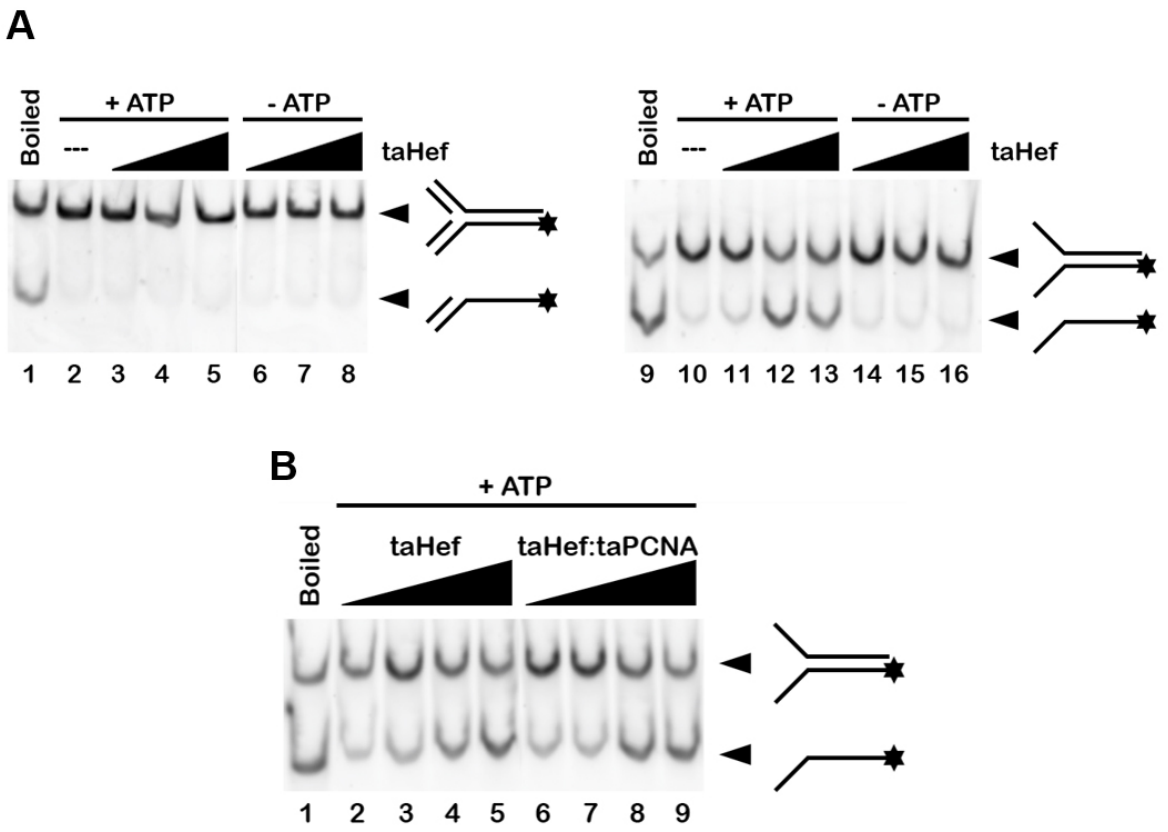


Figure 4.54: Helicase activity of taHef apo and taHef in complex with taPCNA. **(A)** Lanes 1-8: taHef is unable to unwind a fork-structured DNA substrate with double-stranded Y arms. 1) Positive DNA control, 2) no protein but 5 mM ATP were added (negative control), 3)-5) increasing taHef protein concentrations in the presence of ATP. Lanes 3-8 display no difference to the negative control indicating no helicase activity of taHef on the fork DNA substrate. Lanes 9-16: taHef unwinds an open fork DNA substrate with single-stranded Y-arms in the presence of ATP. 9) positive DNA control, 10) no protein but 5 mM ATP were added (negative control), 11)-13) in the presence of ATP taHef separates the open fork DNA, 14)-16) in the absence of ATP taHef is incapable to unwind the DNA substrate. The asterisk at the schematically drawn DNA stem indicates the Cy3 label attached to the 5' end of the bottom strand. Re-annealing of the separated DNA strands was prevented by the addition of a capture oligonucleotide. Concentrations: 50 nM, 500 nM, and 5 µM protein; 200 nM DNA substrate; 2 µM capture oligo; 5 mM ATP. **(B)** Helicase activity of apo taHef compared to the taHef-taPCNA complex on the open fork substrate. In both cases the DNA is unwound at a protein concentration of 200 nM (lanes 4 and 9) indicating no stimulation of taHef in the presence of equimolar taPCNA. Concentrations: 50 nM, 100 nM, 200 nM, and 500 nM protein; 200 nM DNA, 2 µM capture oligo; 5 mM ATP.

taHef was able to unwind the open fork in an ATP-dependent process but failed to dissociate the forked DNA (Figure 4.54 A). Interestingly, taHef required single-stranded DNA overhangs to initiate helicase activity whereas unwinding was blocked in the absence of a ssDNA overhang. pfHef, however, is able to unwind the fork substrate comprising double-stranded Y-arms (Nishino et al., 2005b). This observed variability in substrate specificity might be related to sequence differences in the thumb-like insertion domain. In addition, the open fork led to a moderate ATP hydrolysis reaction by taHef of about 3 nM ATP / nM protein / sec while the presence of the fork substrate elevated the ATPase activity to about 15 nM ATP / nM protein / sec (Figure 4.55).

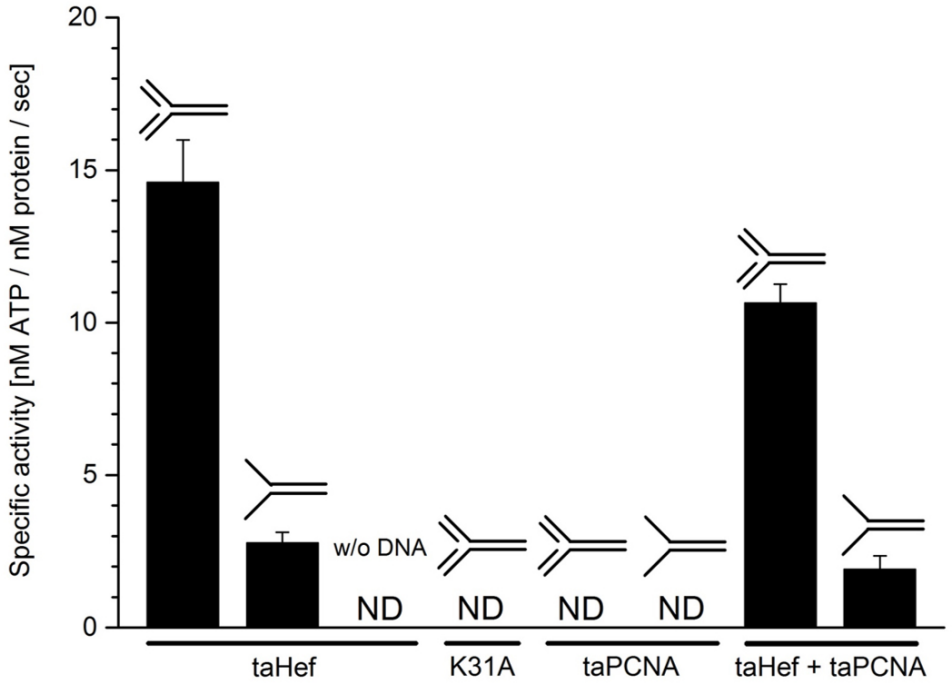


Figure 4.55: ATPase activity of taHef apo and taHef in complex with taPCNA. Specific ATPase activity of taHef, taPCNA and the taHef-taPCNA complex stimulated by fork-structured DNA substrates. ATP hydrolysis by taHef is strongly elevated in the presence of fork DNA with double-stranded Y-arms that cannot be unwound (~15 nM ATP / nM protein / sec). The open fork DNA which can be unwound by taHef clearly displays a lower stimulation effect on ATPase activity (~3 nM ATP / nM protein / sec). In the absence of DNA no ATP hydrolysis can be observed. taHef K31A comprises a mutation in the Walker A motif (helicase motif I) which is essential for ATP binding and subsequently leading to the incapability of ATP hydrolysis. Equimolar taPCNA does not show ATPase activity neither in the presence of fork nor open fork DNA. The taHef-taPCNA complex displays ATP hydrolysis activity that, within the error range, is similar to taHef apo.

To explore the physiological role of the Hef-PCNA interaction, we then compared taHef apo and taHef in complex with taPCNA in helicase and ATPase assays. For the helicase activity assays we focused on the open fork DNA substrate and utilized a protein concentration gradient ranging from 50 to 500 nM taHef in the presence and

absence of equimolar amounts of PCNA. The DNA was already fully separated at a taHef concentration of 200 nM (Figure 4.54 B) irrespective of the presence of PCNA, indicating no stimulation of taHef helicase activity through its interaction with PCNA. Analysis of the ATPase activity of the taHef-taPCNA complex involving fork and open fork DNA substrates showed similar results compared to apo taHef (Figure 4.55).

4.11 Crystal Structure of ctPCNA

In the context of studying the interaction between the FANCM homolog Mph1 from the thermophilic fungus *Chaetomium thermophilum* (ctMph1) and ctPCNA (Chapter 4.12), it was also important to characterize apo ctPCNA. Due to high thermostability eukaryotic proteins derived from *Chaetomium thermophilum* are particularly suitable for crystallization approaches and for an interpretation of the intended ctMph1-ctPCNA complex crystal structure the comparison with the single components is crucial. Thus after the successful purification of ctPCNA in high amounts (Figure 4.13) crystallization experiments were conducted.

4.11.1 Crystallization, Data Collection, and Refinement

Recombinantly expressed and purified ctPCNA crystallized in multiple different conditions within hours and up to a few days at 20 °C. All tested crystals displayed the same space group and unit cell dimensions. The best diffracting crystals (~ 2 Å) were obtained by a sitting drop vapor diffusion setup using a crystallization solution containing 0.2 M sodium malonate and 20% PEG 3350 and mixing equal volumes of 0.3 µl protein and reservoir solution equilibrated against 40 µl reservoir solution (Figure 4.56). ctPCNA was present at a concentration of 70 mg/ml (2.2 mM) in 20 mM Tris/HCl, 200 mM NaCl, 2 mM DTT, pH 7.5. Setups were performed automatically using the Zinsser Honeybee liquid handling system but also manually prepared screens yielded similar crystals. Prior to x-ray diffraction testing and data collection, the crystals were transferred into a cryo-protective solution containing the reservoir solution and 20% PEG 400. The crystals were flash cooled in liquid nitrogen and data collection was performed at 100 K.

High resolution diffraction data were collected at beamline ID29 at the ESRF at a wavelength of 0.920 Å and an oscillation range of 0.2° for a total rotation of 360°. The crystals belonged to the H-centered trigonal space group H3 with unit cell dimensions of a = b = 147.3 Å, c = 42.2 Å and $\alpha = \beta = 90°$, $\gamma = 120°$ and contain one molecule per asymmetric unit. With the help of Florian Sauer the diffraction data were processed using XDS and scaled with XSCALE (Kabsch, 2010). Unfortunately, this data set was severely anisotropic, with diffraction limits of 2.0 Å along the a* and b* directions, but only 3.0 Å along the c* direction. Thus the UCLA MBI *Diffraction Anisotropy Server* was used for ellipsoidal truncation and anisotropic scaling (Strong et al., 2006). This server was used to remove very weak data and thereby the highest resolution

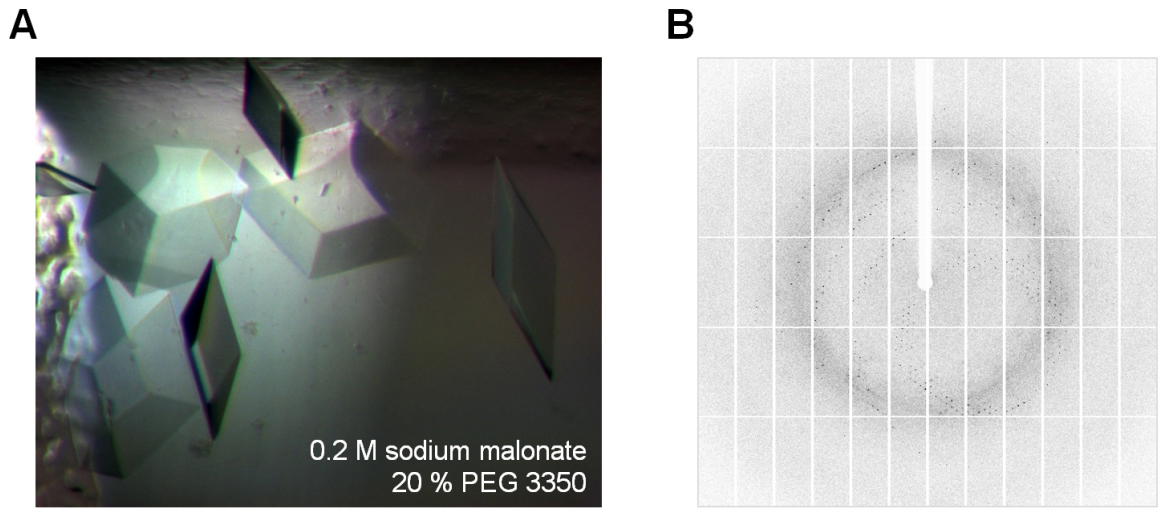


Figure 4.56: ctPCNA protein crystals and an exemplary diffraction image.
(A) Protein crystals of ctPCNA (70 mg/ml; 2.2 mM) that were grown by the sitting drop vapor diffusion method. 0.3 µl protein solution combined with 0.3 µl reservoir solution were equilibrated against 40 µl reservoir solution and revealed crystals of 200 x 200 x 20 μm^3 after hours to a few days. **(B)** One exemplary diffraction image of ctPCNA that was taken at beamline ID29 (ESRF) indicating a resolution range of 2-3 Å.

shell (2.27-2.20 Å) with a completeness of 18.9 % became unacceptable. Therefore the data were truncated at 2.54 Å (completeness of 74.3 % at 2.63-2.54 Å) and only these data processed by the UCLA MBI *Diffraction Anisotropy Server* were further used for refinement leading to a better interpretability of the resulting electron density maps. The structure was solved by MR using human PCNA (PDB code: 1VYM) (Kontopidis et al., 2005) as a search model. The model was built manually in Coot and refinement was carried out against the highest resolution data set up to 2.54 Å which yielded an R factor of 17.4 % (R_{free} = 23.5 %) using phenix.refine (Emsley et al., 2010; Adams et al., 2010). Final structure validation was performed using MolProbity (Chen et al., 2010). Although the initial data itself were poor the corresponding refinement statistics indicate a good model (Table 4.4).

4.11.2 Structure Analysis

The asymmetric unit contains one monomer and the complete ctPCNA homotrimer is generated by symmetry-related molecules. Thereby a typical PCNA quaternary structure with the pseudohexameric symmetry is established. ctPCNA consists of two structurally similar domains (N-terminal domain I and C-terminal domain II) (Mailand et al., 2013) that are joined by β4/β18 interactions and are connected by the interdomain connecting loop (IDCL). The anti-parallel β-sheets are located at the exterior while the α-helices constitute the interior that is responsible for DNA binding (McNally et al., 2010). The IDCL is positioned between β9 of domain I and β10 of domain II.

Table 4.4: Data collection & refinement statistics ctPCNA.

Data set		
Space group	H3	
Wavelength (Å)	0.920	
Cell dimensions		
a,b,c (Å)	147.3, 147.3, 42.2	
α, β, γ (deg)	90, 90, 120	
Resolution (Å)	42.5-2.20 (2.63-2.54) [2.27-2.20]	
	XSCALE	*UCLA anisotropy server*
Completeness (%)	99.6 (99.9) [99.8]	77.4 (74.3) [18.9]
R_{merge}[a]	0.091 (0.459) [2.388]	0.082 (0.374) [0.892]
Mean $I/\sigma I$[b]	10.5 (6.1) [2.5]	13.1 (8.1) [6.8]
Redundancy	6.8 (7.1) [7.0]	5.5 (5.8) [1.6]
No. of unique reflections	17227	10727
Refinement		
Resolution (Å)	42.5-2.54	
R_{cryst} / R_{free}[c]	0.174 / 0.235	
No. of working / test reflections	10200 / 527	
No. of protein / solvent atoms	1942 / 15	
Coordinate error	0.32	
Wilson B-factor (Å^2)	48.94	
Overall average B-factor (Å^2)	67.67	
Rms[d] deviations from ideal values		
Bond length (Å)	0.008	
Bond angles (deg)	1.316	
Dihedral angles (deg)	14.886	
Planar groups (Å)	0.005	
Ramachandran statistics[e] (%)	94.2 / 5.0 / 0.8	

Values in parentheses refer to 2.63-2.54 Å used as the highest resolution shell and values in squared brackets refer to 2.27-2.20 Å, respectively.

[a] $R_{merge} = \Sigma(I - < I >)/\Sigma < I >$

[b] $I/\sigma I$ indicates the average of the intensity divided by its standard deviation.

[c] $R_{cryst} = \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is the same as R_{cryst} for 5 % of the data randomly omitted from the refinement.

[d] Root mean square.

[e] Ramachandran statistics indicate the fraction of residues in the favored, allowed and outlier regions of the Ramachandran diagram as defined by MolProbity (Chen et al., 2010).

Intermolecular β-sheets are formed by β9 and β13 resulting in a 9 β-sheets containing bent fold ranging from β15 to β5 which is comparable to the fold from β6 to β14 (Figure 4.57).

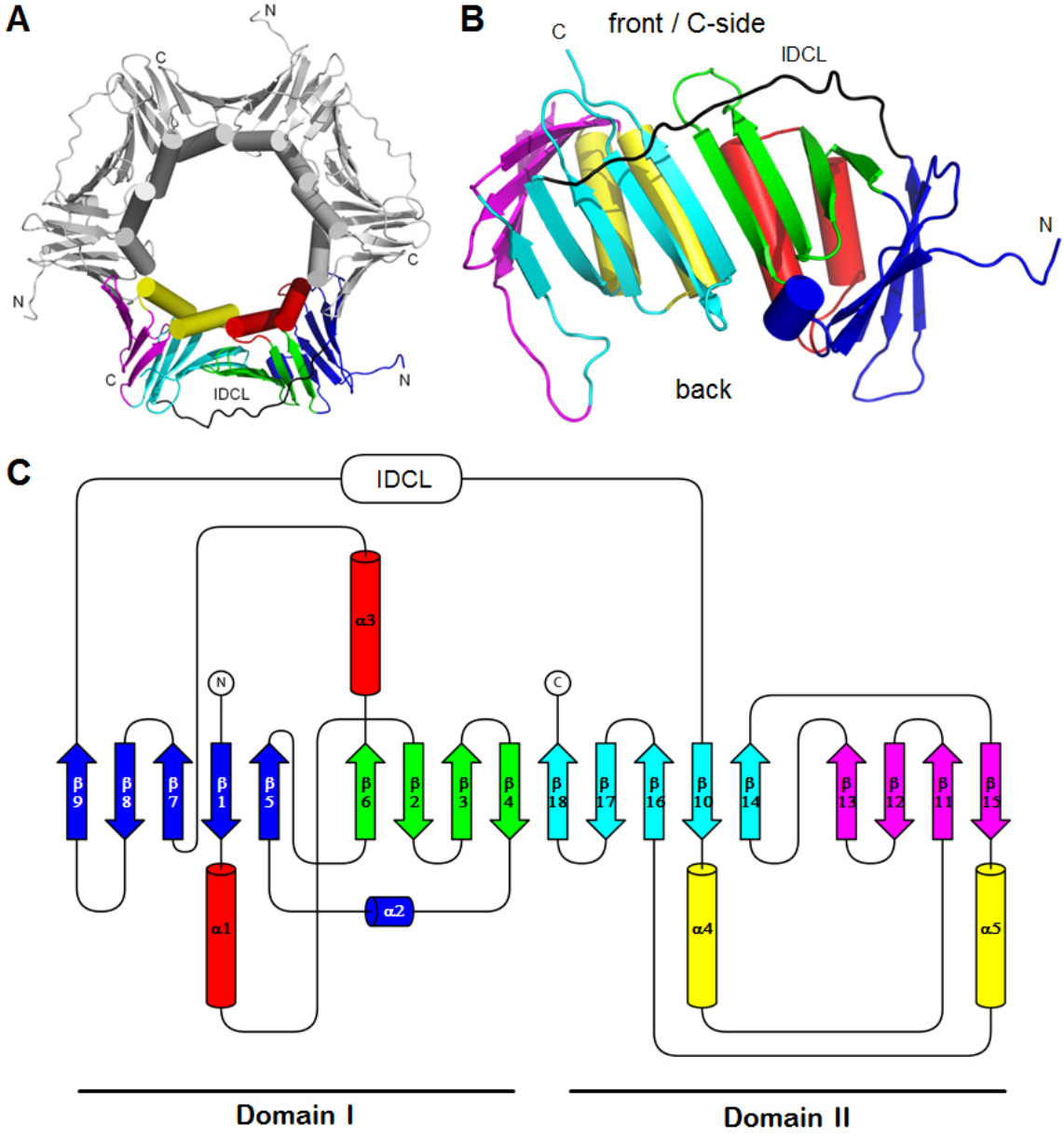


Figure 4.57: Crystal structure and 2D topology of ctPCNA.
(A) The crystal structure of ctPCNA comprises one monomer per asymmetric unit and the homotrimer is generated by symmetry mates of the trigonal space group H3. One monomer is depicted in color (in accordance to the 2D topology illustration), the two other monomers are shown in grey. View from the front / C-side. The N-terminal extension represents the partial TEV-site of the uncleaved His_6-tag. **(B)** One monomeric subunit in a detailed side view (top: front / C-side; bottom: back). **(C)** Schematic representation of the ctPCNA 2D topology. N-/C-termini as well as the interdomain connecting loop (IDCL) are indicated.

The molecular structure as well as the function of DNA sliding clamps is highly conserved among all kingdoms of life (Bruck and O'Donnell, 2001). This observation is also reflected by the ctPCNA crystal structure which, in a superposition with other eukaryotic and also archaeal PCNAs, exhibits a well-fitting overlay (Figure 4.58). One remarkable difference between archaeal and eukaryotic PCNAs in general are the loops between β8-β9 and β13-β14, respectively, which stick out towards the back side of the protein (Figure 4.57). These loops of yet unknown function are not present in archaea and should be addressed in further mutational studies.

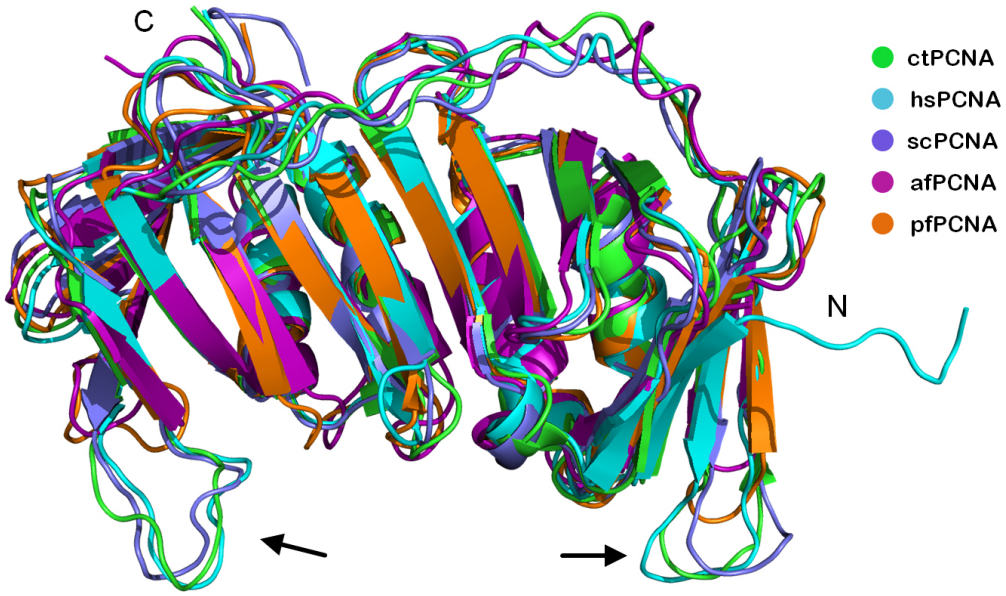


Figure 4.58: ctPCNA superposition with archaeal and eukaryotic PCNAs. Superposition of the ctPCNA monomeric subunit with other archaeal (af = *Archaeoglobus fulgidus* / PDB code: 1RWZ; pf = *Pyrococcus furiosus* / PDB code: 1GE8) and eukaryotic (hs = *Homo sapiens* / PDB code: 1VYM; sc = *Saccharomyces cerevisiae* / PDB code: 1PLQ) apo PCNA crystal structures. Arrows indicate loops that are only present in eukaryotes (left: β13-β14; right: β8-β9).

In the ctPCNA crystal structure an N-terminal protrusion can be observed which represents the partial TEV-site of the uncleaved His_6-tag. Interestingly, this extension together with the loops between β8-β9 and β13-β14 form important crystal contacts (Figure 4.59). The TEV-site specifically binds to the IDCL of a symmetry-related molecule while the loops that are located on the back side of the protein interact with the C-side of another symmetry-related molecule. Remarkably, this crystal packing creates solvent channels that would allow soaking with e.g. PIP-box derived peptides, which can be exploited in further experiments using this crystal form. As shown in Chapter 4.14 additionally to the PIP-box consensus sequence other residues might hold important positions for the regulation of the interaction between PCNA and binding partners. Therefore this ctPCNA crystal form presents an elegant scaffold for possible

structural characterizations of protein-ligand complexes.

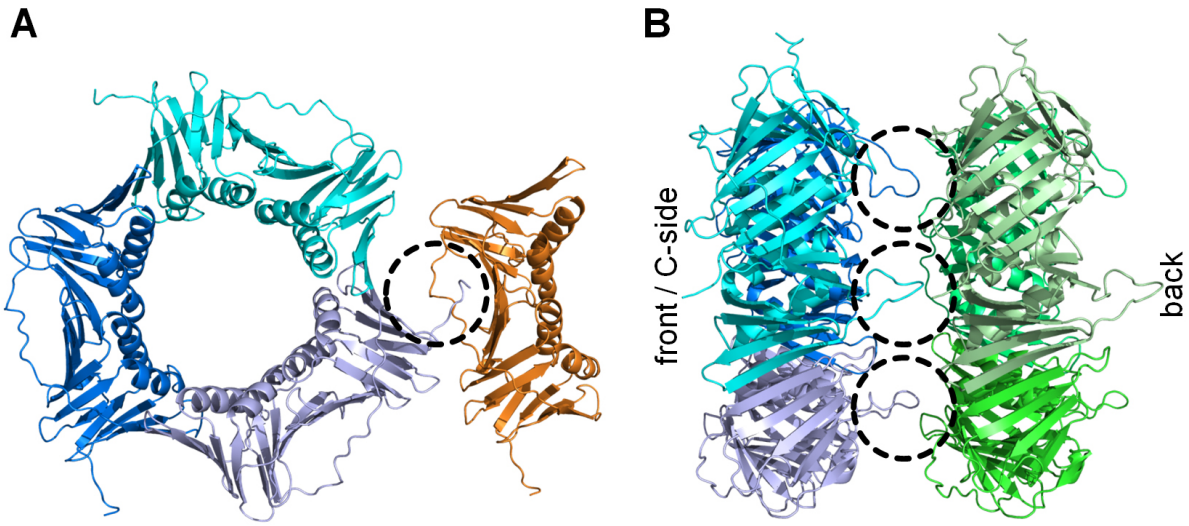


Figure 4.59: ctPCNA crystal contacts.
(A) The N-terminal extension of ctPCNA representing the partial TEV-site of the His_6-tag directly interacts with the IDCL of a symmetry related molecule. View from the front / C-side. **(B)** The loop region between β13 and β14 of the back side forms crystal contacts with the C-side of a symmetry related molecule. Crystal contacts are indicated by dashed circles.

It is therefore recommended to retain the His_6-tag for further co-crystallization or soaking experiments with peptides that comprise either the WT or a mutated sequence of the PIP-box from ctMph1 or any other PCNA interacting *Chaetomium thermophilum* protein. In this particular case, removal of the His_6-tag is likely to change the crystallization properties of the protein and could possibly prevent crystallization. However, for co-crystallization setups of ctPCNA together with full-length ctMph1 or the ctMph1$^{1-763}$ variant both approaches, removal and retention of the His_6-tag, should be tried.

4.12 Interaction between ctMph1 and ctPCNA

It was shown previously in yeast that the RFC-like protein Elg1 physically contacts MHF1/2 and genetically interacts with the FA-like proteins Mph1 (FANCM ortholog) and Chl1 (FANCJ ortholog) (Singh et al., 2013). The authors also speculate about a role of Pol30 (PCNA) in connecting various replication-dependent DNA repair mechanisms including FA repair. Furthermore, it was also proposed that Mph1 and Rad5 function in a common pathway (PRR) in which the poly-ubiquitination of PCNA plays a crucial role (Daee et al., 2012). However, the direct interaction between Mph1 and PCNA has not been proposed yet. Based on our data on complex formation of the archaeal proteins taHef and taPCNA it was thus speculated that this interaction may

also be conserved among eukaryotic systems. Intriguingly, most of the FA proteins as well as the repair pathway itself are only present in higher eukaryotes (Moldovan and D'Andrea, 2009). Our analysis of the fungal Mph1 protein could thus add another building block towards the conservation of functional FANCM homologs from archaea to small multicellular and higher eukaryotes and the interaction with the sliding clamp PCNA would indicate a fundamental role for these proteins.

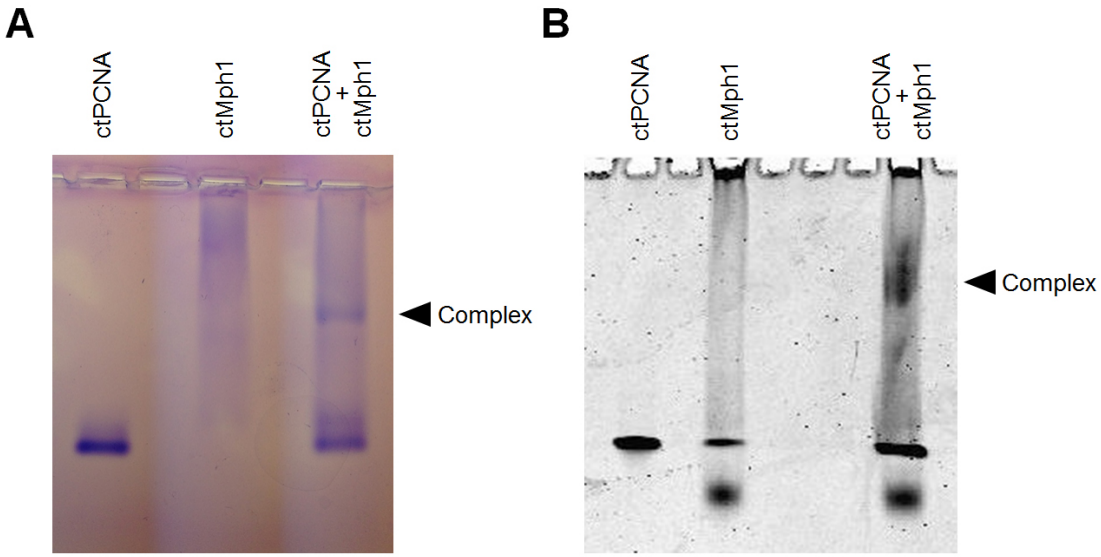


Figure 4.60: ctMph1-ctPCNA complex formation on native gel electrophoresis. 10 µM of purified ctMph1 and ctPCNA as well as an equimolar mixture of both were incubated in standard-buffer for 30 min at RT before loading on the gel. **(A)** 1% standard agarose gel in TEA-buffer. **(B)** 6% native PAA gel (37.5:1 AA:BisAA) in 0.5 x TBE-buffer. Both gels were Coomassie-stained to visualize the protein content. The new appearing bands marked in the mixture indicate complex formation of ctMph1 and ctPCNA.

Sequence analyses revealed that Mph1 is also conserved in the model organism *Chaetomium thermophilum* (ctMph1), a thermophilic fungus that has been proposed to contain proteins with increased stability and could therefore be especially amenable for structural characterization. The genes encoding ctMph1 and ctPCNA were cloned into the pETM-11 bacterial expression vector and separately expressed and purified from *E. coli* RIL cells. Strikingly, in native gel electrophoresis experiments on both agarose and polyacrylamide gels ctMph1 and ctPCNA form a stable complex as indicated by the appearance of a new band (Figure 4.60).

Both running buffers TAE for the agarose gel and TBE for the polyacrylamide (PAA) gel had a basic pH value (pH 8.5 and 8.3, respectively) and thus enabled protein migration towards the same direction. While for ctPCNA a clear single band is visible, ctMph1 produces a rather undefined smear on the gels. However, the complex shows a well-defined band as indicated (Figure 4.60).

To exactly determine the amino acid sequences of the ctMph1 degradation products (Chapter 4.3) or rather to find out if the protein was truncated C- or N-terminally, mass spectrometry (MS) measurements were performed in collaboration with the group of Andreas Schlosser. The degraded ctMph1 protein was first separated and Coomassie-stained by standard SDS-PAGE and then the bands between 80 and 110 kDa were extracted from the gel (Figure 4.61) for subsequent MS analysis utilizing the method for acid-catalyzed ^{18}O-labeling of carboxyl groups (Haaf and Schlosser, 2012), in which the protein samples are digested using a specific protease such as GluC or trypsin in the presence of the isotope ^{18}O containing H_2O. Thus during hydrolysis ^{18}O is incorporated at the newly generated C-termini while the pre-existing C-termini remain unaltered. The fragmented protein is then objected to NanoLC-MS/MS and through the distinction between ^{16}O and ^{18}O C-termini the exact degradation site can be determined. Due to technical issues with respect to the protein size it was not possible to analyze the undigested samples with subsequent mass determination and thus the ^{18}O method was applied constituting a very solid and reliable alternative technique. The MS experiments were conducted by Jens Vanselow (RVZ, Würzburg) and the results clearly show that ctMph1 was degraded C-terminally (Figure 4.62).

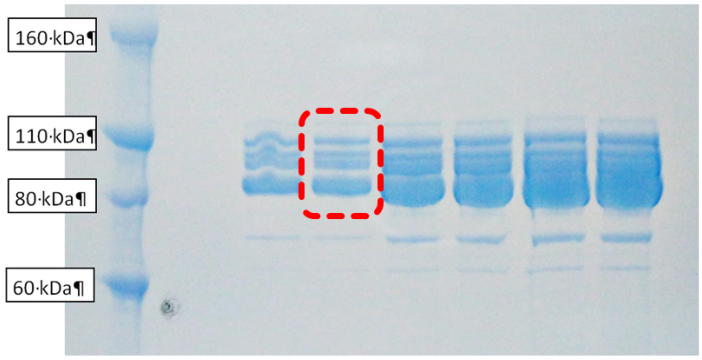


Figure 4.61: ctMph1 degradation products analyzed by mass spectrometry.
Dilution row of ctMph1 degradation products that were separated via standard SDS-PAGE and followed by Coomassie-staining. The indicated area (red frame) was extracted from the gel and subjected to ^{18}O NanoLC-MS/MS analysis.

The main degradation site within ctMph1 (orange dashed line in Figure 4.63) is located approximately between residues 770 and 910 while the N-terminus of the protein is unaffected. Interestingly, the degradation products contain the complete helicase entity which is analogous to taHef including the presumably unstructured N-terminus (Figure 4.3). C-terminally the degradation products comprise extensions of various lengths.

Strikingly, degraded ctMph1 comprises only putative PIP-box 1 (residues 108-114) which is located within the unstructured N-terminus while PIP-box 2 (residues

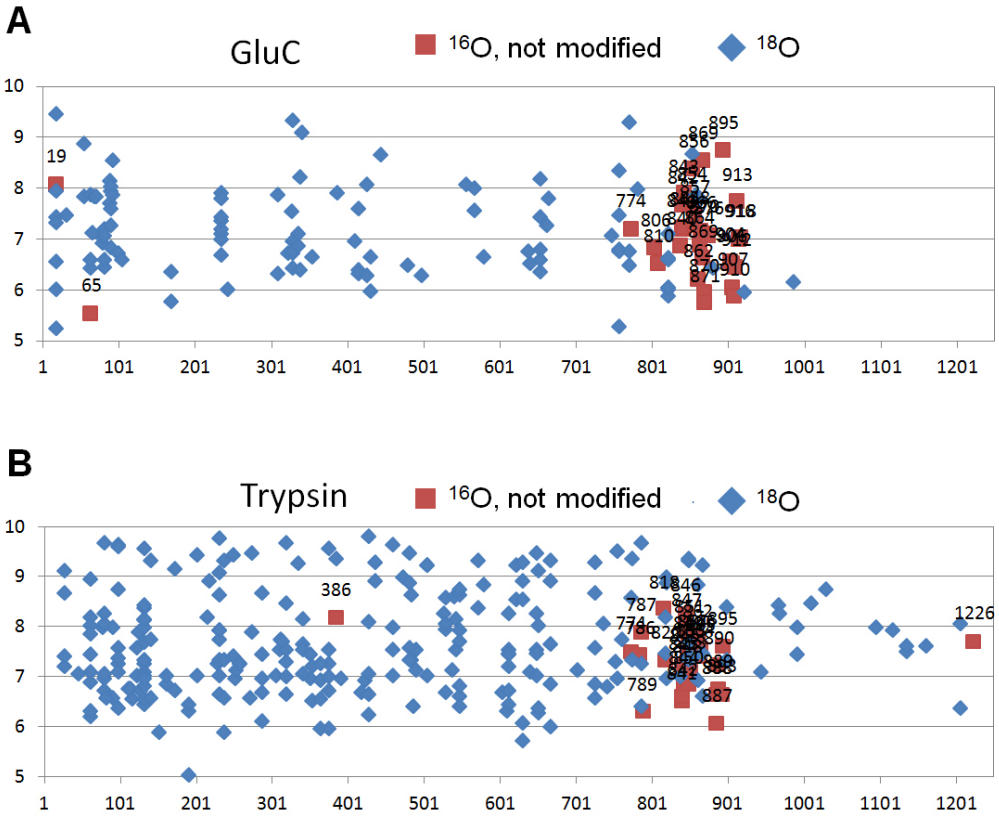


Figure 4.62: C-terminal degradation of ctMph1.
^{18}O isotope incorporation at newly generated C-termini compared to pre-existing ^{16}O C-termini during protease treatment of ctMph1 in combination with a NanoLC-MS/MS analysis reveals C-terminal degradation between residues 770 and 910 while the N-terminus of the protein is not impaired. **(A)** ctMph1 digested with GluC. **(B)** ctMph1 digested with trypsin.

918-925) is lost (Figure 4.63). Consequently, it is very likely that the ctMph1-ctPCNA interaction is mediated by PIP-box 1 assuming that ctPCNA binding depends on a classical PIP-box.

Notably, the N-terminal PIP-box is not conserved and thus not present in *Saccharomyces cerevisiae* Mph1 (UniProt: P40562). There are three possible explanations for this difference: (1) ctMph1 has a different function than its yeast homolog which then would not interact with PCNA, (2) yeast Mph1 interacts with PCNA via a non-canonical PIP-box that could not be identified due to it's strong alteration from the typical consensus sequence or (3) both yeast and *Chaetomium* Mph1 bind PCNA through a different yet not characterized interaction site. Thus it would be very interesting to additionally analyze Mph1 proteins from other organisms, e.g. *Saccharomyces cerevisiae*, for PCNA binding.

Importantly, further experiments are necessary for a precise characterization of this

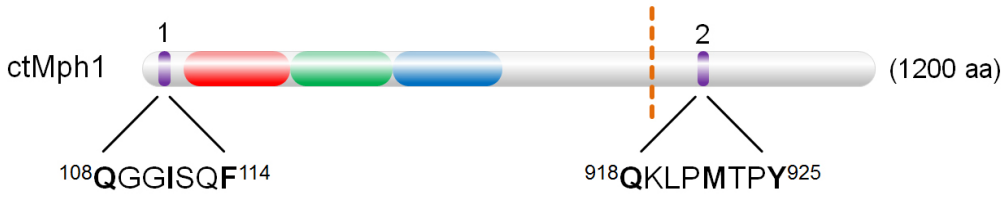


Figure 4.63: Localization of putative PIP-boxes in ctMph1.
Full-length ctMph1 (1200 amino acids; 134 kDa) contains two putative PIP-box motifs (purple; 1+2) and the corresponding amino acids sequence is depicted below. Unstructured short N- and extensive C-termini (grey) embrace the helicase entity (red: Hel1, green: thumb-like insertion, blue: Hel2). The C-terminal degradation is indicated by the orange dashed line.

interaction and also the role of the N-terminal PIP-box 1 of ctMph1. Thus new expression constructs of ctMph1 as well as ctPCNA were cloned. Since apo ctMph1 degrades to a stable final product, ctMph1 was mutated to generate a protein variant that resembles this degradation product. L764 was replaced by a stop codon resulting in S763 being the final C-terminal amino acid in this ctMph1$^{1-763}$ variant (Figure 4.3). This construct could potentially yield pure protein optimal for crystallization and biochemical experiments, e.g. DNA processivity analyses. Furthermore, ctPCNA was cloned into the pCDF-11 expression vector which is compatible with pETM-11. This enables co-transformation and co-expression of ctPCNA together with ctMph1 either full-length or the truncated ctMph1$^{1-763}$ variant. Possibly the co-expression with ctPCNA could stabilize full-length ctMph1 and thereby prevent degradation. Importantly, the corresponding "killPIP" variant ctMph1 I111R/F114S would clearly show if PIP-box 1 is the mediator for ctPCNA binding. This variant has not been generated yet. Another possibility to characterize the interaction between the putative ctMph1 PIP-box 1 to ctPCNA would be ITC experiments with peptides derived from ctMph1 comprising WT as well as the "killPIP" sequences. Furthermore, a short peptide including PIP-box 1 could also be co-crystallized with ctPCNA. The ctPCNA crystal structure shows that the hydrophobic binding pocket is well accessible and connected to solvent channels within the crystal. Thus additionally to co-crystallization soaking experiments could be pursued.

Notably, preliminary DNA binding assays were also conducted using degraded ctMph1 but the results were inconclusive and not reproducible.

4.13 Interaction between hsFANCM and hsPCNA

After the identification and characterization of the interaction between PCNA and the archaeal and eukaryotic FANCM homologs taHef and ctMph1, respectively, the next logical step was the analysis of a putative interaction between the human proteins. It is known that the FA DNA repair pathway removes ICL damages during replication (Moldovan and D'Andrea, 2009). Strikingly, the interaction with PCNA would clearly

explain how FA is connected to replication and furthermore it could elucidate the activation of the downstream repair pathway via FANCM. The analysis of the primary sequence reveals that hsFANCM contains eight potential classical PIP-boxes distributed over the entire protein but only the very N-terminal motif is conserved among FANCM proteins from a broad range of different higher eukaryotes. Thus a peptide that resembles the N-terminus of FANCM was analyzed in a binding assay. Furthermore, to verify this putative interaction between hsFANCM and hsPCNA under physiological conditions, CoIP experiments using human HEK293 cell lines were initiated.

Full-length FANCM as well as FANCM isoform 2 was cloned into pETM-11 and pETM-30 bacterial expression vectors. The only difference between both vectors is the additional N-terminal GST-tag in pETM-30 which is known to support solubility of the expressed protein (Smith and Johnson, 1988). Unfortunately, neither of the bacterial expression constructs yielded soluble protein that could be utilized for *in vitro* experiments (Chapter 4.3).

In contrast, hsPCNA could be easily expressed and purified from *E. coli* in high amounts. Notably, two bacterial expression constructs utilizing the vector pBADM-11 were generated. While in one construct the hsPCNA gene was standardly inserted into the multiple cloning site of pBADM-11, the other construct lacked the N-terminal His−6-tag, linker, and TEV-site of the vector to produce a native expression construct. Only the His_6-tagged version was expressed and purified whereas the native protein was not tested (Chapter 4.3).

4.13.1 Putative PIP-Boxes within hsFANCM

Eight putative canonical PIP-boxes distributed over the entire hsFANCM protein that may possibly mediate PCNA interaction could be identified by sequence analysis. Figure 4.64 shows a schematic representation of the protein with its domain organization and the localization of the eight putative PIP-boxes.

Figure 4.64: Localization of putative PIP-boxes within hsFANCM.
Full-length hsFANCM protein (2048 amino acids; 232 kDa) contains eight putative PIP-box motifs (purple). The N-terminus and the extensive linker region (grey) between the helicase (red: Hel1, green: thumb-like insertion, blue: Hel2) and the C-terminal enzymatically inactive ERCC4-like nuclease domain (yellow/crossed) are predicted to be unstructured by PsiPred.

Due to the unavailability of reliable structural information regarding hsFANCM functional predictions for the putative PIP-boxes are impossible. Only PIP-boxes 2

and 3 are located within the N-terminal helicase entity for which the crystal structures of taHef and pfHef can serve as structural models. Figure 4.2 shows that PIP-box 2 would be situated within helix α12 and PIP-box 3 within helix α18 (according to the taHef model). Therefore motifs 2 and 3 are likely to be buried in secondary structure elements and are not accessible. Hence their contribution to PCNA binding is unlikely. Nevertheless, they were treated as possible candidates in further studies. PIP-boxes 1 and 4-8 are positioned in the structurally mostly uncharacterized N-terminus and extensive linker region of hsFANCM (grey in Figure 4.64). It is unknown if there are secondary structure elements present at these positions. PsiPred does not predict any structural elements for motifs 1 and 4-8 but rather interprets them as unstructured (Figure 4.4).

If hsPCNA and hsFANCM form a complex it will be unclear which of the potential sites is/are used. To identify the potential PCNA interaction mediating PIP-box critical residues of each motif were mutated to abolish potential binding. As shown for taHef (Chapter 4.9) conversion of the hydrophobic residue at position 4 of the PIP-box to the positively charged and large arginine (I505R) in combination with the aromatic residue at position 7 mutated to the small and polar serine (F508S) disrupts PCNA interaction. The same mutations were thus introduced into hsFANCM and termed killPIP mutations.

Sequences of the putative PIP-boxes of hsFANCM and corresponding killPIP mutations are depicted in Table 4.5.

Table 4.5: Putative PIP-boxes in hsFANCM and corresponding killPIP mutations.

PIP-box	*Residues*	*Amino acid sequence*	*killPIP mutation*
1	5 - 12	**Q**RT**L**FQT**W**	L8R / W12S
2	354 - 361	**Q**GI**I**EGE**F**	I357R / F361S
3	525 - 532	**Q**LE**V**V**V**KQ**F**	V528R / V529R / F532S
4	671 - 678	**Q**SS**L**KKD**W**	L674R / W678S
5	730 - 736	**Q**LS**L**SE**W**	L733R / W736S
6	786 - 794	**Q**MED**V**TST**F**	V790R / F794S
7	1130 - 1137	**Q**DES**L**LL**F**	L1134R / F1137S
8	1538 - 1545	**Q**DSS**L**LD**F**	L1542R / F1545S

Putative PIP-boxes within hsFANCM and corresponding killPIP mutations.

It is possible that a putative PCNA interaction is mediated not just by a single PIP-box but rather by multiple PIP-boxes as shown for e.g. DNA Ligase I (Pascal et al., 2006). Under the assumption that two PIP-boxes of hsFANCM would contribute to PCNA binding with a total number of eight motifs the sum of all permutations would result in 28 possible combinations. In this study only mutations of individual PIP-boxes were generated.

As mentioned in Chapter 1.3.2 so far only three controversial pathological mutations have been reported for human FANCM (Meetei et al., 2005; Harutyunyan et al., 2011). Neither of these controversial mutations affect a putative PIP-box indicating that these phenotypes are not caused by the disturbance of the proposed PCNA interaction.

The amino acid sequences of various FANCM proteins were analyzed and solely canonical PIP-box 1 is conserved. Figure 4.65 shows an alignment of the 50 N-terminal residues of FANCM from different higher eukaryotes. Similar to taHef this PIP-box comprises only one aromatic amino acid in its consensus sequence. In taHef the aromatic residue is located at position 7 (F508) which is also the final C-terminal residue of the protein. In contrast, hsFANCM contains a threonine at position 7 (T11) and the aromatic amino acid tryptophane is situated at position 8 (W12).

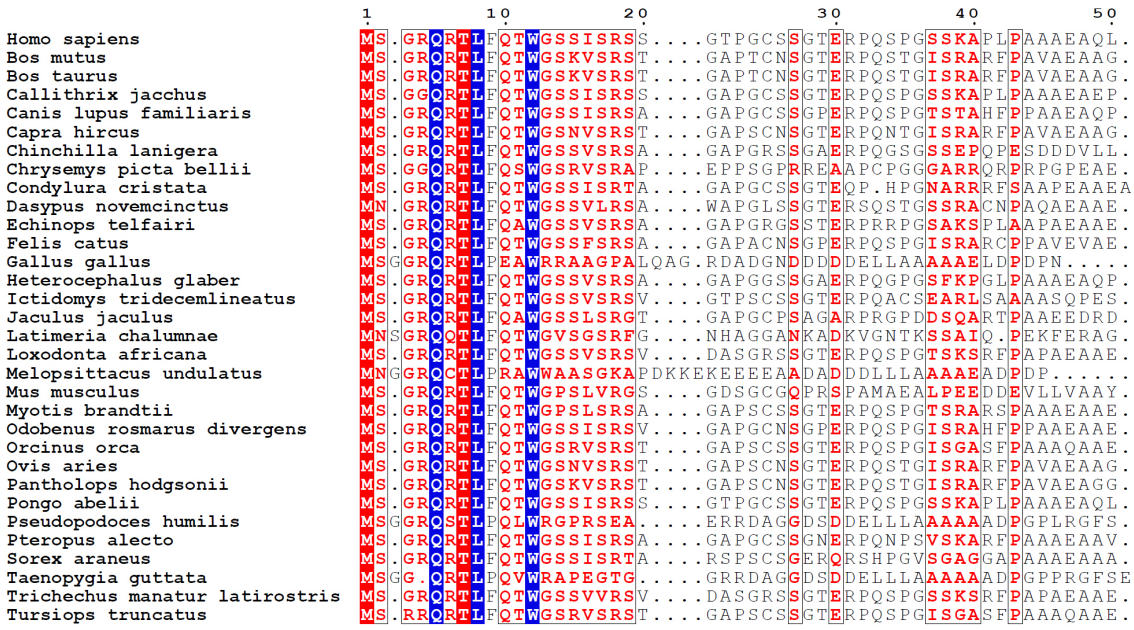


Figure 4.65: Alignment of the N-terminus of various FANCM proteins.
Alignment of the N-terminal 50 amino acids of FANCM proteins from various higher eukaryotes. Conserved residues are shown by red letters and strictly conserved residues in white with a red background. The N-terminally conserved putative canonical PIP-box motif is highlighted by a blue background.

Interestingly, some PCNA interaction partners such as FEN1 and p21 contain an additional binding interface which is located C-terminally immediately adjacent to the PIP-box and forms an intermolecular β-sheet with the IDCL of PCNA (Sakurai et al., 2005; Gulbis et al., 1996). Unfortunately, the sequences are not conserved and show barely any homology. Thus a similar feature cannot be predicted for hsFANCM and must be evaluated experimentally.

Furthermore, non-canonical PIP-boxes as present e.g. in TLS polymerases, eukaryotic clamp loader RFC, endonuclease ApnI, p53 or CDK inhibitors p57 were not considered (Moldovan et al., 2007). From the perspective of sequence analysis the identification of non-canonical PIP-boxes would be too complex because some non-canonical PIP-boxes exhibit massive alterations from the consensus sequence and can only be exclusively determined by experimental approaches.

At position 1505-1509 hsFANCM contains the amino acid sequence KHVAR which resembles a putative AlkB homologue 2 PCNA-interacting motif (APIM). For the human AlkB homologue 2 (hABH2) this APIM was shown to be a functional PCNA interacting motif distinct from the PIP-box. Furthermore, the APIM was predicted to mediate PCNA interaction for a wide range of different DNA processing proteins including transcription factors, E3 ubiquitin ligases, protein kinases and others. The consensus sequence is defined as (KR)-(FYW)-(LIVA)-(LIVA)-(KR) (Gilljam et al., 2009). hsFANCM comprises an H at position 1506 instead of (FYW) which does not ideally fit with the proposed consensus sequence. Nevertheless, a histidine with its aromatic side chain might be a functional substitution at this position. However, due to the low scientific quality of this study and the fact that since the discovery of the APIM only two other publications from the same authors have been published with respect to this particular PCNA binding motif its factual existence and biological function remain disputable (Gilljam et al., 2012; Müller et al., 2013). Especially the data on the putative interaction between XPA and PCNA mediated by the suspected APIM within XPA unmasks this motif as a misinterpretation. In this study only the mutation of the aromatic residue F to A at position 2 within the motif abolishes co-localization of the two proteins. The authors also missed to analyze DNA binding of XPA wild type in comparison to the F164A variant (Gilljam et al., 2012). Nevertheless, it would be fascinating to analyze this particular region in hsFANCM in more detail.

4.13.2 Binding of hsFANCM N-terminal PIP-box to hsPCNA

As described above, it was not possible to recombinantly express and purify full-length hsFANCM nor its isoform 2 for *in vitro* interaction studies with hsPCNA. In contrast, high amounts of correctly folded and functional hsPCNA can easily be expressed and purified from *E. coli* (Kontopidis et al., 2005). The preparation of recombinant hsPCNA is shown in Chapter 4.3. Because the N-terminus of FANCM comprising the putative PIP-box 1 is highly conserved (Figure 4.2) synthetic 17 amino acids long peptides that resemble this part of the protein were analyzed for hsPCNA interaction in ITC experiments (Figure 4.66).

Peptide hsFANCM$^{1-17}$ contains the wild type sequence and binds to hsPCNA with a K_D of 13 µM and a 1:1 stoichiometry. ΔG of this binding reaction has a value of -6.7 kcal mol^{-1} and presents a spontaneous reaction. ΔH comprises -12 kcal mol^{-1} and ΔS of -17.7 cal mol^{-1} K^{-1} indicates the transition from a disordered towards a defined constitution which is comparable to the taHef-taPCNA interaction (Figure 4.46).

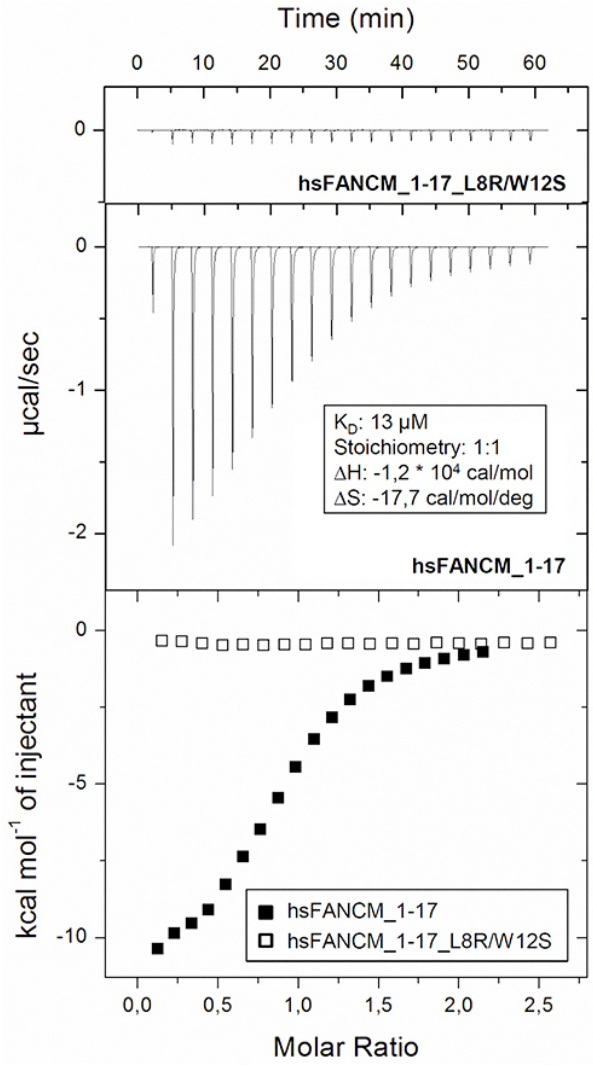


Figure 4.66: Interaction between the N-terminal PIP-box of hsFANCM and hsPCNA. Isothermal titration calorimetry (ITC) measurements of peptides that are based on the N-terminal 17 amino acids of hsFANCM (WT & L5R/W12S "killPIP" variant) and hsPCNA. hsFANCM$^{1-17}$-hsPCNA *in vitro* binding kinetics feature a dissociation constant (K_D) of 13 µM and a 1:1 stoichiometry. Thermodynamic parameters of the binding reaction: ΔH = -12 kcal mol^{-1}, ΔS = -17.7 cal mol^{-1} K^{-1} and a calculated ΔG = -6.7 kcal mol^{-1}. For the hsFANCM$^{1-17}$ L5R/W12S (killPIP) variant no heat release and thus no binding can be detected.

Likewise, the entropy term ($T\Delta S$) with a value of -5.3 kcal mol^{-1} counteracts the overall enthalpy driven reaction.

To verify that this binding event is mediated by a classical PIP-box interaction and to exclude additional binding sites within the N-terminus, a mutated variant of this peptide was also analyzed. The double variant L5R/W12S contains the same mutations as used for taHef to abolish PCNA binding (Chapter 4.9). As expected, this double

variant hsFANCM$^{1-17}$ L5R/W12S completely disrupts the interaction between the peptide and hsPCNA and thus the formation of the 3_{10}-helix which is characteristic for PIP-box is prevented. Notably, these findings indicate that the amino acids C-terminally to the PIP-box 1 motif do not interact with the IDCL of PCNA because no residual binding activity can be observed in the killPIP variant.

Although the K_D of 13 µM for hsFANCM$^{1-17}$-hsPCNA is about ten-fold lower than for taHef-taPCNA (K_D = 1 µM), it is possible that posttranslational modifications such as ubiquitination which does not occur in archaea or additional binding events of full-length FANCM could also contribute to the interaction. Interestingly, the affinities of binding partners towards PCNA show a broad range and are influenced by variations within the PIP-box motif and also posttranslational modifications. p21 is known to exhibit a very strong interaction with PCNA (K_D = 80 nM) while FEN1 binds with a 750-fold lower affinity (K_D = 60 µM). The p66 subunit of Pol δ binds with a K_D of 15 µM. In contrast, due to alterations within the PIP-box motif, Y-family TLS polymerases Pol η, Pol κ, Pol ι, and REV1 display a strongly reduced affinity towards PCNA. Since TLS polymerases contain additional ubiquitin-binding domains, they only bind to PCNA to substitute the replicative polymerases δ/ε upon PCNA mono-ubiquitination with an increased overall affinity. Other modifications such as SUMOylation are necessary for the recruitment of Srs2 or Elg1 to PCNA (Mailand et al., 2013). Importantly, these affinity-based regulations of PCNA interactions are crucial for the finetuned coordination of DNA processing proteins at the replication fork.

Furthermore, the here presented ITC data describe an artificial experimental setup. The utilized peptide represents only a very small portion of the entire hsFANCM protein and its PCNA binding properties are presumably different in the context of full-length hsFANCM. Indeed, the N-terminus of hsFANCM is predicted to be unstructured with an α-helix as the first secondary structure element starting at residue A44 (Figure 4.4). This flexibility in combination with accessibility is a precondition for a functional PIP-box. Buried within a secondary structure element or another rigidly defined area of the protein the PIP-box would be unable to interact with PCNA.

Since there is no experimentally verified structural information with respect to the N-terminus of hsFANCM or the full-length protein the actual conformation of this PIP-box remains ambiguous and this experiment should not be overrated. Nevertheless, these findings provide a starting point and corroborate PIP-box 1 as a strong candidate for mediating PCNA interaction.

4.13.3 Co-Immunoprecipitation hsFANCM-hsPCNA

For the analysis of the proposed interaction between hsFANCM and hsPCNA cell-based *in vivo* and *ex vivo* assays were conducted because it was not possible to recombinantly express and purify hsFANCM from *E. coli* for standard *in vitro* experiments such as ASEC or ITC as performed using archaeal taHef-taPCNA (Chapter 4.9).

The plasmid pIRESneo3 containing full-length hsFANCM and various mutants together with the empty vector as a negative control and also for generating further constructs was kindly provided by Weidong Wang & Yutong Xue (Xue et al., 2008). In addition to the provided plasmids hsFANCM killPIP 1-8 mutants (Table 4.5) were prepared and hsFEN1 and hsPCNA were cloned into the empty pIRESneo3 vector. hsFEN1 served as the positive control and was first cloned into the bacterial expression vector pETM-11 with an additional C-terminal HA-tag and an N-terminal His_6-tag, linker with the TEV-site retained. The entire expression cassette was subsequently extracted from pETM-11 and inserted into pIRESneo3. For hsPCNA cloning the bacterial pBADM-11 expression construct was used and in a first step the linker between the N-terminal His_6-tag and the TEV-site was replaced by an HA-tag. The entire expression cassette was then again extracted and inserted into pIRESneo3. Finally, the HA-tag in the pIRESneo3 construct was also replaced by a Flag-tag to allow CoIP using an α-Flag antibody and hsPCNA as the bait instead of hsFANCM. In this particular case the HA-/Flag-tags were added N-terminally because the C-terminus of PCNA is very close to the PIP-box binding pocket and could impair potential complex formation. For constitutive protein expression stable HEK293 cell lines were generated by Detlev Schindler (Human Genetics, Würzburg) using standard transfection and long-term G418 selection.

Prior to the actual CoIP experiment the monoclonal mouse α-PCNA antibody (PC10) was tested on various substrates (Figure 4.67). The antibody is very specific and detects hsPCNA at low nanomolar concentrations. It does not show any signal for purified taPCNA, *E. coli* whole cell lysate or A/G beads. Notably, the antibody strongly cross-reacts with the HA-antibody as indicated by the signals at 25 and 55 kDa. These two bands represent the light chain and the heavy chain of the α-HA antibody, respectively, when separated on a standard denaturing SDS-PAGE. Fortunately, these bands do not coincide with hsPCNA (~35 kDa) and can clearly be distinguished. Furthermore, this experiment highlights the stability of *E. coli* recombinantly expressed (pBADM-11) and purified hsPCNA indicated by the lanes for hsPCNA that was stored for 1 or 2 weeks (1w, 2w) at 4 °C.

Since CoIP experiments did not belong to our standard methods, the technique itself had to be established. Besides human cell culture this involved the optimization of subsequent cell lysis, extraction and enrichment of nuclear proteins, buffer conditions for antibody binding, controls, and the final Western blot detection. Therefore the positive control, a stably transfected HEK293 cell line overexpressing hsFEN1 which is known to interact with PCNA (Sakurai et al., 2005), was used to co-immunoprecipitate endogenous hsPCNA using an α-HA antibody. For this purpose three different buffer conditions (RIPA, LP, LS-HS; Table 2.19) were tested for both cell lysis and the following interaction analysis. Remarkably, the LP-buffer represented the only condition that sufficiently lysed the cells and nuclei and enabled successful CoIP of hsFEN1 and hsPCNA (red box in Figure 4.68). The LP-buffer was the only tested

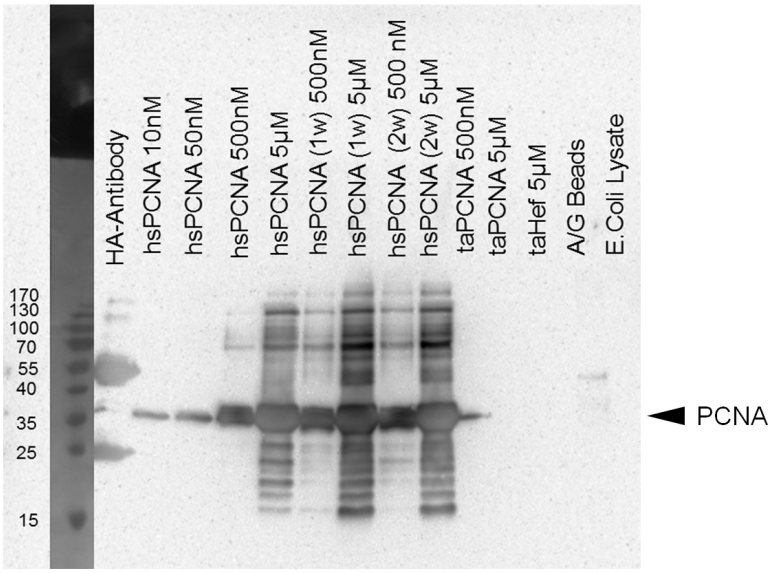


Figure 4.67: α-PCNA antibody specificity test.
Western blot using α-PCNA (PC10) as the primary antibody in a 1:5000 dilution. The secondary antibody was the α-mouse antibody derived from goat in a 1:10000 dilution. Purified hsPCNA (fresh, 1w = 1 week, 2w = 2 weeks stored at 4 °C) and taPCNA at various concentrations as well as other components of the CoIP experimental setup (HA-antibody, A/G beads) and *E. coli* whole cell lysate were tested.

buffer containing a detergent, Triton-X, which is necessary to gain access to nuclear proteins. Furthermore, the beads control (negative control) did not show any unspecific binding of hsPCNA to the A/G beads in the absence of hsFEN1. Interestingly, the α-HA antibody split into the heavy chain and the light chain was also visible on the blot (asterisk).

After the identification of the ideal conditions for CoIP studies of endogenous hsPCNA by HA-tagged hsFEN1, CoIP experiments were conducted using the same stably transfected HEK293 cells expressing HA-tagged hsFEN1 as the positive control, the empty vector as the negative control, and HA-tagged hsFANCM. Unfortunately, hsFANCM could not be visualized and concomitantly no hsPCNA was co-immunoprecipitated. Subsequent expression tests exhibited that neither hsFANCM nor the killPIP 2 variant were expressed although all cell lines were generated by G418 selection (Figure 4.69). In addition to α-HA also the specific α-FANCM was tested to visualize hsFANCM that probably lost the C-terminal HA-tag due to degradation. α-PCNA was applied as a control.

To evaluate if the concentration of hsFANCM in the whole cell lysate was too low for Western blot detection nuclear extraction was performed to enrich hsFANCM (Figure 4.70). Again the empty pIRESneo3 vector and hsFEN1 HEK293 cells lines were used as negative and positive controls, respectively. Additionally, hsFANCM WT and

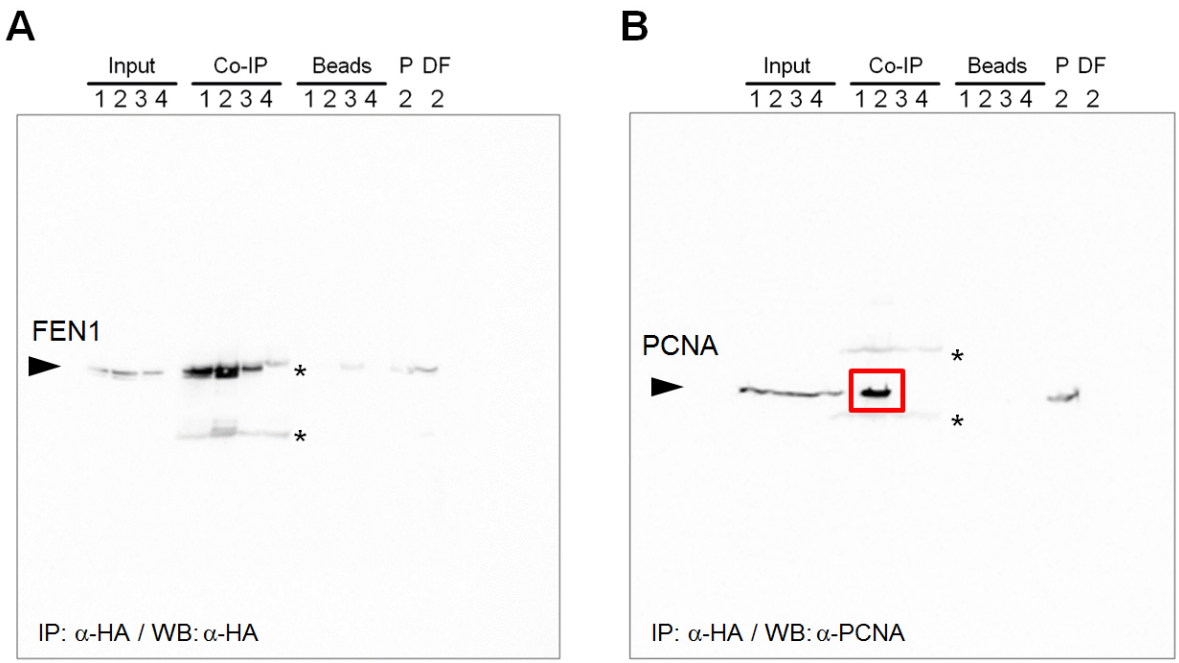


Figure 4.68: CoIP analysis with HA-tagged hsFEN1 and endogenous hsPCNA. CoIP utilizing an HA-antibody. HEK293 cells expressing HA-tagged hsFEN1 (1-3) and cells containing the empty pIRESneo3 vector (4, negative control) were lysed in different buffers (1: RIPA, 2: LP, 3: LS-HS, 4: RIPA). Input: soluble fraction of the lysate prior to the CoIP. CoIP: proteins that are extracted by the HA-antibody which is immobilized on A/G beads. Beads: control to visualize unspecific binding of the tested proteins to the A/G beads in the absence of the HA-antibody. P: pellet after cell lysis. DF: flow-through after CoIP. **(A)** Western blot detection of HA-tagged hsFEN1 using α-HA as the primary antibody. **(B)** Western blot detection of hsPCNA using α-PCNA (PC10) as the primary antibody. Successful hsPCNA CoIP by hsFEN1-HA in LP-buffer is highlighted by the red box. Two slight bands visible at CoIP (1-4) indicated by the asterisk represent the light chain and the heavy chain of the denatured HA-antibody.

hsFEN1 cells were also treated with the ICL-inducing DNA damaging agent mytomycin C (MMC). On the Western blot utilizing the α-HA antibody there is only a clear signal for hsFEN1 visible while MMC stimulation does not influence the expression pattern or distribution within the cell. Unfortunately, this experiment again did not reveal any hsFANCM expression at all. No distinct bands were present as seen for the positive control hsFEN1 and the background of hsFANCM did not differ from the negative control. Neither the α-HA nor the α-FANCM displayed the supposedly overexpressed hsFANCM. Notably, on the α-FANCM blot in all nuclear (N) fractions a signal slightly below 70 kDa was observed which might represent endogenous hsFANCM isoform 2 (75 kDa). This commercial α-FANCM antibody was generated using hsFANCM isoform 2 as the antigen. Interestingly, the same band is also visible on the α-HA blot (both marked by "?" in Figure 4.70).

It was thus shown that the pIRESneo3 hsFANCM cell line did not express detectable

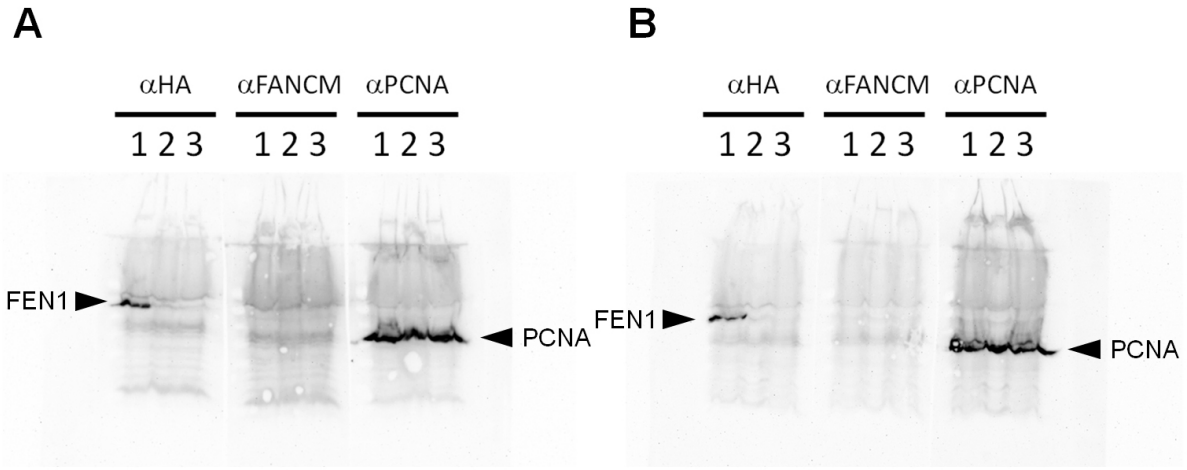


Figure 4.69: Expression analysis for hsFANCM WT, killPIP 2, and hsFEN1.
Western blot of whole cell lysate of two independent expression tests exhibit no expression for hsFANCM WT and the hsFANCM killPIP 2 variant while hsFEN1 is correctly expressed. **(A)** 1: hsFEN1, 2: empty vector, 3: hsFANCM WT. **(B)** 1: hsFEN1, 2: hsFANCM killPIP 2, 3: empty vector.

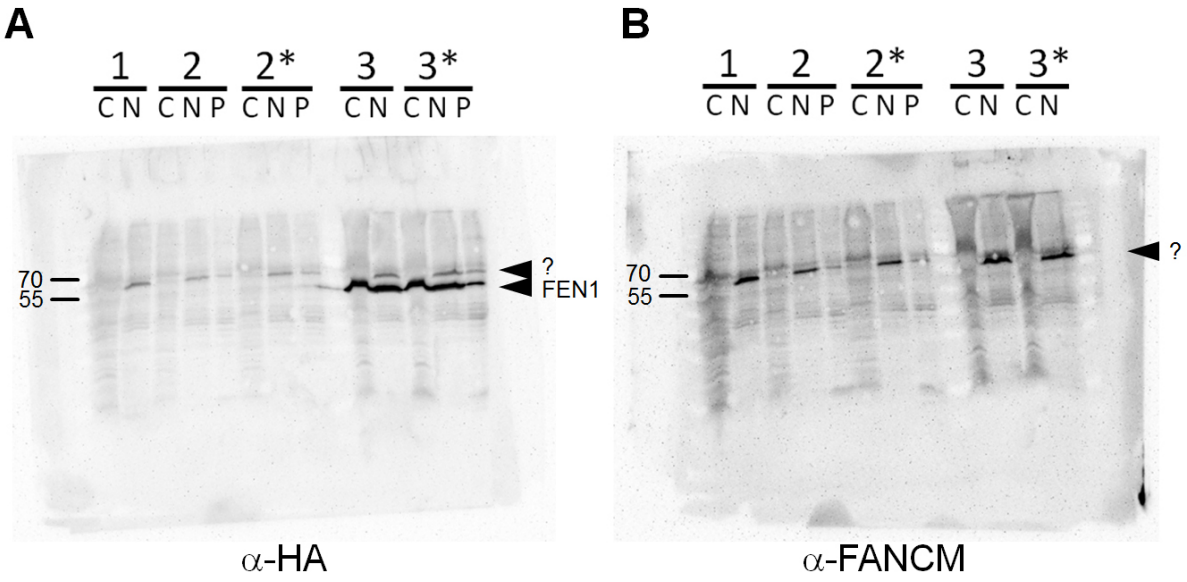
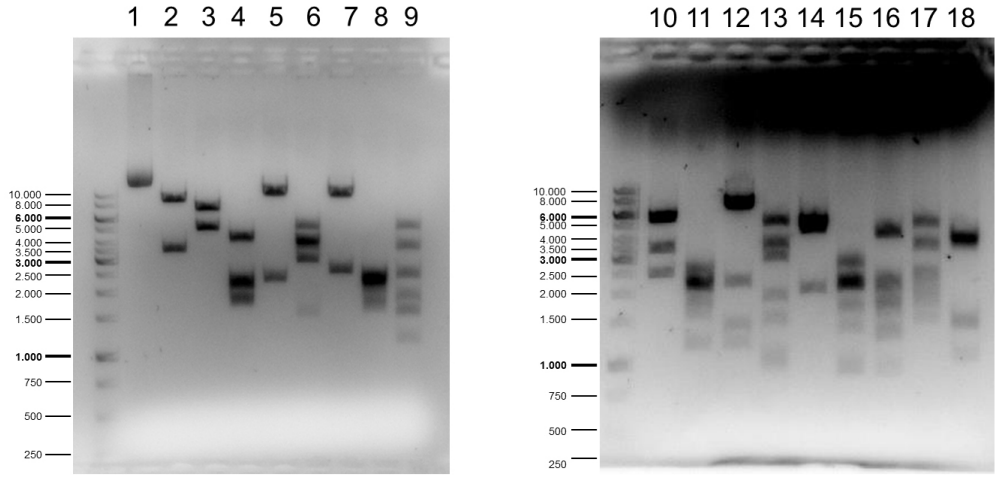


Figure 4.70: Nuclear extraction of hsFANCM and hsFEN1.
Western blot analysis after nuclear extraction separating the cytoplasm, nuclear, and the remaining pellet fractions using **(A)** α-HA **(B)** α-FANCM primary antibodies testing unstimulated and 200 ng/ml MMC-stimulated (*) HEK293 cells. 1: empty vector, 2: hsFANCM WT, 3: hsFEN1, C: cytoplasm, N: nucleus, P: pellet after nuclear extraction. Expression of HA-tagged hsFEN1 is not altered upon MMC stimulation and the protein is also equally distributed between the cytoplasm and the nucleus. The signal marked by "?" slightly below 70 kDa might represent the endogenous hsFANCM isoform 2.

amounts of hsFANCM protein. The vector was transfected again and a new stable hsFANCM cell line was generated but likewise did not show any hsFANCM expression although the correct sequences of all used and newly generated constructs were verified by sequencing. A restriction enzyme test digestion also confirmed the intact overall constitution of the pIRESneo3 hsFANCM plasmid (Figure 4.71). Only the hsFANCM WT construct was analyzed because this plasmid served as the template for all killPIP mutants.



1 (-) 11521
2 (NcoI) 3358, 8171
3 (SmaI) 4723, 6798
4 (BsaI) 1638, 1742, 2085, 2106, 3970
5 (SalI) 38, 2192, 9303
6 (ScaI) 2808, 3761, 4952
7 (StuI) 2429, 9092
8 (NcoI/BsaI) 270, 1476, 1638, 1886, 2085, 2088, 2106
9 (NcoI/ScaI) 1057, 1464, 1755, 2301, 4952

10 (NcoI/StuI) 330, 2429, 3358, 5412
11 (SmaI/BsaI) 512, 1126, 1493, 1742, 2085, 2106, 2477
12 (SmaI/SalI) 38, 1149, 1356, 2192, 6798
13 (SmaI/ScaI) 869, 977, 1831, 2892, 4952
14 (SmaI/StuI) 448, 1981, 4275, 4817
15 (BsaI/SalI) 23, 848, 864, 1332, 1638, 2058, 2106, 2642
16 (BsaI/StuI) 64, 859, 1247, 1574, 1742, 2085, 3970
17 (SalI/ScaI) 38, 445, 1747, 2018, 2333, 4952
18 (ScaI/StuI) 1004, 1383, 1425, 3761, 3948

Figure 4.71: Restriction fingerprint of pIRESneo3 hsFANCM WT.
DNA restriction enzyme digestion of the pIRESneo3 hsFANCM WT expression construct to verify its correct overall constitution. Numbers indicate the sizes of resulting fragments (bp) with a 100% cutting efficiency assumed. Midori-Green stained 1 % agarose gel.

At this point we initiated a collaboration with Weidong Wang (NIA/NIH, Baltimore, USA) who is an expert for cell-based analyses of FA. I provided all constructs and data for further experiments. All further CoIP examinations of the putative hsFANCM-hsPCNA interaction were performed in his laboratory. Notably, in these studies other expression vectors were used than the pIRESneo3 construct which did not yield protein expression. The experimental protocols and details have not been discussed so far but are related to previous publications (Xue et al., 2008). The following CoIP experiment was conducted and the figure prepared by Yutong Xue.

Interestingly, initial tests revealed that hsPCNA is co-immunoprecipitated by Flag-tagged hsFANCM when the cells have been stimulated by hydroxyurea (HU) while unstimulated or MMC-stimulated cells do not display complex formation of the two proteins (Figure 4.72). In this experiment two different α-PCNA antibodies were

evaluated and the mouse α-PCNA (PC10) was much more sensitive than the rabbit α-PCNA antibody which was needed to detect the small fraction of hsPCNA that was associated with Flag-hsFANCM. In further experiments the cells were stimulated with HU and the mouse α-PCNA antibody was utilized.

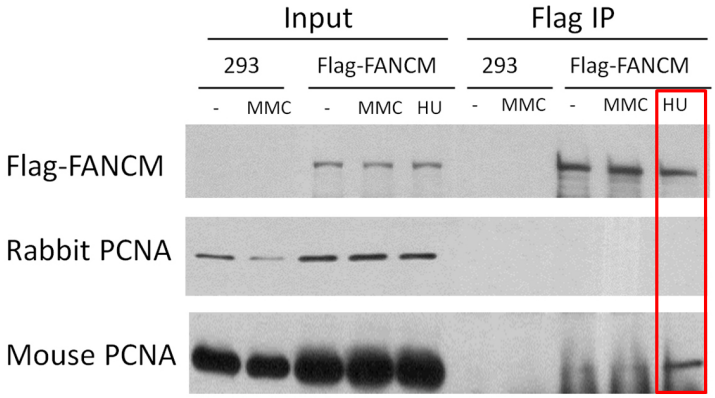


Figure 4.72: Flag CoIP with Flag-tagged hsFANCM and endogenous hsPCNA. Flag CoIP to immunoprecipitate Flag-tagged hsFANCM. hsPCNA was detected by two different (rabbit and mouse) antibodies and HEK293 cells were analyzed either unstimulated or stimulated with MMC or HU. Flag-hsFANCM was visualized using an α-Flag antibody. The CoIP of endogenous hsPCNA by overexpressed Flag-hsFANCM in HU-stimulated HEK293 cells is highlighted by the red box.

To analyze complex formation under native and physiological conditions the co-localization of hsFANCM and hsPCNA was characterized by immunostaining experiments with HeLa cells (Figure 4.73). First, the method itself was optimized and verified. Different cell permeabilization and fixation conditions were tested and for immunostaining of nuclear proteins 1% Triton-X in combination with 0.5% NP-40 was essential. To prevent cross-reactivity the secondary antibodies were highly cross-absorbed.

Endogenous hsFANCM as well as hsPCNA could be immunostained in HeLa cells but a clear co-localization was not observed neither in untreated nor in MMC-treated cells. However, hsPCNA was always present in the nucleus while hsFANCM seemed to be distributed over the entire cell and just migrated into the nucleus upon genotoxic stress. These data are in afreement with the CoIP data from the Wanf laboratory where only hydroxyurea (HU) triggered hsFANCM-hsPCNA complex formation while MMC did not (Figure 4.72). It would thus be very interesting to also test HU-stimulated HeLa cells for co-localisation and foci formation of hsFANCM and hsPCNA. Another elegant approach would be the additional usage of labeled ICL damages. $PtBP_6$ is a cisplatin analog that contains a photoreactive benzophenone moiety and is established for peptide mass fingerprint analysis for protein identification at cisplatin DNA damage sites (Zhu and Lippard, 2009). Notably, there are additional possibilities to induce fluorescently labeled ICL damages into a DNA substrate (Nowak-Karnowska et al., 2014; Skalski et al., 2010).

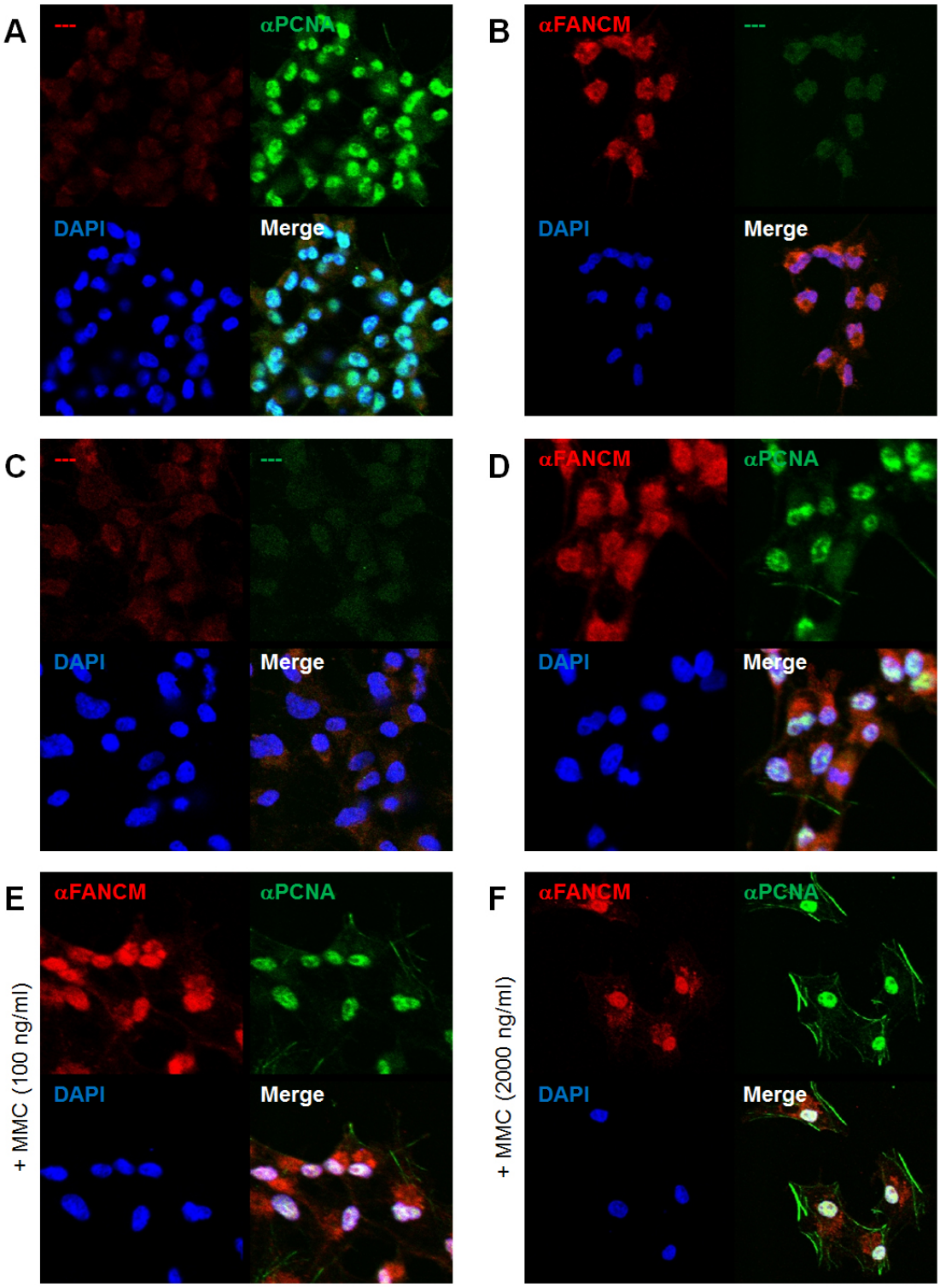


Figure 4.73: hsFANCM and hsPCNA immunostaining.
Immunostaining of endogenous hsFANCM (red) and hsPCNA (green) together with DAPI (blue) in MMC stimulated and unstimulated HeLa cells. **(A)-(C)** Staining controls: absence of either one or both primary antibodies. **(D)** Untreated cells stained with both antibodies. **(E)-(F)** Low and high dose MMC-treated cells (100 ng/ml and 2000 ng/ml, respectively).

4.14 Differences in PIP Binding Pockets

Because (1) DNA sliding clamps and also the PIP-box motif consensus sequence are highly conserved among archaea and eukaryotes and (2) in contrast to taPCNA, ctPCNA yielded much better diffracting crystals (Chapters 4.8 & 4.11), I aimed to co-crystallize a chimeric protein complex consisting of taHef and ctPCNA. Therefore to assure interaction between these proteins initially they were tested in analytical size exclusion chromatography (ASEC) experiments (Figure 4.74). Strikingly, archaeal taHef does not bind eukaryotic ctPCNA or hsPCNA although all proteins are functionally intact as indicated by the ctPCNA crystal structure (Chapter 4.11) as well as the interaction with ctMph1 (Chapter 4.12) and the hsPCNA interaction with hsFEN1 (Figure 4.74 D), respectively.

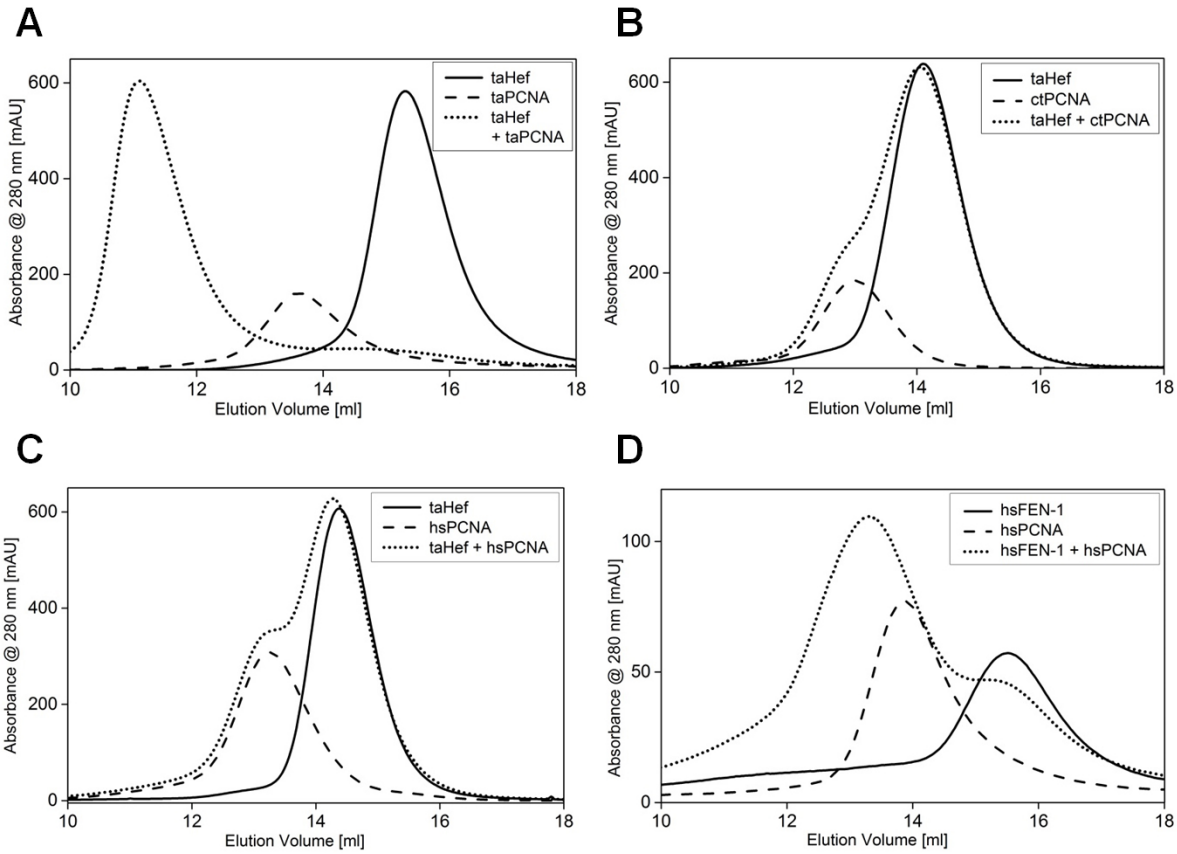


Figure 4.74: Analytical SEC of taHef and PCNA from different organisms. Analytical size exclusion chromatography (ASEC) of taHef (solid line) with **(A)** taPCNA, **(B)** ctPCNA, and **(C)** hsPCNA (apo PCNA: dashed line, mixture of both proteins: dotted line). **(D)** hsFEN1 (solid line) with hsPCNA (apo PCNA: dashed line, mixture of both proteins: dotted line) were analyzed as a positive control for hsPCNA. The usage of a different Sdx200 10/300 column for (D) resulted in a slight shift of the elution volumes.

The chromatogram for the hsFEN-hsPCNA mixture was not accurate enough to

clearly show complex formation but the respective fractions visualized via standard SDS-PAGE explicitly showed an interaction and thereby confirmed the correct function of the hsPCNA PIP-box binding pocket (Figure 4.75). Interestingly, hsFEN1 degraded as indicated by the double band and primarily the full-length protein (upper band) bound to hsPCNA. Since the PIP-box is located at the C-terminus of hsFEN1 it can be concluded that recombinantly in *E. coli* over expressed and purified hsFEN1 was degraded C-terminally. Notably, the partial reduction of the peak for single hsFEN1 upon complex formation with hsPCNA can be explained not only by the C-terminal degradation but also by the lower protein concentrations of 25 µM that were used for this particular experiment. The K_D value of 60 µM for the hsFEN1-hsPCNA complex formation indicates a rather weak interaction and thus only a small fraction of the proteins associated (Bruning and Shamoo, 2004). Nevertheless, this experiment was sufficient to confirm the correct structural constitution of hsPCNA or rather the PIP-box binding pocket.

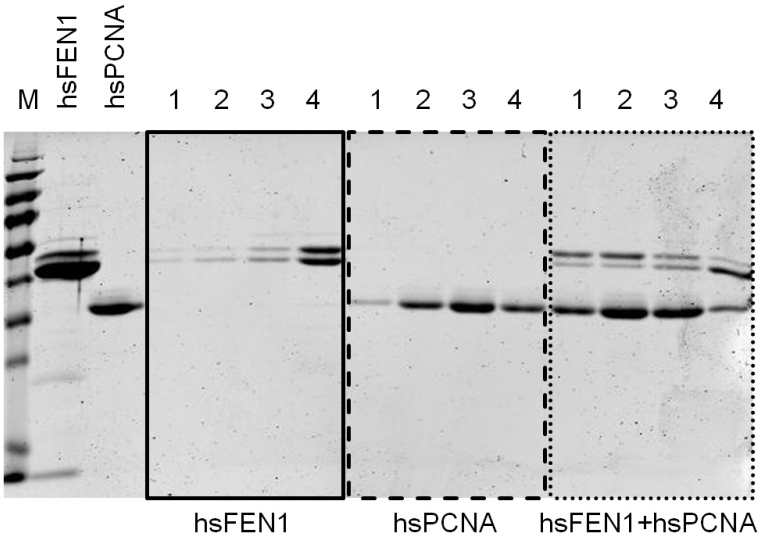


Figure 4.75: hsFEN1-hsPCNA complex formation.
SDS-PAGE analysis of the fractions from the SEC shown in Figure 4.74 D. Elution volumes are indicated by numbers (1: 12-13 ml, 2: 13-14 ml, 3: 14-15 ml, 4: 15-16 ml).

The lack of interaction between taHef and ctPCNA as well as hsPCNA was very unexpected because the C-terminal PIP-box of taHef is the only PCNA interaction site of this protein. Furthermore, due to bacterial expression of all used proteins post-translational modifications can also be excluded. Interestingly, this observation leads to two consecutive conclusions: (1) in addition to the variable PIP-box consensus sequence other factors or residues have an important influence on the interaction with PCNA and therefore (2) this experiment strongly supports the ITC data on the N-terminal PIP-box of hsFANCM (Chapter 4.13.2) because it shows that not only the abundantly present consensus sequence is sufficient for PCNA binding even if it might be unstructured and accessible. To verify this hypothesis I propose PCNA binding

studies (e.g. ITC) with peptides derived from non-PCNA-interacting proteins that contain the wide spread PIP-box consensus sequence. In the context of hsFANCM I suggest peptides containing PIP-boxes 2-8 in addition to the already tested PIP-box 1 and a peptide containing the C-terminal taHef PIP-box as a control.

However, since PIP_{pfRFCL} is very similar to PIP_{taHef} (Figure 4.47) this complex with pfPCNA (Matsumiya et al., 2002) was compared to the eukaryotic PIP_{hsFEN1}-hsPCNA (Sakurai et al., 2005) to unravel differences in the PIP-box binding properties of PCNA in archaea and eukaryotes (Figure 4.76).

Although PCNA and the PIP-box binding pocket are well conserved there are still minor differences that might be responsible for their discrete binding properties. The binding pocket is formed by hydrophobic residues and two tyrosines in eukaryotes (red in Figure 4.76 C). hsPCNA residues L126 and I128 are not visible in the pfPCNA crystal structure because the IDCL is unstructured while in hsPCNA the IDCL is stabilized by an intermolecular β-sheet with the C-terminus of PIP_{hsFEN1}. The alignment indicates that these residues are also conserved in archaea. The glutamine of the PIP-box forms hydrogen bonds with the protein backbone of hsPCNA residues A208 (nitrogen) and A252 (oxygen) via a coordinated water molecule within the Q-pocket. Thus the amino acid side chains especially of residue 208 (referred to hsPCNA) play only a minor role. As mentioned before, the two tyrosines Y133 and Y250 (referred to hsPCNA) seem to be important. Y133 is conserved among *Chaetomium thermophilum*, *Saccharomyces cerevisiae*, and *Homo sapiens* while this position is variable in archaea (Ser in *Thermoplasma acidophilum*, Phe in *Pyrococcus furiosus*, Leu in *Archaeoglobus fulgidus* and *Haloferax volcanii*). Y250 is also part of the hydrophobic binding pocket and differs between eukaryotes and archaea. *Homo sapiens* and *Chaetomium thermophilum* contain a tyrosine while *Saccharomyces cerevisiae* comprises a phenylalanine. In contrast, archaea significantly differ at this position and display hydrophobic side chains (Leu in *Thermoplasma acidophilum* and *Pyrococcus furiosus*, Ile in *Archaeoglobus fulgidus*, Met in H*aloferax volcanii*). Notably, Y250 is located within an elsewise highly conserved patch.

These two residues therefore display very interesting candidates for mutational analysis. Additionally to PCNA mutations, interaction studies with multiple PIP-box variations utilizing ITC and co-crystallization could elucidate the function of residues that surround or are inserted within the PIP-box consensus sequence. Furthermore, other cross species interactions with PCNA such as taHef and PCNA from another archaeal organism (e.g. *Pyrococcus furiosus*) or vice versa eukaryotic PIP-boxes or entire proteins could be tested with archaeal PCNAs.

Notably, there are no pathogenic archaea known (Cavicchioli et al., 2003) and therefore this observation regarding differences in this essential interaction mechanism cannot be exploited in terms of drug development against archaea. However, it was suggested to consider the bacterial PCNA equivalent, the β-clamp, as a potential drug target

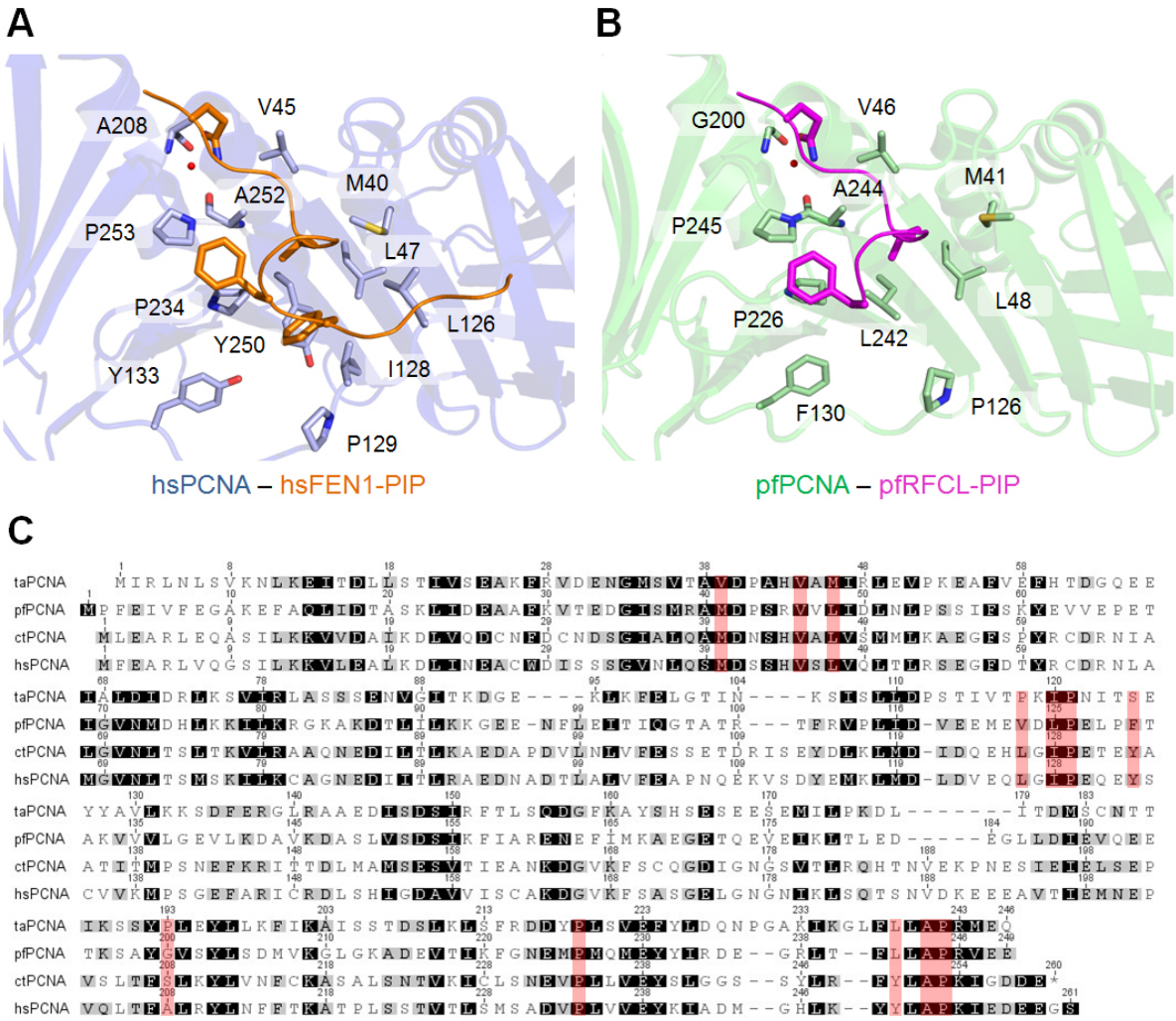


Figure 4.76: Eukaryotic and archaeal PIP-box binding pockets in PCNA. **(A)** Human PCNA with the bound PIP-box peptide derived from hsFEN1 (PDB code: 1U7B) (Bruning and Shamoo, 2004). **(B)** Archaeal PCNA with the bound PIP-box peptide derived from pf RFDL (PDB code: 1ISQ) (Matsumiya et al., 2002). **(C)** Alignment of PCNA amino acid sequences from archaeal and eukaryotic organisms (ta = *Thermoplasma acidophilum*, pf = *Pyrococcus furiosus*, ct = *Chaetomium thermophilum*, hs = *Homo sapiens*). Residues that form the hydrophobic PIP-box binding pocket shown in (A) and (B) are highlighted in red.

(Georgescu et al., 2008; Wolff et al., 2011). Although the β-clamp constitutes a homodimer, the binding mechanism to interaction partners seems to be similar to the archaeal and eukaryotic homologs. The binding pocket is also composed of the C-terminal protrusion and one of the two peripheral extensive loops which is comparable to the IDCL of PCNA. Thus a similar interaction site for PIP-like peptides is formed. Interestingly, the consensus sequence for the β-clamp binding motif is the pentapeptide QL(S/D)LF which is also evolutionary related to the PIP-box motif (Dalrymple et al., 2001). Complex structures of the β-clamp together with bound peptides revealed a

likewise binding behavior of the glutamine at position 1 into the conserved Q-pocket and the leucine at position 4 into a similar hydrophobic pocket (Georgescu et al., 2008). Therefore a deeper understanding of the binding pockets of DNA sliding clamps might not only provide insight into the affinity-dependent recruitment of DNA processing to the replication fork but could also contribute to the development of better antibacterial agents.

Eukaryotic PCNA as a major coordinator of essential cellular functions such as cell growth, death, and maintenance is also discussed as a potential drug target for cancer treatment (Wang, 2014). Thus the here presented data can contribute by providing a deeper insight in PIP-box binding to PCNA which plays an important role in the regulation of PCNA interaction partners at the replication fork.

5 Conclusions

FANCM is a key component in the FA pathway and plays a crucial role in the initiation of replication-dependent ICL repair in higher eukaryotes (Whitby, 2010). The present study illustrates (1) an in-depth structural characterization of the archaeal FANCM homolog Hef from *T. acidophilum* (taHef) and (2) the identification of PCNA as a novel interaction partner of both the archaeal and eukaryotic FANCM homologs Hef and Mph1, respectively. Moreover, it could also be shown that a peptide containing the conserved N-terminal PIP-box of FANCM binds to PCNA while preliminary CoIP experiments for the determination of this putative interaction between the full-length proteins are very promising but still ongoing. Helicase and ATPase assays on taHef revealed that this interaction is most likely restricted to facilitate the approach of Hef to a stalled replication fork but does not influence its enzymatic activity. It was known previously that the FA pathway removes ICL damages from the DNA in a replication-dependent process but a clear connection has been lacking (Moldovan and D'Andrea, 2009). Importantly, the here presented data on the interaction between Hef/Mph1 and PCNA as well as putatively between FANCM and PCNA might provide the yet missing link and explain how the FA repair machinery is recruited to the ICL-stalled replication fork.

These new insights do not only conduce a better understanding of ICL damage removal by the FA pathway but also to the intricate network of DNA repair itself. In the literature, different DNA damages and their corresponding repair mechanisms are mainly discussed separately (Hoeijmakers, 2009) although many of the various pathways might be tightly interwoven. Notably, PCNA is involved in multiple DNA repair mechanisms not only during S phase and also plays a regulatory role which is indicated by triggering the correct genotoxic stress response via its mono- or poly-ubiquitinated state (Moldovan et al., 2007). Thus PCNA serves as a central platform for the coordination of distinct DNA repair mechanisms. Furthermore, the FA pathway also includes components of NER, HRR, and TLS (D'Andrea, 2010) indicating a point of intersection for other repair pathways. Hence, it can be speculated that the FA pathway could also play a more general role in the maintenance of genomic integrity rather than just in ICL repair.

5.1 Hef as a Model for FANCM

taHef holds a unique position among the different Hef proteins since it is the only known Hef protein lacking the C-terminal nuclease domain. In addition to the nuclease activity this domain is also responsible for homo-dimerization, thereby generating a

stable non-globular elongated dimer in solution while taHef is monomeric (Lestini et al., 2013). Although the crystal structures of the isolated helicase and nuclease entities of pfHef have provided a substantial step towards the understanding of Hef, important properties of the full-length protein could not be elucidated so far.

In this study the crystal structure of full-length taHef was solved indicating significant differences to the helicase portion of pfHef. In contrast to pfHef the taHef structure shows a widely open conformation which correlates with the flexibility between the two RecA-like motordomains Hel1 and Hel2 and has not been anticipated before. This motility is an essential prerequisite for the function of a helicase (Lohman et al., 2008). It is known that ATP hydrolysis is coupled to the motion between the two domains but the extent of the motion is still a matter of debate. A likewise extensive flexibility was also suggested for the helicase XPB for proper DNA separation activity (Fan et al., 2006). The data on taHef confirm this assumption and thereby highlight the intensity of helicase dynamics. Since the SAXS model demonstrates an intermediate conformation for taHef in solution, the open conformation as shown by the crystal structure may represent an extreme constitution that is favored by the crystallization conditions. However, another important difference between the pfHef and taHef structures is the presence of helix α21 in taHef, which may directly affect the orientation of the two helicase domains relative to each other. Therefore taHef was analyzed in the presence and absence of this helix and surprisingly, the SAXS data do not suggest a significant conformational difference between the wild type protein and the variant. This leads to the conclusion that helix α21 does not exclusively fulfill a structural function to hold Hef in an open conformation required to enclose the DNA between Hel1 and Hel2 and the extent of flexibility is not an artifact due to the use of different constructs. Helix α21 could rather be involved in downstream DNA processing as it substantially contributes to the positively charged surface potential between Hel1 and Hel2. This hypothesis is supported by the fact that the removal of α21 in pfHef significantly decreases its helicase and ATPase activities (Nishino et al., 2005b). However, the exact functional role of helix α21 remains speculative and requires further investigation.

Furthermore, the data on taHef helicase and ATPase activity on different fork-structured DNA substrates show that taHef requires single-stranded extensions for strand separation, whereas a double-stranded fork substrate constitutes an obstacle that prevents DNA unwinding and simultaneously triggers accelerated ATP consumption. A similar phenomenon was observed for the prokaryotic nucleotide excision repair helicase UvrB (Moolenaar et al., 2000), i.e. when the helicase is unable to move forward and separate the double-stranded DNA, it stalls and hydrolyzes ATP continuously. Intriguingly, the presence of PCNA does not alter the ATPase nor the helicase activity of taHef, thus suggesting that the interaction is entirely dedicated to recruit taHef to the replication fork to fulfill its function.

Importantly, in addition to the sequence homology between Hef and FANCM, both proteins also seem to be functionally related. *In vivo* studies on the euryarchaeon

Haloferax volcanii show that Hef specifically localizes to arrested replication forks (Lestini et al., 2013). Thus the authors concluded that Hef, similarly to FANCM, enhances replication fork stability via a direct interaction with the collapsed replication forks. This observation is in agreement with the previous finding that Hef functions in an alternative pathway to resolve recombination intermediates than the Holliday junction resolvase Hjc (Lestini et al., 2010). In this context, Hef does not only stabilize the DNA but it also clearly processes several fork-like structures an it was shown that the N-terminal helicase portion prepares the DNA for subsequent cleavage by the C-terminal nuclease domain in a cooperative manner (Komori et al., 2004). Hef thus fulfills multiple functions in processing the replication fork while in the FA pathway the cleavage of the DNA is accomplished by a cascade including a multiplicity of different proteins. Furthermore, another recent publication identified binding of PCNA to the intrinsically disordered region (IDR) between the helicase and nuclease entities of Hef from *Thermococcus kodakarensis* with a K_D of 530 nM determined by surface plasmon resonance analysis (Ishino et al., 2014). This value is comparable to the ITC data for taHef and taPCNA but, remarkably, the authors were not able to assign the exact interaction site.

However, although Hef is generally accepted as a model for FANCM which is strongly supported by the interaction with PCNA, the data on Hef must not be overrated and should still be handled with caution with respect to the comparison to FANCM. Massive differences can be observed not only in protein size, but also in other interaction partners, post-translational modifications, and the mode of action. For a precise understanding of ICL repair by the FA pathway the analysis of archaeal homologs is insufficient and a further characterization of the human proteins is essential. Nevertheless, the here presented data on the archaeal FANCM homolog Hef provide a general basis for the understanding of replication-dependent ICL repair, conformational dynamics of helicases, and highlights the role of PCNA in the coordination of various DNA repair mechanisms.

5.2 Interaction with PCNA

Strikingly, this study revealed a direct interaction between Hef and PCNA. It was shown previously that Hef plays a crucial role in replication-linked DNA damage repair by stabilization and processing of the replication fork (Komori et al., 2004; Lestini et al., 2013). The here presented results now provide the basis how this interaction may be established. The direct interaction of the C-terminal PIP-box in taHef with taPCNA permits its recruitment to the replication fork and most likely initiates the subsequent repair process. Importantly, this interaction adds yet another partner to the coordinated binding and release of DNA metabolizing enzymes to the replication fork by PCNA. This is a fundamental and highly regulated mechanism and its significance in DNA maintenance cannot be overestimated (Zheng et al., 2011a). Since the PIP-box is conserved among archaeal Hef proteins, it can be assumed that an interaction

between Hef and PCNA in general takes place. Its position within the disordered region C-terminally adjacent to the helicase entity facilitates its accessibility for PCNA which is in agreement with the observation that PIP-boxes are typically located in flexible areas and only upon binding to PCNA form a 3_{10}-helix (Pascal et al., 2006).

Notably, the *in vitro* data on the taHef-taPCNA complex clearly reveal a 1:1 stoichiometry but *in vivo* this situation could be different due to steric hindrances by other PCNA interacting proteins. The role of potentially three Hef proteins recruited simultaneously to the replication fork therefore remains elusive and might be an *in vitro* artifact. The recruitment and also the release of PCNA binding partners is tightly regulated and thus in the presence of competing proteins the Hef-PCNA stoichiometry is not known. To evaluate if three Hefs associate with PCNA *in vivo* or if another stoichiometry is preferred, *in vivo* studies should be pursued with differently fluorescently labeled proteins for Förster resonance energy transfer (FRET) or confocal microscopy analysis.

Archaeal proteins have been used as models for complex and intricate eukaryotic systems (White, 2003). Due to the sequence homology to hsFANCM and the presence of a conserved accessible PIP-box in FANCM and taHef, I demonstrate that taHef can be explicitly considered as an archaeal model for FANCM. Interestingly, hsFANCM contains several putative canonical PIP-box motifs distributed over the entire protein sequence but only the PIP-box at the very N-terminus is highly conserved throughout a broad range of various higher eukaryotes. Our experiments so far demonstrate that this N-terminal PIP-box binds to PCNA. Although the K_D of 13 µM is about ten-fold lower than for taHef-taPCNA, it is possible that posttranslational modifications such as ubiquitination which does not occur in archaea or additional binding events of full-length FANCM could also contribute to the interaction.

It is thus valid to speculate that the interaction between both proteins is conserved throughout the different kingdoms. Therefore, our studies on the Hef-PCNA interaction might also provide insights into the understanding of ICL damage detection in the Fanconi anemia pathway. The presumable interaction between FANCM and PCNA could be the yet missing link that connects replication and FA repair. In analogy to Hef FANCM, in complex with its binding partners FAAP24 and MHF1/2, also recognizes and protects stalled replication forks (Blackford et al., 2012). It was shown previously that the FA core complex directly interacts with PCNA via FANCL and that PCNA ubiquitination is required for downstream FANCD2 ubiquitination (Geng et al., 2010). The interaction between FANCM via its motif MM1 and the FA core complex is mediated through its integral component FANCF (Deans and West, 2009).

It is tempting to speculate that PCNA could assume a regulatory function within the FA pathway leading to two hypotheses: (1) In addition to the recruitment of the core complex via the MM1 motif of FANCM and FANCF, a simultaneous interaction with PCNA could have a synergistic effect. Located in close proximity to the ICL

damage, PCNA might "hand over" the impaired DNA from FANCM to FANCL thereby triggering subsequent FANCD2 ubiquitination by the core complex. (2) PCNA is also intensively involved in the FA connected classical repair pathways nucleotide excision repair, homologous recombination, and translesion synthesis. Thereby it might also play an important role in the coordination of these other repair components additionally to FANCM with respect to proper ICL damage removal.

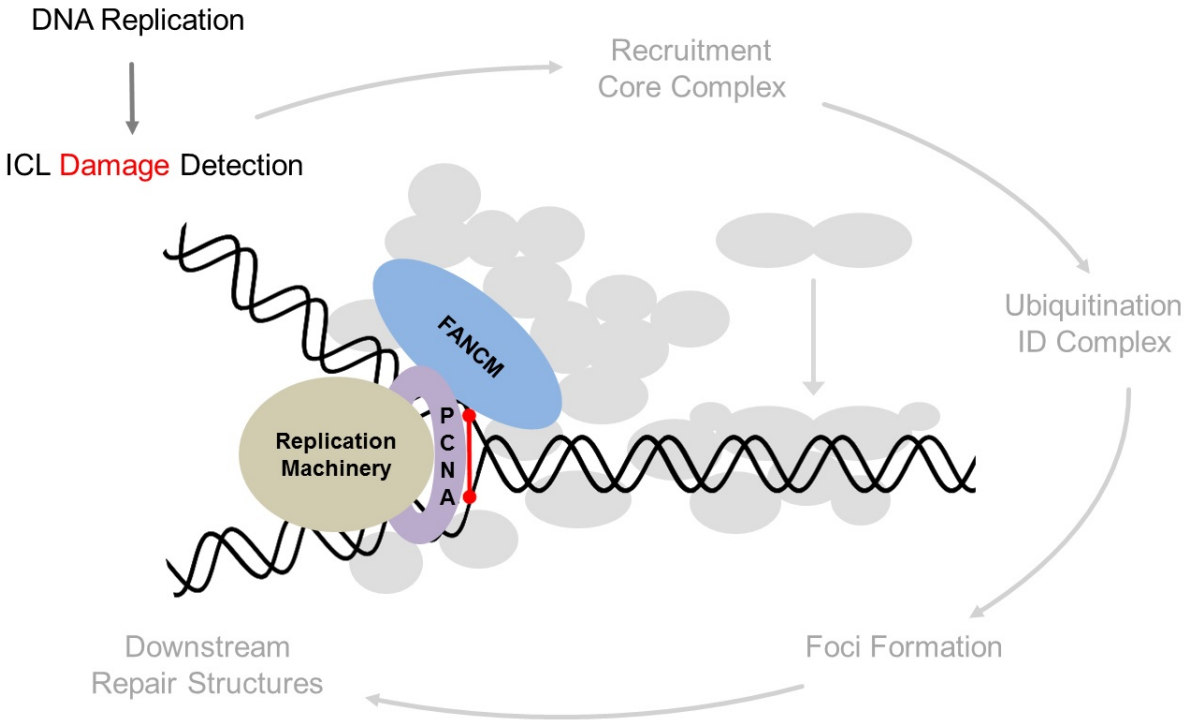


Figure 5.1: Model of the FA pathway including FANCM recruitment by PCNA. Simplified model of the FA pathway (Figure 1.3) highlighting the role of a presumable FANCM-PCNA interaction during ICL damage detection. PCNA might function as the missing link between the replication machinery of which it is a component and regulates various related processes and the replication-dependent FA DNA repair pathway through its initiator FANCM.

Remarkably, the putative interaction between FANCM and PCNA still remains speculative and requires a definite proof. Although this hypothesis is mainly based on the Hef-PCNA interaction, it is strongly supported by the binding of the fungal FANCM homolog ctMph1 to ctPCNA. This observation is of major importance in two ways: (1) it clearly shows that this interaction is not restricted to archaea but also present in eukaryotes and (2) it provides a new and even more significant model particularly regarding structural analysis. Therefore further biochemical and especially crystallization experiments utilizing apo ctMph1 as well as ctMph1 in complex with ctPCNA are highly recommended.

5.3 Outlook

As mentioned before, the presumable FANCM-PCNA interaction remains speculative thus leading to a multiplicity of possible future experiments. With respect to the FA pathway a focus on the characterization of the eukaryotic Mph1, FANCM, and PCNA proteins is suggested rather than a continuation on the archaeal Hef. The present results as well as the already prepared new expression constructs and protocols provide a solid basis for continuative *in vitro* and also *in vivo* experiments.

ctMph1 and ctPCNA can be readily recombinantly expressed and purified from *E. coli* and an interaction between both proteins could be determined. These preliminary data thus provide a solid basis to further characterize this system. Although full-length ctMph1 heavily degrades, the newly generated constructs permit co-expression of ctMph1 and ctPCNA which could further stabilize ctMph1 and prevent degradation. In addition a shortened construct has been generated based on the analysis of the degradation product of ctMph1 which is analogous to the helicase core of the FANCM protein family including the N-terminal extension. Interestingly, these constructs may be meaningful not only for crystallization experiments aiming for the apo as well as the complex structures but also for biochemical and functional analyses, e.g. to evaluate the role of the C-terminally unstructured region of ctMph1. To improve the data quality of the apo ctPCNA crystal structure, additional crystals should be tested and optimized. In the context of co-crystallization, ctMph1-DNA binding studies should be performed to identify a DNA substrate which binds with high affinity and is sufficiently short to permit crystallization. Single-stranded and fork-structured DNA substrates should represent reasonable candidates but also HJs should be considered. Since degraded ctMph1 comprises only one PIP-box within the N-terminal extension which might be responsible for the interaction with ctPCNA, this particular PIP-box should be addressed in future experiments by site-directed mutagenesis. Based on the work described in this thesis a "killPIP" variant (I111R/F114S) should be introduced in ctMph1. Another possibility is the analysis of a short peptide comprising the wild type as well as the mutated amino acid sequence of the ctMph1 PIP-box in the presence of ctPCNA in ITC experiments similarly to the FANCM$^{1-17}$-PCNA experiment (Chapter 4.13.2). If it is not possible to crystallize the complex containing full-length or truncated ctMph1 and ctPCNA, it could still be considered to co-crystallize or soak ctPCNA with such a peptide. After verification of the N-terminal PIP-box of ctMph1 as the mediator for ctPCNA binding and ideally determination of the K_D for this interaction, the PIP-box could be replaced by the p21 PIP-box which is the strongest PCNA interaction partner known so far (Bruning and Shamoo, 2004) to facilitate crystallization. To corroborate the hypothetical ctMph1-ctPCNA crystal structure biochemical analyses with respect to helicase and ATPase activity will also be required.

Regarding the analysis of the putative hsFANCM-hsPCNA interaction, CoIP experiments by our collaboration partner Weidong Wang are ongoing and promising. Notably, a positive CoIP itself does not necessarily show a direct interaction but

only indicates that bait and prey are part of the same complex. This could be a complex consisting of multiple subunits such as the FA core complex and thus it could also present an indirect interaction. As PCNA also interacts with the core complex component FANCL (Geng et al., 2010), a disruption of the hsFANCM-hsPCNA CoIP via the "killPIP" mutation is required to assure that the putative hsFANCM-hsPCNA interaction is directly mediated by a particular PIP-box. Since I could already show that a short peptide that comprises the N-terminal PIP-box of hsFANCM binds to hsPCNA with a K_D of 13 µM, hsPCNA could be co-crystallized with this peptide.

The main question about this putative interaction, however, is its physiological and functional role. This question can only be answered by *in vivo* studies and for this purpose cell survival assays present a good option. Chicken DT40 FANCM knock out cells could be rescued by either wild type or the mutated form of FANCM. Furthermore, the viability of these cells should be analyzed in the presence and absence of various DNA damaging agents. Another important aspect that also needs to be considered during these complementary experiments is the mono-ubiquitination of the FANCI-FANCD2 (ID) complex in response to ICL damages. This modification of the ID-complex is a critical step within the FA pathway and crucial for downstream repair processes. Interestingly, preliminary CoIP results revealed that MMC does not trigger FANCM-PCNA complex formation while hydroxyurea (HU) seems to have a stimulating effect. Therefore, the immunostaining experiments should be repeated with HU stimulated HeLa cells to analyze co-localization of FANCM and PCNA. Similar to the CoIP experiments, immunostaining does not provide a definite proof of a direct interaction. Thus PIP-box variants are also required to potentially abolish the co-localization. Mutations within endogenous genes for *in vivo* experiments can be introduced using the TALEN and CRISPR/Cas systems (Yasue et al., 2014). This technique holds the advantages that the natural protein expression level is not influenced which might cause side effects, there is no need for transfection and selection to generate stable cell lines, and also genes with a lethal knock out phenotype can be modified. Especially regarding cell viability assays it is important to use as little selection pressure as possible such as G418 which is needed for stable cell lines.

In addition to the PIP-box, the APIM motif has been proposed as a new PCNA binding motif (Gilljam et al., 2009). Notably, the APIM consensus sequence is present in hsFANCM at residues 1505-1509 and although the function of this motif itself is controversial, this region of hsFANCM should be investigated in more detail. Even if this sequence does not mediate PCNA interaction it might comprise a different important function within hsFANCM. Therefore, the H1506A mutation should be introduced in hsFANCM to disrupt the APIM motif. Notably, neither of the three known pathogenic FANCM mutations affect this area of the protein (Chapter 1.3.2).

Ubiquitination of PCNA plays an important role in the repair of stalled replication forks (Bergink and Jentsch, 2009) and posttranslational modifications of PCNA could also contribute to the presumable FANCM-PCNA interaction. Thus, if FANCM

and PCNA interact at the ICL-stalled replication fork, a key question will be the ubiquitination state of PCNA. There are multiple possibilities to address this, e.g. through the utilization of specific ubiquitin and poly-ubiquitin antibodies in CoIP experiments. Other options would be a mutation of residue K164 in PCNA to prevent its ubiquitination and SUMOylation or within the ubiquitin molecule itself a mutation of residue K63 to prevent poly-ubiquitination. It was shown previously that Mph1 and Rad5 function in a common pathway. Rad5 is a key component of the repair system for stalled replication forks and poly-ubiquitinates PCNA (Daee et al., 2012). This observation also points towards a putative interaction between Mph1 and PCNA. Interestingly, in humans this task is fulfilled by the Rad5 homolog HLTF and thus in future experiments a possible connection between HLTF and FANCM could be examined.

Finally, PCNA is not the only ring-shaped DNA sliding clamp that is involved in DNA maintenance. The heterotrimeric Rad9-Hus1-Rad1 (9-1-1) complex reveals a structure that is very similar to PCNA and is particularly important for the activation of checkpoint signaling in response to genotoxic stress (Sohn and Cho, 2009). Similarly to PCNA, the 9-1-1 complex also stimulates FEN1 activity (Wang et al., 2004) and thus it can be speculated that the mode of interaction could also be similar and might involve the PIP-box. Importantly, the 9-1-1 complex primarily serves as a platform for ATR-mediated checkpoint kinase 1 (Chk1) activation via TopBP1. Moreover, FANCM is crucial for the ATR-mediated checkpoint signaling by promoting TopBP1 localization to the DNA (Schwab et al., 2010). The 9-1-1 complex should thus be included in CoIP experiments. Commercial antibodies against 9-1-1 are available and in the chosen setup with overexpressed and tagged FANCM as the bait, it would be interesting and also straight forward to check for 9-1-1 in addition to ubiquitin and PCNA.

Bibliography

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213–21.

Alpi, A. F. and Patel, K. J. (2009). Monoubiquitylation in the Fanconi anemia DNA damage response pathway. DNA Repair (Amst) *8*, 430–5.

Amlacher, S., Sarges, P., Flemming, D., van Noort, V., Kunze, R., Devos, D. P., Arumugam, M., Bork, P. and Hurt, E. (2011). Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. Cell *146*, 277–89.

Anjana, R., Vaishnavi, M. K., Sherlin, D., Kumar, S. P., Naveen, K., Kanth, P. S. and Sekar, K. (2012). Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. Bioinformation *8*, 1220–4.

Aravind, L., Walker, D. R. and Koonin, E. V. (1999). Conserved domains in DNA repair proteins and evolution of repair systems. Nucleic Acids Res *27*, 1223–42.

Auerbach, A. D. (2009). Fanconi anemia and its diagnosis. Mutat Res *668*, 4–10.

Banks, D., Wu, M., Higa, L. A., Gavrilova, N., Quan, J., Ye, T., Kobayashi, R., Sun, H. and Zhang, H. (2006). L2DTL/CDT2 and PCNA interact with p53 and regulate p53 polyubiquitination and protein stability through MDM2 and CUL4A/DDB1 complexes. Cell Cycle *5*, 1719–29.

Bergink, S. and Jentsch, S. (2009). Principles of ubiquitin and SUMO modifications in DNA repair. Nature *458*, 461–7.

Bienko, M., Green, C. M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannouche, P., Wider, G., Peter, M., Lehmann, A. R., Hofmann, K. and Dikic, I. (2005). Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. Science *310*, 1821–4.

Blackford, A. N., Schwab, R. A., Nieminuszczy, J., Deans, A. J., West, S. C. and Niedzwiedz, W. (2012). The DNA translocase activity of FANCM protects stalled replication forks. Hum Mol Genet *21*, 2005–16.

Bogliolo, M., Schuster, B., Stoepker, C., Derkunt, B., Su, Y., Raams, A., Trujillo, J. P., Minguillon, J., Ramirez, M. J., Pujol, R., Casado, J. A., Banos, R., Rio, P., Knies, K., Zuniga, S., Benitez, J., Bueren, J. A., Jaspers, N. G., Schärer, O. D., de Winter, J. P., Schindler, D. and Surralles, J. (2013). Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. Am J Hum Genet *92*, 800–6.

Bowman, G. D., O'Donnell, M. and Kuriyan, J. (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature *429*, 724–30.

Broecker, J., Vargas, C. and Keller, S. (2011). Revisiting the optimal c value for isothermal titration calorimetry. Anal Biochem *418*, 307–9.

Brosh, R. M., J. (2013). DNA helicases involved in DNA repair and their roles in cancer. Nat Rev Cancer *13*, 542–58.

Bruck, I. and O'Donnell, M. (2001). The ring-type polymerase sliding clamp family. Genome Biol *2*, REVIEWS3001.

Bruning, J. B. and Shamoo, Y. (2004). Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1. Structure *12*, 2209–19.

Caruthers, J. M. and McKay, D. B. (2002). Helicase structure and mechanism. Curr Opin Struct Biol *12*, 123–33.

Cavicchioli, R., Curmi, P. M., Saunders, N. and Thomas, T. (2003). Pathogenic archaea: do they exist? Bioessays *25*, 1119–28.

Chapados, B. R., Hosfield, D. J., Han, S., Qiu, J., Yelent, B., Shen, B. and Tainer, J. A. (2004). Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. Cell *116*, 39–50.

Chen, V. B., Arendall, W. B., r., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12–21.

Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A. R., Laghmani el, H., Joenje, H., McDonald, N., de Winter, J. P., Wang, W. and West, S. C. (2007). Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. Mol Cell *25*, 331–43.

Clauson, C., Schärer, O. D. and Niedernhofer, L. (2013). Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. Cold Spring Harb Perspect Biol *5*, a012732.

Cole, R. S. (1973). Repair of DNA containing interstrand crosslinks in Escherichia coli: sequential excision and recombination. Proc Natl Acad Sci U S A *70*, 1064–8.

Coulthard, R., Deans, A. J., Swuec, P., Bowles, M., Costa, A., West, S. C. and McDonald, N. Q. (2013). Architecture and DNA recognition elements of the Fanconi anemia FANCM-FAAP24 complex. Structure *21*, 1648–58.

da Silva, M. W., Bierbryer, R. G., Wilds, C. J., Noronha, A. M., Colvin, O. M., Miller, P. S. and Gamcsik, M. P. (2005). Intrastrand base-stacking buttresses widening of major groove in interstrand cross-linked B-DNA. Bioorg Med Chem *13*, 4580–7.

Daee, D. L., Ferrari, E., Longerich, S., Zheng, X. F., Xue, X., Branzei, D., Sung, P. and Myung, K. (2012). Rad5-dependent DNA repair functions of the Saccharomyces cerevisiae FANCM protein homolog Mph1. J Biol Chem *287*, 26563–75.

Dalrymple, B. P., Kongsuwan, K., Wijffels, G., Dixon, N. E. and Jennings, P. A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. Proc Natl Acad Sci U S A *98*, 11627–32.

D'Andrea, A. D. (2010). Susceptibility pathways in Fanconi's anemia and breast cancer. N Engl J Med *362*, 1909–19.

Darland, G., Brock, T. D., Samsonoff, W. and Conti, S. F. (1970). A thermophilic, acidophilic mycoplasma isolated from a coal refuse pile. Science *170*, 1416–8.

Deans, A. J. and West, S. C. (2009). FANCM connects the genome instability disorders Bloom's Syndrome and Fanconi Anemia. Mol Cell *36*, 943–53.

Deans, A. J. and West, S. C. (2011). DNA interstrand crosslink repair and cancer. Nat Rev Cancer *11*, 467–80.

Dionne, I., Nookala, R. K., Jackson, S. P., Doherty, A. J. and Bell, S. D. (2003). A heterotrimeric PCNA in the hyperthermophilic archaeon Sulfolobus solfataricus. Mol Cell *11*, 275–82.

Dronkert, M. L. and Kanaar, R. (2001). Repair of DNA interstrand cross-links. Mutat Res *486*, 217–47.

Elias, J., Dimitrio, L., Clairambault, J. and Natalini, R. (2014). The p53 protein and its molecular network: modelling a missing link between DNA damage and cell fate. Biochim Biophys Acta *1844*, 232–47.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486–501.

Enoiu, M., Jiricny, J. and Schärer, O. D. (2012). Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. Nucleic Acids Res *40*, 8953–64.

Enzlin, J. H. and Schärer, O. D. (2002). The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. EMBO J *21*, 2045–53.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr *62*, 72–82.

Fan, L., Arvai, A. S., Cooper, P. K., Iwai, S., Hanaoka, F. and Tainer, J. A. (2006). Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. Mol Cell *22*, 27–37.

Fanconi, G. (1927). Familiäre infantile perniziosaartige Anämie (perniziöses Blutbild und Konstitution). Jahrb. Kinderh. *117*.

Flechsig, H., Popp, D. and Mikhailov, A. S. (2011). In silico investigation of conformational motions in superfamily 2 helicase proteins. PLoS One *6*, e21809.

Frank, G., Qiu, J., Zheng, L. and Shen, B. (2001). Stimulation of eukaryotic flap endonuclease-1 activities by proliferating cell nuclear antigen (PCNA) is independent of its in vitro interaction via a consensus PCNA binding region. J Biol Chem *276*, 36295–302.

Franke, D. and Svergun, D. I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography *42*, 342–346.

Freyer, M. W. and Lewis, E. A. (2008). Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. Methods Cell Biol *84*, 79–113.

Gari, K., Decaillet, C., Delannoy, M., Wu, L. and Constantinou, A. (2008a). Remodeling of DNA replication structures by the branch point translocase FANCM. Proc Natl Acad Sci U S A *105*, 16107–12.

Gari, K., Decaillet, C., Stasiak, A. Z., Stasiak, A. and Constantinou, A. (2008b). The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. Mol Cell *29*, 141–8.

Geng, L., Huntoon, C. J. and Karnitz, L. M. (2010). RAD18-mediated ubiquitination of PCNA activates the Fanconi anemia DNA repair network. J Cell Biol *191*, 249–57.

Georgescu, R. E., Yurieva, O., Kim, S. S., Kuriyan, J., Kong, X. P. and O'Donnell, M. (2008). Structure of a small-molecule inhibitor of a DNA polymerase sliding clamp. Proc Natl Acad Sci U S A *105*, 11116–21.

Giacovazzo, C., Monaco, H., Artioli, G., Viterbo, D., Milanesio, M., Gilli, G., Gilli, P., Zanotti, G., Ferraris, G. and Catti, M. (2011). Fundamentals of Crystallography. 3 edition, OUP Oxford.

Gilljam, K. M., Feyzi, E., Aas, P. A., Sousa, M. M., Müller, R., Vagbo, C. B., Catterall, T. C., Liabakk, N. B., Slupphaug, G., Drablos, F., Krokan, H. E. and Otterlei, M. (2009). Identification of a novel, widespread, and functionally important PCNA-binding motif. J Cell Biol *186*, 645–54.

Gilljam, K. M., Müller, R., Liabakk, N. B. and Otterlei, M. (2012). Nucleotide excision repair is associated with the replisome and its efficiency depends on a direct interaction between XPA and PCNA. PLoS One *7*, e49199.

Guainazzi, A., Campbell, A. J., Angelov, T., Simmerling, C. and Schärer, O. D. (2010). Synthesis and molecular modeling of a nitrogen mustard DNA interstrand crosslink. Chemistry *16*, 12100–3.

Guinier, A. (1938). The diffusion of X-rays under the extremely weak angles applied to the study of fine particles and colloidal suspension. Comptes Rendus Hebdomadaires Des Seances De L Acad Des Sci *206*, 1374–1376.

Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M. and Kuriyan, J. (1996). Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. Cell *87*, 297–306.

Guo, C., Sonoda, E., Tang, T. S., Parker, J. L., Bielen, A. B., Takeda, S., Ulrich, H. D. and Friedberg, E. C. (2006). REV1 protein interacts with PCNA: significance of the REV1 BRCT domain in vitro and in vivo. Mol Cell *23*, 265–71.

Haaf, E. and Schlosser, A. (2012). Peptide and protein quantitation by acid-catalyzed 18O-labeling of carboxyl groups. Anal Chem *84*, 304–11.

Hall, M. C. and Matson, S. W. (1999). Helicase motifs: the engine that powers DNA unwinding. Mol Microbiol *34*, 867–77.

Harutyunyan, A., Gisslinger, B., Klampfl, T., Berg, T., Bagienski, K., Gisslinger, H. and Kralovics, R. (2011). Rare germline variants in regions of loss of heterozygosity may influence clinical course of hematological malignancies. Leukemia *25*, 1782–4.

Hishiki, A., Hashimoto, H., Hanafusa, T., Kamei, K., Ohashi, E., Shimizu, T., Ohmori, H. and Sato, M. (2009). Structural basis for novel interactions between human translesion synthesis polymerases and proliferating cell nuclear antigen. J Biol Chem *284*, 10552–60.

Ho, T. V. and Schärer, O. D. (2010). Translesion DNA synthesis polymerases in DNA interstrand crosslink repair. Environ Mol Mutagen *51*, 552–66.

Hoadley, K. A., Xue, Y., Ling, C., Takata, M., Wang, W. and Keck, J. L. (2012). Defining the molecular interface that connects the Fanconi anemia protein FANCM to the Bloom syndrome dissolvasome. Proc Natl Acad Sci U S A *109*, 4437–42.

Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G. and Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature *419*, 135–41.

Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. Nature *411*, 366–74.

Hoeijmakers, J. H. (2009). DNA damage, aging, and cancer. N Engl J Med *361*, 1475–85.

Horejsi, Z., Collis, S. J. and Boulton, S. J. (2009). FANCM-FAAP24 and HCLK2: roles in ATR signalling and the Fanconi anemia pathway. Cell Cycle *8*, 1133–7.

Hwang, K. Y., Baek, K., Kim, H. Y. and Cho, Y. (1998). The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. Nat Struct Biol *5*, 707–13.

Indiani, C. and O'Donnell, M. (2006). The replication clamp-loading machine at work in the three domains of life. Nat Rev Mol Cell Biol *7*, 751–61.

Ishino, S., Yamagami, T., Kitamura, M., Kodera, N., Mori, T., Sugiyama, S., Ando, T., Goda, N., Tenno, T., Hiroaki, H. and Ishino, Y. (2014). Multiple Interactions of the Intrinsically Disordered Region between the Helicase and Nuclease Domains of the Archaeal Hef Protein. J Biol Chem *289*, 21627–39.

Jiang, F., Ramanathan, A., Miller, M. T., Tang, G. Q., Gale, M., J., Patel, S. S. and Marcotrigiano, J. (2011). Structural basis of RNA recognition and activation by innate immune receptor RIG-I. Nature *479*, 423–7.

Kabsch, W. (2010). XDS. Acta Crystallogr D Biol Crystallogr *66*, 125–32.

Kee, Y., Kim, J. M. and D'Andrea, A. D. (2009). Regulated degradation of FANCM in the Fanconi anemia pathway during mitosis. Genes Dev *23*, 555–60.

Kelly, S. M., Jess, T. J. and Price, N. C. (2005). How to study proteins by circular dichroism. Biochim Biophys Acta *1751*, 119–39.

Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W. and Steitz, T. A. (1995). Crystal structure of Thermus aquaticus DNA polymerase. Nature *376*, 612–6.

Kim, Y., Lach, F. P., Desetty, R., Hanenberg, H., Auerbach, A. D. and Smogorzewska, A. (2011). Mutations of the SLX4 gene in Fanconi anemia. Nat Genet *43*, 142–6.

Kisker, C., Kuper, J. and Van Houten, B. (2013). Prokaryotic nucleotide excision repair. Cold Spring Harb Perspect Biol *5*, a012591.

Knipscheer, P., Raschle, M., Smogorzewska, A., Enoiu, M., Ho, T. V., Schärer, O. D., Elledge, S. J. and Walter, J. C. (2009). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. Science *326*, 1698–701.

Komori, K., Fujikane, R., Shinagawa, H. and Ishino, Y. (2002). Novel endonuclease in Archaea cleaving DNA with various branched structure. Genes Genet Syst *77*, 227–41.

Komori, K., Hidaka, M., Horiuchi, T., Fujikane, R., Shinagawa, H. and Ishino, Y. (2004). Cooperation of the N-terminal Helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks. J Biol Chem *279*, 53175–85.

Kontopidis, G., Wu, S. Y., Zheleva, D. I., Taylor, P., McInnes, C., Lane, D. P., Fischer, P. M. and Walkinshaw, M. D. (2005). Structural and biochemical studies of human proliferating cell nuclear antigen complexes provide a rationale for cyclin association and inhibitor design. Proc Natl Acad Sci U S A *102*, 1871–6.

Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. and Waksman, G. (1997). Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. Cell *90*, 635–47.

Kottemann, M. C. and Smogorzewska, A. (2013). Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. Nature *493*, 356–63.

Kowalinski, E., Lunardi, T., McCarthy, A. A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D. and Cusack, S. (2011). Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. Cell *147*, 423–35.

Kuper, J. and Kisker, C. (2012). Damage recognition in nucleotide excision DNA repair. Curr Opin Struct Biol *22*, 88–93.

La Touche, C. J. (1948). A Chaetomium-like thermophile fungus. Nature *161*, 320.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–5.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–8.

Lawley, P. D. and Phillips, D. H. (1996). DNA adducts from chemotherapeutic agents. Mutat Res *355*, 13–40.

Lee, K. Y. and Myung, K. (2008). PCNA modifications for regulation of post-replication repair pathways. Mol Cells *26*, 5–11.

Leslie, A. G. W. and Powell, H. R. (2007). Processing Diffraction Data with Mosflm. Evolving Methods for Macromolecular Crystallography *245*, 41–51.

Lestini, R., Duan, Z. and Allers, T. (2010). The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in Haloferax volcanii. DNA Repair (Amst) *9*, 994–1002.

Lestini, R., Laptenok, S. P., Kuhn, J., Hink, M. A., Schanne-Klein, M. C., Liebl, U. and Myllykallio, H. (2013). Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication. Nucleic Acids Res *41*, 10358–70.

Li, M. Z. and Elledge, S. J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods *4*, 251–6.

Li, Y., Korolev, S. and Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. EMBO J *17*, 7514–25.

Liu, H. and Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol *8*, 91.

Liu, T., Ghosal, G., Yuan, J., Chen, J. and Huang, J. (2010). FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. Science *329*, 693–6.

Lohman, T. M., Tomko, E. J. and Wu, C. G. (2008). Non-hexameric DNA helicases and translocases: mechanisms and regulation. Nat Rev Mol Cell Biol *9*, 391–401.

Machida, Y. J., Machida, Y., Chen, Y., Gurtan, A. M., Kupfer, G. M., D'Andrea, A. D. and Dutta, A. (2006). UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation. Mol Cell *23*, 589–96.

Magana-Schwencke, N., Henriques, J. A., Chanet, R. and Moustacchi, E. (1982). The fate of 8-methoxypsoralen photoinduced crosslinks in nuclear and mitochondrial yeast DNA: comparison of wild-type and repair-deficient strains. Proc Natl Acad Sci U S A *79*, 1722–6.

Mailand, N., Gibbs-Seymour, I. and Bekker-Jensen, S. (2013). Regulation of PCNA-protein interactions for genome stability. Nat Rev Mol Cell Biol *14*, 269–82.

Maric, H. M. (2012). Molecular Basis of the Multivalent Glycine and y-Aminobutyric Acid Type A Receptor Anchoring. University of Würzburg.

Matsumiya, S., Ishino, S., Ishino, Y. and Morikawa, K. (2002). Physical interaction between proliferating cell nuclear antigen and replication factor C from Pyrococcus furiosus. Genes Cells *7*, 911–22.

Matsumiya, S., Ishino, Y. and Morikawa, K. (2001). Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from Pyrococcus furiosus. Protein Sci *10*, 17–23.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658–674.

McNally, R., Bowman, G. D., Goedken, E. R., O'Donnell, M. and Kuriyan, J. (2010). Analysis of the role of PCNA-DNA contacts during clamp loading. BMC Struct Biol *10*, 3.

Meetei, A. R., de Winter, J. P., Medhurst, A. L., Wallisch, M., Waisfisz, Q., van de Vrugt, H. J., Oostra, A. B., Yan, Z., Ling, C., Bishop, C. E., Hoatlin, M. E., Joenje, H. and Wang, W. (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. Nat Genet *35*, 165–70.

Meetei, A. R., Medhurst, A. L., Ling, C., Xue, Y., Singh, T. R., Bier, P., Steltenpool, J., Stone, S., Dokal, I., Mathew, C. G., Hoatlin, M., Joenje, H., de Winter, J. P. and Wang, W. (2005). A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. Nat Genet *37*, 958–63.

Mertens, H. D. and Svergun, D. I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J Struct Biol *172*, 128–41.

Miyata, T., Oyama, T., Mayanagi, K., Ishino, S., Ishino, Y. and Morikawa, K. (2004). The clamp-loading complex for processive DNA replication. Nat Struct Mol Biol *11*, 632–6.

Müller, R., Misund, K., Holien, T., Bachke, S., Gilljam, K. M., Vatsveen, T. K., Ro, T. B., Bellacchio, E., Sundan, A. and Otterlei, M. (2013). Targeting proliferating cell nuclear antigen and its protein interactions induces apoptosis in multiple myeloma cells. PLoS One *8*, e70430.

Moarefi, I., Jeruzalmi, D., Turner, J., O'Donnell, M. and Kuriyan, J. (2000). Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. J Mol Biol *296*, 1215–23.

Moldovan, G. L. and D'Andrea, A. D. (2009). How the fanconi anemia pathway guards the genome. Annu Rev Genet *43*, 223–49.

Moldovan, G. L., Pfander, B. and Jentsch, S. (2007). PCNA, the maestro of the replication fork. Cell *129*, 665–79.

Moolenaar, G. F., Herron, M. F., Monaco, V., van der Marel, G. A., van Boom, J. H., Visse, R. and Goosen, N. (2000). The role of ATP binding and hydrolysis by UvrB during nucleotide excision repair. J Biol Chem *275*, 8044–50.

Mosedale, G., Niedzwiedz, W., Alpi, A., Perrina, F., Pereira-Leal, J. B., Johnson, M., Langevin, F., Pace, P. and Patel, K. J. (2005). The vertebrate Hef ortholog is a component of the Fanconi anemia tumor-suppressor pathway. Nat Struct Mol Biol *12*, 763–71.

Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr *53*, 240–55.

Niedzwiedz, W., Mosedale, G., Johnson, M., Ong, C. Y., Pace, P. and Patel, K. J. (2004). The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair. Mol Cell *15*, 607–20.

Niesen, F. H., Berglund, H. and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc *2*, 2212–21.

Nishino, T., Komori, K., Ishino, Y. and Morikawa, K. (2003). X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. Structure *11*, 445–57.

Nishino, T., Komori, K., Ishino, Y. and Morikawa, K. (2005a). Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. Structure *13*, 1183–92.

Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y. and Morikawa, K. (2005b). Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing. Structure *13*, 143–53.

Nowak-Karnowska, J., Chebib, Z., Milecki, J., Franzen, S. and Skalski, B. (2014). Highly Efficient Fluorescent Interstrand Photo-crosslinking of DNA Duplexes Labeled with 5-Fluoro-4-thio-2'-O-methyluridine. Chembiochem *15*, 2045–9.

Oakley, A. J., Prosselkov, P., Wijffels, G., Beck, J. L., Wilce, M. C. and Dixon, N. E. (2003). Flexibility revealed by the 1.85 A crystal structure of the beta sliding-clamp subunit of Escherichia coli DNA polymerase III. Acta Crystallogr D Biol Crystallogr *59*, 1192–9.

Pascal, J. M., Tsodikov, O. V., Hura, G. L., Song, W., Cotner, E. A., Classen, S., Tomkinson, A. E., Tainer, J. A. and Ellenberger, T. (2006). A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. Molecular cell *24*, 279–91.

Patel, S. S. and Donmez, I. (2006). Mechanisms of helicases. J Biol Chem *281*, 18265–8.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L. and McSweeney, S. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. J Synchrotron Radiat *20*, 660–4.

Pernot, P., Theveneau, P., Giraud, T., Nogueira Fernandez, R., Nurizzo, D., Spruce, D., Surr, J., McSweeney, S., Round, A., Felisaz, F., Foedinger, L., Gobbo, A., Huet, J., Villard, C. and Cipriani, F. (2010). New beamline dedicated to solution scattering from biological macromolecules at the ESRF. J. Phys.: Conf. Ser. *247*, 012009.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. and Svergun, D. I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. Journal of Applied Crystallography *45*, 342–350.

Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. and Svergun, D. I. (2007). ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. Journal of Applied Crystallography *40*, s223–s228.

Pfeiffer, H. and Liebhafsky, H. (1951). The origins of Beer's law. Journal of Chemical Education *28*, 123–125.

Pierce, M. M., Raman, C. S. and Nall, B. T. (1999). Isothermal titration calorimetry of protein-protein interactions. Methods *19*, 213–21.

Porath, J., Carlsson, J., Olsson, I. and Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. Nature *258*, 598–9.

Porod, G. (1982). General theory. Small-angle X-ray Scattering, edited by O. Glatter and O. Kratkv. London: Academic Press *0*, 17–51.

Prakash, S., Johnson, R. E. and Prakash, L. (2005). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. Annu Rev Biochem *74*, 317–53.

Putnam, C. D., Hammel, M., Hura, G. L. and Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys *40*, 191–285.

Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N. and Baumeister, W. (2000). The genome sequence of the thermoacidophilic scavenger Thermoplasma acidophilum. Nature *407*, 508–13.

Rupp, B. (2010). Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. Garland Science.

Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H. and Hakoshima, T. (2005). Structural basis for recruitment of human flap endonuclease 1 to PCNA. EMBO J *24*, 683–93.

Sambrook, J., Fritsch, E. and Maniatis, T. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press *2*, 163–170.

Sanchez-Weatherby, J., Bowler, M. W., Huet, J., Gobbo, A., Felisaz, F., Lavault, B., Moya, R., Kadlec, J., Ravelli, R. B. and Cipriani, F. (2009). Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. Acta Crystallogr D Biol Crystallogr *65*, 1237–46.

Sarkar, S., Davies, A. A., Ulrich, H. D. and McHugh, P. J. (2006). DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta. EMBO J *25*, 1285–94.

Scheller, J., Schurer, A., Rudolph, C., Hettwer, S. and Kramer, W. (2000). MPH1, a yeast gene encoding a DEAH protein, plays a role in protection of the genome from spontaneous and chemically induced damage. Genetics *155*, 1069–81.

Schmidt, A., Rothenfusser, S. and Hopfner, K. P. (2012). Sensing of viral nucleic acids by RIG-I: from translocation to translation. Eur J Cell Biol *91*, 78–85.

Schmidt, K. H., Viebranz, E. B., Harris, L. B., Mirzaei-Souderjani, H., Syed, S. and Medicus, R. (2010). Defects in DNA lesion bypass lead to spontaneous chromosomal rearrangements and increased cell death. Eukaryot Cell *9*, 315–24.

Schärer, O. D. (2013). Nucleotide excision repair in eukaryotes. Cold Spring Harb Perspect Biol *5*, a012609.

Schwab, R. A., Blackford, A. N. and Niedzwiedz, W. (2010). ATR activation and replication fork restart are defective in FANCM-deficient cells. EMBO J *29*, 806–18.

Shukla, P., Ghosh, K. and Vundinti, B. R. (2012). Current and emerging therapeutic strategies for Fanconi anemia. The HUGO Journal *6*, 1.

Singh, S., Shemesh, K., Liefshitz, B. and Kupiec, M. (2013). Genetic and physical interactions between the yeast ELG1 gene and orthologs of the Fanconi anemia pathway. Cell Cycle *12*, 1625–36.

Singh, T. R., Saro, D., Ali, A. M., Zheng, X. F., Du, C. H., Killen, M. W., Sachpatzidis, A., Wahengbam, K., Pierce, A. J., Xiong, Y., Sung, P. and Meetei, A. R. (2010). MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. Mol Cell *37*, 879–86.

Singleton, M. R., Dillingham, M. S. and Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem *76*, 23–50.

Skalski, B., Taras-Goslinska, K., Dembska, A., Gdaniec, Z. and Franzen, S. (2010). Photoinduced fluorescent cross-linking of 5-chloro- and 5-fluoro-4-thiouridines with thymidine. J Org Chem *75*, 621–6.

Smith, D. B. and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene *67*, 31–40.

Sohn, S. Y. and Cho, Y. (2009). Crystal structure of the human rad9-hus1-rad1 clamp. J Mol Biol *390*, 490–502.

Somyajit, K., Subramanya, S. and Nagaraju, G. (2012). Distinct roles of FANCO/RAD51C protein in DNA damage signaling and repair: implications for Fanconi anemia and breast cancer susceptibility. J Biol Chem *287*, 3366–80.

Steitz, T. A. (1999). DNA polymerases: structural diversity and common mechanisms. J Biol Chem *274*, 17395–8.

Stiewe, T. (2007). The p53 family in differentiation and tumorigenesis. Nat Rev Cancer *7*, 165–8.

Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. and Eisenberg, D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. Proc Natl Acad Sci U S A *103*, 8060–5.

Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of Applied Crystallography *25*, 495–503.

Svergun, D., Barberato, C. and Koch, M. H. J. (1995). CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography *28*, 768–773.

Tao, Y., Jin, C., Li, X., Qi, S., Chu, L., Niu, L., Yao, X. and Teng, M. (2012). The structure of the FANCM-MHF complex reveals physical features for functional assembly. Nat Commun *3*, 782.

Thompson, L. H. and Jones, N. J. (2010). Stabilizing and remodeling the blocked DNA replication fork: anchoring FANCM and the Fanconi anemia damage response. Mol Cell *37*, 749–51.

Tom, S., Henricksen, L. A., Park, M. S. and Bambara, R. A. (2001). DNA ligase I and proliferating cell nuclear antigen form a functional complex. J Biol Chem *276*, 24817–25.

Tsutakawa, S. E., Hura, G. L., Frankel, K. A., Cooper, P. K. and Tainer, J. A. (2007). Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. Journal of structural biology *158*, 214–23.

Unk, I., Hajdu, I., Blastyak, A. and Haracska, L. (2010). Role of yeast Rad5 and its human orthologs, HLTF and SHPRH in DNA damage tolerance. DNA Repair (Amst) *9*, 257–67.

Vagin, A. and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. J. Appl. Cryst. *30*, 1022–1025.

Van Houten, B. and McCullough, A. (1994). Nucleotide excision repair in E. coli. Ann N Y Acad Sci *726*, 236–51.

Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. and Wigley, D. B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell *97*, 75–84.

Volkov, V. V. and Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. Journal of Applied Crystallography *36*, 860–864.

Wang, A. T., Sengerova, B., Cattell, E., Inagawa, T., Hartley, J. M., Kiakos, K., Burgess-Brown, N. A., Swift, L. P., Enzlin, J. H., Schofield, C. J., Gileadi, O., Hartley, J. A. and McHugh, P. J. (2011). Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. Genes Dev *25*, 1859–70.

Wang, S. C. (2014). PCNA: a silent housekeeper or a potential therapeutic target? Trends Pharmacol Sci *35*, 178–86.

Wang, X., Andreassen, P. R. and D'Andrea, A. D. (2004). Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. Mol Cell Biol *24*, 5850–62.

Wang, Y., Leung, J. W., Jiang, Y., Lowery, M. G., Do, H., Vasquez, K. M., Chen, J., Wang, W. and Li, L. (2013). FANCM and FAAP24 maintain genome stability via cooperative as well as unique functions. Mol Cell *49*, 997–1009.

Warbrick, E. (1998). PCNA binding through a conserved motif. Bioessays *20*, 195–9.

Whitby, M. C. (2010). The FANCM family of DNA helicases/translocases. DNA Repair (Amst) *9*, 224–36.

White, M. F. (2003). Archaeal DNA repair: paradigms and puzzles. Biochem Soc Trans *31*, 690–3.

William Wilson, W. (2003). Light scattering as a diagnostic for protein crystal growth—A practical approach. Journal of Structural Biology *142*, 56–65.

Williams, H. L., Gottesman, M. E. and Gautier, J. (2012). Replication-independent repair of DNA interstrand crosslinks. Mol Cell *47*, 140–7.

Williams, H. L., Gottesman, M. E. and Gautier, J. (2013). The differences between ICL repair during and outside of S phase. Trends Biochem Sci *38*, 386–93.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235–42.

Wolff, P., Olieric, V., Briand, J. P., Chaloin, O., Dejaegere, A., Dumas, P., Ennifar, E., Guichard, G., Wagner, J. and Burnouf, D. Y. (2011). Structure-based design of short peptide ligands binding onto the E. coli processivity ring. J Med Chem *54*, 4627–37.

Wukovitz, S. W. and Yeates, T. O. (1995). Why protein crystals favour some space-groups over others. Nat Struct Biol *2*, 1062–7.

Xia, L., Zheng, L., Lee, H. W., Bates, S. E., Federico, L., Shen, B. and O'Connor, T. R. (2005). Human 3-methyladenine-DNA glycosylase: effect of sequence context on excision, association with PCNA, and stimulation by AP endonuclease. J Mol Biol *346*, 1259–74.

Xue, Y., Li, Y., Guo, R., Ling, C. and Wang, W. (2008). FANCM of the Fanconi anemia core complex is required for both monoubiquitination and DNA repair. Hum Mol Genet *17*, 1641–52.

Yan, Y., Akhter, S., Zhang, X. and Legerski, R. (2010). The multifunctional SNM1 gene family: not just nucleases. Future Oncol *6*, 1015–29.

Yang, H., Zhang, T., Tao, Y., Wang, F., Tong, L. and Ding, J. (2013). Structural insights into the functions of the FANCM-FAAP24 complex in DNA repair. Nucleic Acids Res *41*, 10573–83.

Yang, K., Moldovan, G. L. and D'Andrea, A. D. (2010). RAD18-dependent recruitment of SNM1A to DNA repair complexes by a ubiquitin-binding zinc finger. J Biol Chem *285*, 19085–91.

Yao, N. Y. and O'Donnell, M. (2012). The RFC clamp loader: structure and function. Subcell Biochem *62*, 259–79.

Yasue, A., Mitsui, S. N., Watanabe, T., Sakuma, T., Oyadomari, S., Yamamoto, T., Noji, S., Mito, T. and Tanaka, E. (2014). Highly efficient targeted mutagenesis in one-cell mouse embryos mediated by the TALEN and CRISPR/Cas systems. Sci Rep *4*, 5705.

Zheng, H., Wang, X., Warren, A. J., Legerski, R. J., Nairn, R. S., Hamilton, J. W. and Li, L. (2003). Nucleotide Excision Repair- and Polymerase eta-Mediated Error-Prone Removal of Mitomycin C Interstrand Cross-Links. Molecular and Cellular Biology *23*, 754–761.

Zheng, L., Dai, H., Hegde, M. L., Zhou, M., Guo, Z., Wu, X., Wu, J., Su, L., Zhong, X., Mitra, S., Huang, Q., Kernstine, K. H., Pfeifer, G. P. and Shen, B. (2011a). Fen1 mutations that specifically disrupt its interaction with PCNA cause aneuploidy-associated cancer. Cell Res *21*, 1052–67.

Zheng, X. F., Prakash, R., Saro, D., Longerich, S., Niu, H. and Sung, P. (2011b). Processing of DNA structures via DNA unwinding and branch migration by the S. cerevisiae Mph1 protein. DNA Repair (Amst) *10*, 1034–43.

Zhu, G. and Lippard, S. J. (2009). Photoaffinity labeling reveals nuclear proteins that uniquely recognize cisplatin-DNA interstrand cross-links. Biochemistry *48*, 4916–25.

Appendix

Abbreviations

°C	Degree Celsius
aa	Amino acid
AA	Acrylamide
af	*Archaeoglobus fulgidus*
ADP	Adenosine diphosphate
Amp	Ampicillin
AMP	Adenosine monophosphate
AMPPNP	Adenylyl-imidodiphosphate
APIM	AlkB homologue 2 PCNA-interacting motif
APS	Ammonium peroxydisulfate
Ara	Arabinose
ASEC	Analytical size exclusion chromatography
ATP	Adenosine triphosphate
ATPγS	Adenosine-5'-O-(3-thio triphosphate)
ATR	Ataxia telangiectasia and Rad3 related
ATRIP	ATR interacting protein
β-TRCP	β-transducin repeat containing protein
BAA	Bisacrylamide
BER	Base excision repair
BLM	Bloom helicase
bp	Base pairs
BRCA2	Breast cancer type 2 susceptibility protein
BRIP1	BRCA1-interacting protein 1
BS	Bloom's Syndrome
BSA	Bovine serum albumine
Cam	Chloramphenicol
CARD1/2	Caspase activation and recruitment domain 1/2
CD	Circular dichroism
CDK	Cyclin-dependent kinase
CS(A/B)	Cockaine syndrome protein (A/B)
ct	*Chaetomium thermophilum*
CV	Column volume
ΔG	Gibb's free energy
ΔH	Free enthalpy

ΔS	Entropy
d	Day
Da	Dalton
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
DSB	Double-strand break
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
E. coli	*Escherichia coli*
e.g.	For example (exempli gratia)
EDTA	Ethylenediaminetetraacetic acid
ERCC1/4	Excision repair cross-complementation group 1/4
F	Fluorescein
FA	Fanconi anemia
FAAP20/24/100	Fanconi anemia-associated protein of 20/24/100 kDa
FAN1	Fanconi-associated nuclease 1
FANC(A-Q)	Fanconi anemia complementation group (A-Q)
Fe-S	Iron-sulfur
FEN1	Flap endonuclease 1
Fml1	FANCM-like protein 1
g	Gram
GG	Global genome
GST	Glutathione S-transferase
h	Hour
HA	Hemagglutinin
HCLK2	Human homolog of the *C.elegans* biological clock protein CLK2
Hef	Helicase-associated endonuclease for fork-structured DNA
HEK293	Human embryonic kidney 293 cells
Hel1/2	Helicase domain 1/2
HeLa	Henrietta Lacks cells
HhH	Helix-hairpin-helix
HIRAN	HIP116, Rad5p N-terminal
His_6	Hexahistidine
HJ	Holliday junction
HLTF	Helicase-like transcription factor
HRP	Horseradish peroxidase
HRR	Homologous recombination repair
hs	*Homo sapiens*
HU	Hydroxyurea
ICL	Interstrand crosslink
IDCL	Interdomain connecting loop
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl-β-thiogalactoside
IR	Ionizing radiation

IRES	Internal ribosome entry site
ITC	Isothermal titration calorimetry
l	Liter
LDH	Lactate dehydrogenase
Kan	Kanamycine
K_D	Dissociation constant
LB	lysogeny broth
m	Meter
M	Molar
MCS	Multiple cloning site
MHF1/2	FANCM-associated histone-fold-containing protein 1/2
min	Minute
MM1/2	Protein-protein interaction motif in FANCM 1/2
MMC	Mitomycin C
MMR	Mismatch repair
Mms2	Ubiquitin-conjugating enzyme variant MMS2
MPG	DNA-3-methyladenine glycosylase
Mph1	Mutator phenotype 1
MR	Molecular replacement
MS	Mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut-off
NADH	Nicotinamide adenine dinucleotide
NER	Nucleotide excision repair
NMR	Nuclear magnetic resonance
OD	Optical density
Ori	Origin of replication
ORF	Open reading frame
PAA	Polyacrylamide
PAGE	Polyacrylamide gel electrophoresis
PALB2	Partner and localizer of BRCA2
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylenglycol
PEP	Phosphoenolpyruvate
pf	*Pyrococcus furiosus*
PIP	PCNA-interacting peptide
PK	Pyruvate kinase
Plk1	Polo-like kinase 1
Pol	Polymerase
PRR	Post-replication repair
PSO2	Sensitivity to psoralen 2
Rad	Radiation sensitive

RecA	Recombination protein A
RFC	Replication factor C
RIG-I	Retinoic acid inducible gene I
RING	Really interesting new gene
RIR	Replication-independent ICL repair
RNA	Ribonucleic acid
RT	Room temperature
s	Second
SAXS	Small-angle x-ray scattering
SCF	Skp, Cullin, F-box containing
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SF2	Superfamily 2
SLX4	Structure-specific endonuclease subunit
SNM1A	Sensitivity to nitrogen mustard 1 A
ssDNA	Single-stranded DNA
T	Temperature
ta	*Thermoplasma acidophilum*
TCR	Transcription-coupled repair
TEMED	Tetramethylethylenediamine
TEV	Tobacco etch virus
TLS	Translesion synthesis
TopBP1	DNA topoisomerase 2-binding protein 1
Ubc13	Ubiquitin-conjugating enzyme 13
Ube2T	Ubiquitin-conjugating enzyme E2 T
UV	Ultraviolet
WT	Wild type
XP(A-G)	Xeroderma pigmentosum complementation group (A-G)
XRCC9	X-ray repair cross-complementation group 9

Amino Acids

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamic acid	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophane	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

Protein Sequences

taHef

MIIQPREYQLNVFKNSIGQNTLIVMPTGLGKTIIAAMFIEKFYGEKKRSL 50
MVAPTKPLVLQHARTIADSTGISENEIGVFTGEIDAEERELIWVTKKIFV 100
STPQVVFNDMRSGILDITKFDLIIFDEAHRAVGNYAYVDIAQEYLKYKKK 150
LIIGLTASPGSSREKIDEIMANLGISKVIAKTEMDEDVRKYVKTIDVKLI 200
KIREPDGTKEIVNMIDAVMSKLLEPVKSSNFSIGRSRKDLVATMQKVINS 250
AKEDRSLFQLVRRITAAIRLDYLKEYIETQGIDVALHYLEEMEKSQDPGI 300
RRAISMLSSLDEFATLRTKLMLAADGYTNPKMEKTLSIMESNVHGNARAI 350
VFTHYRITSDLLMDYIQKNGTSLKPVRFIGQADRGSDEGLSQDQQKKIIE 400
DFKNNVYNVLIATSIAEEGLDIPDTDFVIFYEAVPSEIRYIQRKGRTGRS 450
RNGQVYILVFENSRDMAYYYSSIRKVSRMTQTIEDLRQESGSDEKEEKSD 500
VQKTIFDF 508

ctMph1

MGKDFDEYDDFSDEELLVALTQAVDAAAEAALTGNETRGNSSNPLAVQPP 50
PSQFSRRTSPPRPVSQFSQPVQPRKPGQFSQFTQPAQSGQAGQPSQFSQF 100
NQTAQTKQGGISQFSRGPARTLSNTSNRSYQSNPTSKSPTTSPERPRQPA 150
APRTLQRHPSHTQTYRQTTLFGGTLREDPGRITASQTLNNRPYLADLPLE 200
PPTHHKLNESELGKWVYPLNLGPIRDYQFSIVKNALFHNTLVALPTGLGK 250
TFIAATVMLNFFRWTKDAKIVFAAPTKPLVSQQVEACLNVAGIPRSQTTL 300
LTGETSPSLRREEWETKRLFFMTPQTLMNDLSIGSADPKSIVLLVIDEAH 350
RATGDYAYVKVIEFMRRFNKSFRVLALTATPGSTVEGVQEVIDNLCISHV 400
EIRTEESIDIRQYVHSRNIETVTFDPSDEMEEVKSLFSKALKPLVDKLSS 450
QNIYYGRDPMSLTTYGLLKSRNEWLAGPGRHANQGLKFMIMSVFGILQSV 500
AHSIKLLNFHGIKPFYYAMAEFRSTEEAKGERGSKLKRQLLNDESFQKMM 550
TMIEKWMKLKDFHGHPKLTYLCQVLVNHFADAGEGSNTRAIVFAEYRDSA 600
EEIVRLLNTQPMVSATVFVGQADSKRSEGMKQKQQIETIEKFKSGQFNVL 650
VATSIGEEGLDIGQVDLIICYDASSSPIRMLQRMGRTGRKRAGNIVLLLM 700
KGKEEEKFAEAKDNYQKMQQLICDGASFTFRHELSARIIPRDIKPEVDKR 750
HIDIPIENTQDKSLPEPKKTPLRGKKKPLKKKFHMPDGVITGFIKASEFG 800
QHSGKNQAGNAPKEPSELEELADIPRLDRVFLDKDKINQLNRRYKTLPRG 850
QNGKTEEIDASTLTASPMALVALTRTSKIKHSDLTKRCVQLFQKLKERRE 900
KSWAYEDPCPDFDPKDWQKLPMTPYADDTDGGQDVDDHDESSKKSPTRPR 950
QTKRGRVPSSGASSPAGKVTKKRRIDHNWLDDCAKVVDDDDEEAAASATT 1000
TPAHGRGHNGKGGRTPSSIRGRIKRVGDHLEDIGDDCTRTSDMAETDGSD 1050
SGGDLADFIVDDDVPSSALRVRSSGDEDGFSSSLPASSGLSLRLGKQEKD 1100
EVEEKPFWEPTKFDPTQDSGSDDEMPDLDQIVESAKKEKQKQKEDKGEWK 1150
EKATSRVARTRTTTRNGGDEENEDEEPRPSPRKPLGKRRKQILTDSEDDE 1200

hsFANCM

MSGRQRTLFQTWGSSISRSSGTPGCSSGTERPQSPGSSKAPLPAAAEAQL 50
ESDDDVLLVAAYEAERQLCLENGGFCTSAGALWIYPTNCPVRDYQLHISR 100
AALFCNTLVCLPTGLGKTFIAAVVMYNFYRWFPSGKVVFMAPTKPLVTQQ 150
IEACYQVMGIPQSHMAEMTGSTQASTRKEIWCSKRVLFLTPQVMVNDLSR 200
GACPAAEIKCLVIDEAHKALGNYAYCQVVRELVKYTNHFRILALSATPGS 250
DIKAVQQVITNLLIGQIELRSEDSPDILTYSHERKVEKLIVPLGEELAAI 300
QKTYIQILESFARSLIQRNVLMRRDIPNLTKYQIILARDQFRKNPSPNIV 350
GIQQGIIEGEFAICISLYHGYELLQQMGMRSLYFFLCGIMDGTKGMTRSK 400
NELGRNEDFMKLYNHLECMFARTRSTSANGISAIQQGDKNKKFVYSHPKL 450
KKLEEVVIEHFKSWNAENTTEKKRDETRVMIFSSFRDSVQEIAEMLSQHQ 500
PIIRVMTFVGHASGKSTKGFTQKEQLEVVKQFRDGGYNTLVSTCVGEEGL 550
DIGEVDLIICFDSQKSPIRLVQRMGRTGRKRQGRIVIILSEGREERIYNQ 600
SQSNKRSIYKAISSNRQVLHFYQRSPRMVPDGINPKLHKMFITHGVYEPE 650
KPSRNLQRKSSIFSYRDGMRQSSLKKDWFLSEEEFKLWNRLYRLRDSDEI 700
KEITLPQVQFSSLQNEENKPAQESTTGIHQLSLSEWRLWQDHPLPTHQVD 750
HSDRCRHFIGLMQMIEGMRHEEGECSYELEVESYLQMEDVTSTFIAPRNE 800
SNNLASDTFITHKKSSFIKNINQGSSSSVIESDEECAEIVKQTHIKPTKI 850
VSLKKKVSKEIKKDQLKKENNHGIIDSVDNDRNSTVENIFQEDLPNDKRT 900
SDTDEIAATCTINENVIKEPCVLLTECQFTNKSTSSLAGNVLDSGYNSFN 950
DEKSVSSNLFLPFEEELYIVRTDDQFYNCHSLTKEVLANVERFLSYSPPP 1000
LSGLSDLEYEIAKGTALENLLFLPCAEHLRSDKCTCLLSHSAVNSQQNLE 1050
LNSLKCINYPSEKSCLYDIPNDNISDEPSLCDCDVHKHNQNENLVPNNRV 1100
QIHRSPAQNLVGENNHDVDNSDLPVLSTDQDESLLLFEDVNTEFDDVSLS 1150
PLNSKSESLPVSDKTAISETPLVSQFLISDELLLDNNSELQDQITRDANS 1200
FKSRDQRGVQEEKVKNHEDIFDCSRDLFSVTFDLGFCSPDSDDEILEHTS 1250
DSNRPLDDLYGRYLEIKEISDANYVSNQALIPRDHSKNFTSGTVIIPSNE 1300
DMQNPNYVHLPLSAAKNEELLSPGYSQFSLPVQKKVMSTPLSKSNTLNSF 1350
SKIRKEILKTPDSSKEKVNLQRFKEALNSTFDYSEFSLEKSKSSGPMYLH 1400
KSCHSVEDGQLLTSNESEDDEIFRRKVKRAKGNVLNSPEDQKNSEVDSPL 1450
HAVKKRRFPINRSELSSSDESENFPKPCSQLEDFKVCNGNARRGIKVPKR 1500
QSHLKHVARKFLDDEAELSEEDAEYVSSDENDESENEQDSSLLDFLNDET 1550
QLSQAINDSEMRAIYMKSLRSPMMNNKYKMIHKTHKNINIFSQIPEQDET 1600
YLEDSFCVDEEESCKGQSSEEEVCVDFNLITDDCFANSKKYKTRRAVMLK 1650
EMMEQNCAHSKKKLSRIILPDDSSEEENNVNDKRESNIAVNPSTVKKNKQ 1700
QDHCLNSVPSGSSAQSKVRSTPRVNPLAKQSKQTSLNLKDTISEVSDFKP 1750
QNHNEVQSTTPPFTTVDSQKDCRKFPVPQKDGSALEDSSTSGASCSKSRP 1800
HLAGTHTSLRLPQEGKGTCILVGGHEITSGLEVISSLRAIHGLQVEVCPL 1850
NGCDYIVSNRMVVERRSQSEMLNSVNKNKFIEQIQHLQSMFERICVIVEK 1900
DREKTGDTSRMFRRTKSYDSLLTTLIGAGIRILFSSCQEETADLLKELSL 1950
VEQRKNVGIHVPTVVNSNKSEALQFYLSIPNISYITALNMCHQFSSVKRM 2000
ANSSLQEISMYAQVTHQKAEEIYRYIHYVFDIQMLPNDLNQDRLKSDI 2048

taPCNA

MIRLNLSVKNLKEITDLLSTIVSEAKFRVDENGMSVTAVDPAHVAMIRLE 50
VPKEAFVEFHTDGQEEIALDIDRLKSVIRLASSSENVGITKDGEKLKFEL 100
GTINKSISLLDPSTIVTPKIPNITSEYYAVLKKSDFERGLRAAEDISDSI 150
RFTLSQDGFKAYSHSESEESEMILPKDLITDMSCNTTIKSSYPLEYLLKF 200
IKAISSTDSLKLSFRDDYPLSVEFYLDQNPGAKIKGLFLLAPRMEQ 246

ctPCNA

MLEARLEQASILKKVVDAIKDLVQDCNFDCNDSGIALQAMDNSHVALVSM 50
MLKAEGFSPYRCDRNIALGVNLTSLTKVLRAAQNEDILTLKAEDAPDVLN 100
LVFESSETDRISEYDLKLMDIDQEHLGIPETEYAATITMPSNEFKRITTD 150
LMAMSESVTIEANKDGVKFSCQGDIGNGSVTLRQHTNVEKPNESIEIELS 200
EPVSLTFSLKYLVNFCKASALSNTVKICLSNEVPLLVEYSLGGSSYLRFY 250
LAPKIGDDE 259

hsPCNA

MFEARLVQGSILKKVLEALKDLINEACWDISSSGVNLQSMDSSHVSLVQL 50
TLRSEGFDTYRCDRNLAMGVNLTSMSKILKCAGNEDIITLRAEDNADTLA 100
LVFEAPNQEKVSDYEMKLMDLDVEQLGIPEQEYSCVVKMPSGEFARICRD 150
LSHIGDAVVISCAKDGVKFSASGELGNGNIKLSQTSNVDKEEEAVTIEMN 200
EPVQLTFALRYLNFFTKATPLSSTVTLSMSADVPLVVEYKIADMGHLKYY 250
LAPKIEDEEGS 261

taFEN1

MGTDISDIVVSHETSLKDQGNQTFSIDTYNILYQLLSNVRQYDGMPLMDS 50
HGNVTSHLYGIFYRTINLLENRIRPVYVFDGKPSPLKNRTISERQMMKEK 100
AKAELEEAIERGEEDLRQYYSRINYITPQIVDDTKKLLDYMGIPYVDAPS 150
EGEAQASYMTKKNVDGVISQDYDCLLFGARKILRNFAIYGRRKVPRKNIY 200
KTVYPEYIILDEVLSANQINQDQLIGIGILVGTDFNEGIKGIGAKKALAL 250
IKKEGDIKSVLKHIGKNIENLDEIIDFFKNPPVVDYDFKFRKPDTDAIEH 300
FLCDEHDFSRERIRDHLESLRKNDQASTQFRLDSFD 336

hsFEN1

MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGG 50
DVLQNEEGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKR 100
SERRAEAEKQLQQAQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI 150
PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAK 200
KLPIQEFHLSRILQELGLNQEQFVDLCILLGSDYCESIRGIGPKRAVDLI 250
QKHKSIEEIVRRLDPNKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSE 300
PNEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLDDFFKVTGSL 350
SSAKRKEPEPKGSTKKKAKTGAAGKFKRGK 380

Affidavit / Eidesstattliche Erklärung

Affidavit

I hereby confirm that my thesis entitled

The Intricate Network of
Replication-dependent Interstrand Crosslink DNA Repair

is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, ...
(Date) (Signature)

Eidesstattliche Erklärung

Hier mir erkläre ich an Eides statt, die Dissertation

Das komplexe Netzwerk der
replikationsabhängigen Reparatur von DNA-Quervernetzungen

eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, ...
(Datum) (Unterschrift)

This page is intentionally left blank.

Publications

Research Article

Florian Rohleder, Jochen Kuper, Adam Round, and Caroline Kisker (2014). *A new member within the PCNA interacting network: The archaeal FANCM homolog Hef.* Nucleic Acids Res *(manuscript under revision).*

Posters

06/2014	FASEB Conference "Machines on Genes": *A New Member within the PCNA Interacting Network: the Archaeal FANCM Homolog Hef.*
10/2013	RVZ Annual Retreat & GSLS Students Symposium SciX: *FANCM Protein Family - Structural Insights & Interaction with PCNA.*
10/2012	RVZ Annual Retreat & GSLS Students Symposium Epos: *The Archaeal FANCM Homolog Hef - Interaction with PCNA.*
10/2011	RVZ Annual Retreat & GSLS Students Symposium Bio Bang: *The Archaeal FANCM Homolog Hef - Structure & Interactions.*
10/2010	GSLS Students Symposium Chiasma: *DNA Damage Detection in Fanconi Anemia.*
10/2009	RVZ Annual Retreat: *Towards the Crystal Structure of Human Helicase-like Transcription Factor.*
07/2009	EMBL Workshop on Characterization of Protein Complexes: *Towards the Crystal Structure of the FEN1-PCNA Complex.*
09/2008	RVZ Annual Retreat: *Protein Assemblies in Nucleotide Excision Repair.*

Oral Presentations

02/2014	RVZ Symposium: *A New Member within the PCNA Interacting Network: the Archaeal FANCM Homolog Hef.*
08/2013	Biotechnology Symposium, Bremen: *Structural and Functional Analysis of DNA Damage Detection in Fanconi Anemia.*
05/2011	RVZ Seminar: *The Archaeal FANCM Homolog Hef: Structure & Interactions.*
09/2010	RVZ Annual Retreat: *DNA Damage Detection in Fanconi Anemia.*
07/2010	Biotechnology Symposium, Bad Neustadt a.d. Saale: *Structure and Function of a Helicase.*
01/2010	RVZ Seminar: *Interaction between Archaeal FEN1 and XPD & Crystallization of Human HLTF.*
07/2009	EMBL Workshop on Characterization of Protein Complexes: *X-ray Crystallography.*
06/2009	Biotechnology Symposium, Würzburg: *Eukaryotic DNA Repair.*
07/2008	Biotechnology Symposium, Hanau: *Nucleotide Excision Repair.*

Acknowledgments

Primarily, I am highly delighted to thank my supervisor and mentor Caroline Kisker for the opportunity to conduct my PhD research in her working group as well as for her great scientific guidance, support, and encouragement.

Special thanks also to my advisory committee: Roland Benz for great (not only) scientific discussions and the amazing annual Biotechnology Symposium; Detlev Schindler as the real Fanconi anemia expert for his advices and help especially with the human cell lines. I would also like to express my gratitude to Hermann Schindelin for his support and explanations concerning protein crystallography. Furthermore, I owe Bennett van Houten many open-minded, fruitful and vivid discussions particularly about DNA repair.

It is a great pleasure to thank my postdoc Jochen Kuper for the extraordinary human and professional supervision. I really enjoyed the sublime teamwork inside and the fun activities outside the lab.

I would also like to thank my collaboration partners Adam Round from the EMBL in Grenoble (France) for the SAXS data analysis and Weidong Wang & Yutong Xue from the NIA/NIH in Baltimore (USA) for the Co-IP and human cell culture experiments.

Moreover, thanks to Heidi Roth and Stefanie Wolski for cordially accepting me in the DNA repair group and countless constructive discussions; Bodo Sander for introducing me to SAXS and ITC as well as for the good atmosphere in the lab; Hans Maric also for help with ITC; Agnes Elias for cloning new FANCM variants; Florian Sauer for the great help with the ctPCNA structure. I also thank Bernhard Fröhlich and Christian Weinberger for technical support concerning all kinds of IT issues.

A remarkable quality of the RVZ is the scientific exchange across lab borders. So I would like to thank Annegrit Seifried for help with the CoIP experiments; Simon Stritt for hands-on explanations concerning immunostaining; Mike Friedrich and Jürgen Pinnecker for help with the STED microscope for confocal imaging; Heike Hermanns for providing human cell lines and corresponding protocols; Jens Vanselow for the mass spectrometry analysis of ctMph1.

I would also like to thank my present and past colleagues and friends at the Department of Structural Biology for the very friendly, collaborative, and productive working atmosphere. Finally, special thanks also to the Rudolf Virchow Center itself and the Graduate School of Life Sciences for providing an outstanding scientific research environment.

Mein ganz besonderer Dank gilt natürlich meinen Eltern Ilona und Wolfgang Rohleder für ihr Vertrauen und ihre Unterstützung. Ihnen möchte ich diese Arbeit widmen.