

Structural Insights into the Fanconi Anaemia Pathway

Colette Lipp
DPhil Candidate
Trinity Term 2020

Department of Biochemistry
St. Peter's College



Under the supervision of:

Dr. Martin Cohn

Acknowledgements

I would like to thank my parents, as without their continued love, support and encouragement, I would not be where I am today. Their frequent trips to Oxford have also allowed Oxford to feel like a home away from home. I hope that I have made you proud and I know that you are excited about where my life may lead me.

I would also like to thank my supervisor, Dr. Martin Cohn, for the opportunity to study at Oxford and to complete my DPhil project. In addition, I would like to thank my secondary supervisor, Dr. Paul Elliot for his extensive guidance and advice, especially when the road became rocky.

Throughout my DPhil I have received extensive additional guidance. I would like to thank my thesis committee members, Dr. Tim Nott, Professor Neil Brockdorff and Associate Professor Maike Bublitz, as without their help and guidance I would not have been able to complete my DPhil. I would also like to thank Dr. Krzystof Kus for his guidance and troubleshooting surrounding size-exclusion chromatography. Further to this, I would like to thank the Cryo-EM facility at the Dunn school. In particular, I would like to thank Dr. Errin Johnson and Adam Costin for training me, for their great help and advice, as well as for always making me feel welcome with their Aussie accents and offerings of Aussie snacks. In addition, I would also like to thank Dr. Joseph Caesar and Dr. Justin Deme for their advice with regards to my cryo-EM data processing.

I would also like to thank the members of the Cohn laboratory for making the lab environment an incredibly enjoyable one. In particular, I would like to thank a former member, David Lopez-Martinez, for his constant advice, assistance and perseverance with my never-ending questions and concerns. I would not have made it through my DPhil without your guidance! In addition, I would also like to thank Ganesha Pitchai for assisting with my initial training in the lab. I would also like to thank Kelvin Yaprianto and Di Yang for their helpful discussions, as well as for their morale and enthusiasm. We started our DPhils together and you have been an immense source of motivation and support over the last three years.

Finally, I would like to thank Stacey, as without her unconditional love, support, motivation, reassurance and ability to make me forget all my worries, I never would have made it through my DPhil. You always know how to pick me up when I am down. Meeting you and making unforgettable memories has been the best thing about my time in Oxford and is what I cherish most. I only hope that I have made you proud and that I can now do the same for you.

Abstract

The human genome is subject to numerous endogenous and exogenous stresses that cause a variety of DNA damage and has thus evolved a complex DNA damage response (DDR). Within this DDR are a number of DNA repair pathways, which have evolved to repair different types of DNA damage. One such mechanism is the Fanconi Anaemia (FA) pathway, which repairs interstrand crosslinks (ICLs). Crucial to the activation of this pathway is the recognition of the ICL, and the subsequent recruitment and ubiquitination of the FANCD2/FANCI heterodimeric complex. This complex is site-specifically monoubiquitinated on each protomer which is essential for the removal of the ICL and the subsequent repair of DNA. In addition, the deubiquitination of the monoubiquitinated FANCD2/FANCI has been shown to be necessary for the FA pathway to function. The FA pathway is a complex pathway and many open questions remain to be answered, including the mechanism of recognition of the ICL, how the FANCD2/FANCI complex is ubiquitinated, and whether it is ubiquitinated as a heterodimeric complex or sequentially as single substrates, as well as whether ubiquitination and deubiquitination occur on or off DNA *in vivo*. In addition, it has been postulated that upon ubiquitination of the complex, FANCD2/FANCI undergo a conformational change locking onto DNA, perhaps necessary for the recruitment of downstream repair factors.

Here, *in vitro* investigation of the key FANCD2/FANCI complex was undertaken, including studies on whether monoubiquitination is stimulated by DNA, and whether deubiquitination of the monoubiquitinated FANCD2/FANCI occurs on or off DNA. Work towards the structures of FANCD2/FANCI and possible ubiquitination intermediates, monoubiquitinated-FANCD2/FANCI and monoubiquitinated-FANCI/FANCD2 via negative staining electron microscopy (EM) and cryo-EM is also described. Furthermore, the investigation of filament formation upon ubiquitination of FANCD2/FANCI was carried out. Additionally, a previously identified ICL sensor protein, UHRF1, was also investigated. Specifically, the effect of UHRF1 on the monoubiquitination of FANCD2/FANCI was explored *in vitro* as well as progress towards the first full-length UHRF1 structure via negative staining EM and cryo-EM is described.

Table of Contents

Acknowledgements **3**

Abstract **5**

Table of Contents **7**

Abbreviations **10**

1. Introduction **15**

1.1 DNA Damage Response **17**

1.1.1 DNA Damaging Agents 17

1.1.2 DNA Repair Pathways 19

1.1.3 Fanconi Anaemia Pathway 27

1.1.3.1 Sources of Interstrand Crosslinks 27

1.1.3.2 Repair of Interstrand Crosslinks 31

1.1.3.3 Alternative Roles of the FA Pathway 45

1.1.4 Regulation of the DNA Damage Response by Post-Translational Modifications 47

1.1.5 Translational Impact of the DNA Damage Response: Novel Cancer Therapeutics 52

1.2 Determining Protein Structure: an Overview **54**

1.2.1 Transmission Electron Microscopy 54

1.2.1.1 Negative Staining 56

1.2.1.2 Cryo-Electron Microscopy 58

1.3 Aims of the Project **62**

2. Expression and Purification of Proteins and Optimisation of Ubiquitination Assay **63**

2.1 Experimental Design **63**

2.2 Purification of UBA1 (E1), UBE2T (E2), FANCL (E3), Ubiquitin, FANCD2/FANCI, FANCD2, FANCI and scDNA **65**

2.3 Optimisation of Ubiquitination Assay: Improvement in the Efficiency of Ubiquitination 65

2.4 Optimisation of the Sample Concentration for Structural Studies 67

3. Structural Studies of the Monoubiquitinated FANCD2/FANCI Complex 71

3.1 Experimental Design 72

3.2 Ubiquitination of FANCD2/FANCI and FANCI alone is DNA-Dependent.... 73

3.3 Deubiquitination of Monoubiquitinated-FANCD2/FANCI and Monoubiquitinated-FANCI Alone is DNA-Independent..... 77

3.4 Towards an Improved Cryo-EM Structure of FANCD2/FANCI 80

3.5 Towards a Cryo-EM Structure of Monoubiquitinated-FANCD2/FANCI 91

3.6 Towards a Cryo-EM Structure of Monoubiquitinated-FANCI/FANCD2 97

3.7 The Monoubiquitinated-FANCD2/FANCI Intermediate Complex does not Appear to form Filamentous Arrays 103

3.8 Discussion 116

4. Structural Studies of UHRF1.....135

4.1 Experimental Design.....135

4.2 UHRF1 hinders the Ubiquitination of FANCD2/FANCI 136

4.3 Towards a Cryo-EM Structure of UHRF1 146

4.4 Discussion 159

5. Discussion.....164

5.1 UHRF1.....164

5.2 FANCD2/FANCI 166

5.3 Future Work and Conclusions.....168

6. Materials and Methods 171

6.1 UBA1 (E1) Protein Expression and Purification.....171

6.2 UBE2T (E2) Protein Expression and Purification 171

6.3 FANCL (E3) Protein Expression and Purification.....172

6.4 FANCD2/FANCI Protein Expression and Purification 173
6.5 FANCI Protein Expression and Purification 173
6.6 FANCD2 Protein Expression and Purification 174
6.7 Ubiquitin Protein Expression and Purification 175
6.8 USP1/UAF1 Protein Expression and Purification 175
6.9 UHRF1 Protein and UHRF1 Mutant Protein Expression and Purification 176
6.10 scDNA Transformation and Purification 177
6.11 Preparation of Biotinylated DNA Substrate for Filament Formation 177
6.12 FANCD2/FANCI Ubiquitination and Deubiquitination Assays 177
6.13 FANCI Ubiquitination and Deubiquitination Assays 178
6.14 Purification of the Monoubiquitinated-FANCD2/FANCI complex 178
6.15 Purification of the Monoubiquitinated-FANCI/FANCD2 complex 178
6.16 Purification of Filamentous Arrays 179
6.17 Filamentous Array Negative Control Sample Preparation 179
6.18 Negative Stain Electron Microscopy 180
6.19 Negative Stain Electron Microscopy Processing 180
6.20 Cryo-Electron Microscopy 181
6.21 Cryo-Electron Microscopy Processing 181
7. References 182

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
53BP1	p53 binding protein 1
γH2AX	Phosphorylated H2AX
AGT	Alkylguanine transferase
AML	Acute myeloid leukaemia
Alt-EJ	Alternative end joining
AMP	Adenosine monophosphate
AP	Apurinic/apyrimidinic
APE1	Apurinic/apyrimidinic endonuclease 1
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
ATR	Serine/threonine-protein kinase
BER	Base-excision repair
BLM	Bloom syndrome protein
BOD1L	Biorientation of chromosomes in cell division protein 1-like
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 2 susceptibility protein
CCD	Charged coupled device
Cdc	Cell division cycle
CDK1	Cyclin-dependent kinase 1
CHK1	Serine/threonine-protein kinase 1
CK2	Casein kinase 2
CMG	Cdc45, MCM, Go-Ichi-Ni-San
CSA	Cockayne syndrome WD repeat protein A
CSB	Cockayne syndrome WD repeat protein B
CTD	C-terminal domain
CtIP	C-terminal binding protein-interacting protein
dCMPs	Deoxycytidine monophosphates
DDR	DNA damage response
DEB	Diepoxybutane
DED	Direct electron detectors

DNA2	DNA2 nuclease-helicase
DNA-PKcs	DNA dependent protein kinase catalytic subunit
DNMT1	DNA methyltransferase 1
DSB	Double-strand break
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ELF	E74-like ETS transcription factor
EM	Electron microscopy
EMDB	Electron microscopy data bank
EME1	Essential meiotic structure-specific endonuclease 1
ERCC	Excision repair cross-complementation group
Exo	Exonuclease
FA	Fanconi Anaemia
FAAP	Fanconi Anaemia-associated protein
FAN1	FANCD2/FANCI-associated nuclease 1
FANC	Fanconi Anaemia complementation group
FEN1	Flap endonuclease 1
GG-NER	Global genome nucleotide excision repair
GINS	Go-Ichi-Ni-San
H2AX	H2A histone variant X
H3K9me3	Trimethylated histone H3 lysine 9
HDAC	Histone deacetylase
HECT	Homologous to the E6-AP carboxyl terminus
HR	Homologous recombination
ICL	Interstrand crosslink
LB	Luria Broth
LIG	DNA ligase
IR	Ionising radiation
LP-BER	Long-patch base-excision repair
MCM	Mini-chromosome maintenance
MDA	Malondialdehyde
MLH	Mutator L homologue
MMC	Mitomycin C

MMR	Mismatch repair
MRN	MRE11/RAD50/Nijmegen breakage syndrome 1
MSH	Mutator S homologue
Mut	Mutator
NBS1	Nijmegen breakage syndrome 1
NEIL3	Endonuclease 8-like 3
NER	Nucleotide-excision repair
NHEJ	Non-homologous end joining
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
PARP	Poly(ADP-ribose) polymerase
PCNA	Proliferating cell nuclear antigen
PDB	Protein data bank
PHD	Plant homeodomain
PIAS	Protein inhibitor of activated STAT
PNK	Polynucleotide kinase
POL	DNA polymerase
RAD23B	RAD23 homologue B
RAD51C	RAD51 homologue C
RAD51D	RAD51 homologue D
RAD54B	RAD54 homologue B
RFC	Replication factor C
RING	Really interesting new gene
RPA	Replication protein A
ROS	Reactive oxygen species
SAE	SUMO activating enzyme
scDNA	Supercoiled plasmid DNA
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SENP	Sentrin-specific protease
SLX	Scaffolding protein for nucleases
SMX	SLX1-SLX4/MUS81-EME1/XPF-ERCC1 tri-nuclease
SP-BER	Short-patch base-excision repair
SRA	SET and RING-associated
SSA	Single-strand annealing

SSB	Single-strand break
ssDNA	Single-stranded DNA
SUMO	Small ubiquitin-like modifier
TC-NER	Transcription coupled nucleotide-excision repair
TEM	Transmission electron microscopy
TFIIH	Transcription factor IIH complex
TLS	Translesion synthesis
TMP	Trimethylpsoralen
TTD	Tandem tudor domain
UAF1	USP1-associated factor 1
UBA	Ubiquitin-like modifier activating enzyme
UBE2T	Ubiquitin-conjugating enzyme E2 T
UBL	Ubiquitin-like
UBZ	Ubiquitin-binding zinc finger
UHRF	Ubiquitin-like containing PHD and RING finger domains
URD	Ubiquitin-dependent recruitment domain
USP1	Ubiquitin carboxyl-terminal hydrolase 1
UV	Ultra-violet
WRN	Werner protein
XLF	Xeroderma Pigmentosum complementation group 4-like factor
XP	Xeroderma Pigmentosum complementation group
XRCC	X-ray repair cross-complementing protein

1. Introduction

Fanconi Anaemia (FA) is a rare genetic disorder that is caused by defects in one of the 22 currently identified FA genes (Berger et al., 2018; Liang et al., 2019; Liu et al., 2020; Lopez-Martinez et al., 2016). It is a disorder that is both genetically and phenotypically heterogeneous, which is characterised by a diverse array of clinical symptoms. These include progressive bone marrow failure, progressive pancytopenia (deficiency of the red blood cells, white blood cells, and platelets), a predisposition to haematological malignancies and solid tumours, and chromosomal instability (Ciccia and Elledge, 2010; Jo and Kim, 2015; Moraes et al., 2012; Sims et al., 2007). Despite the varied clinical presentation of the disorder, FA can be accurately diagnosed by studying chromosomal breakage induced by DNA crosslinking agents, due to the hypersensitivity of FA cells to interstrand crosslinks (ICLs). The clastogenic effect of these agents on FA cells leads to a unique cellular marker of FA (Adamo et al., 2010; Auberbach, 2009; Schwab et al., 2015).

FA was first described in 1927 by Guido Fanconi (Altieri et al., 2008; Bhattacharjee and Nandi, 2017; Duxin and Walter, 2015). It is thought to have an incidence of 1 in 200-400,000 in the general population, however the incidence varies among different ethnicities (Lopez-Martinez et al. 2016; Walden and Deans, 2014). For example, in the Ashkenazi Jewish population, FA has an incidence of around 1 in 89. In 1982, the International Fanconi Anaemia Registry was established at The Rockefeller University for patients who have had their FA diagnosis confirmed by chromosomal breakage studies. The database is a centralised repository for clinical and genetic information on patients with FA (Auberbach, 2009). In 1998, the Fanconi Anaemia Mutation Database was established to compile the mutations identified in patients with FA (Ceccaldi et al., 2016).

FA patients usually develop bone marrow failure within the first two decades of their life (Ceccaldi et al., 2016). Pancytopenia contributes to cancer development, usually at a much earlier age than the general population. Due to the nature of the disorder, FA patients are particularly susceptible to developing haematological malignancies, such as acute myeloid leukaemia (AML) and solid tumours such as squamous cell carcinomas (Bhattacharjee and Nandi, 2017). Due to the increased success of bone marrow transplants, the average lifespan of an FA patient is now 33 years. However, FA patients must still be monitored for the occurrence of cancer (Ciccia and Elledge, 2010; Moraes et al., 2012).

FA patients have a predisposition to cancer due to a mutational defect in a DNA repair pathway, called the FA repair pathway, which leads to aberrant chromosomal breakage (Fang et al., 2020).

The detection and repair of DNA damage requires a comprehensive cellular response that is collectively known as the DNA Damage Response (DDR) (Almedia et al., 2020). The FA pathway is part of this response and mediates the repair of ICLs, which are highly toxic lesions that block DNA replication and transcription, although recent evidence has revealed that DNA synthesis can restart on the distal side to the ICL (Huang et al., 2019; Moraes et al., 2012; Velimezi et al., 2018). Patients with FA are hypersensitive to ICL-inducing agents, including many chemotherapeutic agents (Adamo et al., 2010; Joo et al., 2011; Moraes et al., 2012). Loss of function mutations in a number of genes involved in the DDR leads to a predisposition to a variety of cancers. However, defects in DDR are not only causative of cancer but can be exploited therapeutically (Hosoya and Miyagawa, 2014; Martin et al. 2008). For example, the cellular toxicity of ICLs is exploited in cancer chemotherapy, as tumour cells may have the inability to activate the appropriate DDR (Martin et al., 2008; Moraes et al., 2012). Therefore, fundamental analysis and scientific research of the DDR, and specifically the FA pathway, is necessary to be able to fully understand and exploit this response, to enable the development of targeted cancer therapies.

1.1 DNA Damage Response

The DDR involves a highly complex network of interacting and complementary DNA repair pathways (Altieri et al., 2008; Cussiol et al., 2020; Pearl et al. 2015). A DDR is initiated following DNA damage, which leads to either repair of the damaged lesion or cell death via apoptosis (Gavande et al., 2016; Li et al., 2016; O'Grady et al., 2014; Tian et al., 2014). If the DDR is dysfunctional, DNA lesions may remain unrepaired or become improperly repaired, which has serious consequences for the cell, including genomic instability and deregulation of cellular functions (Chatterjee and Walker, 2017; Liu et al., 2020; O'Connor, 2015; Smogorzewska et al., 2007). Cells are constantly exposed to genotoxic stress, and to maintain genome integrity have acquired specialized systems that detect and repair lesions (Hakem, 2008; Houtgraaf et al., 2006). Activation of this DDR is a complex cascade of reactions, involving recognition factors, transducers and effectors and these systems are tightly regulated by post-translational modifications. Upon recognition of the damaged DNA, the DDR activates the DNA damage checkpoint, which blocks the cell cycle until the damaged DNA is repaired, and also stimulates the DNA repair pathway that is necessary to deal with the specific type of DNA damage (Campos and Clemente-Blanco, 2020; Colombo et al., 2020; Hosoya and Miyagawa, 2014). Each DNA repair pathway is specific for a particular type of DNA damage, however, may be compensated for by the function of another DNA repair pathway (Curtin, 2012; Hakem, 2008). Many DNA damaging agents give rise to numerous types of DNA lesions and damage, which are summarized below.

1.1.1 DNA Damaging Agents

Cells are exposed to up to 10^5 DNA lesions per day. These lesions can be induced by a number of stresses, including from endogenous and exogenous sources (Campos and Clemente-Blanco, 2020; Colombo et al., 2020; Gavande et al., 2016). Under physiological conditions DNA is continuously damaged, for example through replication errors, programmed double-strand breaks (DSBs) and through the presence of reactive species produced by a normal metabolism. The major exogenous sources of DNA damage include ultra-violet (UV) radiation and ionizing radiation (IR) (Hakem, 2008; Houtgraaf et al., 2006; Tian et al., 2014).

Reactive Species

A major source of endogenous DNA damage is through oxidation of DNA components. This oxidation process may involve hydroxyl radicals, singlet oxygen or hydrogen peroxide, and each lead to different types of DNA modifications, including oxidized DNA bases, abasic sites, chain breaks and DNA-protein crosslinks (Altieri et al., 2008; Houtgraaf et al., 2006). Reactive oxygen

species (ROS) can also cause DNA damage when ROS exceed cellular antioxidant defences, known as oxidative stress. Excessive ROS causes chemical modifications to bases, chain breaks, adducts and crosslinks. Base-Excision Repair (BER) is the repair pathway that repairs most oxidatively induced DNA lesions, whilst Non-Homologous End Joining (NHEJ) repairs DSBs induced by ROS (Curtin, 2012; Davis and Lin, 2011; Friedberg, 2003; Jacq et al., 2013).

Chemical Species

Many chemical species cause a multitude of DNA damage. Free radicals generated by environmental pollutants and cigarette smoke can cause bulky chemical adducts, whilst exposure to reactive nitrogen species can deaminate DNA bases. Bases can also be methylated endogenously, although not all methylated products are cytotoxic (Davis and Lin, 2011; Jacq et al., 2013). Hypochlorous acid reacts with DNA inducing DNA-protein crosslinks and chlorination of DNA bases, whilst platinum compounds are capable of covalent binding leading to DNA-protein crosslinks, intrastrand crosslinks and ICLs. The Nucleotide-Excision Repair (NER) pathway repairs bulky chemical adducts, whilst the FA repair pathway in conjunction with Homologous Recombination (HR) and NER repairs ICLs (Altieri et al., 2008; Curtin, 2012).

Physical Sources

Major exogenous sources of DNA damage include UV radiation and IR. IR produces a wide variety of damage due to the generation of radical species. These include abasic sites and strand breaks (Altieri et al., 2008; Friedberg, 2003; Houtgraaf et al., 2006). UV radiation mainly induces photoproducts, the most common and hazardous of which are bulky lesions, for example cyclobutane pyrimidine dimers. IR oxidative damage is mainly repaired by BER, whilst strand breaks are repaired by NHEJ and HR. UV-related bulky lesions are mainly repaired by NER (Chatterjee and Walker, 2017; Ciccia and Elledge, 2010; Curtin, 2012; Jacq et al., 2013).

1.1.2 DNA Repair Pathways

Base-Excision Repair

Base-excision repair allows for the specific recognition and excision of a damaged DNA base. This damage includes many different base modifications that cause distortions in the DNA helix structure, including loss of a base (an apurinic or apyrimidinic site (AP site)), attack by ROS, base hydrolysis, single-strand breaks (SSBs), deamination and damage resulting from alkylating agents (Campos and Clemente-Blanco, 2020; Hosoya and Miyagawa, 2014; Madhusudan and Hickson, 2005; Tian et al., 2014). BER consists of five basic steps, all catalyzed by different enzymes. First, damaged bases are removed by DNA *N*-glycosylases to yield AP sites and the resulting AP sites are then processed by AP-specific endonucleases (or an AP-lyase activity associated with some *N*-glycosylases) to leave single-strand interruptions. Some breaks may require end-processing to generate unblocked 5' and 3' ends. Following this, the gaps are then filled in by a DNA polymerase and the resulting nicked strand is ligated together. Each of these steps is detailed below.

The first and most critical step in BER is recognition and excision of the damaged base. This is achieved by a class of enzymes called DNA *N*-glycosylases. These enzymes are classified by their three-dimensional (3D) structure and involve three different subgroups: helix-hairpin-helix, helix-two-turn-helix and uracil DNA *N*-glycosylases. Most DNA *N*-glycosylases are specific for a certain type of damaged base, however some exhibit broad substrate specificity. DNA *N*-glycosylases may be monofunctional or bifunctional. Monofunctional enzymes are capable of only excising the base from the DNA, resulting in an AP site, whilst bifunctional enzymes possess AP lyase activity and are capable of excising the base from the DNA and cleaving the DNA backbone on the 3' side of the AP site. However, both monofunctional and bifunctional DNA *N*-glycosylases require further endonuclease activity to fully cleave the damaged base (Altieri et al., 2008; Houtgraaf et al., 2006; Madhusudan and Hickson, 2005).

The key enzyme employed to generate AP sites with unblocked 5' and 3' ends is the apurinic/apyrimidinic endonuclease 1 (APE1). This enzyme generates unblocked 3' and 5' ends to enable DNA polymerization and ligation. For sites recognized by monofunctional DNA *N*-glycosylases, APE1 cleaves the DNA backbone on either side of the AP site. For sites recognized by bifunctional DNA *N*-glycosylases, APE1 cleaves the DNA backbone on the 5' side of the AP site (Altieri et al., 2008; Curtin, 2012, Hakem, 2008). An APE1-independent pathway has also been identified in mammals, whereby some bifunctional DNA *N*-glycosylases work with polynucleotide

kinases/phosphatases to generate SSBs (Chatterjee and Walker, 2017; Madhusudan and Hickson, 2005).

Following AP site and SSB generation, the resulting gap needs to be filled via DNA polymerase activity, followed by ligation to seal the remaining nick in the DNA strand. If the remaining gap is a single nucleotide, it is repaired by short-patch-BER (SP-BER), whilst if the gap extends beyond a single nucleotide, the gap is repaired by long-patch-BER (LP-BER) (Davis and Lin, 2011; Hakem, 2008). In SP-BER, the DNA polymerase that fills in the nucleotide gap is DNA polymerase β (POLβ) and the final ligation step is mediated by DNA ligase 3 (LIG3) associated with X-ray repair cross-complementing protein 1 (XRCC1). In LP-BER, the DNA polymerase that fills in a gap of between 2 – 20 nucleotides is DNA polymerase δ (POLδ) or DNA polymerase ε (POLε), with flap endonuclease 1 (FEN1) and proliferating cell nuclear antigen (PCNA) (Curtin, 2012; Houtgraaf et al., 2006). The final ligation step is mediated by DNA ligase 1 (LIG1), also associated with XRCC1. Poly(ADP-ribose) polymerase 1 (PARP1) is a key player in BER, as XRCC1 can only be recruited upon the activation of PARP1, achieved by the binding of PARP1 to SSBs (Altieri et al., 2008; Chatterjee and Walker, 2017; Madhusudan and Hickson, 2005).

Mismatch Repair

Mismatch repair (MMR) is essential to maintaining genome stability, as it recognizes mismatched base pairs and extra-helical loops introduced during DNA replication that have evaded the proofreading activity of DNA polymerases (Ciccia and Elledge, 2010; O'Grady et al., 2014). MMR has the specific ability to recognize the newly synthesised strand, resulting in preservation of genetic integrity. This results in an increased fidelity of up to 1000-fold in eukaryotes. If these DNA mismatches were to be left unrepaired, it would result in a permanent mutation, and thus a loss of genetic information (Altieri et al., 2008; Bouwman and Jonkers, 2012). In addition, MMR can repair DNA lesions resulting from a variety of stresses, including from intracellular metabolism and insults from exogenous sources. Defects in MMR have been implicated in drug resistance, particularly in cancer chemotherapeutics such as platinum agents and Temozolomide (Curtin, 2012; Li et al., 2016; O'Grady et al., 2014). MMR was originally elucidated in *Escherichia coli* (*E. coli*) but is highly conserved across prokaryotes and eukaryotes (Altieri et al., 2008; Curtin, 2012). The detailed steps of this repair pathway are outlined below.

In *E. coli*, three proteins are essential to MMR: Mutator S, L and H (MutS, MutL and MutH) (Hakem, 2008; O'Grady et al., 2014; Tian et al., 2014). Functioning as a homodimer, MutS acts as the sensor for the mismatched base, and actively recruits other proteins, such as MutL and MutH,

following a conformational change. In mammals, there are five Mutator S Homologues (MSH2 – MSH6) which all act as sensors for the DNA damage (Houtgraaf et al., 2006; Li et al., 2016). MSH2 is required for all mismatch correction, the heterodimer MSH2-MSH6 recognises mismatched nucleotides, the heterodimer MSH2-MSH3 recognises insertion and deletion loops, whilst MSH4 and MSH5 are involved in meiotic recombination mismatches (Altieri et al., 2008; Chatterjee and Walker, 2017; Colombo et al., 2020; Curtin, 2012).

Following recognition of the DNA damage by MutS, MutL is recruited, which acts as a link between the recognition of the damage and the activation of MutH, an endonuclease which creates a nick by the mismatched nucleotide. MutL also allows the recruitment of DNA helicase II , which unwinds the DNA from the nick generated by MutH (Altieri et al., 2008; Chatterjee and Walker, 2017; Houtgraaf et al., 2006). Once the DNA duplex is unwound, one of four (redundant) single-strand DNA-specific exonucleases (Exonucleases I, V and VII (ExoI, ExoV and ExoVII) and RecJ) excises the newly synthesized strand between the nick and the mismatch site. In mammals, there are three Mutator L Homologues (MLH1 – MLH3) which are necessary for downstream processing of the mismatch. These act as heterodimers, in concert with postmeiotic segregation increased 2. In mammals there is no homologue for MutH, however endonuclease I has been implicated in MMR. There is also no evidence of any helicase activity in mammals (Curtin, 2012; Hakem, 2008; Li et al., 2016; O'Grady et al., 2014).

Following excision of the mismatched base, the gap is filled in by replicative DNA polymerase III (POLIII), and the remaining nick sealed by DNA ligase. In mammals, the exact mechanism of the final step of repair is unknown, however POLδ, POLε and LIG1 have been implicated in DNA synthesis and sealing of the gap, together with a number of other repair proteins, including Replication protein A (RPA), PCNA, Replication factor C (RFC) and FEN1 (Altieri et al., 2008; Curtin, 2012; Houtgraaf et al., 2006; O'Grady et al., 2014).

Nucleotide-Excision Repair

Nucleotide-excision repair recognizes a wide variety of abnormal structural lesions, including those bulky adducts generated by UV radiation and other environmental carcinogens. It is a flexible damage-repair mechanism and recognizes lesions based on their bulky structure, such as cyclobutane pyrimidine dimers (Ciccia and Elledge, 2010; Davis and Lin, 2011; Hosoya and Miyagawa, 2014; O'Connor, 2015). NER utilizes 20-30 proteins in successive steps, involving the recognition of the bulky lesion, double incision of the damaged strand either side of the lesion, removal of the damaged oligonucleotide, and DNA polymerase and ligase activity to fill in and

seal the resulting gap (Colombo et al., 2020; Davis and Lin, 2011; Friedberg, 2003). There are two classes of NER, global genome repair (GG-NER), functional throughout the entire genome, and transcription coupled repair (TC-NER), functional on DNA strands of actively transcribed genes. Both pathways are similar, with differences in the damage recognition step. NER can also remove minor base damages which do not distort the double helix, as a back-up system for BER (Altieri et al., 2008; Moraes et al., 2012; Mouw and D'Andrea, 2014).

For GG-NER, recognition of the distorted DNA requires Xeroderma Pigmentosum complementation group C (XPC)/RAD23 homologue B (RAD23B). This heterodimeric complex is needed to recruit further proteins to the site of damage, such as DNA-damage-binding proteins 1 and 2 for repair of UV-induced damage, and allows the binding of additional factors which unwind the DNA double helix to signal the presence of the lesion (Altieri et al., 2008; Houtgraaf, et al., 2006). For TC-NER, RNA polymerase II recognises the lesion and stalls, promoting the recruitment of Cockayne syndrome WD repeat protein A and B (CSA and CSB), which in turn promote the dissociation of RNA polymerase II from the DNA strand to allow repair to continue. In both GG-NER and TC-NER, Xeroderma Pigmentosum complementation group A (XPA), RPA and transcription factor IIH complex (TFIIH) facilitate in unwinding the DNA and stabilizing the open DNA structure. TFIIH contains DNA helicases (Xeroderma Pigmentosum complementation groups B and D (XPB and XPD)) that unwind a stretch of 20-30 nucleotides around the damaged site (Chatterjee and Walker, 2017; Curtin, 2012; Hakem, 2008). The final step in forming the pre-incision structure involves RPA which is recruited via its interaction with XPA and is essential in stabilizing the open structure by binding to the single, undamaged DNA strand, protecting it from nuclease attack (Mouw and D'Andrea, 2014; Wang et al., 2018).

Once the pre-incision structure is formed, Xeroderma Pigmentosum complementation group G (XPG) and Xeroderma Pigmentosum complementation group F (XPF)/excision repair cross-complementation group 1 (ERCC1) excise the damaged strand on either side of the lesion (Chatterjee and Walker, 2017; Curtin, 2012). XPG first cuts the DNA strand a few nucleotides away from the lesion on the 3' side. Following XPG binding, XPF associates with ERCC1 to form a functional complex, and cuts the DNA strand 15-24 nucleotides away from the lesion on the 5' side (Hakem, 2008). Once the damaged lesion is removed, DNA repair is completed by DNA synthesis of the remaining gap by POLδ or POLε in complex with RFC and PCNA, followed by ligation of the resulting nick by LIG3 (Altieri et al., 2008; Houtgraaf et al., 2006).

Translesion Synthesis

Translesion Synthesis (TLS) is a mechanism whereby DNA damage may be bypassed so as not to stall replication forks (Curtin, 2012). As this mechanism is not a repair pathway, it is considered a DNA damage tolerance mechanism (Altieri et al., 2008). TLS uses specialized DNA polymerases called TLS polymerases, which lack the ability to proofread, thus resulting in error-prone DNA synthesis (Chatterjee and Walker, 2017). However, they allow the cell to survive, by allowing replication to continue. In *E. coli* there are three specific TLS polymerases, DNA polymerase II, IV and V (POLII, POLIV and POLV). In mammals, PCNA plays a central role in recruiting TLS polymerases. Multiple TLS polymerases have been identified including DNA Polymerase η, ι and κ (POLη, POLι and POLκ) and Rev1 (O'Connor, 2015; Tian et al., 2014). Defects in TLS contribute to carcinogenesis and can increase sensitivity to DNA-damaging agents (Friedberg, 2003; Lehmann, 2006).

Non-Homologous End-Joining

Two distinct but interconnected pathways repair most DSBs. DSBs can be induced endogenously, for example by ROS, or exogenously, for example by IR. DSBs can also arise as a result of unrepaired SSBs or unrepaired base lesions. One of the pathways repairing DSBs is Non-Homologous End-Joining. This pathway is functional throughout the entire cell cycle but acts primarily in the G0 and G1 phase (Altieri et al., 2008; Curtin, 2012; Zhao et al., 2017). It is the simplest mechanism to repair DSBs, by joining blunt DNA ends by direct ligation (Campos and Clemente-Blanco, 2020; Ciccia and Elledge, 2010). However, as no sequence homology is needed to join these ends together, it results in error-prone repair and can lead to sequence alterations at the break point or lead to genome rearrangements that can ultimately lead to apoptosis (Altieri et al., 2008; Curtin, 2012; Davis and Lin, 2011). Various proteins are necessary for NHEJ and which one depends on the type of end resulting from the DSB (Hosoya and Miyagawa, 2014; Houtgraaf et al., 2006; O'Connor, 2015).

If the DSB is a clean DSB with blunt ends, or with small complementary overhangs, only the core NHEJ machinery is required to repair the damage. The heterodimeric Ku70/Ku80 complex specifically binds to the DNA ends, which recruits the DNA dependent protein kinase catalytic subunit (DNA-PKcs) (Campos and Clemente-Blanco, 2020; Hakem, 2008). The binding of DNA-PKcs causes the Ku70/Ku80 dimer to rotate inward along the DNA helix, facilitating access of further proteins to the break site (Ciccia and Elledge, 2010; Madhusudan and Hickson, 2005). DNA-PKcs facilitate the synapsis of the broken DNA ends, and also undergoes

autophosphorylation, making the DNA ends available for the ligation step. Ligation is completed by DNA ligase 4 (LIG4), stabilized by X-ray repair cross-complementing protein 4 (XRCC4) and XRCC4-like factor (XLF) (Chatterjee and Walker, 2017; Tian et al., 2014). For DSBs resulting from a breakage, an additional protein complex is required to sense the damage. This complex is known as the MRN complex, and contains MRE11, RAD50 and Nijmegen breakage syndrome 1 (NBS1) (MRN complex) (Altieri et al., 2008; Colombo et al., 2020). The MRN complex also plays a role in repairing the DNA and may facilitate in tethering the two DNA ends together (Curtin, 2012; Houtgraaf et al., 2006; Wu et al., 2019).

For DSBs with so-called "dirty" DNA ends, end processing is required before NHEJ can take place. For example, IR-induced DSBs are not directly ligatable as they are not proper substrates for LIG4. These "dirty" ends can be processed by polynucleotide kinase (PNK), which interacts with XRCC4. Another subset of incompatible DNA ends include longer overhangs, hairpins, flaps and gaps. These DNA structures contain a single-strand DNA (ssDNA) double-strand DNA (dsDNA) transition and are specifically processed by the Artemis nuclease. The Artemis nuclease is recruited via DNA-PKcs. Several other proteins are implicated in NHEJ, including the exonuclease Werner protein (WRN) and the DNA polymerases, terminal deoxynucleotidyl-transferase, Polymerase λ (POLλ) and Polymerase μ (POLμ) (Hakem, 2008). For particularly difficult to join breaks, the MRN complex, ataxia telangiectasia mutated (ATM) serine/threonine kinase and phosphorylation of H2A histone variant X (H2AX) are involved (Altieri et al., 2008; Ciccia and Elledge, 2010). In addition, an alternative NHEJ pathway has been identified where PARP1 competes with the Ku complex for binding to the DNA break ends. Once PARP1 binds to the DSB ends, LIG3/XRCC1 seals the DNA break (Campos and Clemente-Blanco, 2020; Curtin, 2012; Madhusudan and Hickson, 2005).

Homologous Recombination

A large proportion of DSBs do not result from direct breakage of the DNA duplex, but instead occur as a result of stalled replication. An example of this is when a replication fork collides with an SSB, or when replication forks stall when arriving at base lesions. These highly toxic DSBs require complex processing in order to be accurately repaired, and this is termed Homologous Recombination (Campos and Clemente-Blanco, 2020; Hakem, 2008; O'Connor, 2015; Wu et al., 2010). This pathway uses sister chromatids or the homologous chromosome as a repair template, and thus is highly accurate. It is predominantly active in the S and G2 phase of the cycle. HR in eukaryotes is highly complex due to the large number of proteins involved in the pathway (Ciccia and Elledge, 2010; Hosoya and Miyagawa, 2014; Houtgraaf et al., 2006).

HR is initiated by the degradation of the 5'-end of the DNA molecule, to produce a resected single-strand. First, the MRN complex with C-terminal binding protein-interacting protein (CtIP) produces a short 3' overhang of ssDNA. MRN recruitment is facilitated by the breast cancer type 1 susceptibility protein (BRCA1) and promoted through the binding of NBS1 to phosphorylated H2AX (γH2AX) (Chatterjee and Walker, 2017; Ciccia and Elledge, 2010). Then, the MRN-CtIP complex together with ExoI extend the resection to give rise to a longer strand of ssDNA (Colombo et al., 2020; Houtgraaf et al., 2006). γH2AX aids in the recruitment of p53 binding protein 1 (53BP1) and BRCA1. This ssDNA strand is swiftly coated by RPA to protect the ssDNA from degradation by other nucleases (Campos and Clemente-Blanco, 2020; Tian et al., 2014; Wu et al., 2010).

HR then continues with the ssDNA-RPA invading the complementary sister chromatid, forming heteroduplex DNA (Campos and Clemente-Blanco, 2020; Houtgraaf, et al., 2006). This process requires the breast cancer type 2 susceptibility protein (BRCA2) and RAD51. BRCA2 helps in unloading RPA from the ssDNA and assists in loading RAD51 onto the ssDNA. This RAD51-coated DNA is termed the nucleofilament. RAD51 directs the 3' ssDNA tail of the nucleofilament to find, invade and pair with the undamaged homologous sequence, whilst BRCA2 controls the recombinase activity of RAD51 (Ciccia and Elledge, 2010; Davis and Lin, 2011). A number of additional protein factors assist in this process including RAD52, RAD54, RAD54 homologue B (RAD54B), RAD51 homologue B, C, D (RAD51B, RAD51C, RAD51D) and X-ray repair cross-complementing proteins 2 and 3 (XRCC2 and XRCC3). This strand invasion and formation of the heteroduplex DNA displaces a DNA strand of the sister chromatid, forming a D-loop (Bhat and Cortez, 2018; Hakem, 2008; Wu et al., 2010).

After D-loop formation, the newly annealed 3' end is extended through repair synthesis past the original break point to restore the missing genetic information. As the template used in this repair is the sister chromatid, the resulting repair is error-free (Ciccia and Elledge, 2010). Opposite the D-loop, an X-structure is formed, termed a Holliday junction. This junction is formed at the border between the hetero- and homoduplex DNA. Depending on which proteins bind at the junction, the junction may slide in either direction (Campos and Clemente-Blanco, 2020; Davis and Lin, 2011).

If the Holliday junction moves in the same direction as replication, it releases the newly synthesized strand, once repair synthesis is complete. Finally, the two broken ends must be rejoined, a process facilitated by RAD52 and completed by annealing of the complementary sequences. This process may result in flaps or gaps depending on how long the 3' end was extended during the repair

synthesis. Flaps are removed by specific endonucleases, for example the XPF/ERCC1 complex, whilst gaps can be filled in and sealed by PCNA-dependent POLδ or POLε and LIG1 (Campos and Clemente-Blanco, 2020; Ciccia and Elledge, 2010; Davis and Lin, 2011).

In some instances, a double Holliday junction is formed. This scenario describes when both DNA ends invade the homologous template and form a double Holliday junction. For repair to be completed, the DNA strands must be uncrossed by Bloom syndrome protein (BLM) helicase and topoisomerase, or cut by structure-specific nucleases, called Holliday-junction resolvases, such as XRCC3 and RAD51C (Campos and Clemente-Blanco, 2020; Ciccia and Elledge, 2010). Recently an additional resolvase has been identified, a tri-nuclease complex containing three endonucleases (Scaffolding protein for nucleases 1 and 4 (SLX1-SLX4), MUS81-Essential Meiotic Structure-Specific Endonuclease 1 (EME1) and XPF-ERCC1), known as the SMX tri-nuclease (Wyatt et al., 2017).

In addition to NHEJ and HR, there are other alternative pathways for repairing DSBs, including alternative end joining (alt-EJ) (also known as microhomology-mediated end joining) and single-strand annealing (SSA). Both of these pathways result in a loss of information due to the need for DNA resection. Alt-EJ can be promoted in certain instances and uses regions of microhomology exposed through end resection to mediate the ligations of DSB ends (Ceccaldi et al., 2016; Ciccia and Elledge, 2010; Howard et al., 2015). SSA uses homologous sequences to repair DSBs, however does not require strand invasion and involves annealing of the DNA strands after resection of the DSB in a RAD51 independent manner (Ciccia and Elledge, 2010; Gudmundsdottir and Ashworth, 2006; Madhusudan and Hickson, 2005).

1.1.3 Fanconi Anaemia Pathway

The Fanconi Anaemia pathway is the DNA repair pathway responsible for repairing ICLs. This complex pathway requires the combined functions of the FA proteins, HR, TLS and NER (Li et al., 2016; Moraes et al., 2012). To date, 22 Fanconi Anaemia genes have been identified (Table 1-1), whereby a mutation in one of these genes results in the FA disorder. The FA pathway functions primarily during S phase of the cell cycle, via four main steps, and requires converging replication forks (Räschle et al., 2008; Sareen et al., 2012; Tian et al., 2014). However, there is also growing evidence for replication independent repair of ICLs that takes place in G1 (MacKay et al., 2010; Walden and Deans, 2014; Wang et al., 2014; Zhang et al., 2015).

The first, and most important step in the FA pathway involves the recognition of the ICL (Bhattacharjee and Nandi, 2017; Ciccia and Elledge, 2010; Madhusudan and Hickson, 2005; Moraes et al., 2012). This is then followed by incision on either side of the lesion via HR (Ceccaldi et al., 2016; Deans and West, 2013; Joo et al., 2011). The ICL is then bypassed using TLS, and finally repaired through NER (Hosoya and Miyagawa, 2014; Li et al., 2016; Walden and Deans, 2014). Further details of each of these steps are described below. In addition to ICL repair, the FA pathway has also been implicated in the promotion of replication fork stability and in the control of chromosome segregation during mitosis (Ceccaldi et al., 2016; Ciccia and Elledge, 2010).

1.1.3.1 Sources of Interstrand Crosslinks

ICLs may arise from a variety of DNA damage, from both endogenous and exogenous sources, and result in toxic lesions that inhibit essential processes such as replication and transcription (Ceccaldi et al., 2016; Chatterjee and Walker, 2017; Moraes et al., 2012). This is due to the covalent bond formed between the strands of DNA, preventing the DNA duplex from separating (Bhattacharjee and Nandi, 2017; Liu et al., 2020). Depending on the type of DNA damage, the resulting ICL may cause different distortions of the DNA helix (Gibbs-Seymour et al., 2015; MacKay, 2010; Mouw and D'Andrea, 2014).

Table 1-1: Identified Fanconi Anaemia Proteins.

The 22 identified FA proteins, synonyms and functions of each are exhibited.

FA Protein	Synonym	Function
FANCA	-	FA Core Complex
FANCB	-	FA Core Complex
FANCC	-	FA Core Complex
FANCD1	BRCA2	HR
FANCD2	-	Recruitment of downstream effector proteins
FANCE	-	FA Core Complex
FANCF	-	FA Core Complex
FANCG	XRCC9	FA Core Complex
FANCI	-	Recruitment of downstream effector proteins
FANCJ	BRIP1, BACH1	HR, helicase
FANCL	-	FA Core Complex, E3 ubiquitin ligase
FANCM*	-	FA Core Complex, DNA translocase
FANCN	PALB2	HR, BRCA2 partner
FANCO	RAD51C	HR
FANCP	SLX4	Scaffolding protein for nucleases
FANCQ	ERCC4, XPF	ERCC1 partner, nuclease
FANCR	RAD51	HR
FANCS	BRCA1	HR, removes CMG
FANCT	UBE2T	FANCL partner, E2 conjugating enzyme
FANCU	XRCC2	HR
FANCV	REV7	TLS
FANCW	RFWD3	E3 ligase, homologous recombination

** There is some controversy surrounding whether FANCM is a true FA gene. The patient identified with mutations in FANCM also had mutations in FANCA that remained unidentified at the time. In addition, individuals with a homozygous loss of function of FANCM do not display FA symptoms (Lim et al., 2014; Singh et al., 2009).*

Endogenous Crosslinking Agents

Reactive aldehydes have been shown to be an endogenous source of DNA ICLs. These may arise through the metabolism of environmental agents, such as alcohol, cigarette smoke or dietary fat (Ceccaldi et al., 2016; Liu et al., 2020; Mouw and D'Andrea, 2014). For example, acetaldehyde which is a metabolite of ethanol, can form ICLs as a result of a condensation reaction between DNA and acetylaldehyde (Schärer, 2005). The main adduct formed however is DNA-protein crosslinks, resulting from acetaldehyde reacting with guanine. ICLs formed by acetaldehyde are located in the minor groove and do not disturb the base pairing of the bases involved (Figure 1-1a) (Lopez-Martinez et al., 2016). The production of malondialdehyde (MDA) and unsaturated aldehydes such as acrolein and crotonaldehyde as a result of lipid peroxidation is another example

of metabolically producing endogenous reactive aldehydes (Deans and West, 2013; Renaud et al., 2016). MDA can form ICLs between guanines in CpG sequences, whilst unsaturated aldehydes can react with nitrogen bases, to also form ICLs in CpG sequences. These ICLs also do not distort the structure of the double helix. However, most FA patients do not have defects in the metabolic pathways that process aldehydes, so it is unlikely that aldehyde induced ICLs can cause FA phenotypes (Lopez-Martinez et al. 2016).

Another endogenous source of ICLs is nitric oxide (or in its hydrated form, nitrous acid), which has also been shown to induce ICLs through formation of a diazonium ion. These ICLs are also usually present between two guanine residues in CpG sequences and are bound by a common N2 amine group (Figure 1-1b) (Schärer, 2005).

Exogenous Crosslinking Agents

Many exogenous crosslinking agents are anti-tumour agents. The first discovered exogenous crosslinking agents were the nitrogen mustards (Deans and West, 2013; Madhusudan and Hickson, 2005). These form ICLs through the alkylation of the N7 positions of guanine, preferentially formed in complementary GNC sequences (Schärer, 2005). The first of these to be used as a chemotherapeutic agent was mechlorethamine, which is a bifunctional alkylating agent. The chloroethyl moieties form an adduct between the two bases on opposite strands. The generated ICL causes a distortion of the DNA double helix due to the shortening of the N7-N7 bond. This distortion causes a bend of around 10° per ICL and causes an unwinding of 2 – 6° (Figure 1-1c) (Lopez-Martinez et al., 2016). However, many other adducts are formed from nitrogen mustards, with ICLs only making up 5% of the lesion products. This class of exogenous crosslinking agents have also been engineered in recent times into activatable prodrugs (Lopez-Martinez et al., 2016; Schärer, 2005).

A second class of alkylating crosslinking agents are the nitrosoureas, such as BCNU. Nitrosoureas have been shown to also form ICLs between N7s of two guanines, between the N3 of cytosine and N1 of guanine, and between the O6 of guanine and N2 of cytosine, and cause minor distortions of the DNA double helix. BCNU attacks the O6 of guanine, resulting in the conversion to form 1,O^6-ethanoguanine, followed by slow rearrangement to form an ICL between the O6 of guanine and N2 of cytosine (Figure 1-1d). These ICLs are minor products, as BCNU can also form intrastrand crosslinks between adjacent guanines. Nitrosoureas are also used in anti-tumour therapy, and it has been shown that removal of the ICL by the DNA-repair enzyme alkylguanine transferase (AGT) contributes to the resistance of tumour cells to treatment with nitrosoureas.

Therefore, inhibitors of AGT are being developed for use in combination therapy with nitrosoureas (Lopez-Martinez et al., 2016; Schärer, 2005).

Another class of exogenous crosslinking agents often used in chemotherapy are platinum compounds (Bouwman and Jonkers, 2012; Deans and West, 2013; Liu et al., 2020; Madhusudan and Hickson, 2005; Mouw and D'Andrea, 2014). Most of the ICLs formed by platinum compounds are formed in the major groove of the DNA and cause a large distortion of the DNA double helix (Schärer, 2005). Cisplatin was the first described platinum compound and reacts with purine residues to form mainly intrastrand crosslinks (90%) and ICLs (5 – 8%). ICLs are formed at the N7 position of guanine at GpC sequences and induce the extrusion of two cytosines, causing the major distortion, whilst the platinum is located in the minor groove. This large distortion results in the DNA helix bending 47° towards the minor groove and unwinding 110° (Figure 1-1e). Other platinum compounds include carboplatin and transplatin (Lopez-Martinez et al., 2016).

Several naturally occurring compounds can also react to form ICLs. One example of these compounds is mitomycin C (MMC), a natural product produced by *Streptomyces caespitosus*. Before MMC can form ICLs, it must be reduced from the quinone moiety into its phenol form. Once this reduction cascade takes place, MMC can react with N2 groups of adjacent and opposite guanine residues to form an ICL. These ICLs form in the minor groove of the DNA helix and cause only a minor distortion of the DNA helix (Figure 1-1f). MMC can also react to form monoadducts through reaction with the N7 of guanines. Monoadducts are the major product of MMC, making up 50% of products, whilst intrastrand crosslinks make up 35%, and ICLs the remaining 15%. MMC is a favoured chemotherapeutic agent, as the need for it to be reduced results in it being more selective for cancer cells, as the tumour microenvironment is hypoxic, in comparison to normal tissues (Deans and West, 2013; Lopez-Martinez et al., 2016; Schärer, 2005).

Another naturally occurring compound is psoralen, which is produced by at least eight families of plants. Psoralens form ICLs with two thymine residues at adjacent and opposite positions after irradiation with UV light (Figure 1-1g). Psoralens are planar and hydrophobic molecules, allowing them to penetrate cells and intercalate with DNA bases. Upon UV irradiation they efficiently form ICLs (40% of products). A derivative of psoralen is trimethylpsoralen (TMP) which can form up to 90% of ICLs following irradiation. ICLs induced by psoralens only create a minor distortion in the DNA helix, resulting in an unwinding of around 25°. As psoralens are somewhat controllable, they are popular for use in biological studies into ICLs. They are also popular clinically due to the ability to target the formation of ICLs to a particular area using local UV irradiation (Lopez-Martinez et al. 2016; Schärer, 2005).

A final crosslinking agent is diepoxybutane (DEB). This crosslinking agent is often used in chromosomal breakage studies to diagnose FA. DEB is the product of the transformation of 1, 3-butadiene, produced in the plastic and rubber industry. This is an exogenous alkylating agent, and reacts to form a multitude of products, including monoadducts, ICLs, SSBs and DNA-protein crosslinks. DEB mainly reacts with the N7 and N1 of guanine, however ICLs are formed through the reaction of N7-N7 at GpCpC sequences causing a major distortion of 34° towards the major groove of the DNA helix (Figure 1-1h) (Lopez-Martinez et al., 2016).

1.1.3.2 Repair of Interstrand Crosslinks

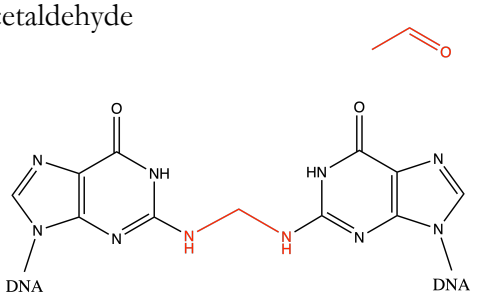
No matter the source or type of ICL, the resulting lesion is toxic and creates a covalent bond between the opposite strands of the DNA helix. These lesions need to be removed to allow replication and transcription to continue. The FA pathway mediates the repair of ICLs and is detailed below.

Recognition of the Interstrand Crosslink

The first step in the FA pathway is recognition of the ICL. FANCM was first identified as an ICL lesion sensing protein and binds to DNA (Ceccaldi et al., 2016; Ciccia and Elledge, 2010; Niraj et al., 2017). It was found that FA-associated protein (FAAP) 24 was essential for FANCM chromatin loading, and together with histone fold proteins MHF1 and MHF2, the complex is recruited to replication forks stalled by ICLs (Figure 1-4, Step 2) (Chatterjee and Walker, 2017; Deans and West, 2013; Liu et al., 2020). Once the complex is bound to chromatin, FANCM is thought to function as a landing pad for the FA core complex and has been shown to specifically interact with FANCF (Figure 1-4, Step 4) (Duxin and Walter, 2015; Huang et al., 2014; Mouw and D'Andrea, 2014). In an alternative ICL recognition mechanism, the non-FA protein Ubiquitin-like containing PHD and RING finger domains 1 (UHRF1) has been found to recognise psoralen induced ICLs, and to promote repair through recruitment of FANCD2 to ICL sites (Figure 1-4, Step 1) (Liang et al., 2015b). In addition, several proteins involved in the MMR pathway have been implicated in ICL sensing. Specifically, MutSβ can sense psoralen induced ICLs, and MutLα and MSH2/MLH1 interact with the FA proteins FANCJ and FANCD2 respectively. MutSα has also been shown to recognise ICLs causing major distortions in the DNA helix in an FA independent manner. Therefore, depending on the type of distortion resulting from the ICL, different recognition factors are necessary to sense the damaged lesion (Ceccaldi et al., 2016; Ciccia and Elledge, 2010; Walden and Deans, 2014).

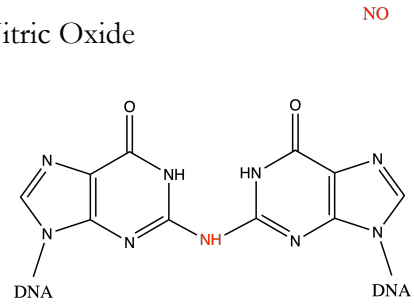
a

Acetaldehyde



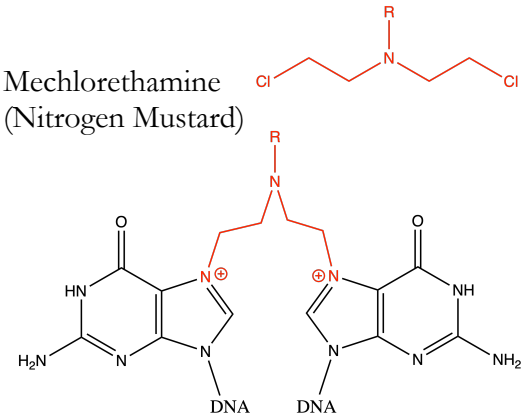
b

Nitric Oxide



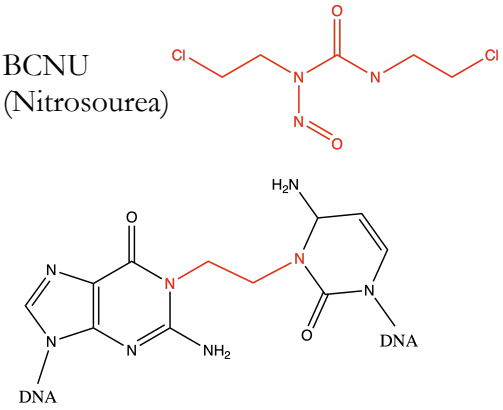
c

Mechlorethamine
(Nitrogen Mustard)



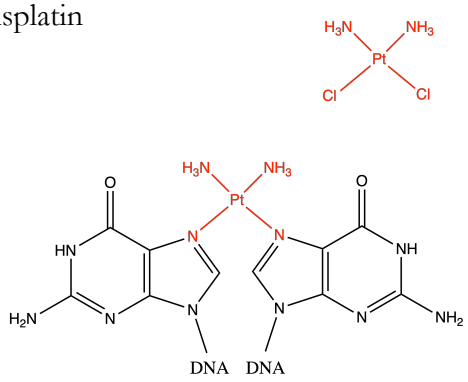
d

BCNU
(Nitrosourea)



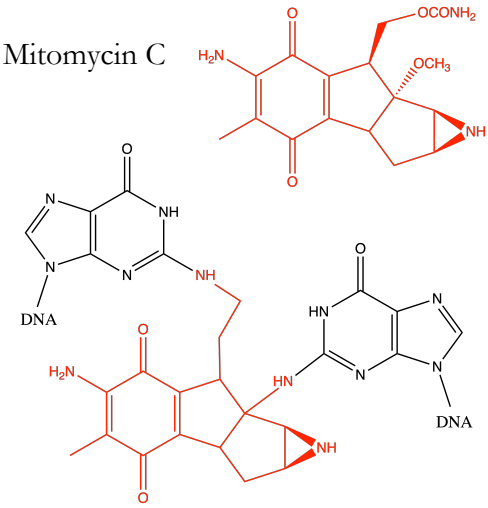
e

Cisplatin



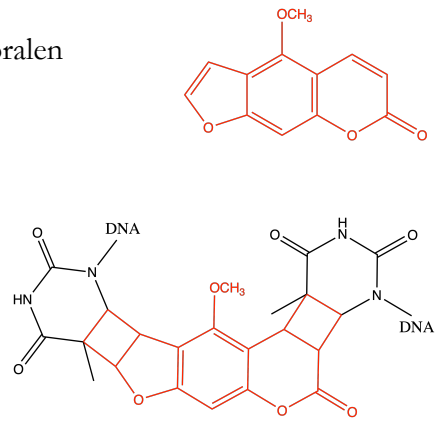
f

Mitomycin C



g

Psoralen



h

Diepoxybutane

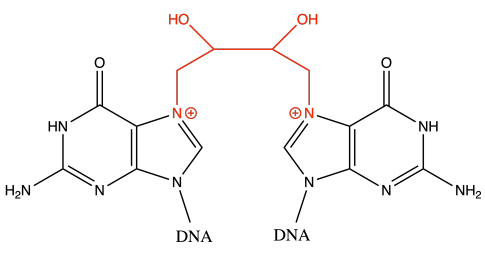


Figure 1-1: Schematic representation of the chemical structures of the key crosslinking agents and the corresponding ICLs formed. a) Acetaldehyde. b) Nitric Oxide. c) Mechlorethamine (Nitrogen Mustard). d) BCNU (Nitrosourea). e) Cisplatin. f) Mitomycin C. g) Psoralen. h) Diepoxybutane. (Adapted from Lopez-Martinez et al., 2016).

UHRF1 in the DDR

UHRF1 (also known as ICBP90, NP95 and RNF106) was first identified as a factor that binds the inverted CCAAT box in the topoisomerase IIα promoter that regulates its expression. However, subsequent studies have not demonstrated that UHRF1 functions as a transcription factor, but that its main function is to regulate gene expression through epigenetic regulation (Tien et al., 2011). In 2002, however, UHRF1 was implicated in the DDR. Muto et al. described that UHRF1 plays a role in the DDR in mice, as a knockout of UHRF1 in mouse embryonic stem cells were more sensitive to X-rays, UV light and hydroxyurea (Muto et al., 2002). Recent evidence has also further implicated UHRF1 in the DDR. In 2011, Tien et al. reported that UHRF1 plays a role in maintaining genomic integrity, as UHRF1 is recruited to DNA damage at discrete sites (Tien et al., 2011). In 2016, Zhang et al. reported that phosphorylated UHRF1 is recruited to DSBs via an interaction with BRCA1. This interaction promotes polyubiquitination of RIF1 resulting in its dissociation from DSB sites. Therefore, UHRF1 interacts with BRCA1 which in turn promotes HR (Zhang et al., 2016). Further to this study, Yin et al. reported that targeting the UHRF1/BRCA1 protein complex through using a histone deacetylase (HDAC) inhibitor to target UHRF1 in combination with a PARP inhibitor, resulted in synthetic lethality in prostate cancer cells. This study revealed the synthetic lethality of co-administering HDAC and PARP inhibitors due to promoting DNA damage and inhibiting HR, and thus the potential for this drug combination in future prostate cancer clinical trials (Yin et al., 2018). In another report, Hahm et al. described in 2018 that UHRF1 is methylated in response to DNA damage. This methylation promotes the interaction between PCNA and UHRF1 and results in the polyubiquitination of PCNA. As polyubiquitination of PCNA is required to induce HR, UHRF1 therefore plays an essential role in promoting the HR pathway (Hahm et al., 2018). In addition, deletion of UHRF1 in mice is embryonic lethal and knockout embryonic stem cells show a loss of DNA methylation, enhanced DNA replication arrest, impaired maintenance of higher-order chromatin structure and increased sensitivity to DNA damaging agents (Gelato et al., 2014; Tian et al., 2015).

In 2015, Liang et al. and Tian et al. both reported that UHRF1 is recruited to DNA ICLs following DNA damage (Liang et al., 2015b, Tian et al., 2015). Tian et al. reported that UHRF1 is recruited to DNA lesions *in vivo* and binds directly to ICL-containing DNA. In addition, UHRF1-deficient

cells result in increased sensitivity to DNA damage, and have reduced recruitment of ICL repair nucleases, ERCC1 and MUS81. Further to this, they also reported that depletion of UHRF1 and components of the FA pathway resulted in increased DNA damage sensitivity, in comparison to each depletion alone, and that activation of the FA pathway by monoubiquitination of FANCD2 was not affected by UHRF1 depletion. Their results suggest that UHRF1 plays a role in ICL recognition, independent of FA pathway components (Tian et al., 2015). In contrast to this, Liang et al. reported that not only is UHRF1 recruited to ICLs, but that UHRF1's recruitment to ICLs is required for the subsequent recruitment of FANCD2 to ICL lesions. Therefore, Liang et al. suggest that UHRF1 and FANCD2 have an epistatic relationship (Liang et al. 2015b). The discrepancies between the two papers could be due to the different methods utilized, as Tian et al. produced a hypomorphic mutant of UHRF1, whilst Liang et al. established a knockdown of UHRF1 via small hairpin RNA. In addition to these two papers describing UHRF1 as an ICL sensor protein, a further study claims that UHRF1 and/or UHRF2 (a paralogue of UHRF1) stimulates FANCD2 monoubiquitination (Motnenko et al., 2018). However, in either case, it is clear that UHRF1 is recruited to ICL sites and may act as one of the sensors for this damage. In addition, it is clear that UHRF1 plays a role in other DDR pathways, and that its mechanistic role in these pathways is yet to be elucidated. Therefore, structural studies of UHRF1 would be helpful to gain mechanistic insight into how UHRF1 plays a role in the DDR and could potentially be helpful in designing inhibitors for UHRF1 as cancer therapeutics.

Despite the recent evidence for UHRF1's role in the DDR, UHRF1 is most well-known as a protein that is required for cellular proliferation and epigenetic regulation (Arita et al., 2008; Cheng et al., 2013). It is considered a unique factor that links DNA maintenance methylation and histone modifications which allows it to regulate eukaryotic chromatin structure and gene activity. Specifically, UHRF1 binds to hemi-methylated DNA at CpG dinucleotide sites and helps to recruit DNA methyltransferase 1 (DNMT1) for DNA maintenance methylation (Bostick et al., 2007; Qian et al., 2008; Sharif et al., 2007). In addition, UHRF1 also interacts with histone H3 in a methylated and unmethylated state (Arita et al., 2012; Gelato et al., 2014). UHRF1 also possesses E3 ubiquitin-ligase activity through its really interesting new gene (RING) domain and has been shown to ubiquitinate histone H3 (Cheng et al., 2013; Harrison et al. 2016; Nishiyama et al., 2013). UHRF1 has also been implicated in cancer progression and is over-expressed in various tumour types, including bladder, prostate and ovarian cancer, and is thus a target for cancer therapeutics (Achour et al., 2008; Felle et al., 2011; Hahm et al., 2018; Lallous et al, 2011; Yin et al., 2018). All

structural studies to date of UHRF1 have been surrounding its role in DNA maintenance methylation and histone modifications and are discussed below.

Structural Studies of UHRF1 Domains

UHRF1 is a 90 kDa multi-domain protein. It is made up of five subdomains: an N-terminal ubiquitin-like (UBL) domain, a tandem tudor domain (TTD), a plant homeodomain (PHD), a SET and RING-associated (SRA) domain and a C-terminal RING domain (Figure 1-2). Various reports have attributed different domains to the different reported functions of UHRF1. For example, the TTD has been reported to bind the trimethylated histone H3 lysine 9 (H3K9me3) (Cheng et al., 2013). The PHD domain has also been reported to recognise H3K9me3 (Qian et al., 2008). In addition, the PHD domain has been reported to bind to the unmodified histone H3 N-terminal tail (Lallous et al., 2011). However, Arita et al. and Cheng et al. reported in 2012 and 2013 that the TTD-PHD, tethered by a linker, binds to the N-terminal tail of histone H3 not through a TTD or PHD interface, but through the linker that joins the two domains (Arita et al., 2012; Cheng et al., 2013). The SRA domain has been shown to selectively recognise hemi-methylated CpG dinucleotides through flipping the 5-methylcytosine out of the DNA helix and binding specifically to a pocket in the SRA domain (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008; Qian et al., 2008; Sharif et al., 2007). In 2014, Berkyurek et al. also showed that the SRA domain directly interacts with DNMT1, promoting release of the hemi-methylated DNA from the SRA domain and allowing the DNA substrate to access the catalytic site of DNMT1 (Berkyurek et al., 2014). The RING domain has been shown to ubiquitinate histone H3 at lysine 23 (H3K23), which is essential for the recruitment of DNMT1 to DNA replication sites (Nishiyama et al., 2013). Recently, there have been numerous subdomain structures published supporting the hypothesis that UHRF1 adopts a closed structure whilst unbound, but when bound to an amino acid ligand, or to DNA, that UHRF1 then adopts an open structure. Key to this conformational change are smaller regions of UHRF1 between the subdomains which mediate intramolecular interactions, including the linker between TTD-PHD domains and a C-terminal polybasic region (Fang et al., 2016; Gao et al., 2018; Gelato et al., 2014; Harrison et al., 2016; Vaughan et al., 2018). An example of each subdomain structure is discussed below.

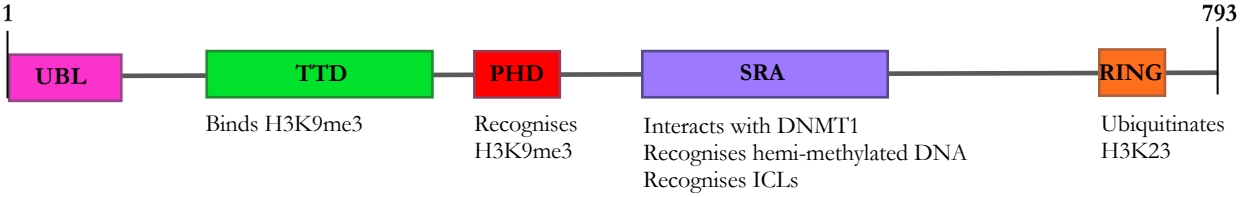


Figure 1-2: Schematic of the UBL, TTD, PHD, SRA and RING domains of the UHRF1 protein.

The structure of the UBL domain of UHRF1 was deposited into the PDB in 2005 by Walker et al. under PDB reference 2FAZ. Curiously, a paper describing the structure was never published. However, the entry shows the crystal structure of the UBL domain to 2.00 Å. The structure shows a similar fold to ubiquitin, showing a 3.5 turn α-helix, a smaller 3_{10} helix and a 5-stranded β-sheet (Figure 1-3a) (Walker et al., 2005).

The highest resolution isolated TTD domain of human UHRF1 was deposited into the PDB by Kori et al. in 2019 under the PDB reference 5YYA. Although not the first structure of the TTD, it was the first isolated human TTD to be solved. Solved by X-ray crystallography to 1.70 Å, the structure adopts a similar fold to previously solved structures (TTD-PHD, TTD-polybasic region). Each Tudor domain is tightly packed against each other and is made up of the characteristic "Royal" fold containing a 5-stranded β-barrel (Figure 1-3b) (Kori et al., 2019). Previous structures have identified an aromatic cage that interacts with the trimethyl moiety of H3K9me3, and also that the TTD undergoes a conformational change to accommodate the histone H3 tail, highlighting the importance of the inter-sub-Tudor domain peptide binding groove (Nady et al., 2011). In addition, intramolecular interactions have been identified between a polybasic region and the TTD (Gao et al., 2018; Gelato et al., 2014). Several structures have also been published of the TTD-PHD together and outlines the importance of the linker between the two domains which maintains a specific spatial arrangement of the TTD-PHD domains to allow interaction with histone H3 (Arita et al., 2012; Cheng et al., 2013).

Lallous et al. published the X-ray crystal structure of the human PHD domain in 2011 to 1.45 Å under reference 3ZVZ. The structure showed that the PHD domain was made up of a fold containing a pair of anti-parallel β-strands flanked by two short α-helices and two long loops (Figure 1-3c) (Lallous et al., 2011). The structure of this PHD domain was very similar to other structures also published around the same time, including in the TTD-PHD domain structures (Arita et al., 2012; Cheng et al., 2013).

The SRA domain was published in 2008 by multiple groups, highlighting how this domain contributes to the hemi-methylated-DNA specificity (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008; Qian et al., 2008). Avvakumov et al. published the human SRA domain structure to 1.70 Å under PDB reference 3BI7. The domain has a compact architecture constructed around an 8-stranded flared β-barrel core covered in α-helices (Figure 1-3d). The overall architecture is a concave surface that generates a hydrophobic core and contains a binding

pocket that accommodates the 5-methylcytosine flipped out of the DNA duplex, as well as two loops that interact with the major and minor grooves of the DNA. Interestingly, Fang et al. found that a C-terminal polybasic region located after the SRA domain binds to the TTD domain, facilitating SRA-PHD intramolecular interactions, resulting in a closed UHRF1 structure. However, upon the SRA domain binding to hemi-methylated DNA, UHRF1 loses its intramolecular interactions and adopts an open conformation (Fang et al., 2016).

Finally, the RING domain of UHRF1 was deposited into the PDB in 2008 by Walker et al. under reference 3FL2. Once again, however, this structure was never discussed in a published report. Nevertheless, the X-ray crystal structure is solved to 1.75 Å, and displays a core RING domain architecture of one central α-helix and two anti-parallel β-strands, as well as two long loops (Zhang et al., 2013). One loop and the two anti-parallel β-strands are located on the N-terminal side of the central α-helix, with the second loop on the C-terminal side. Both loops are stabilized by the coordination of a zinc ion. In addition, there are three additional α-helices in the N-terminus, and one additional α-helix at the C-terminus (Figure 1-3e).

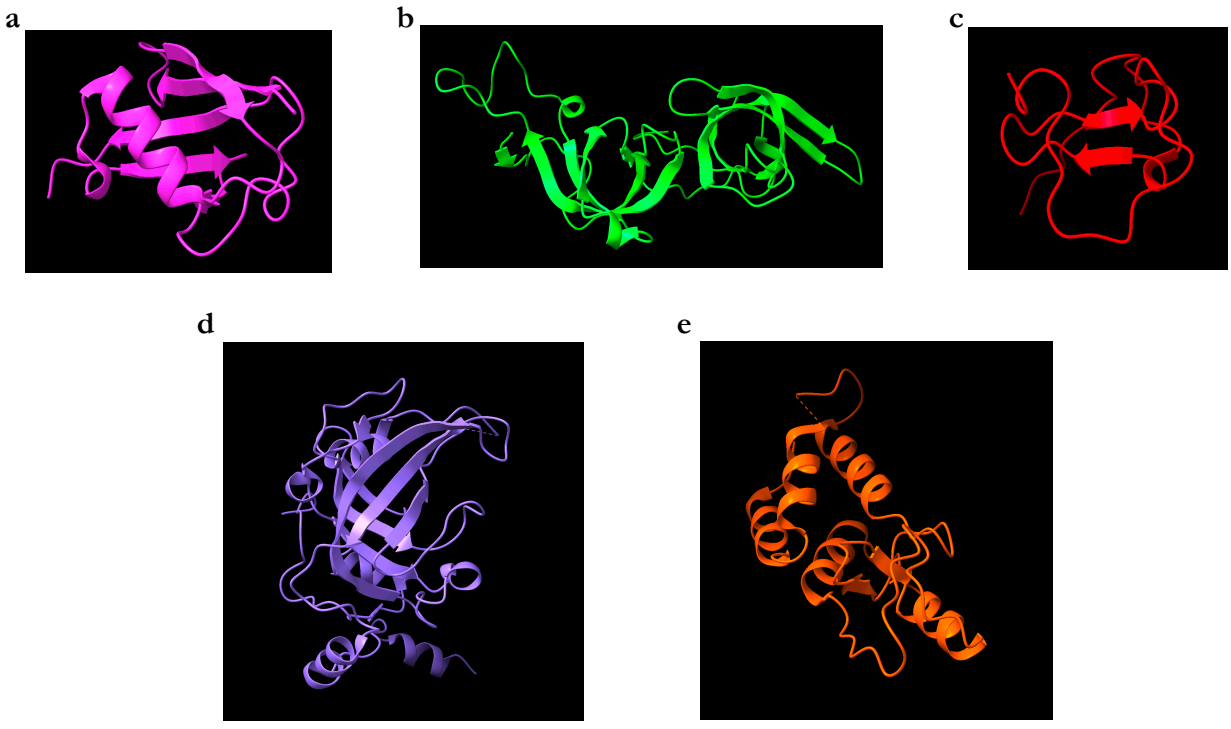


Figure 1-3: Previously solved structures of the different domains of the UHRF1 protein. a) Structure of the UBL domain (pink) (Walker et al., 2005; PDB: 2FAZ). b) Structure of the TTD domain (green) (Kori et al., 2019; PDB: 5YYA). c) Structure of the PHD domain (red) (Lallous et al., 2011; PDB: 3ZVZ). d) Structure of the SRA domain (purple) (Avvakumov et al., 2008; PDB: 3BI7). e) Structure of the RING domain (orange) (Walker et al., 2008; PDB: 3FL2).

It is clear that there is extensive structural data available for UHRF1. However, despite the fact that there are many individual domains solved, there is no structural information on the whole protein, nor how these domains interact. All of the domain structures have clearly showed that there are intramolecular contacts that are crucial for mediating the conformation, and thus the function of UHRF1. Therefore, there is a clear need for structural knowledge of the interdomain organisation of UHRF1. Further structural insight into the UHRF1 protein as a whole will assist in understanding the true mechanism of UHRF1, not only in DNA maintenance methylation, but also in DNA repair. In addition, the conformation of isolated UHRF1 in comparison to UHRF1 bound to an ICL would elucidate how the ICL is detected and provide insight into UHRF1's role in the FA pathway.

Recruitment of the FA Core Complex

Following recognition of the ICL by a sensor protein, the next phase in the FA pathway involves the recruitment of the FA core complex. The FA core complex is made up of eight FA proteins, FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL and FANCM, as well as two FA associated proteins, FAAP100 and FAAP20, and is recruited to DNA ICLs (Ceccaldi et al., 2016; Ciccia and Elledge, 2010; Rajendra et al., 2014; Smogorzewska et al., 2007). The main purpose of the FA core complex is to act as a multi-subunit ubiquitin ligase which monoubiquitinates the two key proteins in the FA pathway, FANCD2 and FANCI (Figure 1-4, Step 6) (Chatterjee and Walker, 2017; Duxin and Walter, 2015; Liu et al., 2020; Sims et al., 2007). FANCL contains the RING-finger domain which acts as the E3 ubiquitin ligase, ubiquitin-conjugating enzyme E2 T (UBE2T, FANCT) acts as the E2 conjugating enzyme and ubiquitin-like modifier activating enzyme 1 (UBA1) acts as the E1 activating enzyme (Chaugule et al., 2020; Longerich et al., 2009; Mouw and D'Andrea, 2014; Walden and Deans, 2014).

Previously it has been shown that FANCB-FANCL-FAAP100 form a catalytic module that has the same activity as the entire core complex *in vitro*. However, this study showed that DNA was dispensable for the complete monoubiquitination of FANCD2 in complex with FANCI *in vitro* with the catalytic core, but that with the full core complex, the monoubiquitination of FANCD2 was dependent on the presence of DNA (Rajendra et al., 2014). In 2017, it was suggested that the catalytic module made up of FANCB-FANCL-FAAP100 forms a symmetric dimer of trimers, and that the FANCC-FANCE-FANCF proteins act as a bridge between the catalytic module and the FANCD2/FANCI complex. In this model, it was proposed that the FANCC-FANCE-FANCF proteins orient the FANCD2/FANCI complex to allow simultaneous monoubiquitination of each component (Swuec et al., 2017; van Twest et al., 2017). In addition, in 2015, it was elucidated that

the E74-like ETS Transcription Factor (ELF) domain of FANCL interacts with ubiquitin, which is crucial for the monoubiquitination of FANCD2 in cells, but not *in vitro* (Miles et al., 2015). However, despite these numerous insights, the exact structural make-up of the core complex remained to be revealed.

Recently, it has been elucidated for the avian core complex via cryo-electron microscopy (cryo-EM) and mass spectrometry studies that FANCB and FAAP100 form two central dimers, flanked by two copies of FANCL. FANCL has an ELF domain, a ubiquitin-dependent recruitment domain (URD) and a RING finger domain. In the lower copy of FANCL, the three domains (ELF, URD and RING finger) are visible, whilst in the top copy of FANCL, only the ELF domain could be fitted to the structural map. Therefore, the central catalytic module forms an extended asymmetrical dimer of heterotrimers. This central core acts as a scaffold to assemble the remaining five proteins. In the study it is suggested that FANCC-FANCE-FANCF form a base that interacts with the basal FANCL. Previously it was suggested that FANCA and FANCG act as a chromatin-targeting module (Huang et al., 2014). In the Shakeel et al. study, FANCA was not visible in the reconstruction, whilst FANCG was likely to fit a disordered arm, and is in proximity to the top FANCL (Shakeel et al., 2019).

Shakeel et al. also show that if any of the subunits in the catalytic module are deleted (FANCB, FANCL or FAAP100), the monoubiquitination activity of the core complex is abrogated. However, if one of the peripheral subunits is deleted (FANCA, FANCC, FANCE, FANCF or FANCG), residual monoubiquitination activity is retained. This provides a mechanistic explanation as to the mutation rates found in the FA population. Over 60% of mutations are found in FANCA, over 15% of mutations are found on FANCC and approximately 10% of mutations are found on FANCG. The rest of the core complex proteins account for only around 5% of mutations, and mutations found on FANCD2 or FANCI account for around 3% of mutations (Auberbach, 2009; Deans and West, 2013). Therefore, this study confirmed that mutations on the periphery of the core complex are tolerated and result in partial monoubiquitination activity, whilst mutations in the catalytic core are likely to be lethal and result in a non-functioning FA pathway (Alpi et al., 2008; Huang et al., 2014; Shakeel et al., 2019).

Despite the significant advances this study provides regarding the core complex, it remains unclear how the core complex specifically monoubiquitinates FANCD2 and FANCI. It has been suggested that FANCL, like several other E3 ligases, may need dimerization for E3 ligase activity, explaining the presence of two copies of FANCL in the core complex. In addition, FANCL may promote FANCD2 substrate binding in cooperation with FANCE. However, it remains unclear

whether the monoubiquitination happens sequentially, or whether one copy of FANCL monoubiquitinates FANCD2 and the other monoubiquitinates FANCI, either simultaneously or sequentially (Swuec et al., 2017).

Monoubiquitination of the FANCD2/FANCI Complex

Once the core complex is recruited, this allows the monoubiquitination of FANCD2 and FANCI at lysines 561 and 523 in humans respectively (Figure 1-4, Step 6) (Garcia-Higuera et al., 2001; Longerich et al., 2009; Sims et al., 2007; Smogorzewska et al., 2007). This is considered to be the most important step in the FA pathway. FANCD2 and FANCI are paralogues and form a saxophone-shaped heterodimer (Ceccaldi et al., 2016; Joo et al., 2011; Smogorzewska et al., 2007). The crystal structure of the ~300 kDa truncated non-ubiquitinated mouse FANCD2/FANCI complex was solved in 2011 to 3.4 Å resolution via X-ray crystallography. This structure revealed that the two ubiquitination sites lie embedded at the interface of the two proteins (Joo et al., 2011). This initial structure was followed in 2016 with a low resolution (22 Å) cryo-EM structure of the full-length non-ubiquitinated human FANCD2/FANCI complex (Liang et al., 2016). The two structures are in good agreement, revealing a 90 Å wide and 160 Å long complex, with a central hollow section, which is presumably where the complex binds to the DNA (Liang et al., 2016). The main difference between the two structures is the existence of a "Tower" domain in the human complex, identified in the C-terminus of the FANCD2 protein, from residue 1146-1451. This Tower domain is horizontal in the cryo-EM structure, but lies vertically in the X-ray crystallographic structure. This was reasoned to be due to either the deletions in the mouse sequences (which indeed deleted part of the sequence of this Tower domain) or was attributed to the different methods of obtaining the structures. In any case, the Tower domain was shown to be critical for the recruitment of FANCD2/FANCI to ICLs, as well as to the monoubiquitination of FANCD2, and thus crucial to the repair of ICLs (Liang et al., 2016).

Several other factors are also involved in the activation of FANCD2/FANCI. For example, it is clear that post-translational modifications are essential for the progression of the FA pathway. Several phosphorylation events occur both before and after ubiquitination of FANCD2/FANCI which are crucial to a functioning FA pathway (which will be further discussed in Section 1.1.4) (Figure 1-4, Step 5) (Ceccaldi et al., 2016; Jackson and Durocher, 2013; Lopez-Martinez et al., 2019; Rajendra et al., 2014). In addition, PCNA is ubiquitinated by RAD18 in response to ICL lesions and has been shown to promote FANCD2 ubiquitination via a direct physical interaction with FANCD2 and also through promotion of FANCL recruitment (Bhattacharjee and Nandi, 2017; Ciccia and Elledge, 2010). Interestingly, FANCI but not FANCD2 has also been shown to

be essential for the efficient recruitment of the FA core complex, suggesting an upstream function to the core complex recruitment (Castella et al., 2015; Liu et al., 2020). Another post-translational modification that is essential to ICL repair via the FA pathway, is the deubiquitination of the monoubiquitinated FANCD2/FANCI complex, which is mediated by the ubiquitin carboxyl-terminal hydrolase 1 (USP1)/USP1-associated factor 1 (UAF1) complex (Figure 1-4, Step 7) (Cohn et al., 2007; Cohn et al., 2009; Nijman et al., 2005; Oestergaard et al., 2007; van Twest et al., 2017).

Activation of the FANCD2/FANCI complex by monoubiquitination is an early and critical event that enables the FA pathway to function. However, despite the absolute requirement of the monoubiquitination of FANCD2/FANCI, the timing, location and roles of this ubiquitination and deubiquitination are not well understood (Ceccaldi et al., 2016; Sareen et al., 2012; Walden and Deans, 2014). It has been speculated that perhaps FANCD2/FANCI undergo a conformational change upon binding to DNA and the ICL, thus enabling the FA pathway to repair the DNA damage, or alternatively that the FANCD2/FANCI complex dissociates upon FA pathway activation (Liang et al., 2016; Sareen et al., 2012; Sato et al., 2012b). It is clear that further structural and mechanistic questions remain to be answered. These include how can the core complex access the lysines at the interface of the heterodimer? Do FANCD2 and FANCI arrive at the ICL individually or as a preformed heterodimer? Is there a conformational change upon the FANCD2/FANCI complex binding to DNA or resulting from ubiquitination? In addition, how is the dimer monoubiquitinated: sequentially, simultaneously, and in what order?

In late 2019 and 2020 numerous papers addressing many of the open questions mentioned above were released on BioRxiv and/or published. These include the cryo-EM structures of the avian and human non-ubiquitinated FANCD2/FANCI, the intermediate of the avian and human monoubiquitinated-FANCD2/FANCI and the human monoubiquitinated-FANCD2/ monoubiquitinated-FANCI, as well as the human non-ubiquitinated FANCD2/FANCI in complex with the core complex and UBE2T E2 conjugating enzyme. Unfortunately, these papers addressed the major goals of this DPhil project. Due to the nature and initial goals of this project, these papers will be further discussed in the Results Section 3.8.

Incision of the DNA at the Interstrand Crosslink

The next step in the repair of ICLs by the FA pathway involves the incision of the DNA to release the ICL. Most studies of ICL repair have been conducted in a replication dependent manner. In this scenario, two replication forks need to converge for the DNA to be incised (Chatterjee and Walker, 2017; Duxin and Walter, 2015). Once FANCD2/FANCI have been recruited to

chromatin and ubiquitinated, the monoubiquitinated FANCD2/FANCI controls the unhooking of the ICL from one of the two parental strands via nucleolytic incision. In order for this to occur, the cell division cycle 45 (Cdc45)-mini-chromosome maintenance (MCM)-Go-Ichi-Ni-San (GINS) (CMG) replicative helicase needs to be unloaded, potentially through the interaction of FANCD2 with the MCM replicative helicase of the complex. This process is also dependent on BRCA1 (Bhattacharjee and Nandi, 2017; Ceccaldi et al., 2016; Zhang et al., 2016). In addition, it has recently been shown that unloading of the CMG helicase leads to fork reversal of one of the converged replication forks (Amunugama et al., 2018; Zhang et al., 2015). The ubiquitinated FANCD2 recruits the nuclease scaffold protein SLX4 (FANCP), whose ubiquitin-recognising ubiquitin-binding zinc finger 4 (UBZ4) domain interacts with the ubiquitinated FANCD2. SLX4 then recruits several nucleases, including XPF (FANCQ, ERCC4)-ERCC1, MUS81/EME1 and SLX1 (Figure 1-4, Steps 8 and 9) (Rajendra et al., 2014; Räschle et al., 2008; Walden and Deans, 2014). ERCC4 has been shown to have important nuclease activity in unhooking ICL lesions. SLX4 may assist in this unhooking through stimulating the nuclease activity of ERCC4 (Figure 1-4, Step 10) (Liu et al., 2020; MacKay et al., 2010). Another nuclease that has been shown to be important in ICL repair is the FANCD2/FANCI-associated nuclease 1 (FAN1). FAN1 interacts directly with FANCD2 and FANCI and also interacts directly with ubiquitinated FANCD2/FANCI through its ubiquitin-binding zinc finger (UBZ) domain. It has been elucidated that FAN1 forms a homodimer to locate the ICL, orients itself on the DNA, unwinds a 5' flap and incises the dsDNA opposite the ICL (MacKay et al. 2010; Smogorzewska et al., 2010). However, FAN1 nuclease activity appears to be redundant in the FA pathway, as FAN1 depletion does not affect ICL-induced DSBs and does not lead to FA. Alternatively, perhaps FAN1 is recruited to ICLs independent of the FA pathway (Duxin and Walter, 2015). In addition, SNM1A has been shown to have endonuclease and exonuclease activity at ICL sites, and potentially plays a role in the FA pathway (Baddock et al., 2020; Buzon et al., 2018). Finally, endonuclease 8-like 3 (NEIL3) has been identified to unhook psoralen induced ICLs independently of the FA pathway (Li et al., 2020).

Bypass of the Interstrand Crosslink

Following unhooking of the ICL from one of the DNA strands, the DNA lesion is bypassed using TLS polymerases (Figure 1-4, Step 11). TLS polymerases enable lesion bypass due to their ability to accommodate large lesions such as ICLs. This process involves bypassing the lesion, inserting nucleotides opposite the ICL and extending the nascent strand. The TLS polymerase REV1-Polζ is recruited through the FA core complex to sites of ICLs and inserts deoxycytidine

monophosphates (dCMPs) opposite the ICL (Mouw and D'Andrea, 2014; Räschle et al., 2008). However, the main role of REV1 appears to be its ability to aid polymerase exchange, forming a TLS polymerase complex involving REV1-Polζ, Pol ξ and Pol ϰ (Deans and West, 2013; Liu et al., 2020). It has been suggested that Polη and Polν insert nucleotides opposite the ICL, however, the polymerases that carry out this insertion are still unknown (Chatterjee and Walker, 2017; Niraj et al., 2017). Following insertion, extension is carried out by REV1-Polζ, leading to the formation of one undamaged DNA duplex (Figure 1-4, Steps 11 and 12). This DNA duplex can now be used by HR to repair the broken DNA duplex (Bhattacharjee and Nandi, 2017; Ceccaldi et al., 2016; Ciccia and Elledge, 2010).

Repair of the Interstrand Crosslink

Once the ICL has been bypassed, the DSB generated through ICL repair is repaired via HR. First the DSB is resected, through the DSB end-resection factor CtIP (Chatterjee and Walker, 2017; Ciccia and Elledge, 2010). FANCD2, BRCA1 and MRE11 recruit CtIP to stalled replication fork sites. The MRN complex may also play a role in DSB end resection, as well as another resection factor, the DNA2 nuclease-helicase (DNA2) (Figure 1-4, Step 13) (Deans and West, 2013; Liu et al., 2020). Once the DSB is resected, RAD51 ssDNA nucleofilaments form. RAD51 is required for HR in ICL repair and is promoted via the BRCA2/PALB2 complex. These single-stranded filaments catalyse strand invasion of the homologous DNA template (Figure 1-4, Step 14) (Duxin and Walter, 2015; Mouw and D'Andrea, 2014). Biorientation of chromosomes in cell division protein 1-like (BOD1L) has also recently been shown to stabilise these RAD51 nucelofilaments and genetic disruption of BOD1L causes ICL hypersensitivity (Ceccaldi et al., 2016). In addition, it has been shown that the FANCD2/FANCI complex is recruited to these RAD51 nucleofilaments and protect the nucleofilaments from degradation by FAN1 (Sato et al., 2016). FANCJ (BRIP1) (independent of BRCA1 and its role in HR) is also essential for ICL repair via the FA pathway. Once strand invasion and repair synthesis has been carried out, ligation of the remaining nicks and removal of the lesion via NER lead to an intact dsDNA duplex (Figure 1-4, Step 15). This completes the repair of ICLs via the FA pathway (Ceccaldi et al., 2016; Niraj et al., 2017).

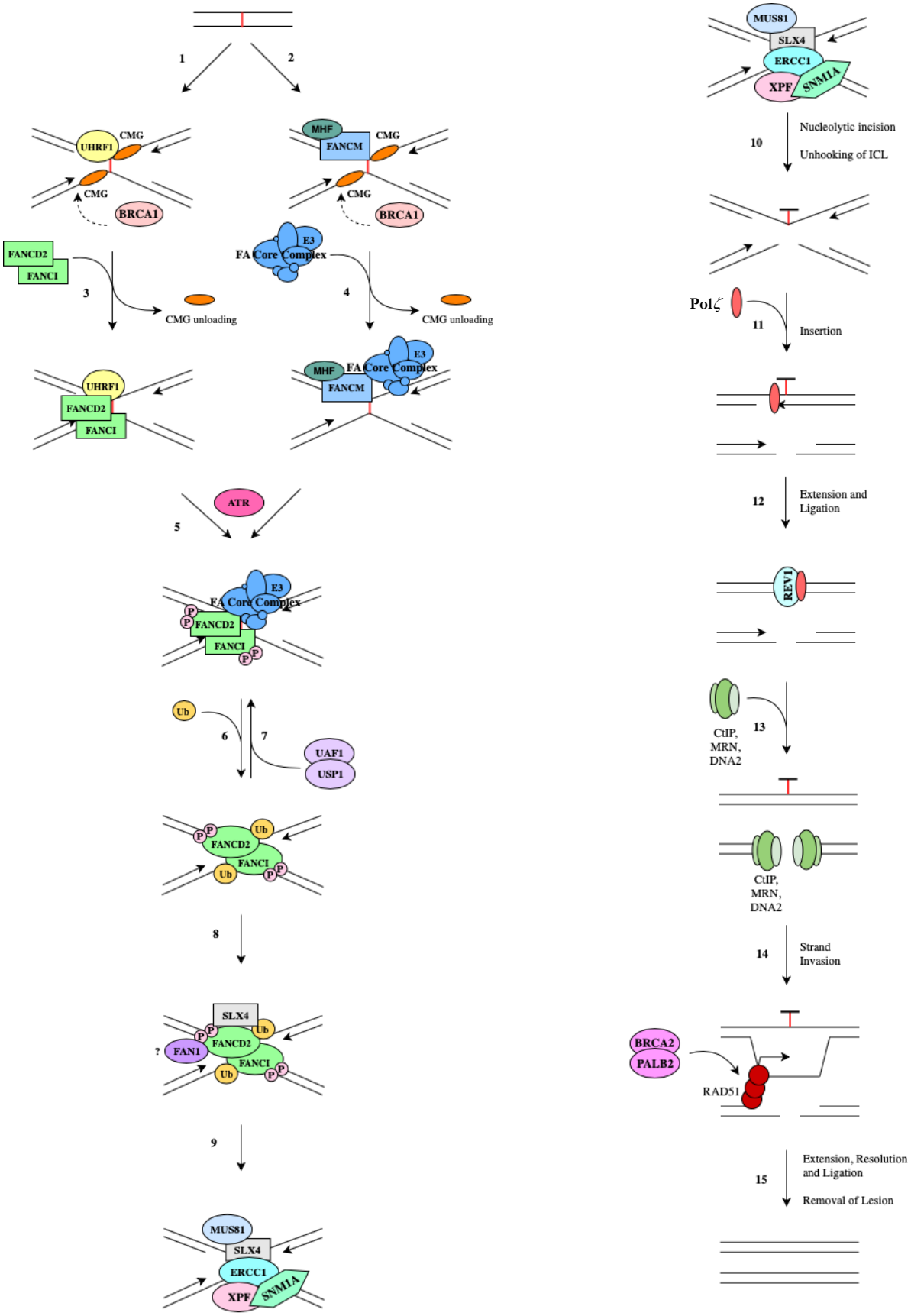

1
2
CMG
UHRF1
CMG
BRCA1
MHF
FANCM
CMG
CMG
BRCA1
FANCD2
FANCI
3
CMG unloading
FA Core Complex
E3
4
CMG unloading
UHRF1
FANCD2
FANCI
MHF
FANCM
FA Core Complex
E3
5
ATR
FA Core Complex
E3
P
FANCD2
FANCI
P
Ub
6
7
UAF1
USP1
P
FANCD2
Ub
FANCI
P
Ub
8
SLX4
Ub
P
?
FAN1
FANCD2
FANCI
P
Ub
9
MUS81
SLX4
ERCC1
XPF
SNM1A
MUS81
SLX4
ERCC1
XPF
SNM1A
10
Nucleolytic incision
Unhooking of ICL
Polζ
11
Insertion
12
Extension and
Ligation
REV1
CtIP,
MRN,
DNA2
13
CtIP,
MRN,
DNA2
14
Strand
Invasion
BRCA2
PALB2
RAD51
15
Extension, Resolution
and Ligation
Removal of Lesion

Figure 1-4: Schematic of ICL repair via the FA pathway. Step 1: Replication forks stall at a DNA ICL and UHRF1 may sense the ICL lesion. Step 2: Alternatively, replication forks stall at a DNA ICL resulting in recruitment of the FANCM-FAAP24-MHF1-MHF2 complex to the ICL lesion. Step 3: BRCA1 facilitates CMG unloading and UHRF1 aids in the recruitment of the FANCD2/FANCI complex. Step 4: Alternatively, BRCA1 facilitates CMG unloading and the FA Core Complex is recruited via FANCM. Step 5: ATR phosphorylates the FANCD2/FANCI complex at multiple sites. Step 6: This phosphorylation allows the site specific monoubiquitination of the FANCD2/FANCI complex at K561 and K523 respectively by the FA core complex. Step 7: The USP1/UAF1 complex may deubiquitinate the monoubiquitinated FANCD2/FANCI complex. Step 8: The monoubiquitinated FANCD2/FANCI complex aids in the recruitment of SLX4 and possibly the FAN1 nuclease. Step 9: SLX4 allows the recruitment of MUS81, XPF-ERCC1 and SNM1A. Step 10: SLX4, MUS81, XPF-ERCC1 and SNM1A assist in the nucleolytic incision of the ICL. Step 11: Unhooking of the ICL results in the crosslinked nucleotide being tethered to the complementary strand, which is bypassed by TLS Pol ζ. Step 12: REV1 and Pol ζ extend the complementary strand resulting in one intact DNA duplex. Step 13: Several factors including CtIP, the MRN complex and DNA2 resect the DSB. Step 14: Formation of RAD51 nucleofilaments catalysing strand invasion. Step 15: Extension, resolution and ligation of the remaining nick, and removal of the lesion via NER leads to an intact dsDNA duplex.

1.1.3.3 Alternative Roles of the FA Pathway

Aside from the FA pathway's main role in repairing ICL lesions, FA proteins have been implicated in many other crucial roles in replication, transcription and repair. For example, they have been shown to protect replication forks, and thus promote their stability (Schwab et al., 2015). The FA pathway is also activated by hydroxyurea, which does not cause ICL damage, but causes replication fork stalling. This raises the possibility that the activation of the pathway suppresses genomic instability and participates in the replication stress response. In addition, another role for the FA pathway includes contributing to the control of cytoplasmic division, a process known as cytokinesis. It is implicated in cytokinesis through facilitating fork restart and replication. Sister chromatid exchange is a common occurrence at chromosomal fragile sites, and the FA pathway acts to promote efficient separation of the sister chromatids through protection of these fragile sites. It has also recently been elucidated that the FA pathway is activated upon the formation of DNA-RNA hybrids, resulting from impaired transcription, as well as from replication-transcription collisions (Liu et al., 2020; Schwab et al., 2015). In addition, FANCD2/FANCI may facilitate nucleosome reorganization during DDR (Sato et al., 2012a). Finally, non-ubiquitinated

FANCD2 may play a role outside of the FA pathway in cooperation with the BLM helicase complex to protect aphidicolin-stalled replication forks (Raghunandan et al., 2015).

In addition to other non-repair roles of the FA pathway outside of ICL repair, there is also evidence for crosstalk between other repair pathways. For example, the FA pathway seems to antagonise the NHEJ pathway and thus promotes HR, alt-EJ and SSA (Adamo et al., 2010; Chatterjee and Walker, 2017; Howard et al., 2015; Renaud et al., 2016; Richardson et al., 2018). There is also evidence that the FA pathway is involved in directing the repair of CRISPR-Cas9 generated DSBs, directing the repair away from NHEJ and toward SSA (Richardson et al., 2018). In addition, as already mentioned above, the FA pathway also has significant crosstalk with the TLS, HR and NER pathways. The crosstalk between DDR pathways is an important aspect that is exploited translationally when designing new therapeutic approaches to target cancer (Deans and West, 2013).

1.1.4 Regulation of the DNA Damage Response by Post-Translational Modifications

The DDR is a complex signaling response involving many different factors. In order to regulate this response, the cell has developed multiple mechanisms including phosphorylation, ubiquitination and SUMOylation. These methods are known as post-translational modifications (Bhattacharjee and Nandi, 2017; Gibbs-Seymour et al., 2015; Hakem, 2008; Jacq et al., 2013; Oestergaard et al., 2007). Each of these modifications are reversible to ensure fine-tuning and tight regulation of the DDR and will be discussed in more detail below.

Phosphorylation and Dephosphorylation

Phosphorylation is a chemical process whereby a phosphate group (PO_3^{2-}) is attached to another molecule. Protein phosphorylation involves the terminal phosphate group of adenosine triphosphate (ATP) being transferred to the amino acid side chain via nucleophilic attack (Figure 1-5). This process requires magnesium (Mg^{2+}) which chelates two of the phosphate groups on ATP to lower the threshold for nucleophilic attack. In eukaryotes phosphorylation occurs on serine, threonine and tyrosine, which provide the nucleophilic group (OH^-). Despite the fact that phosphorylation only occurs on these three amino acids, it is the most abundant mechanism of regulation in the cell, affecting not only the DDR, but also many other cellular processes including the cell cycle, cell growth and signal transduction pathways. Protein phosphorylation and dephosphorylation are mediated by enzymes known as kinases and phosphatases respectively (Ceccaldi et al., 2016: Kim and Cole; 1997; Madhusudan et al., 1994).

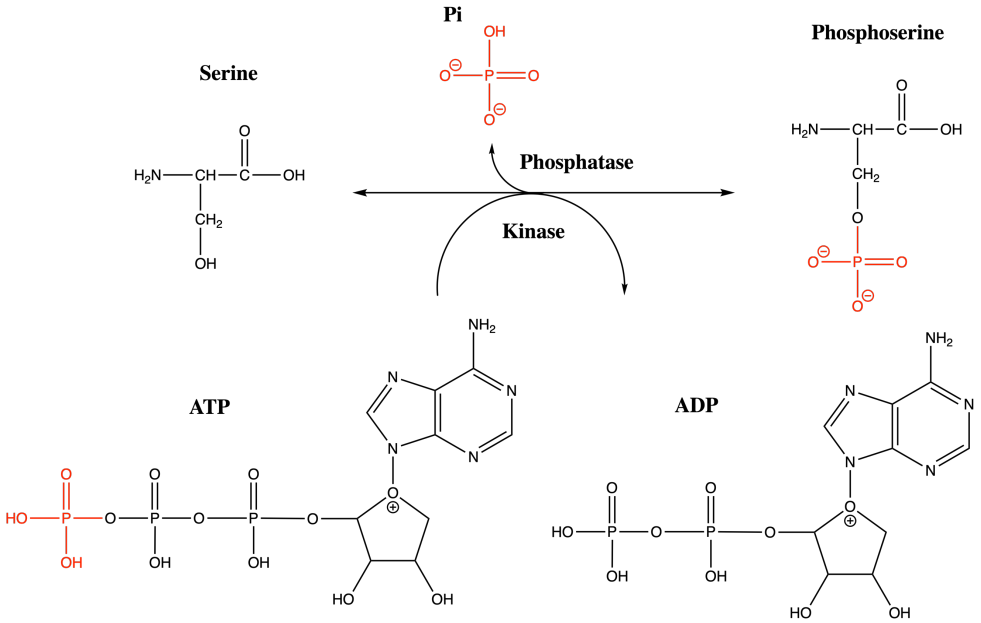


Figure 1-5: Schematic representation of phosphorylation of serine and dephosphoryation of phosphoserine.

Protein phosphorylation regulates protein activity and function through conformational changes of the protein as well as promoting or inhibiting certain protein interactions. These conformational changes may regulate the catalytic activity of the protein itself or may contribute to the recruitment of other proteins which have specific phosphomotifs. Thus, many pathways are tightly controlled by the activation and inactivation capabilities of phosphorylation and dephosphorylation.

Phosphorylation and dephosphorylation have been shown to be very important in the DDR. For example, in the FA pathway, multiple proteins are phosphorylated. This includes FANCG, which is phosphorylated by cyclin-dependent kinase 1 (CDK1) at S383 and S387, which promotes its dissociation from chromatin, FANCA, which is phosphorylated on S1449 by serine/threonine-protein kinase (ATR) specifically after DNA damage, and FANCE, which is phosphorylated on T346 and S374 by serine/threonine-protein kinase 1 (CHK1), which promotes degradation (Ceccaldi et al., 2016). In addition, phosphorylation of the FANCD2/FANCI heterodimer has been shown to be crucial to its monoubiquitination. This monoubiquitination is promoted by ATR and CHK1 which phosphorylate both proteins at multiple sites, including S331, T691and S717 on FANCD2, and S556, S559, S565, S596, S617, S629 on FANCI (Lopez-Martinez et al, 2016; Smogorzewska et al., 2007; Walden and Deans, 2014). Recently it has also been shown that there is a cluster of sites on FANCD2 (between residues 882 – 898) that are phosphorylated by casein kinase 2 (CK2) that prevent its monoubiquitination through inhibition of DNA binding, and that dephosphorylation of these sites is necessary for effective monoubiquitination of FANCD2 (Lopez-Martinez et al., 2019). In addition, FANCI has recently been shown to be directly phosphorylated by ATR on S557, S560 and S565 to stabilize its association with DNA and FANCD2. The authors also postulate that these sites may need to be dephosphorylated to enable effective deubiquitination (Tan et al., 2020b). In addition, it has been shown that certain phosphorylation of FANCI triggers dissociation of the heterodimer from chromatin (Ceccaldi et al., 2016).

Ubiquitination and Deubiquitination

Ubiquitination of a protein results in ubiquitin being added to the substrate. Ubiquitin is a small regulatory protein of around 8.5 kDa and is highly conserved throughout eukaryotes (Chaugule et al., 2020). It has a C-terminal tail and consists of 76 amino acids, of which seven are lysine residues. Four genes encode ubiquitin in the human genome: UBB, UBC, UBA52 and RPS27A (Kedves et al., 2017). Addition of ubiquitin to a substrate can mark the protein for degradation by the proteasome, change the protein's cellular location, affect the activity of the protein, or can also promote or prevent certain protein interactions (Jackson and Durocher, 2013). Ubiquitin is

attached to a protein substrate through three steps: activation, conjugation and ligation, and is an ATP-dependent reaction. This cascade of events results in ubiquitin being attached most commonly to lysines via an isopeptide bond. The isopeptide bond forms between the carboxyl group (COO^-) of the ubiquitin's glycine 76 (the final amino acid of ubiquitin) and the ε-amino group ($\varepsilon\text{-}NH_3^+$) of the substrate lysine (Figure 1-6, Step 5). Less commonly, ubiquitin can also be attached to cysteine residues through a thioester bond, to serine and threonine residues through an ester bond, and to the N-terminal amino group of a protein via a peptide bond. Ubiquitination can take two forms, either a single ubiquitin moiety can be added to a protein, resulting in monoubiquitination, or multiple ubiquitin molecules can be added in a chain, resulting in polyubiquitination (Chaugule et al., 2020; Jackson and Durocher, 2013; Morreale et al., 2017a).

The first step in ubiquitination is activation, achieved by the E1-activating enzyme. E1 binds to ATP and ubiquitin and catalyses the acyl-adenylation of ubiquitin to form a ubiquitin-adenylate intermediate (Figure 1-6, Step 1). The ubiquitin is then transferred to the active site of the E1 cysteine residue, releasing adenosine monophosphate (AMP), forming a thioester bond between the carboxyl group of ubiquitin and the E1 cysteine sulfhydryl group (Figure 1-6, Step 2). Humans contain two genes encoding E1-activating enzymes, UBA1 and UBA6. The second step in ubiquitination is conjugation, achieved through the E2-conjugating enzyme. E2 catalyses the transfer of ubiquitin from E1 to the active site cysteine on E2. E2 binds both the activated ubiquitin and the E1-activating enzyme and transfers the ubiquitin through a trans(thio)esterification reaction (Figure 1-6, Step 3). Humans have 35 different E2-conjugating enzymes. The final step in the ubiquitination cascade is achieved by the E3-ligase enzyme. E3 ubiquitin ligases catalyse the formation of the isopeptide bond between the C-terminal glycine residue of ubiquitin and a lysine residue of the target protein. These enzymes have substrate recognition modules as well as the ability to interact with the E2-conjugating enzyme (Figure 1-6, Steps 4, 5 and 6). E3s have one of two characteristic domains central to their enzymatic activity, either the homologous to the E6-AP carboxyl terminus (HECT) domain or the RING domain. E3 enzymes with HECT domains transiently bind ubiquitin in the transfer, whilst RING domains catalyse the direct transfer from the E2 enzyme to the substrate. To date, in humans at least 377 catalytic E3 ligases have been identified. (Chaugule et al., 2020; Jackson and Durocher, 2013; Jacq et al., 2013; Medvar, et al., 2016).

Deubiquitination is also an important regulatory mechanism. Deubiquitination results in the cleavage of the isopeptide bond between the ubiquitin glycine and the substrate lysine. This deubiquitination is achieved by deubiquitinating enzymes, known as DUBs. DUBs are highly

specific, with usually only a few substrates per DUB enzyme, and allow the cleavage and recycling of ubiquitin (Figure 1-6, Step 7) (Arkinson et al., 2018; Jackson and Durocher, 2013; Jacq et al., 2013).

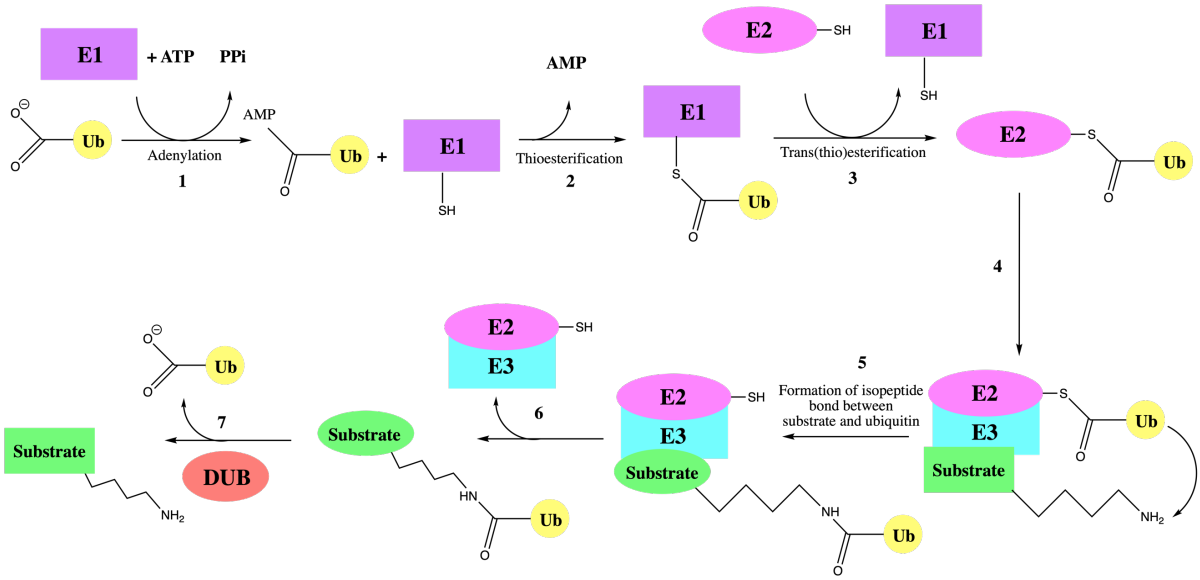


Figure 1-6: Schematic representation of ubiquitination by a RING-E3 ligase and deubiquitination by DUB. Step 1: Adenylation of ubiquitin. Step 2: Thioester bond formed between E1 and ubiquitin. Step 3: Ubiquitin is transferred to E2 from E1 via a trans(thio)esterification reaction. Step 4: RING-E3 ligase binds to E2-Ubiquitin and the substrate. Step 5: RING-E3 ligase catalyses formation of isopeptide bond between ubiquitin and the substrate. Step 6: Release of E2 and E3 resulting in the ubiquitinated substrate. Step 7: Deubiquitination of the substrate by a DUB.

Ubiquitination and deubiquitination are important post-translational modifications in the DDR. As already described above, the monoubiquitination of the FANCD2/FANCI heterodimer at lysines 561 and 523 respectively, is the activating step in the FA pathway, achieved by the E3-ligase FANCL and the core complex (Jackson and Durocher, 2013; Jacq et al., 2013). It has recently been elucidated that these monoubiquitination events cause a conformational change of the heterodimer, allowing the complex to lock onto the DNA (Rennie et al., 2020; Wang et al., 2020). It has also recently been shown that the monoubiquitination of FANCI serves to protect the monoubiquitinated-FANCD2 from deubiquitination by the USP1/UAF1 complex (Rennie et al., 2020). The deubiquitination of both monoubiquitinated FANCD2/FANCI and its dissociation from chromatin is also key to a functional FA pathway repairing ICLs (Jackson and Durocher, 2013; Oestergaard et al., 2007).

SUMOylation and deSUMOylation

SUMOylation of a protein results in the attachment of a Small Ubiquitin-like Modifier (SUMO) protein to the substrate. Most SUMO proteins are small, of around 12 kDa in size, and despite no sequence homology, fold in a nearly identical way to ubiquitin. Four different SUMO isoforms exist in humans, SUMO-1, SUMO-2 SUMO-3 and SUMO-4 (Wang and Dasso, 2009). SUMO proteins are attached to the substrate protein in a similar manner to ubiquitin, with the addition of a precursor step. This precursor step removes a C-terminal peptide from the SUMO precursor, cleaved by sentrin-specific proteases (SENP) to reveal a glycine-glycine motif (the hallmark of SUMO proteins). The resulting SUMO protein is then activated by the equivalent E1 enzyme, a heterodimeric SUMO Activating Enzyme (SAE). The resulting bound SUMO protein is transferred to the E2 equivalent SUMO conjugating enzyme, UBC9. Finally, the SUMO protein is attached to its substrate by one of a number of E3 SUMO ligating proteins, resulting in the formation of an isopeptide bond between the C-terminal glycine of SUMO and the specific lysine on the substrate protein (Hay, 2005; Jackson and Durocher, 2013; Pichler et al., 2004). E3 SUMO ligases can be divided into RING-finger ligases, including Protein Inhibitor of Activated STAT (PIAS) proteins, and HECT domain ligases (Pichler et al., 2004). SUMOylation plays many roles in the cell including assisting in protein stability, nuclear-cytosolic transport and transcriptional regulation. Most SUMOylation is quickly reversed by deSUMOylating enzymes (Hay, 2005; Jackson and Durocher, 2013).

SUMOylation has also been shown to play a role in many DDR pathways, particularly in facilitating the assembly of large protein complexes to sites of DNA damage (Hay, 2005). SUMOylation has also been implicated in the FA pathway. A fraction of the key complex, FANCD2/FANCI, is SUMOylated by the E2 and E3 SUMO enzymes UBC9 and PIAS1/4 upon replication fork stalling, which is reversed by the SUMO protease SENP6. Following SUMOylation, the FANCD2/FANCI complex can be polyubiquitinated by RNF4 which promotes its dissociation from chromatin via the DVC1-p97 ubiquitin segregase complex. Therefore, this post-translational modification regulates the level of the FANCD2/FANCI heterodimer localized to sites of DNA damage (Gibbs-Seymour et al., 2015).

1.1.5 Translational Impact of the DNA Damage Response: Novel Cancer Therapeutics

The DDR is a crucial signaling cascade that provides resistance to damaging agents, whether they be produced endogenously or exogenously. However, the dysregulation of the DDR leads to genomic instability which in turn promotes tumorigenesis (Bhattacharjee and Nandi, 2017; Ciccia and Elledge, 2010; Curtin, 2012). Tumours often exhibit deficiencies in the signaling or repair of DNA damage, and thus the DDR represents an attractive target for cancer therapeutics. Exploiting these deficiencies in DDR allows for a more targeted approach to cancer treatment, but also allow for a reduction in therapeutic resistance (Bouwman and Jonkers, 2012; Davis and Lin, 2011; Hosoya and Miyagawa, 2014; Moraes et al., 2012; Morreale et al., 2017a; O'Grady et al., 2014). This is due to the fact that downregulated DDR pathways make the tumour more sensitive to cytotoxic agents and DDR inhibitors, whilst upregulated DDR pathways contribute to therapeutic resistance. If a tumour has a defect in a particular DDR pathway, it then relies on an alternative pathway to repair damage. However, if this compensatory pathway is targeted, DNA damage results in cytotoxicity for the cancer cells, but survival for normal tissue. This concept is termed synthetic lethality (Gavande et al., 2016; Jackson and Durocher, 2013; Liu et al., 2020; Madhusudan and Hickson, 2005; Mouw and D'Andrea, 2014; O'Connor, 2015; Pearl et al. 2015; Tian et al., 2014).

A synthetic lethal relationship between two pathways exists when there is a loss of function mutation or epigenetic silencing of a particular DNA repair gene resulting in a defect in one DDR pathway, causing a dependence on a second pathway that is used as a therapeutic target (Bhattacharjee and Nandi, 2017; Martin et al., 2008). The most well-known synthetic lethal relationship is between BRCA-deficient tumours and PARP inhibitors (Bouwman and Jonkers, 2012; Hosoya and Miyagawa, 2014; Moraes et al., 2012; Tian et al., 2014). BRCA-deficient tumours are deficient in HR, and thus are highly sensitive to therapeutics that cause replication fork stalling. Upon treatment with PARP inhibitors, high levels of SSBs remain unrepaired due to inhibition in BER and this leads to stalled replication forks (Bhattacharjee and Nandi, 2017; Martin et al., 2008). These stalled replication forks are erroneously repaired in tumour cells leading to cell death (Ciccia and Elledge, 2010; Curtin, 2012; Wu et al., 2010). Therefore, when a patient is treated with PARP inhibitors, normal cells heterozygous for BRCA mutations are able to repair the resulting stalled replication forks, whilst the tumour cells that are BRCA-deficient cannot due to HR-deficiency, thus leading to tumour-cell death (Davis and Lin, 2011; Madhusudan and Hickson, 2005; Martin et al., 2008; O'Connor, 2015).

Many tumours are defective in a DDR pathway, like BRCA-deficient tumours. For example, whilst BRCA1 and BRCA2 mutations causing disruption of HR predispose to breast and ovarian cancer, other DSB repair mutations, including ATM and ATR loss of function mutations lead to a predisposition to breast and stomach cancer. In addition, mutations in the MMR machinery can lead to colorectal cancer, whilst defects in the NER pathway predispose to testicular cancer. There is also evidence for defects in TLS in a number of cancers (Altieri et al., 2008, Friedberg, 2003; Gavande et al., 2016; Hosoya and Miyagawa, 2014; Jacq et al., 2013). A synthetic lethal relationship between checkpoint signaling and the FA pathway is also beginning to emerge, including inhibition of ATM in FA pathway-deficient cells, and CHK1 inhibition increasing cisplatin sensitivity in FA pathway-deficient cells ultimately leading to tumour-cell death. In addition, UBE2T is overexpressed in several tumours, including breast and prostate cancer, and offers another synthetic lethal therapeutic target in the FA pathway (Bhattacharjee and Nandi, 2017; Curtin, 2012; Liu et al., 2020; Morreale et al., 2017b). These discoveries provide promise for the use of synthetic lethal approaches for novel tumour-specific therapeutics. Therefore, continued research is necessary for a robust and comprehensive understanding of the entire DDR to enable advances in cancer treatments (Auberbach, 2009; Ciccia and Elledge, 2010; Gavande et al., 2016; Hosoya and Miyagawa, 2014).

1.2 Determining Protein Structure: an Overview

Structural biology has been revolutionized in recent years. This field studies proteins and nucleic acids at the molecular level and allows many biological questions to be investigated. It is possible to solve the 3D structure of proteins to near atomic resolution and thus to elucidate the spatial orientation of the atoms that make up a protein (Wang and Wang, 2016). This in turn enables the investigation of structure-function relationships, conformational changes and how these affect protein activity, and to investigate disease-causing mutations and what effects these have on protein function. Structural biology has become a powerful tool in recent times due to major technological advances (Bai et al., 2015; Nwanochie and Uversky, 2019). Multiple methodologies have been developed based on entirely different physical principles, including X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and transmission electron microscopy (TEM). Each of these methods have unique advantages and caveats for solving protein structures. In the 1970s, the Protein Data Bank (PDB) was also established, which still serves as a repository for all solved protein structures.

X-ray crystallography was born in 1912 and is based on the principle of X-ray diffractions produced by passing X-rays through crystals. These diffraction patterns are unique to the crystal and can be used to solve the 3D structure. X-ray crystallography has transformed our understanding of biological molecules and has proven to be the backbone of the field of structural biology (Brooks-Bartlett and Garman, 2015; McPherson and Gavira, 2013; Woolfson, 2018). NMR spectroscopy joined X-ray crystallography around 40 years ago as a technique to determine protein structure. NMR involves applying an external magnetic field to a sample and looking at the resulting different resonance frequencies (known as chemical shifts) of different atomic nuclei and their proximity to each other. Therefore, NMR may be used to determine protein structure. NMR enables the determination of structures of around 10 kDa and below and gave the first fundamental insight into the dynamic nature of proteins (Markwick et al., 2008; Takeuchi et al., 2019; Wuethrich, 1990). Electron microscopy (EM) has existed as an imaging technique for over 100 years. However, recent advances in cryo-EM have made this a powerful technique for solving protein structures (Wang and Wang, 2016).

1.2.1 Transmission Electron Microscopy

The earliest electron microscope was developed in the first half of the 20th century by the Ruska brothers (Bai et al., 2015). EM involves passing a beam of high energy electrons through a specimen of interest to form an image. This is achieved as the scattered diffracted waves pass

through an electromagnetic lens to form a regular arrangement of diffraction spots on the back focal plane. This space is termed reciprocal space. When the transmitted and diffracted beams coincide at the image plane, a magnified image of the specimen can be observed. The specimen and the image plane are termed real space. A Fourier transform is required to transform from real space to reciprocal space (Bendersky and Gayle, 2001; Wang and Wang, 2016).

A major concern when looking at biological samples using electron microscopy is the degradation of the specimen. On one side, this is due to passing the high energy electron beam through the sample, resulting in radiation damage through the energy deposited into the specimen. On the other side, EM requires a high vacuum so that the electrons cannot interact with air molecules, and this vacuum would compromise the preservation of liquid aqueous samples. Therefore, EM samples must be fixed to enable imaging. Early EM studies fixed samples via dehydration. A popular method that developed was negative staining, whereby the aqueous solution a specimen is surrounded by is replaced by a dried solution of a heavy-metal salt. This technique will be further discussed below. However, dehydrating samples introduces artifacts into the structures of biological specimens and thus another method of fixing the sample in a closer-to-native state was needed. Following multiple different attempts, it was established that samples could be imaged following preservation by freezing the specimen in a thin layer of non-crystalline solid water, termed amorphous or vitreous ice. As these specimens need to be maintained at liquid nitrogen temperatures to maintain the vitreous ice layer, the technique was termed "cryo-EM". This technique will also be further discussed below (Bai et al., 2015; Thompson et al., 2016).

When using electron microscopy, often the choice of electron microscope is based on the type of imaging experiment, as well as practical availability of the instrument. Each electron microscope has an electron gun which is the source of the electrons. This gun extracts and accelerates electrons, and depending on the type of imaging experiment planned, different guns are used for different techniques. For example, for lower resolution studies, a thermionic electron source is sufficient (for example for negative staining). This source may be tungsten filaments or lanthanum hexaboride crystals (LaB_6) and operates at voltages of between 80 – 200 kV. For higher resolution studies, including that of cryo-EM, brighter and more coherent sources are needed. Therefore, a field emission gun (FEG) is usually used. This is a very fine tungsten filament coated in zirconium oxide, and operates between 200 – 300 kV at 1800 K. Each electron microscope must also have some form of electron detector. Historically, photographic film was used, however this severely limited the amount of data acquisition in a day to approximately 100 micrographs. Nowadays, the most common detectors include charged coupled device (CCD) cameras and direct electron

detectors (DED). DEDs are preferred for high resolution cryo-EM work as they reduce the added noise from the detector, which can cause a significant problem due to the poor signal-to-noise ratio associated with cryo-EM images (Thompson et al., 2016).

Once the specimen has been imaged via EM, the collection of two-dimensional (2D) projections from different orientations can be used to reconstruct the 3D structure of the sample (Bai et al., 2015). Negative staining is capable of obtaining a 3D model of around 20Å for protein structures. However, due to significant technological advances in recent years, cryo-EM is now a powerful technique which can be used to enable the structural determination of protein structures to atomic resolution (Carlo and Harris, 2010; Thompson et al., 2016). The number of deposited structures in the PDB solved via cryo-EM is increasing exponentially year on year. Cryo-EM also has its own dedicated database called the Electron Microscopy Data Bank (EMDB) (Abbott et al., 2018; Alnabati and Kihara, 2019).

1.2.1.1 Negative Staining

Negative staining is a method of fixing the sample of interest to a grid to enable EM analysis. This technique was first developed in the 1950s and was extremely popular during the 1960s and 1970s (Bai et al., 2015). Despite its shortcomings regarding dehydration of the sample, and thus introducing potential artifacts, as well as its resolution limitation, negative staining remains an incredibly popular technique, particularly in virology and microbiology, cell biology, polymer science and protein and lipid biochemistry. This popularity stems from the simplicity of the technique and its success in a multitude of systems. In terms of protein structural biology, negative staining remains an important screening technique, often used to assess the quality and homogeneity of the protein before progressing to cryo-EM (Carlo and Harris, 2010; Thompson et al., 2016).

Negative staining entails applying the sample of interest to an EM grid. EM grids are usually made from copper mesh, but can also be made from gold, nickel, molybdenum or rhodium. They are 3.05 mm in diameter and are specified in the units of metal bars per inch (eg. grids of 200 mesh have 200 lines per inch and also have 20 squares in each direction). These metal grids are covered in a foil layer that usually contains a regular array of repeating holes for cryo-EM, however for negative staining this is usually a continuous carbon foil. This foil layer may also be made of different materials, including copper, nickel, silicon, aluminium and gold. The most common foil layer is amorphous carbon due to it being inert, relatively electron transparent and easily manufactured. The resulting supporting surface on the grid is hydrophobic, which prevents the

application of the aqueous sample solution onto the grid. Therefore, to reduce hydrophobicity, the grids are treated with low energy plasmas, termed "glow discharge" (Carlo and Harris, 2010; Sgro and Costa, 2018; Thompson et al., 2016).

To prepare a sample for negative staining, the EM grid must be glow discharged before the sample of interest is applied to the grid. The excess sample is blotted off, and the grid is stained with a solution of heavy metal salt, most commonly 1 – 2% uranyl acetate. The excess stain is then also blotted off to ensure a thin layer of stain. The stain dehydrates and envelopes the sample, resulting in a negatively stained specimen, which can then be visualized by the absence of stain (Figure 1-7). It is also possible to positively stain samples, whereby the specimen is visualized by the presence of stain. Normally, across the grid there is some variation in stain depth and the grid must be imaged at the optimal stain depth, where it is not too deep or too shallow. The resulting 2D image is generated due to the difference in electron scattering between the sample and the surrounding stain. The resolution of negative staining is limited by the grain size of the stain and its ability to completely envelope the sample on the grid. Due to the nature of the technique, negative staining can at most give the solvent-excluded surface and shape of the sample. Despite the limitations of negative staining, many groups utilize both negative staining and cryo-EM in parallel when determining the structure of a protein (Carlo and Harris, 2010; Sgro and Costa, 2018; Thompson et al., 2016).

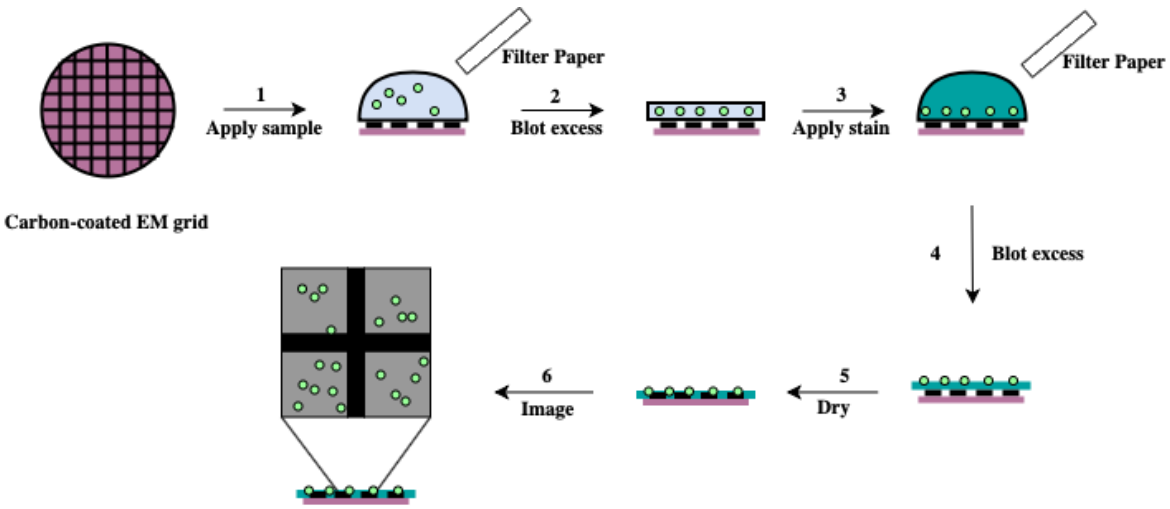


Figure 1-7: Schematic representation of negative staining. Step 1: Apply the sample to grid. Step 2: Blot off excess sample. Step 3: Apply the stain to the grid. Step 4: Blot off excess stain. Step 5: Dry grid. Step 6: Place grid into microscope to visualise sample.

1.2.1.2 Cryo-Electron Microscopy

Dehydration techniques such as negative staining remove the sample from its native aqueous environment. To overcome this disadvantage, cryo-immobilisation was developed, which enables the preservation of the aqueous environment, and also decreases the radiation damage of EM. The key factor of this technique involves the rapid freezing of the sample, at a rate of around 10^6°C per second, in liquid ethane (-188°C). Due to the speed at which the sample is frozen, ice crystals cannot form and the water surrounding the specimen is fixed in a vitreous state. If the sample is frozen too slowly, or the sample is allowed to warm above around -137°C, crystalline ice begins to form, and the water devitrifies. Ice contamination compromises the sample as ice crystals draw water molecules from the hydration shell of the sample, or from the sample itself. In an ideal world, a thin vitreous ice layer forms containing the sample of interest, where the thickness of the ice is approximately the same thickness as the sample. Having a thin layer minimizes diffraction scattering events due to crystalline ice, as well as maximizing sample contrast on the microscope. In addition to crystalline ice, hexagonal ice can also contaminate the sample. This is usually reduced by working in humidity-controlled environments and minimizing ice contamination in the liquid nitrogen (Nwanochie and Uversky, 2019; Sgro and Costa, 2018; Thompson et al., 2016).

When preparing a sample for cryo-EM, grid choice is an important factor. As with negative staining, the grid is made up of metal mesh coated in a metal foil. Usually the first type of grids trialled for cryo-EM are Quantifoil grids, where the foil layer is a regular array of circular holes. This regular pattern facilitates automated data collection. Many types of grids exist however, including ones with irregular patterns of the foil layer, including Lacey carbon grids. Copper grids with a holey carbon foil layer are popular, as copper is cheap, and the carbon foil layer is easily manufactured onto the mesh. However, many different materials are utilized, for both grid support and for foil coatings, due to a variety of shortcomings of these types of grids. For example, when a traditional copper/carbon grid is cooled in liquid ethane, due to the dual nature of the materials and difference in the coefficient of thermal expansion, the metal support shrinks more than the carbon foil, causing "cryo-crinkling" of the carbon foil, resulting in a loss of tension and moving of the sample. Therefore, by ensuring that the mesh of the grid and the foil have similar coefficients of thermal expansion, this effect can be minimized. Another shortcoming of carbon coated grids is that carbon's conductivity decreases as the temperature decreases, and thus charges accumulate on the foil surface. Upon radiation, these charges induce movement of the sample, and thus result in loss of resolution. In recent years gold mesh supported by gold foil has become a popular grid of choice as it removes the possibility of "cryo-crinkling", as the mesh and foil shrink at the same

rate. In addition, it is also more conductive as temperature decreases, reducing the accumulation of charges, and thus reducing sample movement. Using these gold/gold grids can in fact reduce particle movement during imaging by more than an order of magnitude, ultimately increasing resolution (Russo and Passmore, 2014; Sgro and Costa, 2018; Thompson et al., 2016).

In a similar manner to negative staining grids, cryo-EM grids must also be treated to enable the aqueous sample to be applied to the grid. For carbon coated grids that are inherently hydrophobic, this entails glow discharging the grids. Often upon application of the sample, cryo-EM grids are not completely covered, and thus different plasma conditions must be trialed to achieve a more uniform spreading of the solution onto the grid. In addition, it can be a challenge to promote the protein particles into the holes of the foil coating. It can assist to add an additional surface on top of the foil to provide further physical support for the protein particles to adsorb to. However, addition of amorphous carbon surfaces increases background noise, and thus for small proteins of less than 150 kDa they cannot be used. In recent times graphene has become a popular alternative support film, as it is a one atom thick hexagonal lattice of carbon atoms with excellent conductivity properties and is essentially invisible under the electron microscope. However, graphene supports are extremely difficult to make, and also render the grid hydrophobic. Therefore, adding an extra film of graphene oxide instead has become popular as it is more easily added to grids, and is also hydrophilic in nature, eliminating the need to glow discharge the grid. However, even coverage of the grid with graphene oxide is difficult and can contribute to background noise (Sgro and Costa, 2018).

Once the grid is chosen and prepared appropriately to ensure a hydrophilic surface, 3 – 4 μL of aqueous sample is applied to the grid. The excess solution is blotted off to form a thin suspension of the protein solution across the grid. The grid is then rapidly plunged into liquid ethane that has been cooled by liquid nitrogen, known as plunge freezing (Figure 1-8). As mentioned above, the vitreous ice layer should be as thin as possible whilst fully coating the particles, which maximises contrast upon image acquisition. Consistency with blotting and plunging is extremely difficult, especially between different batches of grids prepared. Several commercial plunge freezing devices have been developed for more reproducible results. These may have different features, including single-sided blotting, double-sided blotting and humidity and temperature control. In addition, optimal ice thickness is difficult to reproduce due to different surface properties of the blotting paper. Therefore, several blot-free methods have also been developed. To continue to improve the problem of reproducibility, further automated systems have been developed which reduce the amount of sample required, including a spray-plunging system and a system that uses acoustic

waves to deliver femtolitre droplets onto the EM grid (Arnold, et al., 2017; Danev et al., 2019; Nwanochie and Uversky, 2019; Sgro and Costa, 2018; Thompson et al., 2016).

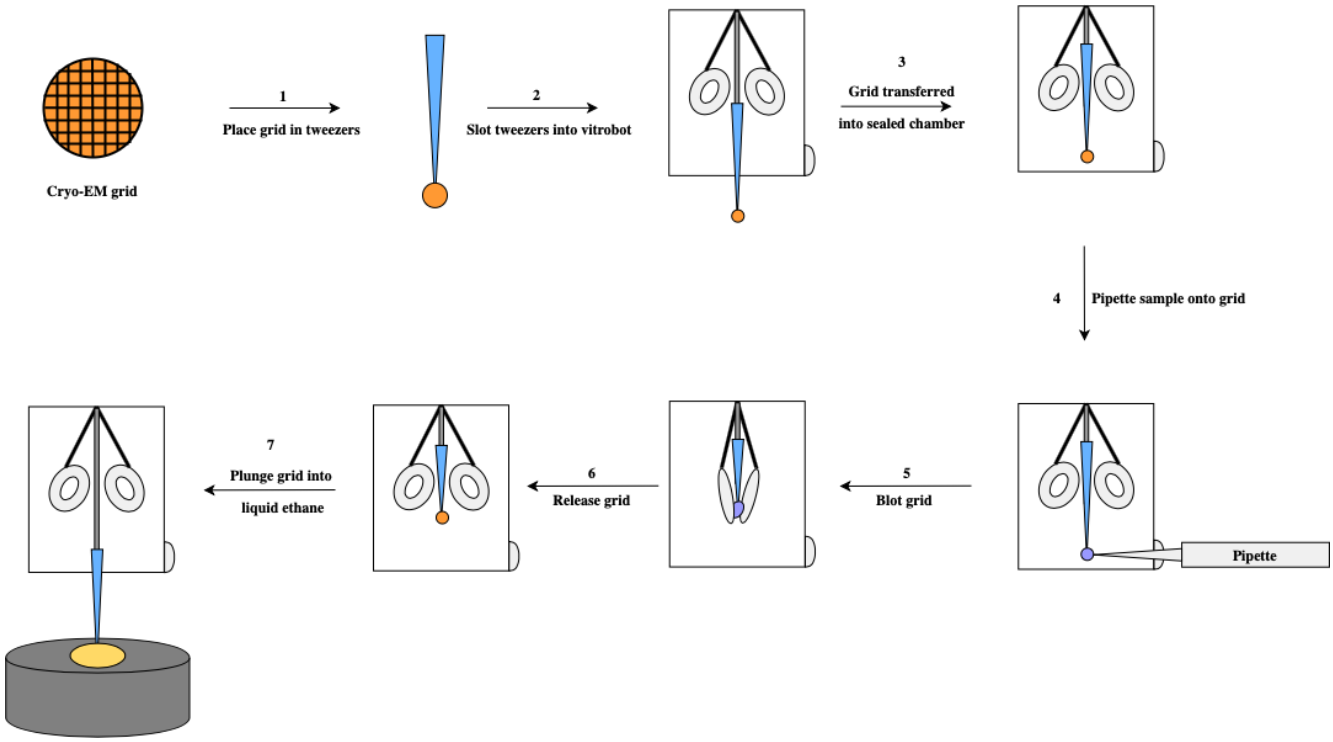


Figure 1-8: Schematic representation of plunge-freezing via Vitrobot. Step 1: Place grid into Vitrobot tweezers. Step 2: Slot tweezers holding grid into Vitrobot. Step 3: Tweezers and grid taken up into a sealed chamber. Step 4: Apply sample through small opening directly onto grid. Step 5: Automatically blot sample with preset conditions with Vitrobot blotting pads. Step 6: Automatically release grid from blotting pads. Step 7: Immediately plunge into liquid ethane for vitrification. (NB: the grid is then transferred to liquid nitrogen for storage until being clipped and inserted into the microscope for imaging).

In addition to the challenges presented above, there are further problems that may arise resulting in the failure of grid preparation, or irreproducibility of grids. For example, when proteins are applied to the cryo-EM grid they can often aggregate or denature. This could be due to the fact that the physical conditions have changed from the original solution droplet to the grid surface, and in particular due to the change in surface area to volume ratio upon grid application. Many factors contribute to the distribution of the particles onto the grid, including the type of grid, type of foil coating and the air-water interface, and all of these play a role in grid preparation (Danev et al., 2019; Sgro and Costa, 2018). The air-water interface creates particularly difficult challenges to overcome, as it has become clear that protein particles can adsorb to the air-water interface and

denature within one second. Therefore, the speed at which a cryo-EM grid is prepared is extremely important (Glaeser, 2018). In addition, depending on the particle's ability to absorb to the grid surface, the concentration of the protein may change, for example, the density may be higher if the particles absorb more easily, or may be lower if they are repelled from the surface or have a higher affinity to the support structure. Interactions of the particles with the air water interface may also cause preferential orientations of the particles. There are many techniques that have been developed to assist with these challenges, including using a glycerol gradient, using chemical crosslinking agents (such as GraFix), adding small amounts of detergents, and many more (Danev et al., 2019; Sgro and Costa, 2018).

Once the cryo-EM grids have been prepared, the grids are screened as an initial evaluation to assess sample and grid quality. This atlas screening provides a collection of grid tiles that show an overview of the grid and support quality, as well as the ice thickness across the grid, and allows identification of areas with intact support and appropriate ice thickness. Once these areas have been identified, data is acquired in these areas. This involves the acquisition of thousands of micrographs and is an automated process. The particles in these micrographs are then aligned and grouped to generate a 2D classification of the particles in different orientations. A 3D model is then generated and refined, followed by 3D classification and refinement of each 3D class (Danev et al., 2019; Nwanochie and Uversky, 2019).

Cryo-EM has existed as a structural biology technique for more than 40 years. However, in the last 5 – 10 years the technique has seen rapid progress, resulting in its ability to achieve atomic resolution structures. The explosion of this method for high-resolution protein structures has been termed a revolution. Recent developments have included improved grid preparation procedures, a new generation of electron detectors, as well as drastically improved image processing procedures. This technique allows unparalleled insight into many biological questions and therefore represents a powerful tool in a new age of structural biology. It is clear that the cryo-EM revolution has only just begun (Bai et al., 2015; Danev et al., 2019; Nwanochie and Uversky, 2019; Thompson, et al., 2016).

1.3 Aims of the Project

FANCD2/FANCI:

1. Investigate the DNA-dependence of the monoubiquitination of FANCD2/FANCI and FANCI alone *in vitro*
2. Investigate the DNA-dependence of the deubiquitination of monoubiquitinated-FANCD2/FANCI and monoubiquitinated-FANCI *in vitro*
3. Study the structures of the human full-length FANCD2/FANCI, the monoubiquitinated-FANCD2/FANCI and the monoubiquitinated-FANCI/FANCD2 complexes by negative staining EM and cryo-EM, and elucidate whether there is a conformational change in the complex's structure upon monoubiquitination
4. Establish whether the monoubiquitinated-FANCD2/FANCI intermediate forms filamentous arrays

UHRF1:

1. Investigate whether UHRF1 promotes monoubiquitination of FANCD2/FANCI *in vitro*
2. Study the structure of human full-length UHRF1 by negative staining EM and cryo-EM

2. Expression and Purification of Proteins and Optimisation of Ubiquitination Assay

In order to investigate the monoubiquitinated FANCD2/FANCI complex, both for biochemical assays as well as for structural characterisation, the ubiquitination assay components needed to be purified in large quantities, and the ubiquitination assay needed to be optimized. In addition, the concentration of the resulting ubiquitinated FANCD2/FANCI complex needed to be optimized to enable structural analysis.

This chapter describes the purification of the ubiquitination assay components, as well as an alteration to the expression and purification procedure previously used in the Cohn laboratory for the E3 ligase, FANCL, resulting in an improved ubiquitination efficiency. Finally, the chapter describes the optimisation of increasing the concentration of the resulting ubiquitinated complex for structural studies.

2.1 Experimental Design

The monoubiquitination of FANCD2 and FANCI *in vivo* requires the full core complex. *In vitro* only the E3 ligase FANCL is required to fully monoubiquitinate FANCD2 in complex with FANCI (Figure 2-1), or to fully monoubiquitinate FANCI alone (Figure 2-2). Therefore, this *in vitro* ubiquitination assay is utilised to prepare the complexes for structural analysis (Alpi et al., 2008; Rennie et al., 2020). The assay involves the E1 enzyme UBA1, the E2 enzyme UBE2T, the E3 ligase FANCL, supercoiled plasmid DNA (scDNA) as the DNA substrate, the co-purified FANCD2/FANCI complex or purified FANCI alone, ATP, $MgCl_2$ and ubiquitin to produce the monoubiquitinated-FANCD2/FANCI complex or monoubiquitinated-FANCI. To form the monoubiquitinated-FANCI/FANCD2 complex, FANCD2 is added post monoubiquitination of FANCI.

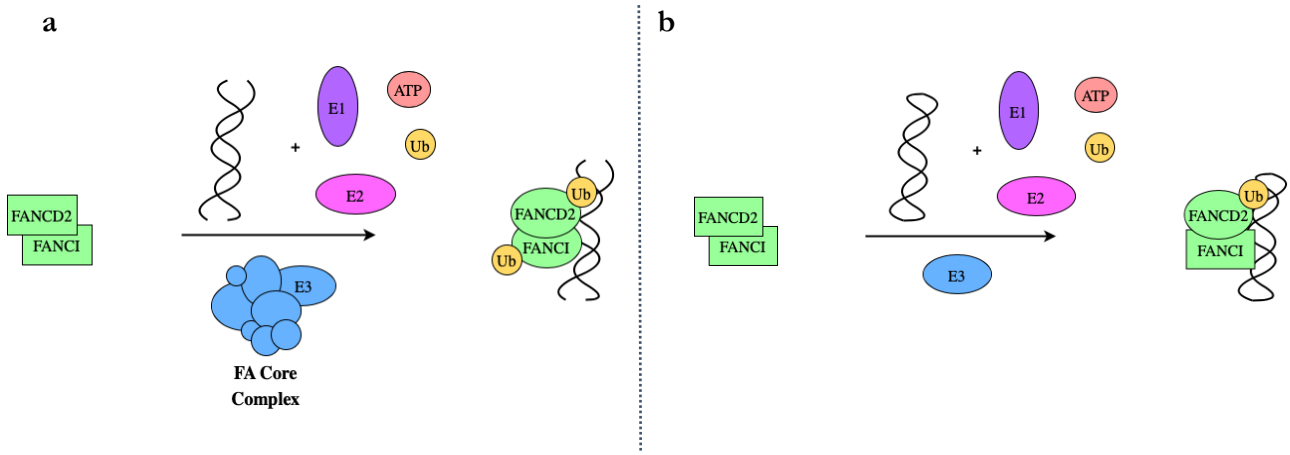


Figure 2-1: Ubiquitination of the FANCD2/FANCI complex. a) in vivo monoubiquitination of FANCD2/FANCI requiring the full core complex. b) in vitro monoubiquitination of FANCD2/FANCI requiring only the E3 ligase, FANCL.

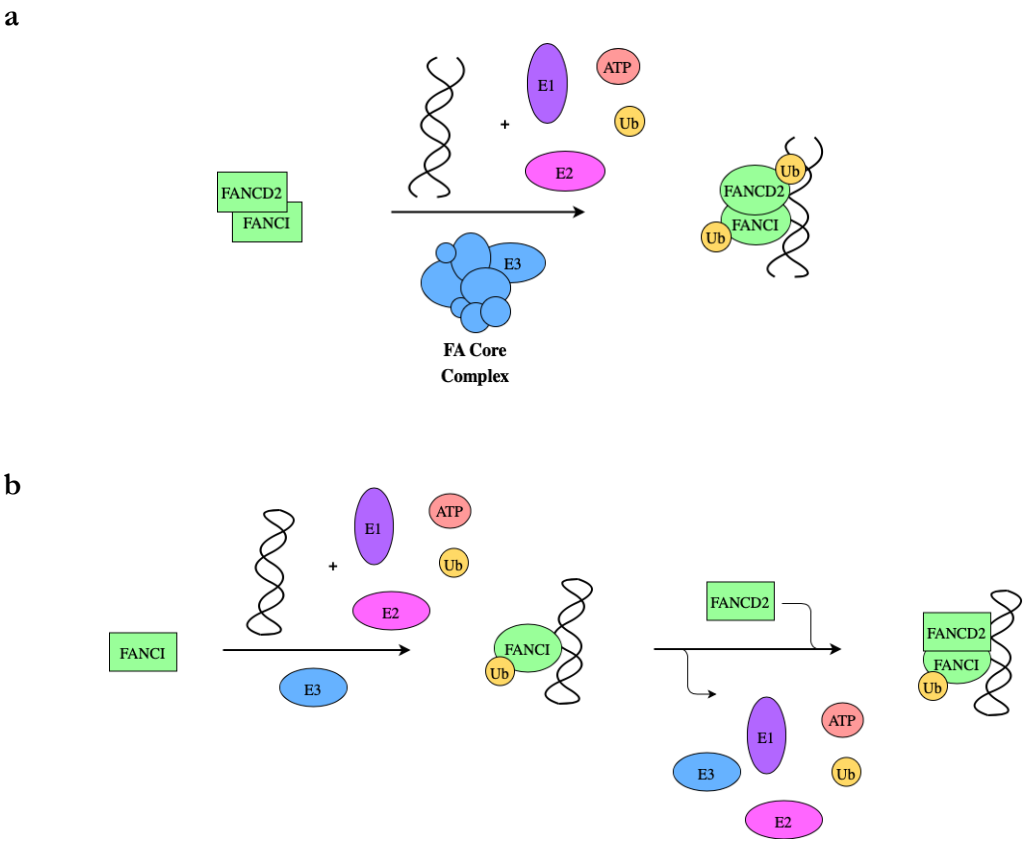


Figure 2-2: Ubiquitination of the FANCD2/FANCI complex. a) in vivo monoubiquitination of FANCD2/FANCI requiring the full core complex. b) in vitro monoubiquitination of FANCI requiring only the E3 ligase, FANCL, followed by incubation with FANCD2.

2.2 Purification of UBA1 (E1), UBE2T (E2), FANCL (E3), Ubiquitin, FANCD2/FANCI, FANCD2, FANCI and scDNA

In order to conduct the ubiquitination assay for structural studies, the ubiquitination assay components have to be expressed and purified in large quantities. FLAG-HA-tagged UBA1 was expressed in Sf9 insect cells and purified via flag affinity purification (Figure 2-3a). 6xHis-Smt3-tagged UBE2T was expressed in *E. coli* BL21(DE3) cells and purified via nickel affinity purification. The 6xHis-Smt3-tagged UBE2T was further purified using size exclusion chromatography. The resulting purified 6xHis-Smt3-tagged UBE2T was incubated with Ulp1 protease to cleave the 6xHis-Smt3 tag. The mixture was then purified via nickel affinity purification, keeping the flow through portion containing the untagged UBE2T. UBE2T was further purified using size exclusion chromatography (Figure 2-3a). FLAG-HA tagged FANCL was expressed in Sf9 cells and purified via flag affinity purification (Figure 2-3a). 6xHis-tagged ubiquitin was expressed in *E. coli* BL21(DE3) cells and purified via nickel affinity purification. The 6xHis-tagged ubiquitin was further purified using size exclusion chromatography (Figure 2-3a). FLAG-HA tagged FANCD2 and 6xHis-tagged FANCI were co-expressed in Sf9 insect cells, using limiting FLAG-HA tagged FANCD2, and purified via flag affinity purification to a one-to-one ratio (Figure 2-3c). FLAG-HA tagged FANCD2 was expressed in Sf9 insect cells and purified via flag affinity purification (Figure 2-3c). 6xHis-tagged FANCI was expressed in Sf9 insect cells and purified via nickel affinity purification (Figure 2-3d). The DNA used for the ubiquitination assay was scDNA: pBlueScript SKII (+) purified using the QIAGEN Plasmid Midiprep Kit (Figure 2-3b).

2.3 Optimisation of Ubiquitination Assay: Improvement in the Efficiency of Ubiquitination

To determine the ubiquitination efficiency for monoubiquitinated-FANCD2/FANCI, the ubiquitination assay was prepared by incubating in reaction buffer the appropriate concentrations of UBA1, UBE2T, FANCL, FANCD2/FANCI, ubiquitin, ATP and 1 μg of scDNA at room temperature. A sample was taken at 6 hours and analysed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 2-4a). It is clear that there is a significant proportion of FANCD2 that is not being ubiquitinated in the assay. To ensure a pure sample for the subsequent structural studies of the monoubiquitinated-FANCD2/FANCI, the efficiency of this ubiquitination assay was optimised as described below.

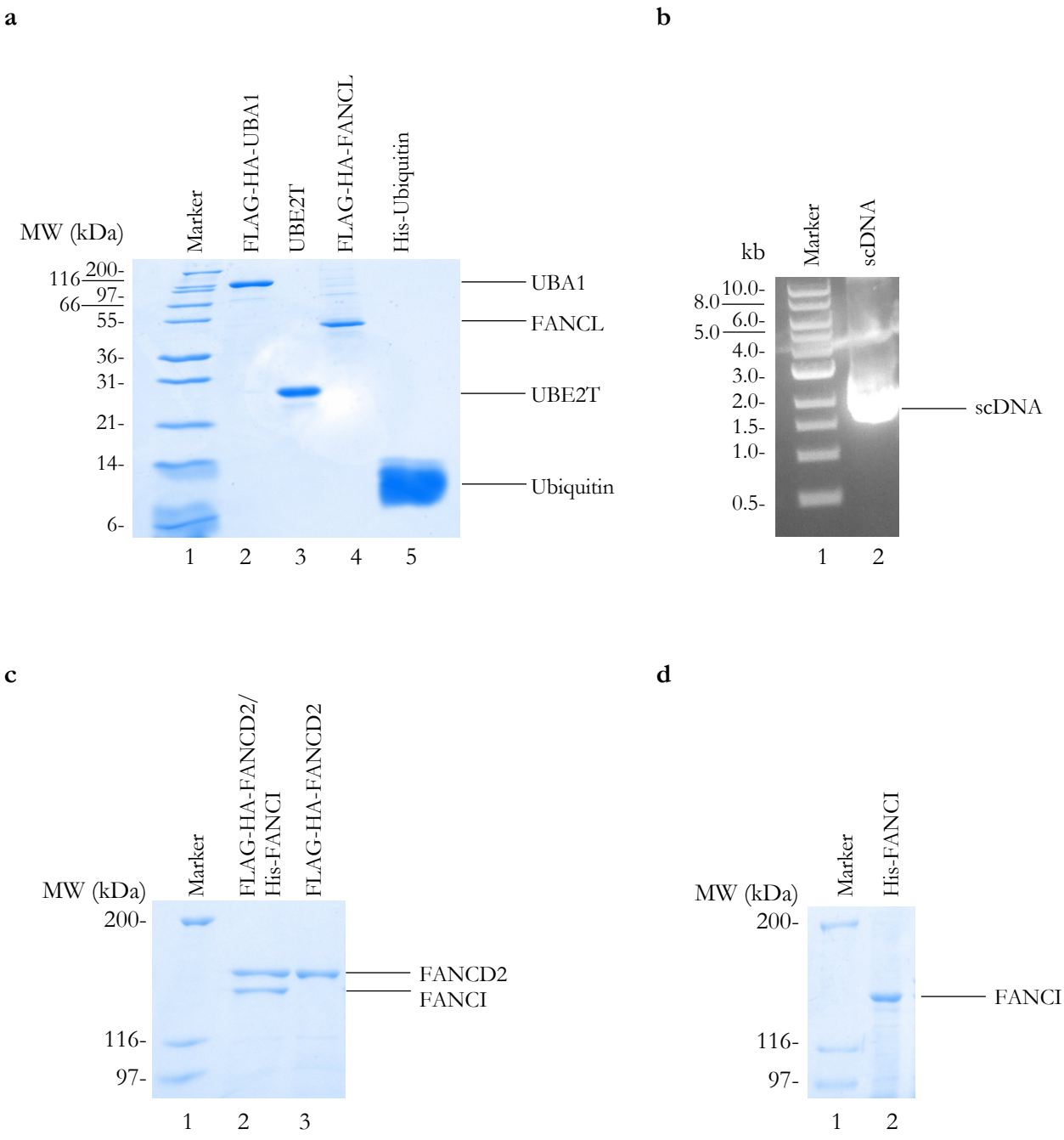


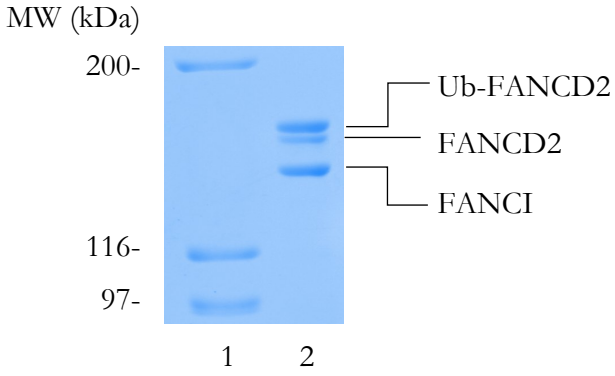
Figure 2-3: Purification of Ubiquitination Assay Components. a) Coomassie blue stain of the purification of UBA1 (E1), UBE2T (E2), FANCL (E3) and Ubiquitin. Lane 1: marker, lane 2: FLAG-HA-UBA1, lane 3: UBE2T, lane 4: FLAG-HA-FANCL, lane 5: His-Ubiquitin. b) Agarose gel of the purification of scDNA. Lane 1: 1kb marker, lane 2: scDNA. c) Coomassie blue stain of the purification of FANCD2/FANCI and FANCD2. Lane 1: marker, lane 2: FLAG-HA-FANCD2/His-FANCI, lane 3: FLAG-HA-FANCD2. d) Coomassie blue stain of the purification of FANCI. Lane 1: marker, lane 2: His-FANCI.

The expression and purification protocol established in the Cohn laboratory for the E3 ligase, FANCL, was three days incubation at 27 ºC. However, as demonstrated in Figure 2-4a the efficiency of the ubiquitination assay could be improved. Therefore, the expression time for the E3 ligase was optimised to ensure maximal enzymatic activity for FANCL. This expression was found to be four days incubation at 27 ºC. It is clear that four days expression reduces the level of non-ubiquitinated FANCD2 in comparison to three days expression (Figure 2-4b).

2.4 Optimisation of the Sample Concentration for Structural Studies

After optimisation of the ubiquitination assay, the purified monoubiquitinated-FANCD2/FANCI was prepared for structural studies. The sample was concentrated on a 1 MDa concentrator and the top fraction was removed and treated with benzonase to release the complex from the DNA. The sample was then purified using size exclusion chromatography. However, as displayed in Figure 2-5a, the sample was not eluted from the column at a concentration visible by Coomassie blue stain (expected in fractions A11, A12 and B12). It can be clearly seen in lane 1 that most of the sample was stuck on the membrane of the concentrator. In order to release this sample from the membrane, it was hypothesized that the DNA could be entangled in the membrane, and thus most of the sample cannot be harvested after concentration. Therefore, following concentration on the 1 MDa concentrator, rather than removing the sample before benzonase treatment, the sample was treated with benzonase whilst sitting on the membrane. Following benzonase treatment, the top fraction was removed and purified using size exclusion chromatography. In Figure 2-5b it can be seen that despite there still being a large proportion of the complex stuck on the membrane (lane 1), the amount of sample injected onto the column had increased (lane 2). Therefore, the sample was visible following gel filtration and can be faintly seen in the A12 and B12 fractions. However, this sample was still at a very low concentration, and thus not suitable for structural studies. Therefore, the volume of the ubiquitination assay was doubled, from a starting material of 100 µg of FANCD2/FANCI (reaction volume of 1.0 mL) to 200 µg of FANCD2/FANCI (reaction volume of 2.0 mL). Figure 2-5c displays the Coomassie blue stain of the resulting gel filtration, showing a dramatic increase in concentration of the eluted fractions A12 and B12 (lanes 9 and 10).

a



b

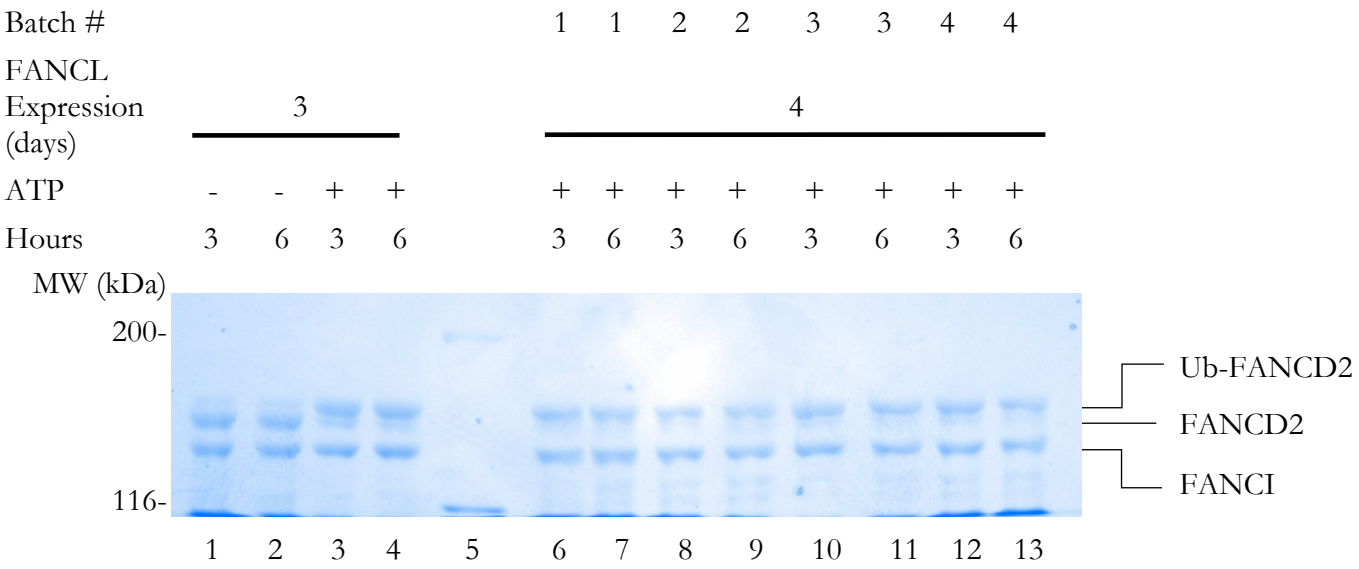


Figure 2-4: Optimisation of E3 ligase efficiency in ubiquitination assay. a) Coomassie blue stain of the purification of monoubiquitinated-FANCD2/FANCI after 6 hours showing the non-ubiquitinated FANCD2. Lane 1: marker, lane 2: monoubiquitinated-FANCD2/FANCI. b) Coomassie blue stain of the optimisation of E3 ligase efficiency showing monoubiquitinated-FANCD2/FANCI. Lanes 1 – 2: negative control, lanes 3 – 4: 72 hours FANCL expression, lane 5: marker, lanes 6 – 13: 96 hours FANCL expression.

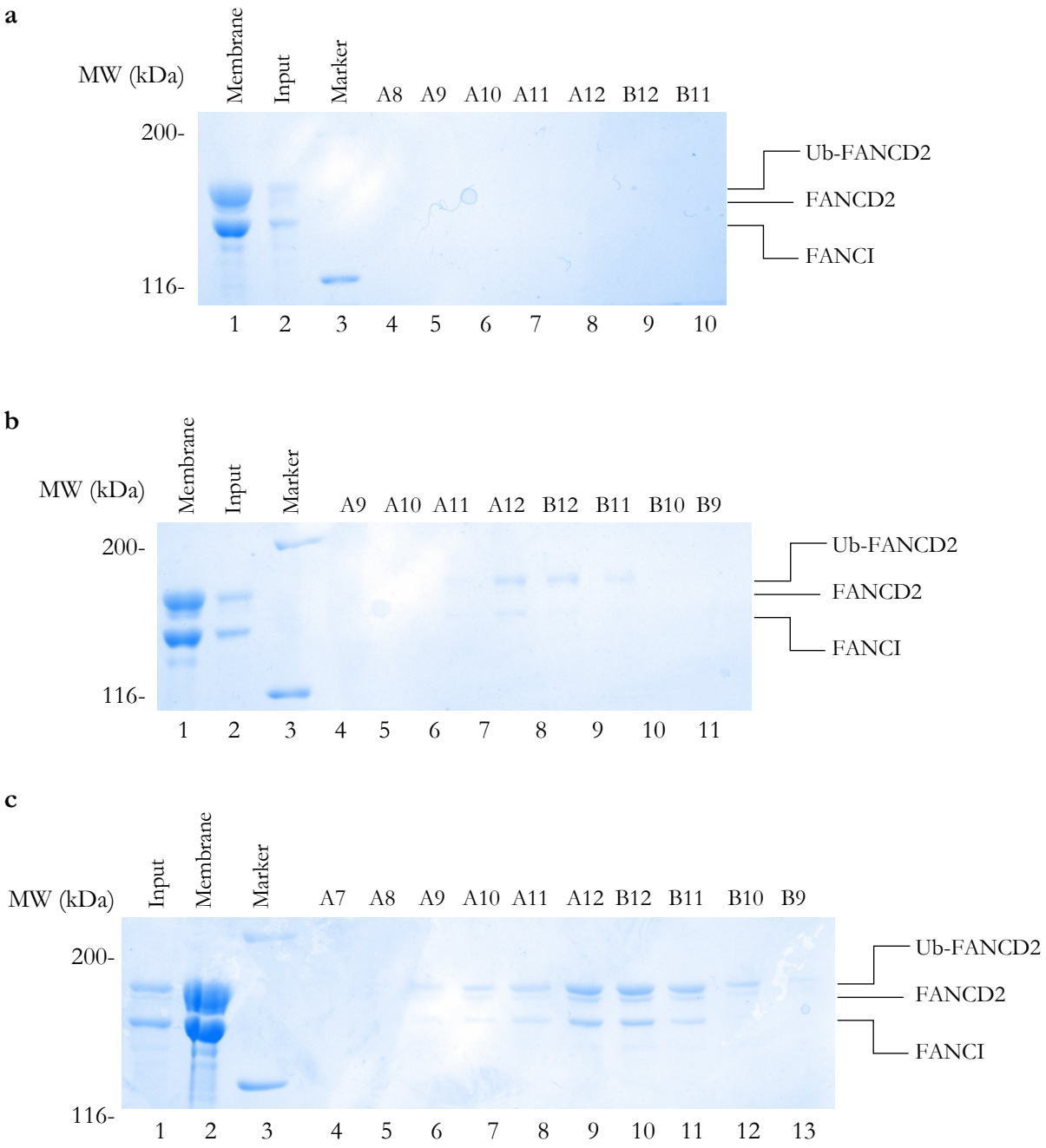


Figure 2-5: Optimisation of Concentration of Monoubiquitinated-FANCD2/FANCI. a) Coomassie blue stain of the purification of low concentration monoubiquitinated-FANCD2/FANCI. Lane 1: membrane fraction, lane 2: top fraction, lane 3: marker, lanes 4 – 10: gel filtration peak fractions (lane 4: A8 peak fraction, lane 5: A9 peak fraction, lane 6: A10 peak fraction, lane 7: A11 peak fraction, lane 8: A12 peak fraction, lane 9: B12 peak fraction, lane 10: B11 peak fraction). b) Coomassie blue stain of the purification of mid concentration monoubiquitinated-FANCD2/FANCI. Lane 1: membrane fraction, lane 2: top fraction, lane 3: marker, lanes 4 – 11: gel filtration peak fractions (lane 4: A9 peak fraction, lane 5: A10 peak fraction, lane 6: A11 peak fraction, lane 7: A12 peak fraction, lane 8: B12 peak fraction, lane 9: B11 peak fraction, lane 10: B10 peak fraction, lane 11: B9 peak fraction). c) Coomassie blue stain of the purification of higher concentration

monoubiquitinated-FANCD2/FANCI. Lane 1: top fraction, lane 2: membrane fraction, lane 3: marker, lanes 4 – 13: gel filtration peak fractions (lane 4: A7 peak fraction, lane 5: A8 peak fraction, lane 6: A9 peak fraction, lane 7: A10 peak fraction, lane 8: A11 peak fraction, lane 9: A12 peak fraction, lane 10: B12 peak fraction, lane 11: B11 peak fraction, lane 12: B10 fraction, lane 13: B9 peak fraction).

With the ubiquitination assay components purified in large quantities, the ubiquitination efficiency dramatically improved, and the concentration of the resulting sample increased, the monoubiquitinated-FANCD2/FANCI could then be prepared for structural studies. In addition, biochemical assays could also be conducted. These biochemical assays and structural studies are further described in Results Chapter 3.

3. Structural Studies of the Monoubiquitinated FANCD2/FANCI Complex

The monoubiquitination of the FANCD2/FANCI complex is crucial for a functioning FA pathway. Both ubiquitination sites lie at the heterodimer interface, raising questions on how the complex is monoubiquitinated, and if indeed whether it is the complex or the individual units prior to heterodimer formation that are monoubiquitinated. The deubiquitination of the heterodimeric complex is also instrumental in a functioning FA pathway. Despite being extensively studied since their identification, the detailed mechanism of how FANCD2 and FANCI are monoubiquitinated and deubiquitinated remains unclear (Joo et al., 2011; van Twest et al., 2017).

At the conception of this project, there was little to no structural insight into the monoubiquitinated complex. As mentioned above, the precursor to the monoubiquitinated FANCD2/FANCI, the non-ubiquitinated FANCD2/FANCI complex, was solved via X-ray crystallography in 2011 (Joo et al., 2011). This structure of the mouse proteins gave the first insight into how the heterodimeric complex forms, and that the two monoubiquitination sites lie at the interface of the dimer. The mouse structure also showed pseudo two-fold symmetry. Following this, the human FANCD2/FANCI structure was solved via cryo-EM to around 22 Å (Liang et al., 2016). This full-length human structure showed many similarities to the original mouse structure, however with a loss of the two-fold symmetry. An additional domain was also identified (the "Tower" domain) in the human structure which was shown to be important for DNA binding. This domain was partially deleted in the mouse structure, presumably due to the need to form good quality crystals for X-ray crystallography and may have resulted in a different orientation of the Tower domain in the mouse complex.

Both non-ubiquitinated structures share a central hollow channel, likely to be where the DNA binds. It has been speculated that upon monoubiquitination of the FANCD2/FANCI complex, the complex undergoes a conformational change which activates the FA pathway to allow the repair of the ICL. This conformational change could alter the DNA binding properties of the complex, as well as allow the binding of downstream repair factors. However, whether this conformational change actually occurs, and if it does occur, how much of the conformational change is attributed to the monoubiquitination of FANCD2 versus the monoubiquitination of FANCI had yet to be elucidated during the formation of this project.

This chapter discusses the progress towards determining the structures of the monoubiquitinated FANCD2/FANCI complexes via cryo-EM. Specifically, the importance of DNA in both ubiquitination and deubiquitination will be discussed, as well as progress towards possible complex structures in the monoubiquitination sequence of events. This includes progress towards an improved cryo-EM structure of the human non-ubiquitinated FANCD2/FANCI complex, the monoubiquitinated-FANCD2/non-ubiquitinated-FANCI intermediate complex and the monoubiquitinated-FANCI/non-ubiquitinated-FANCD2 intermediate complex. The recent structural developments revealed in late 2019 will also be discussed in Section 3.8.

3.1 Experimental Design

As the mechanism for monoubiquitination of the FANCD2/FANCI complex is unknown, there are multiple possible ubiquitination routes from the non-ubiquitinated FANCD2/FANCI to the monoubiquitinated-FANCD2/monoubiquitinated-FANCI (Figure 3-1). This project investigates the starting construct, the non-ubiquitinated FANCD2/FANCI as well as the two possible intermediate complexes, monoubiquitinated-FANCD2/FANCI, and monoubiquitinated-FANCI/FANCD2.

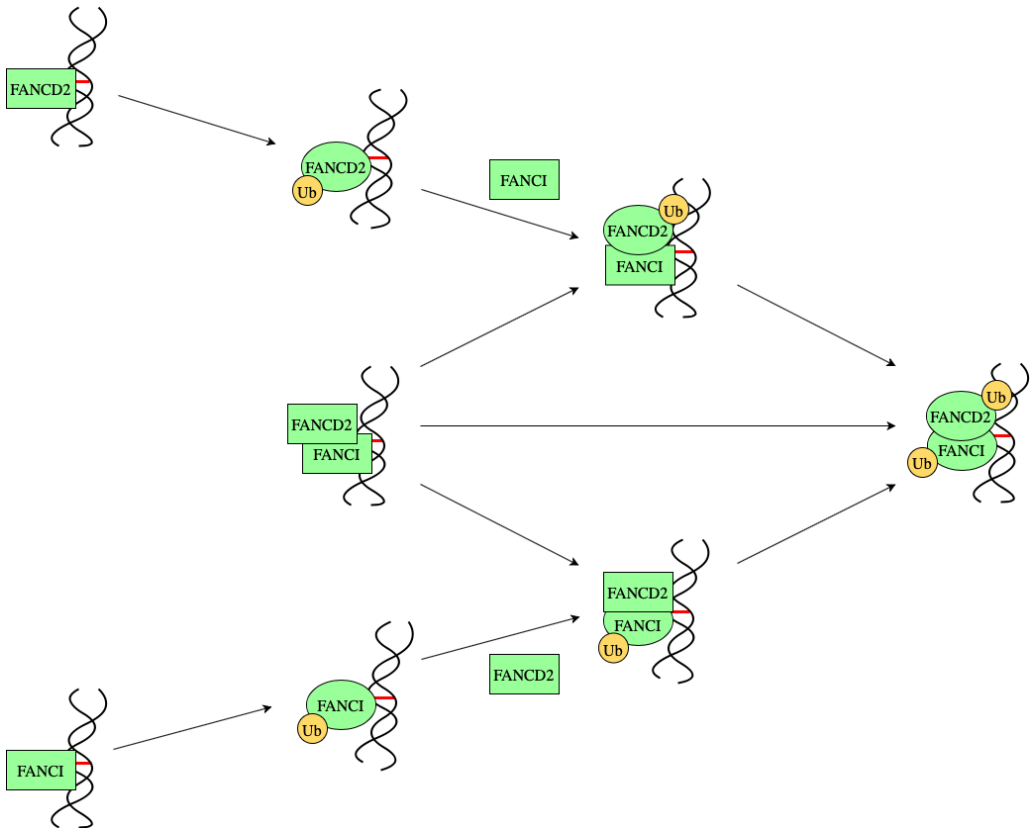


Figure 3-1: Possible ubiquitination routes from non-ubiquitinated FANCD2/FANCI to monoubiquitinated-FANCD2/monoubiquitinated-FANCI.

The approach for determining the structure of these complexes via cryo-EM is described in Figure 3-2. This includes design of the constructs, purification of the complex, assessment of the complex's purity and homogeneity using EM through negative staining analysis and obtaining the structure via cryo-EM.

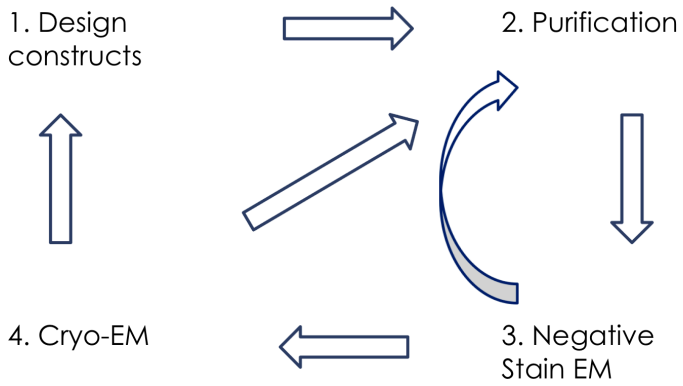


Figure 3-2: Experimental design. Step 1: design the constructs (obtain the complex); step 2: purification (purify the complex); step 3: negative stain EM (to ensure sample homogeneity and purity); step 4: cryo-EM (obtain the structure of the complex).

3.2 Ubiquitination of FANCD2/FANCI and FANCI Alone is DNA-Dependent

To enable structural analysis of the monoubiquitinated intermediates, it is important to ensure that maximal ubiquitination occurs so as to reduce non-ubiquitinated impurities. Previously there has been a discrepancy surrounding the monoubiquitination event regarding whether the event occurs on or off DNA in the cell (Garcia-Higuera et al., 2001; Matsushita et al., 2005; Sareen et al., 2012; Sato et al., 2012b; Smogorzewska et al., 2007). Whether the complex indeed has to be recruited to the site of damage before or after monoubiquitination has a profound impact on the mechanism of the FA pathway. Recently there has been mounting evidence *in vivo* that the complex is recruited to DNA prior to ubiquitination, as deletion of DNA binding residues on FANCD2 (a Tower domain truncation, FANCD2ΔTower) and phosphorylation that inhibits DNA binding of FANCD2 has been shown to block ubiquitination (Liang et al., 2016; Lopez-Martinez et al., 2019). Despite the discrepancies *in vivo*, it has been demonstrated *in vitro* that DNA increases the efficiency of ubiquitination (Chaugule et al., 2020; Liang et al., 2016; Liang et al., 2019; Longerich et al., 2009; Niraj, et al., 2017; Sareen et al., 2012; Sato et al., 2012b; van Twest et al., 2017). Therefore, the ubiquitination assay was conducted with varying DNA concentrations to determine maximum ubiquitination efficiency.

To establish the maximal ubiquitination efficiency for monoubiquitinated-FANCD2/FANCI, the ubiquitination assay was prepared by incubating in reaction buffer the appropriate concentrations of UBA1, UBE2T, FANCL, FANCD2/FANCI, ubiquitin, ATP and between 0.1 – 2 μg of scDNA at room temperature. Samples were taken at 3 hours and 6 hours and analysed by SDS-PAGE. The gel bands corresponding to the 6 hour time points were quantified using Image J (Figure 3-3). It is clear that increasing scDNA concentrations increases the efficiency of monoubiquitination of FANCD2 in complex with FANCI *in vitro*. Therefore, for further structural studies of monoubiquitinated-FANCD2/FANCI, 2 μg of scDNA is used in ubiquitination assays.

To determine the maximal ubiquitination efficiency for monoubiquitinated-FANCI, the ubiquitination assay was prepared by incubating in reaction buffer the appropriate concentrations of UBA1, UBE2T, FANCL, FANCI, ubiquitin, ATP and between 0.1 – 2 μg of scDNA at room temperature. Samples were taken at 1 hour and 3 hours and analysed by SDS-PAGE. The gel bands corresponding to the 3 hour time points were quantified using Image J (Figure 3-4). It is clear that 1 μg is the optimal scDNA concentration that results in the maximal efficiency of monoubiquitination of FANCI *in vitro*. Therefore, for further structural studies, 1 μg of scDNA is used in ubiquitination assays.

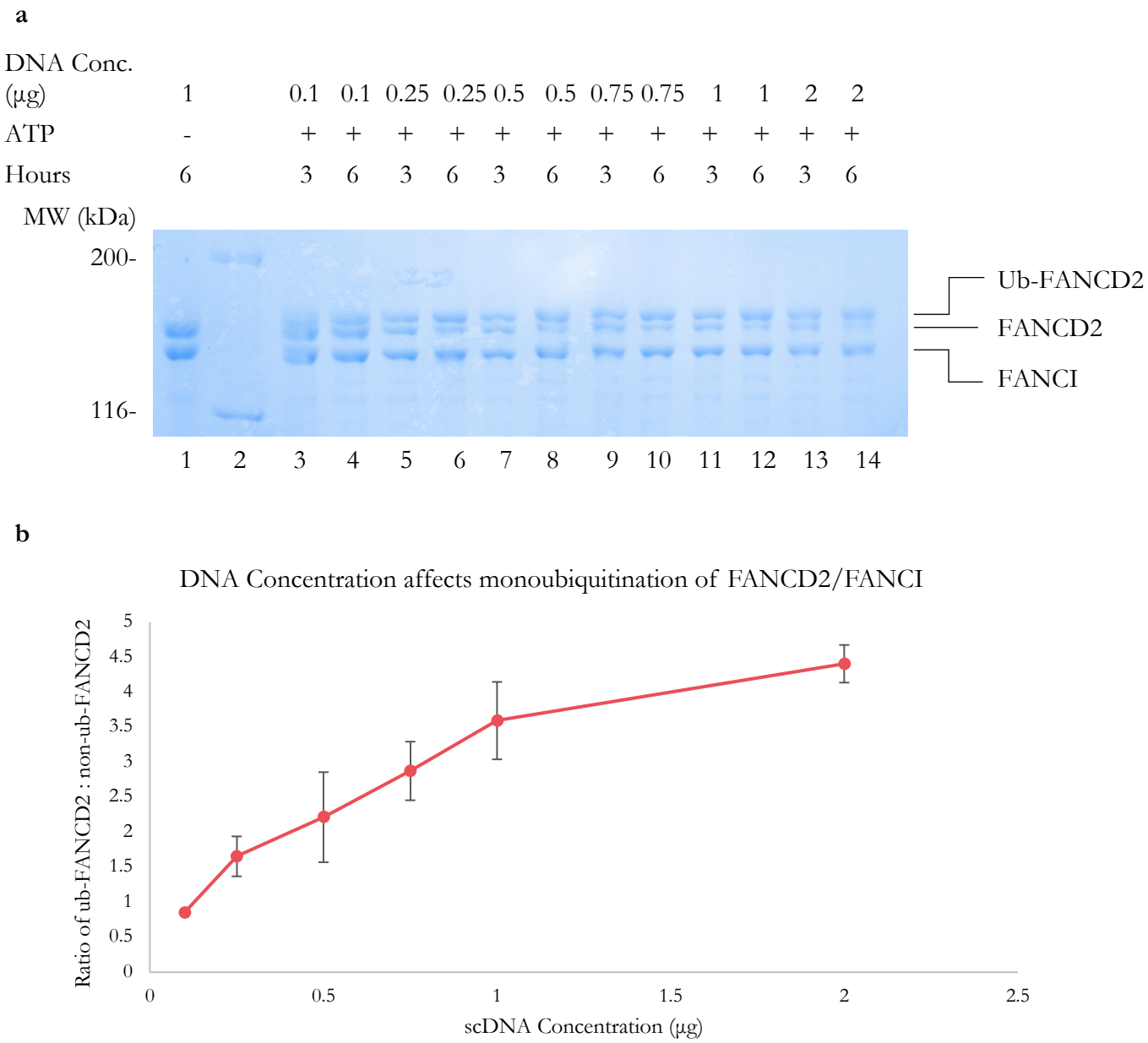
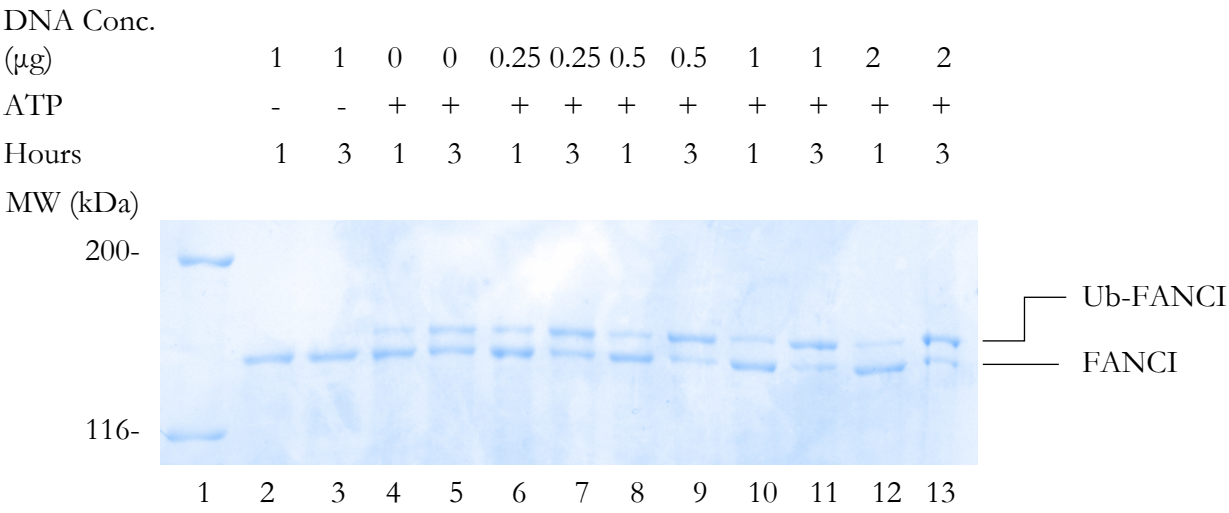


Figure 3-3: Ubiquitination of FANCD2/FANCI depends on DNA. a) Representative Coomassie blue stain of monoubiquitination of FANCD2/FANCI with increasing DNA concentration. Lane 1: negative control (no ATP), lane 2: marker, lane 3: 3 hours post ubiquitination with 0.10 µg scDNA, lane 4: 6 hours post ubiquitination with 0.10 µg scDNA, lane 5: 3 hours post ubiquitination with 0.25 µg scDNA, lane 6: 6 hours post ubiquitination with 0.25 µg scDNA, lane 7: 3 hours post ubiquitination with 0.50 µg scDNA, lane 8: 6 hours post ubiquitination with 0.50 µg scDNA, lane 9: 3 hours post ubiquitination with 0.75 µg scDNA, lane 10: 6 hours post ubiquitination with 0.75 µg scDNA, lane 11: 3 hours post ubiquitination with 1.0 µg scDNA, lane 12: 6 hours post ubiquitination with 1.0 µg scDNA, lane 13: 3 hours post ubiquitination with 2.0 µg scDNA, lane 14: 6 hours post ubiquitination with 2.0 µg scDNA, b) Quantification of monoubiquitination of FANCD2/FANCI with increasing DNA concentration at time = 6 hours (Mean±SD, n=3).

a



b

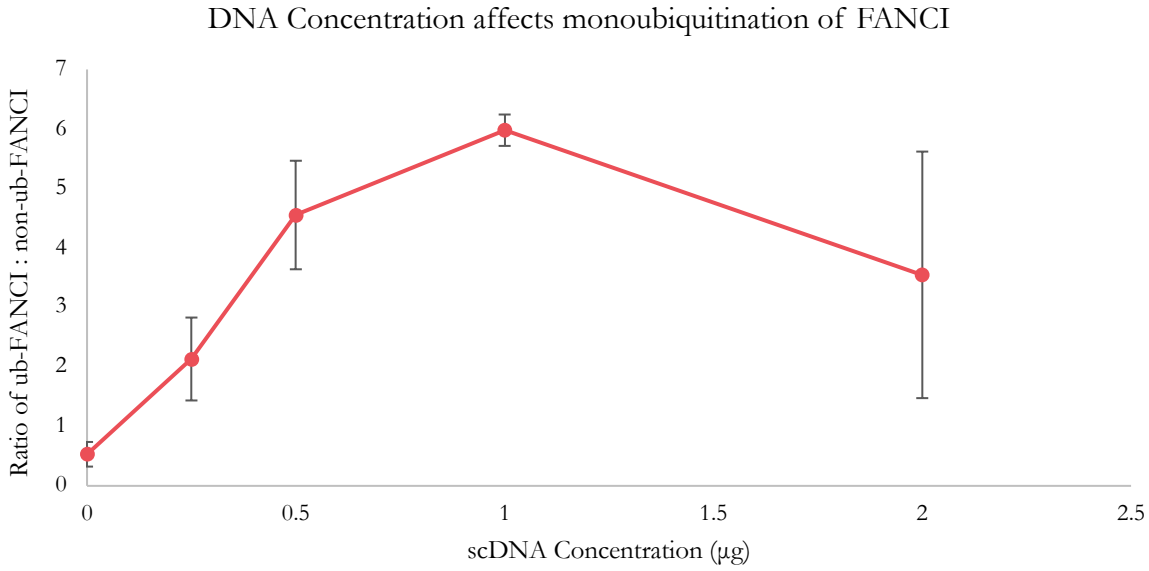


Figure 3-4: Ubiquitination of FANCI depends on DNA. a) Representative Coomassie blue stain of monoubiquitination of FANCI with increasing DNA concentration. Lane 1: marker, lane 2: 1 hour negative control (no ATP), lane 3: 3 hours negative control (no ATP), lane 4: 1 hour post ubiquitination with 0.0 µg scDNA, lane 5: 3 hours post ubiquitination with 0.0 µg scDNA, lane 6: 1 hour post ubiquitination with 0.25 µg scDNA, lane 7: 3 hours post ubiquitination with 0.25 µg scDNA, lane 8: 1 hour post ubiquitination with 0.50 µg scDNA, lane 9: 3 hours post ubiquitination with 0.50 µg scDNA, lane 10: 1 hour post ubiquitination with 1.0 µg scDNA, lane 11: 3 hours post ubiquitination with 1.0 µg scDNA, lane 12: 1 hour post ubiquitination with 2.0 µg scDNA, lane 13: 3 hours post ubiquitination with 2.0 µg scDNA, b) Quantification of monoubiquitination of FANCI with increasing DNA concentration at time = 3 hours (Mean±SD, n=3).

3.3 Deubiquitination of Monoubiquitinated-FANCD2/FANCI and Monoubiquitinated-FANCI Alone is DNA-Independent

In a similar manner to the monoubiquitination event in the FA pathway, there is also a discrepancy in the FA field as to whether deubiquitination of the monubiquitinated FANCD2/FANCI complex occurs on or off DNA (Liang et al., 2019; van Twest et al., 2017). Therefore, to gain insight into this event in the pathway, deubiquitination assays were conducted for the monoubiquitinated-FANCD2/FANCI complex and monoubiquitinated-FANCI.

Monoubiquitinated-FANCD2/FANCI and monoubiquitinated-FANCI were prepared as described above. Following the ubiquitination assays, the ubiquitination was quenched by adding EDTA. Each reaction was then split into two samples, and either benzonase or MilliQ water added and incubated at room temperature for one hour. FLAG-HA tagged USP1 and Strep-tagged UAF1 were co-expressed in Sf9 insect cells, using limiting FLAG-HA tagged USP1, and purified via flag affinity purification (Figure 3-5). USP1/UAF1 was then added to each reaction mixture and incubated at room temperature. Samples were taken at 0, 5, 30 and 60 minute intervals and analysed by SDS-PAGE. The gel bands were quantified using Image J (Figure 3-6, Figure 3-7). It is clear that the deubiquitination of monoubiquitinated-FANCD2/FANCI and the deubiquitination of monoubiquitinated-FANCI is independent of DNA *in vitro*, with both being efficiently deubiquitinated regardless of the presence or absence of DNA. It can be seen however, that the kinetics of deubiquitination for the monoubiquitinated-FANCD2/FANCI are slightly faster when the complex is on DNA in comparison to off DNA.

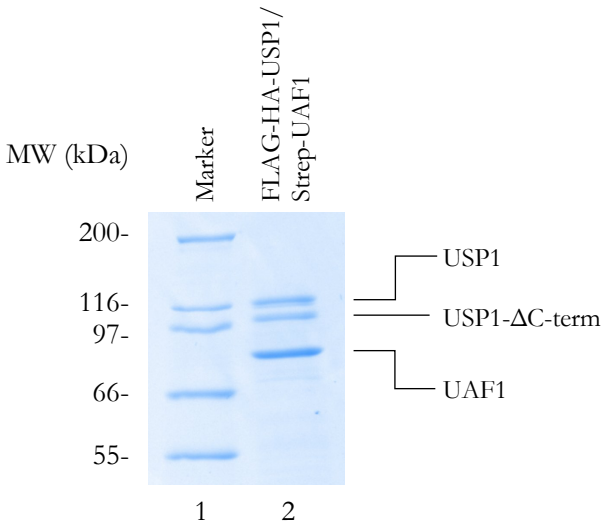
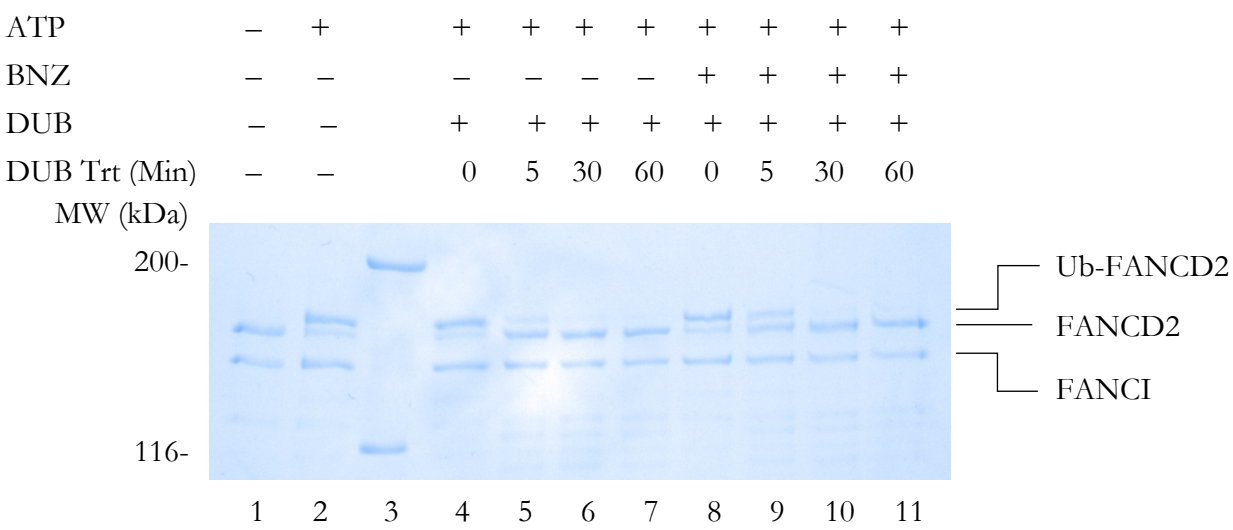


Figure 3-5: Purification of USP1/UAF1 complex. Coomassie blue stain of the purification of USP1/UAF1 complex. Lane 1: marker, lane 2: FLAG-HA-USP1/Strep-UAF1.

a



b

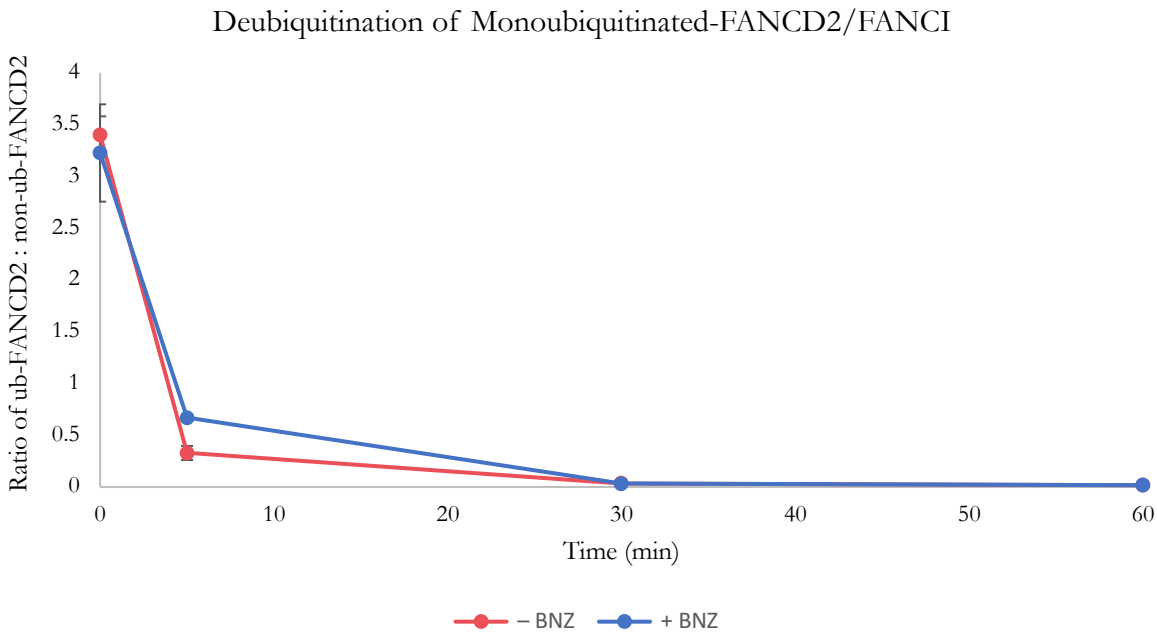
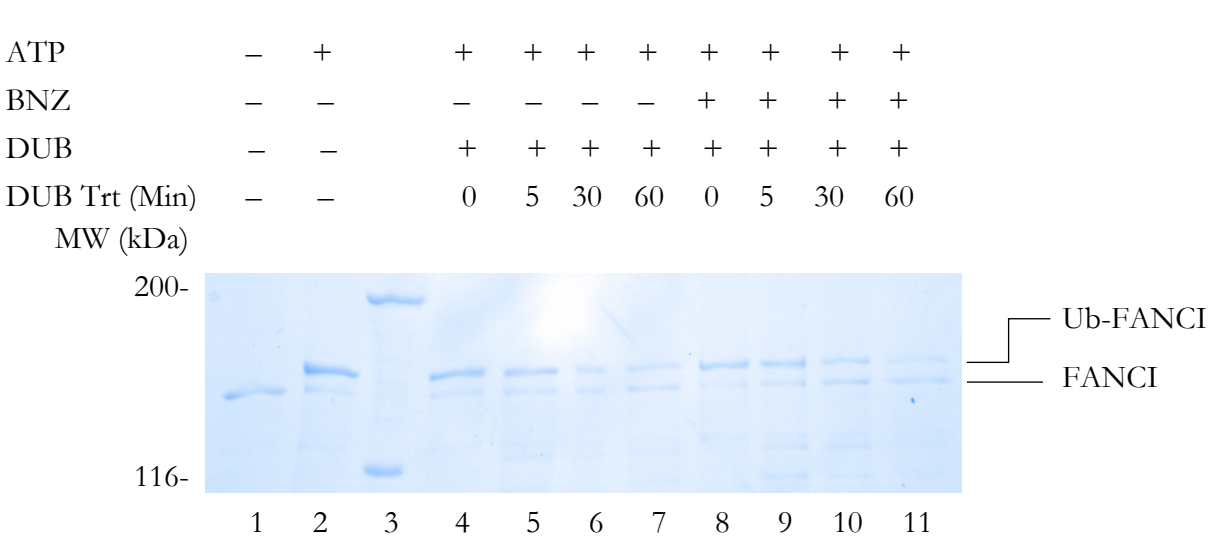


Figure 3-6: Deubiquitination of monoubiquitinated-FANCD2/FANCI is not greatly enhanced by DNA. a) Representative Coomassie blue stain of deubiquitination of monoubiquitinated-FANCD2/FANCI over time. Lane 1: 3 hours negative control (no ATP), lane 2: 3 hours post ubiquitination with 0.0 nM USP1/UAF1, lane 3: marker, lanes 4 – 11: deubiquitination of monoubiquitinated-FANCD2/FANCI after addition of 192 nM USP1/UAF1; lanes 4-7: no benzonase; lanes 8-11: benzonase added; lanes 4 and 8: 0 minutes after DUB addition (DUB Trt), lanes 5 and 9: 5 minutes after DUB addition, lanes 6 and 10: 30 minutes after DUB addition, lanes 7 and 11: 60 minutes after DUB addition. b) Quantification of deubiquitination of monoubiquitinated-FANCD2/FANCI over time (Mean±SD, n=3).

a



b

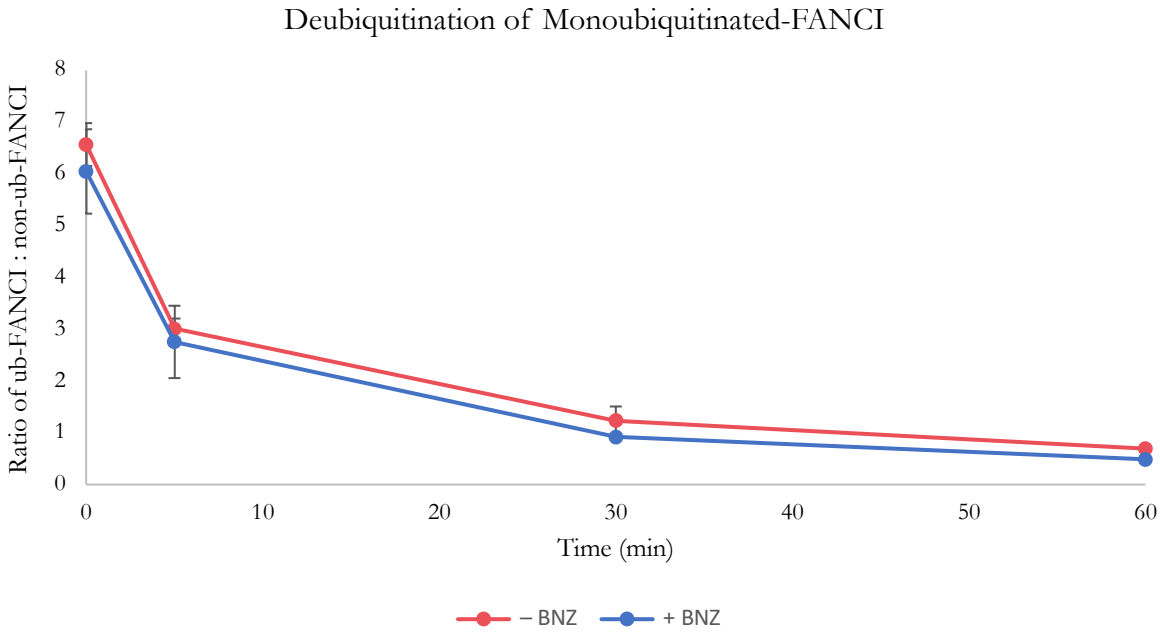


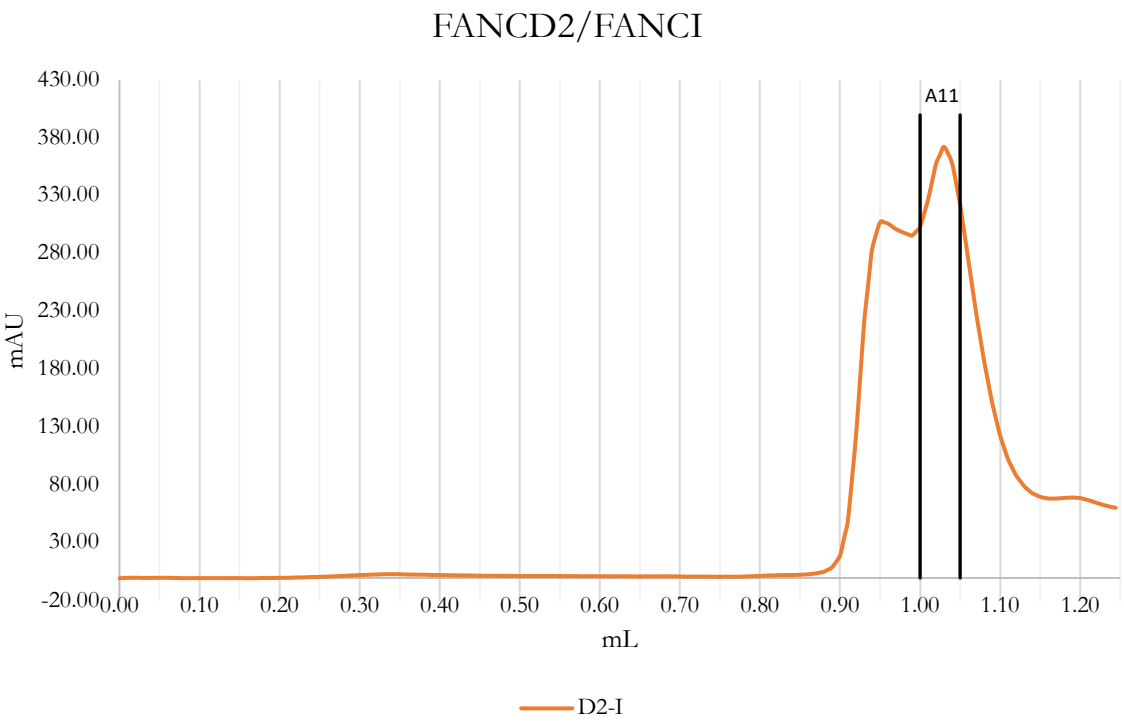
Figure 3-7: Deubiquitination of monoubiquitinated-FANCI is not enhanced by DNA. a) Representative Coomassie blue stain of deubiquitination of monoubiquitinated-FANCI over time. Lane 1: 3 hours negative control (no ATP), lane 2: 3 hours post ubiquitination with 0.0 nM USP1/UAF1, lane 3: marker, lanes 4 – 11: deubiquitination of monoubiquitinated-FANCI after addition of 192 nM USP1/UAF1; lanes 4-7: no benzonase; lanes 8-11: benzonase added; lanes 4 and 8: 0 minutes after DUB addition (DUB Trt), lanes 5 and 9: 5 minutes after DUB addition, lanes 6 and 10: 30 minutes after DUB addition, lanes 7 and 11: 60 minutes after DUB addition. b) Quantification of deubiquitination of monoubiquitinated-FANCI over time (Mean±SD, n=3).

3.4 Towards an Improved Cryo-EM Structure of FANCD2/FANCI

To be able to effectively draw conclusions from any structural changes resulting from the monoubiquitination of FANCD2/FANCI, a higher resolution structure of the precursor, non-ubiquitinated FANCD2/FANCI, also needs to be produced. Therefore, FLAG-HA tagged FANCD2 and 6xHis-tagged FANCI were co-expressed in Sf9 insect cells, using limiting FLAG-HA tagged FANCD2, and purified via flag affinity purification to a one-to-one ratio. The FANCD2/FANCI complex was purified using size exclusion chromatography (Figure 3-8). The resulting pure complex was applied to carbon coated grids, which were then negatively stained, and subjected to microscopy to ensure good sample homogeneity and purity. A 2D classification of the FANCD2/FANCI complex and a 3D model were generated using EMAN2.2 software (Figure 3-9). This model looks similar to the previously solved cryo-EM structure of FANCD2/FANCI, albeit without a visible Tower domain (Figure 3-29b).

As the resulting negative staining micrographs, 2D classification and 3D model look promising, the sample was prepared for cryo-EM. The sample was prepared as per the negative staining sample above, except the complex was concentrated prior to the final purification on size exclusion chromatography due to the need for a higher concentration for cryo-EM. As a first step, it was necessary to master grid preparation and handling as well as blotting conditions for the sample. In particular, it was necessary to be able to move the grid from liquid ethane into liquid nitrogen and slot the grid into the grid box. This is a delicate procedure that if done incorrectly can puncture holes in the grid squares (Figure 3-10a). Once this grid handling was mastered, it was necessary to optimise the grid preparation and blotting conditions for the sample. Initially, the sample pooled in the centre of the grid squares (Figure 3-10b). This phenomenon was solved by incorporating a "wait" time into the blotting procedure, to allow the sample to evenly distribute across the grid. In addition, the level of blot force was optimised, so that there was a gradient of blotting across the grid. When the sample is too thick and the blot force is too weak, the grid squares cannot be easily visualised across the whole grid (Figure 3-10c). The gradient across the grid was optimised to allow for maximum grid squares at the appropriate sample thickness (Figure 3-10d).

a



b

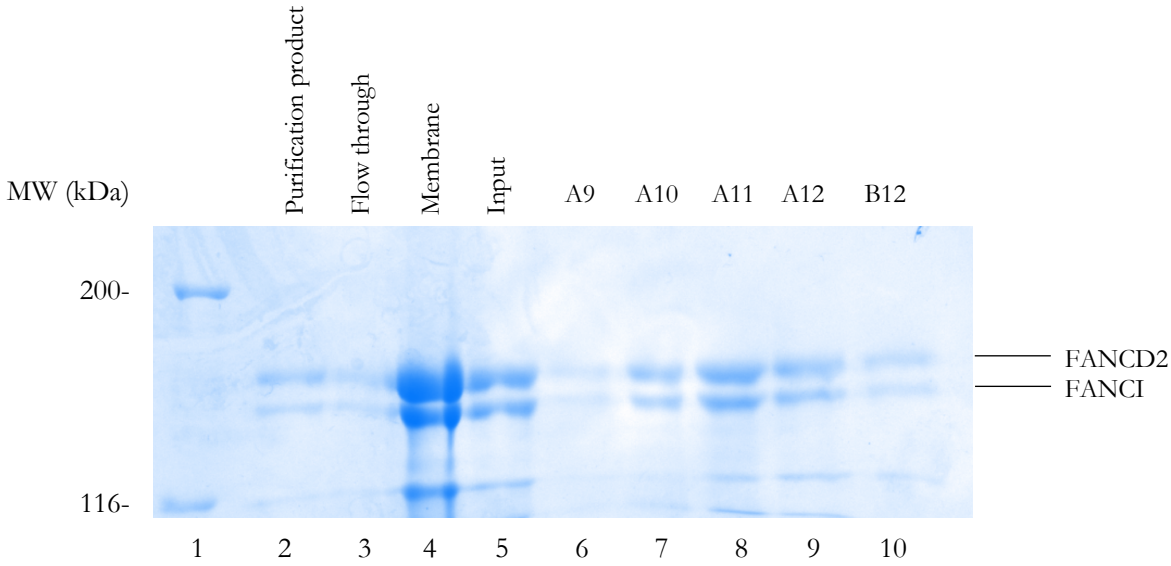
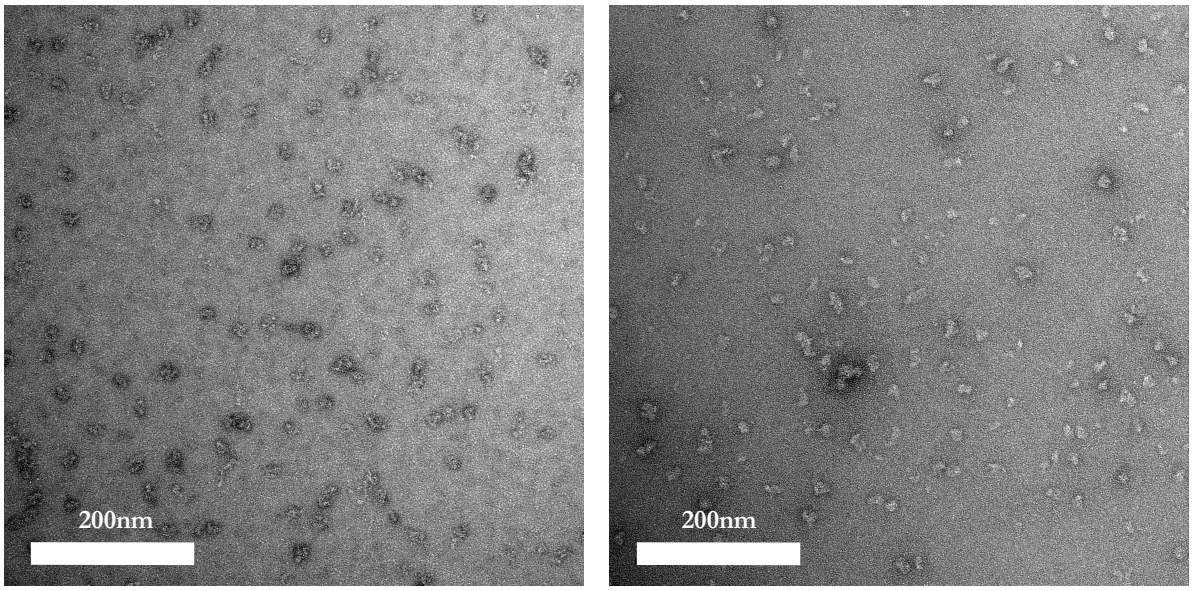
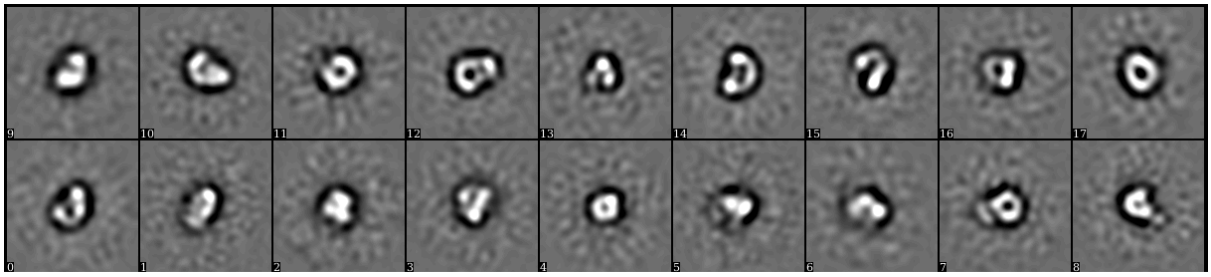


Figure 3-8: Purification of FANCD2/FANCI. a) Chromatogram following purification on a Superdex 200 3.2/300. FANCD2/FANCI identified in peak fraction A11. b) Coomassie blue stain of the purification procedure showing FANCD2/FANCI. Lane 1: marker, lane 2: crude purification product, lane 3: flow through from 10 kDa cut-off concentrator, lane 4: sample stuck on the membrane, lane 5: sample loaded onto the gel filtration column, lanes 6 – 10: gel filtration peak fractions (lane 6: A9 peak fraction, lane 7: A10 peak fraction, lane 8: A11 peak fraction, lane 9: A12 peak fraction, lane 10: B12 peak fraction).

a



b



c

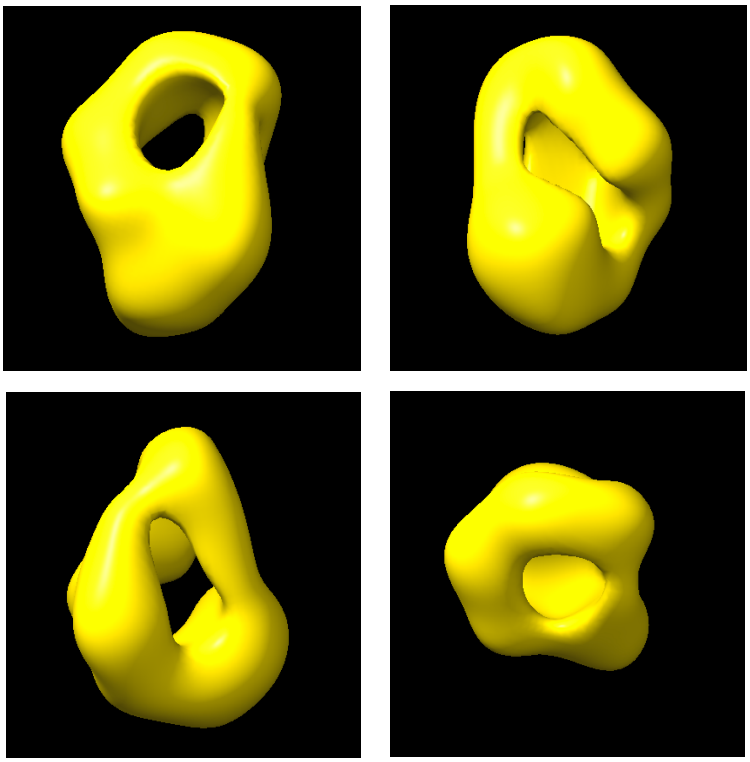
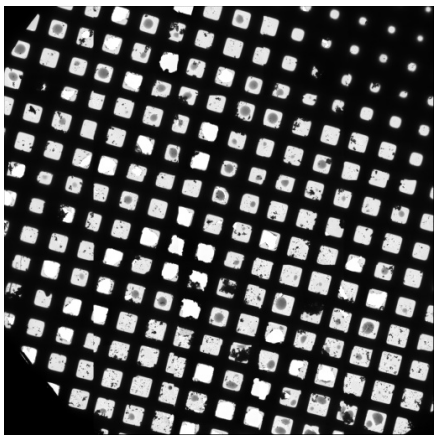
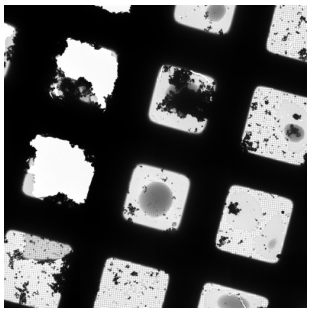
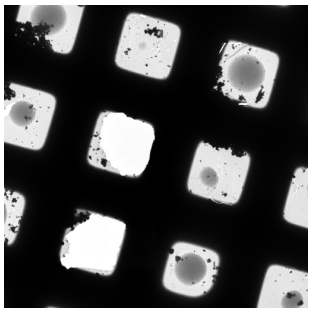


Figure 3-9: Negative staining of FANCD2/FANCI at 0.010 µg/µl. a) Representative micrographs showing the homogeneity of the FANCD2/FANCI sample from the A11 peak fraction from Figure 3-8 at 68 000x magnification, 1.727 Å/pixel. b) 2D classification of FANCD2/FANCI from 7536 particles using EMAN2.2. c) Initial model generated from 2D classification using EMAN2.2.

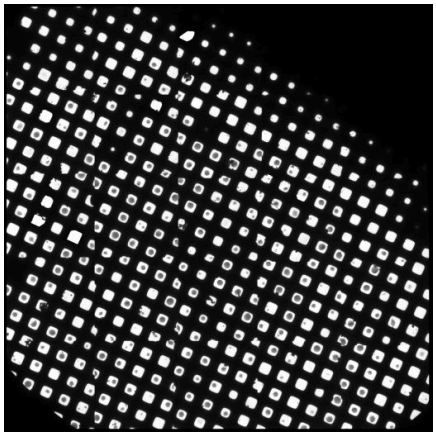
a

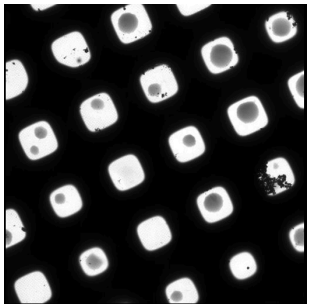


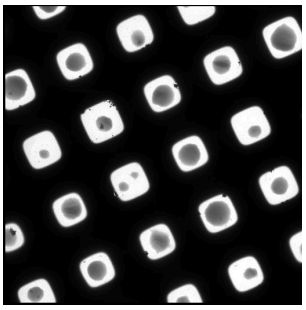




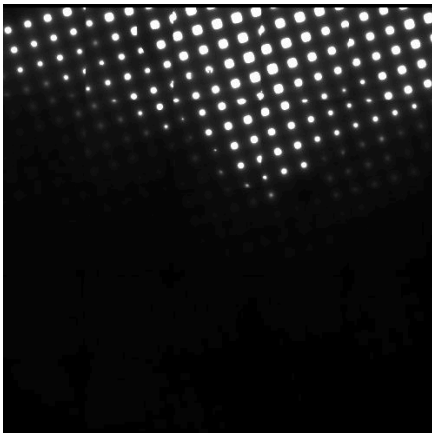
b

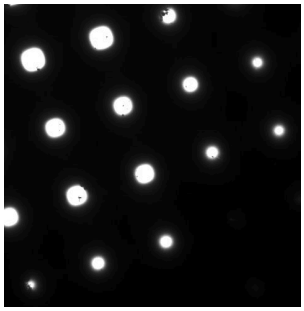


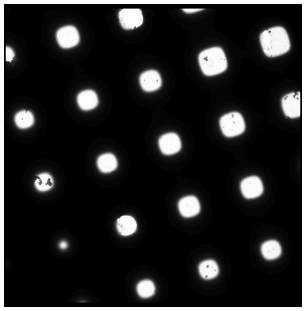




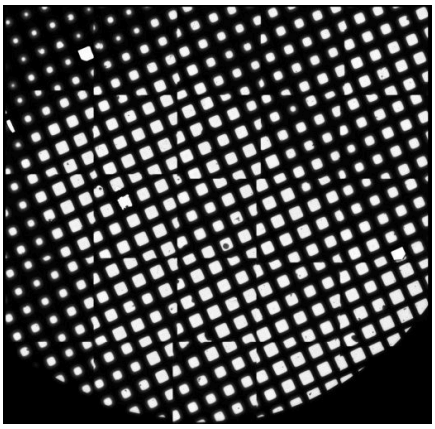
c

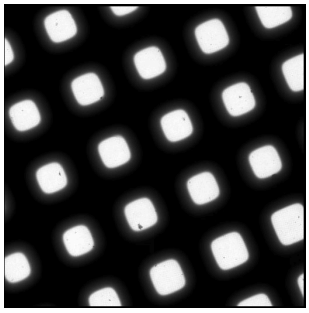






d





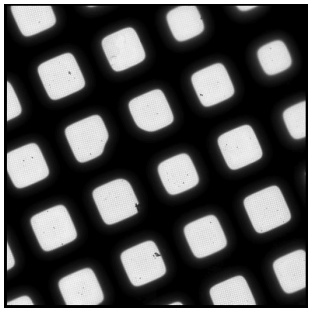
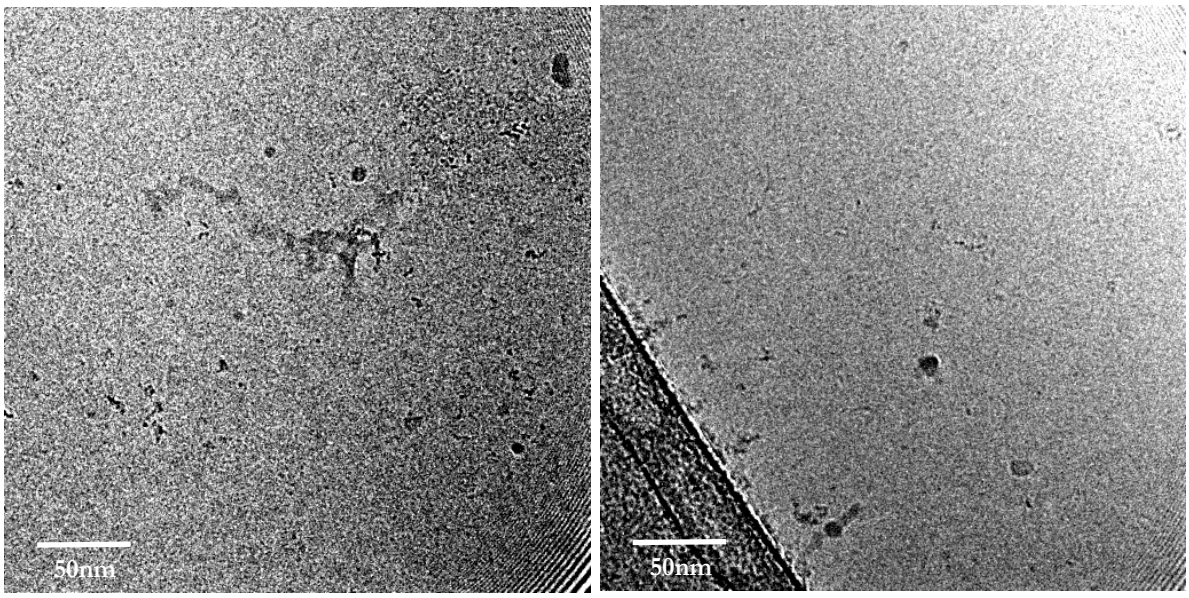


Figure 3-10: Cryo-EM Grid Handling and Blotting Optimisation for Quantifoil 1.2/1.3 300 Mesh Cu. a) Initial grid handling showing puncture holes and ice contamination. b) Optimisation of blotting conditions for sample, showing pooling of the sample at the centre of grid squares. c) Optimisation of blotting conditions showing a sample that is too thick. d) Optimised handling and blotting conditions for grid showing no puncture holes, low ice contamination and a good blotting gradient. Left panel: entire grid; right panel: individual tiles showing close up of grid squares.

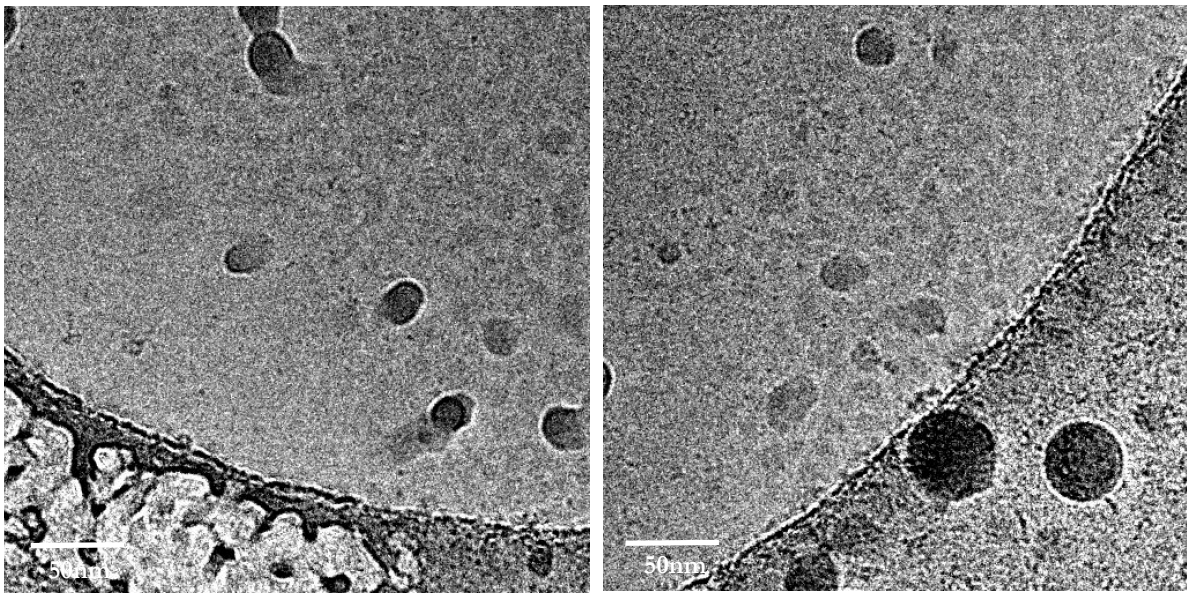
Once the grid handling was mastered and the blotting conditions optimised, a variety of grids were prepared, including Lacey Carbon copper, Quantifoil copper, Quantifoil gold and graphene oxide grids. For Lacey Carbon grids, the protein sample could not really be discerned and seemed to show what appeared to be denatured protein (Figure 3-11a). For the Quantifoil copper grids, the sample seemed to aggregate at the edges of the holes as well as having a high level of ethane contamination (Figure 3-11b). The most promising grids included graphene oxide grids and Quantifoil gold grids. However, for the graphene oxide grids aggregation of the protein can also be seen, as well as ethane contamination (Figure 3-11c). For the Quantifoil gold grids, the protein seemed to aggregate less, however also included some ethane contamination (Figure 3-11d).

On one Quantifoil copper grid, an unusual phenomenon occurred, whereby the FANCD2/FANCI complex seemed to form an array of 2D crystals (Figure 3-12b). The entire grid was imaged, and the 2D crystals seemed to only make up approximately 10% of the grid. Nevertheless, a 2D classification was produced using SIMPLE (Figure 3-12c). Unfortunately, the 2D crystals appeared to have a preferential orientation and thus a 3D model could not be generated.

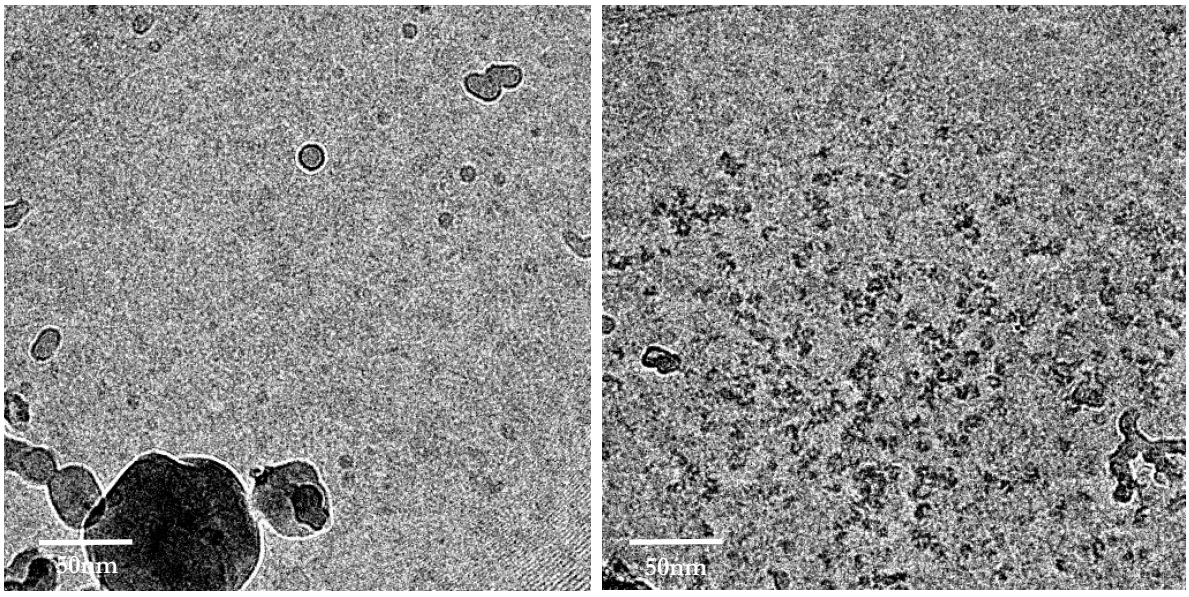
a



b



c



d

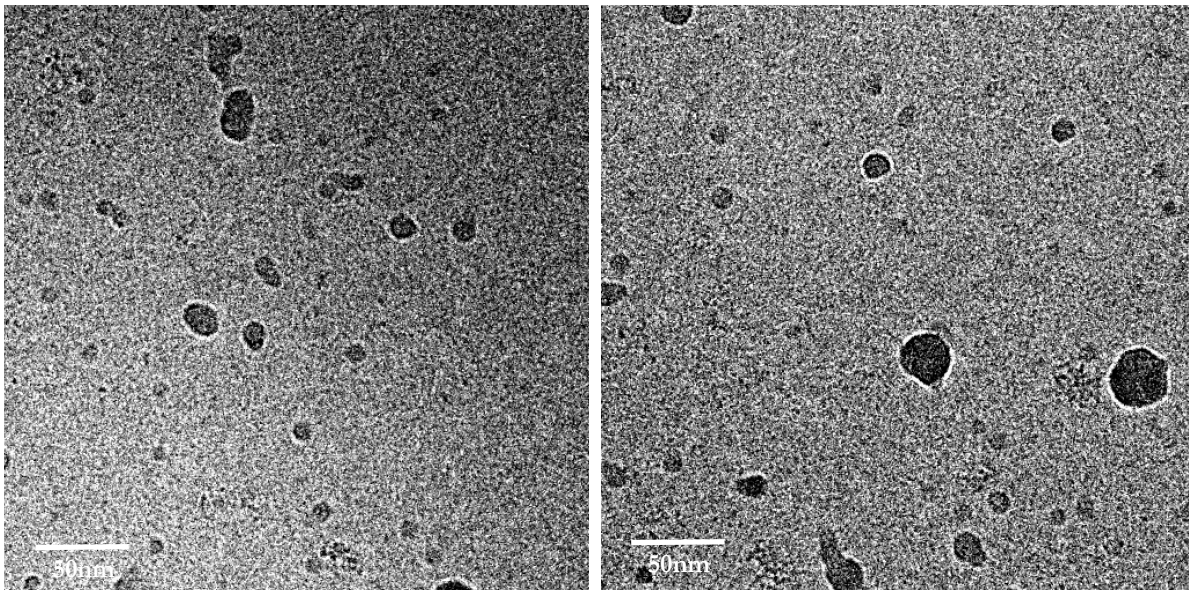
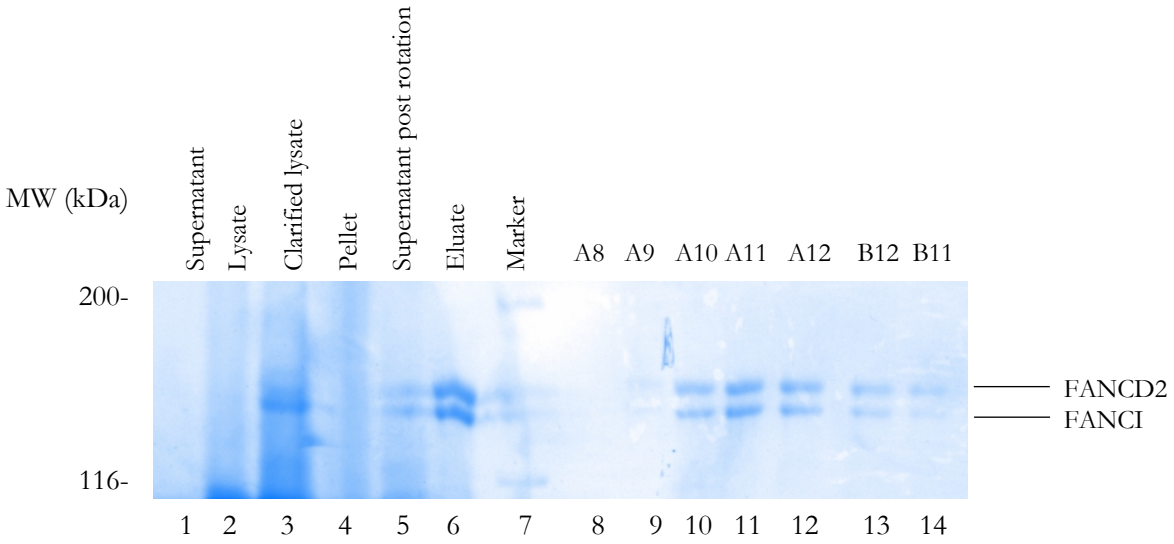
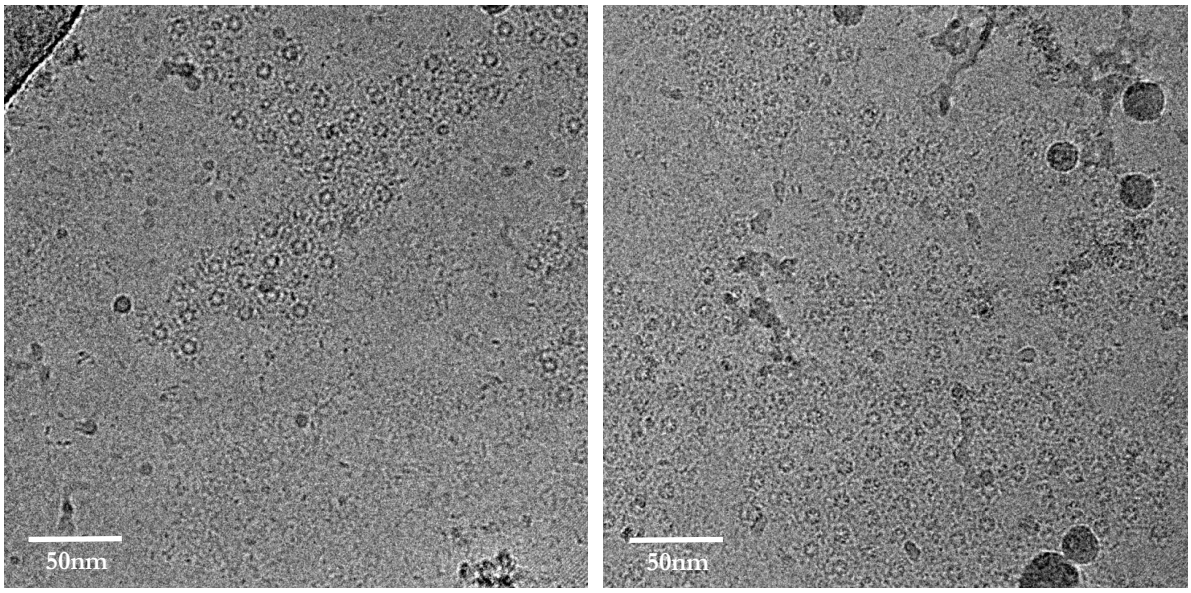


Figure 3-11: Optimisation of cryo-EM conditions for FANCD2/FANCI at 190 000x, 0.78 Å/pixel. a) Representative micrographs of Lacey Carbon Film 300 Mesh Cu grids; protein concentration 0.10 µg/µl. b) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Cu; protein concentration 0.10 µg/µl. c) Representative micrographs of homemade Graphene Oxide Support Film 2/1 300 Mesh Cu; protein concentration 0.30 µg/µl d) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Au; protein concentration 0.30 µg/µl.

a


MW (kDa)
Supernatant
Lysate
Clarified lysate
Pellet
Supernatant post rotation
Eluate
Marker
A8
A9
A10
A11
A12
B12
B11
200-
116-
FANCD2
FANCI
1
2
3
4
5
6
7
8
9
10
11
12
13
14

b


50nm
50nm

c

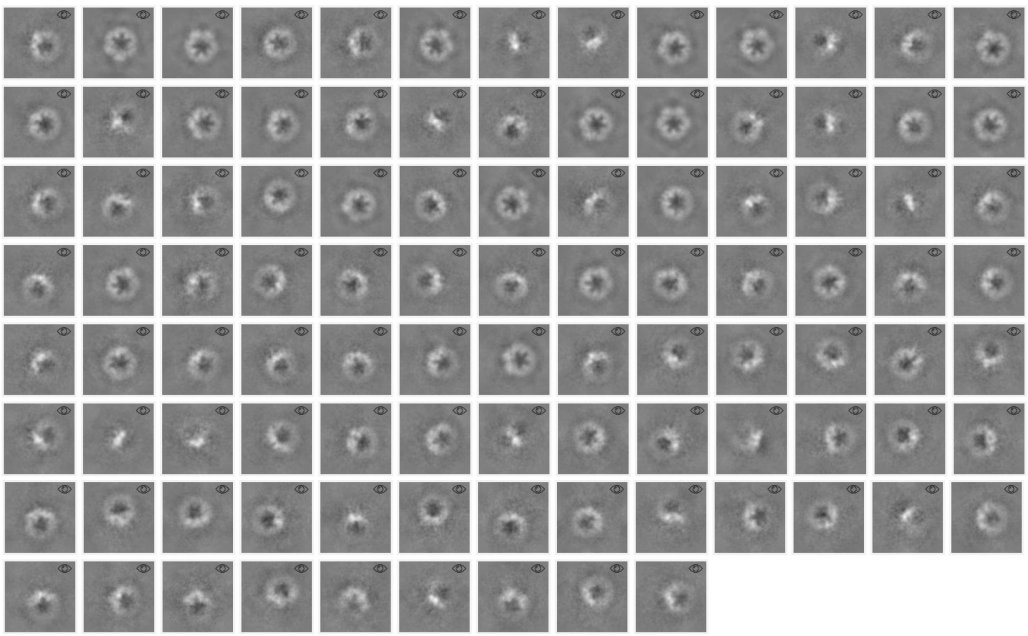
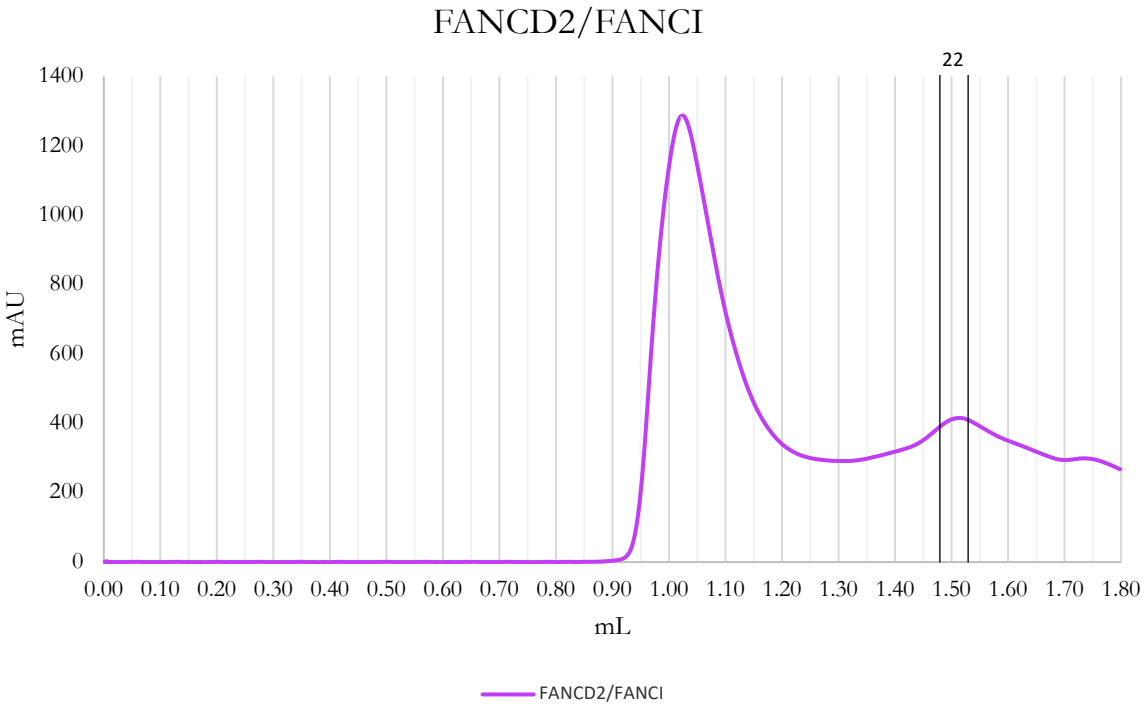


Figure 3-12: Cryo-EM 2D crystals of FANCD2/FANCI at 190 000x, 0.78 Å/pixel. a) Coomassie blue stain of the purification procedure showing FANCD2/FANCI. Lane 1: supernatant, lane 2: cell lysate, lane 3: clarified cell lysate, lane 4: cell pellet, lane 5: supernatant post rotation, lane 6: crude purification product, lane 7: marker, lanes 8 – 14: gel filtration peak fractions (lane 8: A8 peak fraction, lane 9: A9 peak fraction, lane 10: A10 peak fraction, lane 11: A11 peak fraction, lane 12: A12 peak fraction, lane 13: B12 peak fraction, lane 14: B11 peak fraction). b) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Cu; protein concentration 0.10 µg/µl. c) 2D classification of monoubiquitinated-FANCI/FANCD2 from 5770 particles using SIMPLE.

The quality of a sample from size exclusion chromatography increases with the sharpness and separation of the peak. In addition, any molecules that are larger than the beads in the column (the maximum molecular weight resolution) will be eluted in the first peak known as the void volume, at approximately 30% of the total column volume. As the sample applied to the grids above had significant overlap with this void volume peak, and thus was not a symmetrical separate peak, there is a chance that some higher molecular weight contaminations from the void volume were present in the sample peak. However, as can be seen from the fraction elutions, no high molecular weight species can be identified via SDS-PAGE (Figure 3-8b and Figure 3-12a). In any case, the sample was therefore applied to a column with a higher selectivity (molecular weight separation limit) in an attempt to separate the void peak from the sample peak, and to ensure that the sample peak is a symmetrical resolved peak. Therefore, the sample was applied to a Superose 6 3.2/300. As can be seen from Figure 3-13a, the sample was eluted in a separate peak from the void volume centred around elution fraction 22. Figure 3-13b shows that some FANCI aggregates were identified in the void volume (fractions 11-13), whilst the sample can be identified in fractions 20-24.

a



b

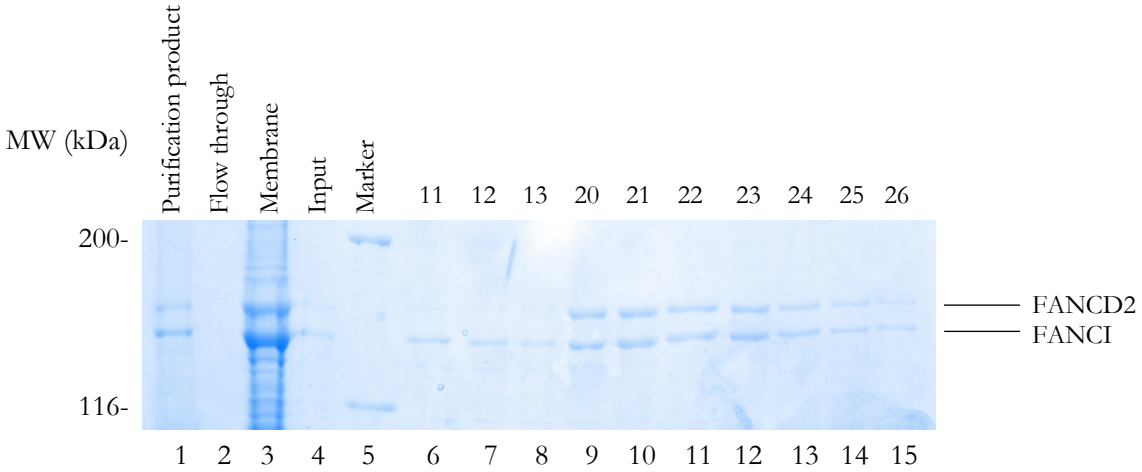


Figure 3-13: Purification of FANCD2/FANCI complex on a Superose 6 3.2/300. a) Chromatogram following purification on a Superose 6 3.2/300. FANCD2/FANCI complex identified in peak fraction 22 (0.26 µg/µl). b) Coomassie blue stain of the purification procedure showing FANCD2/FANCI. Lane 1: crude purification product, lane 2: flow through from 10 kDa cut-off concentrator, lane 3: sample stuck on the membrane, lane 4: sample loaded onto the gel filtration column, lane 5: marker, lanes 6 – 15: gel filtration peak fractions (lane 6: peak fraction 11, lane 7: peak fraction 12, lane 8: peak fraction 13, lane 9: peak fraction 20, lane 10: peak fraction 21, lane 11: peak fraction 22, lane 12: peak fraction 23, lane 13: peak fraction 24, lane 14: peak fraction 25, lane 15: peak fraction 26).

3.5 Towards a Cryo-EM Structure of Monoubiquitinated-FANCD2/FANCI

Monoubiquitinated-FANCD2/FANCI was prepared using the ubiquitination assay as described above. The reaction mixture was then applied to a 1 MDa concentrator, so that the monoubiquitinated-FANCD2/FANCI complex bound to scDNA remained above the membrane, whilst the other ubiquitination reaction components passed through to the flow through. Following concentration, the monoubiquitinated-FANCD2/FANCI was treated with benzonase whilst on the membrane to release the complex from DNA. The complex was then finally purified using size exclusion chromatography (Figure 3-14, Figure 3-15). The resulting pure monoubiquitinated-FANCD2/FANCI complex was applied to carbon coated grids, which were then negatively stained, and subjected to microscopy to ensure good sample homogeneity and purity. A 2D classification of the complex and a 3D model were generated using EMAN2.2 software (Figure 3-16).

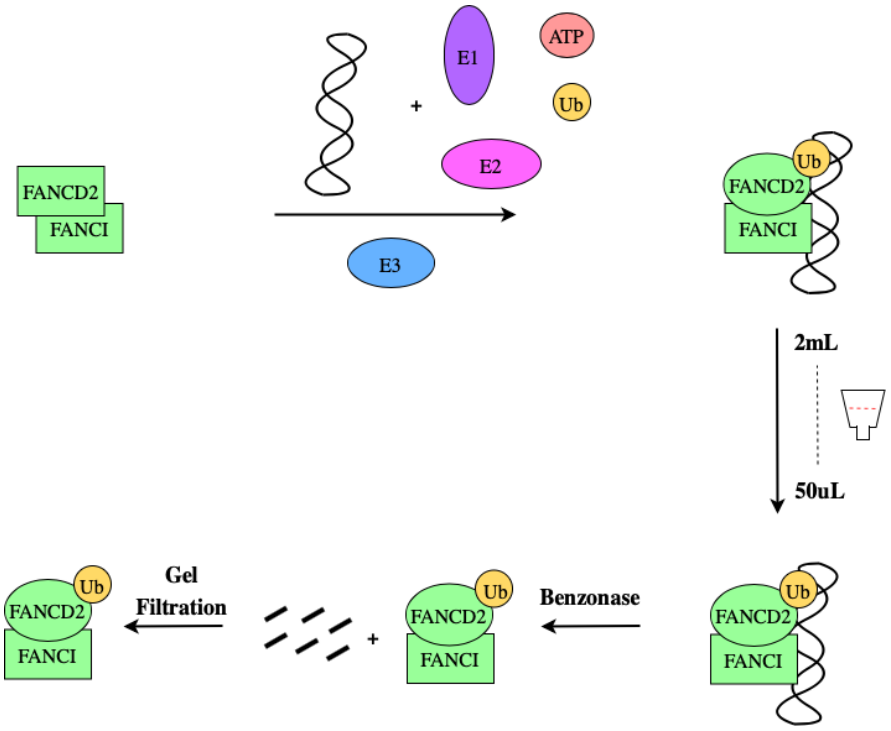


Figure 3-14: Purification schematic of monoubiquitinated-FANCD2/FANCI.

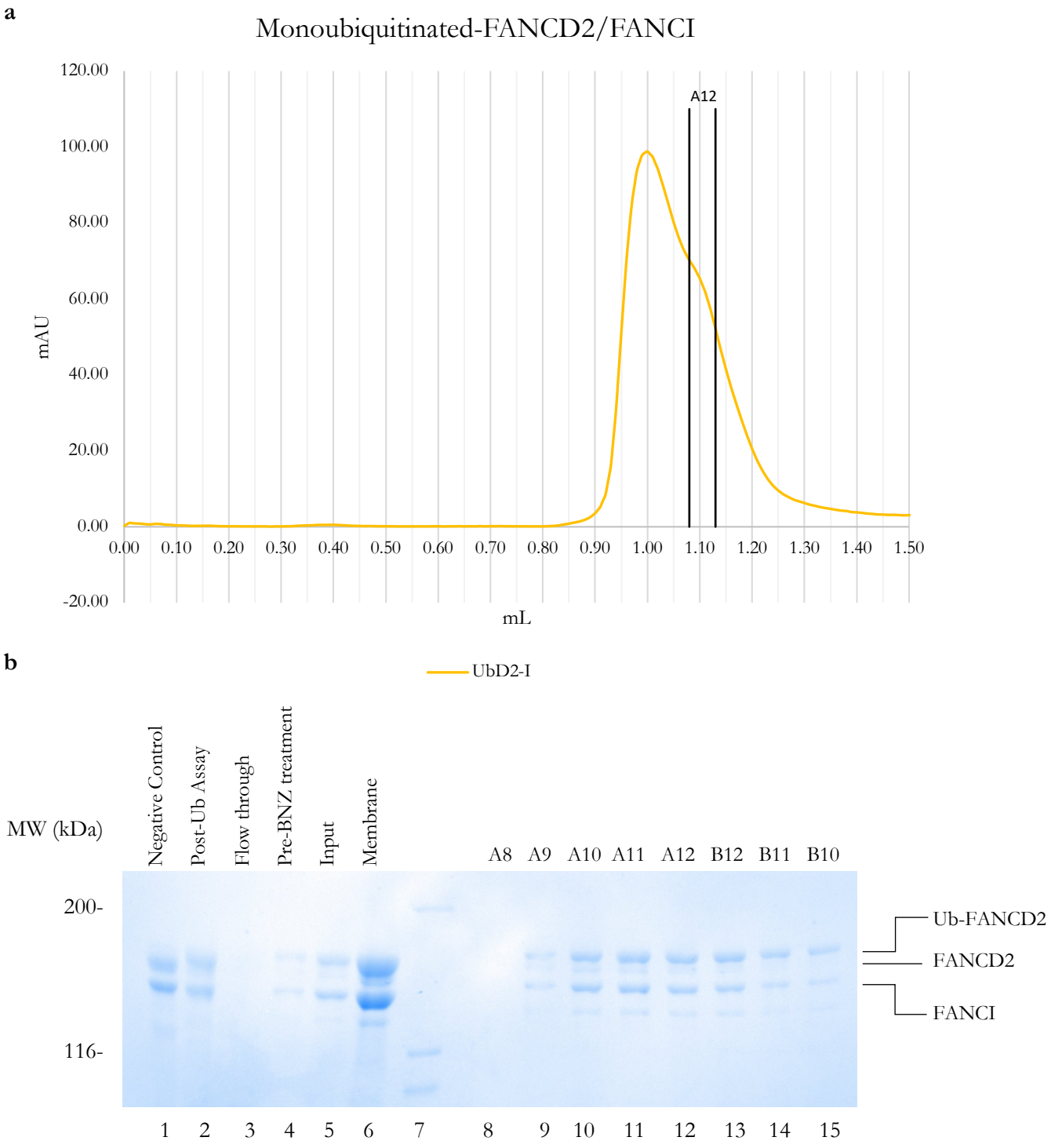
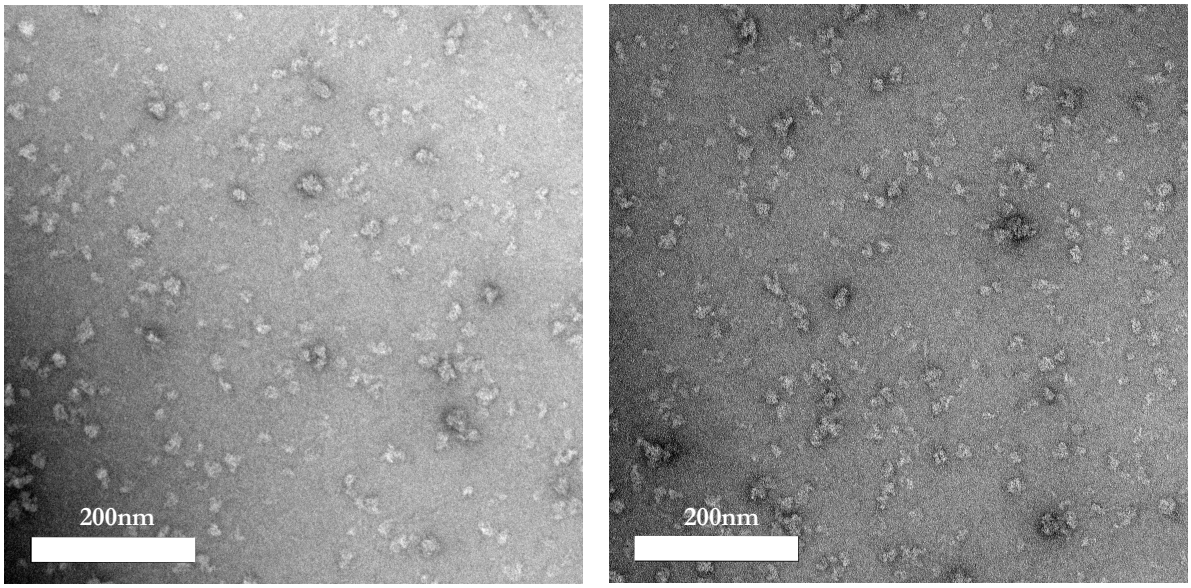


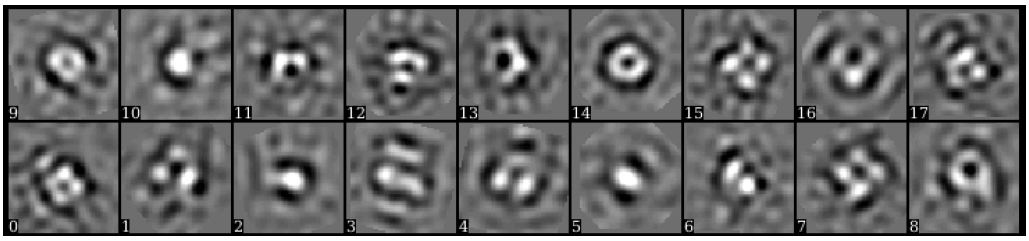
Figure 3-15: Purification of monoubiquitinated-FANCD2/FANCI. a) Chromatogram following purification on a Superdex 200 3.2/300. Monoubiquitinated-FANCD2/FANCI identified in peak fraction A12. b) Coomassie blue stain of the ubiquitination assay and the purification procedure showing monoubiquitinated-FANCD2/FANCI. Lane 1: negative control (no ATP), lane 2: 6 hours post ubiquitination, lane 3: flow through from 1 MDa cut-off concentrator, lane 4: sample before benzonase treatment, lane 5: sample loaded onto the gel filtration column, lane 6: sample stuck on the membrane, lane 7: marker, lanes 8 – 15: gel filtration peak

fractions (lane 8: A8 peak fraction, lane 9: A9 peak fraction, lane 10: A10 peak fraction, lane 11: A11 peak fraction, lane 12: A12 peak fraction, lane 13: B12 peak fraction, lane 14: B11 peak fraction, lane 15: B10 peak fraction).

a



b



c

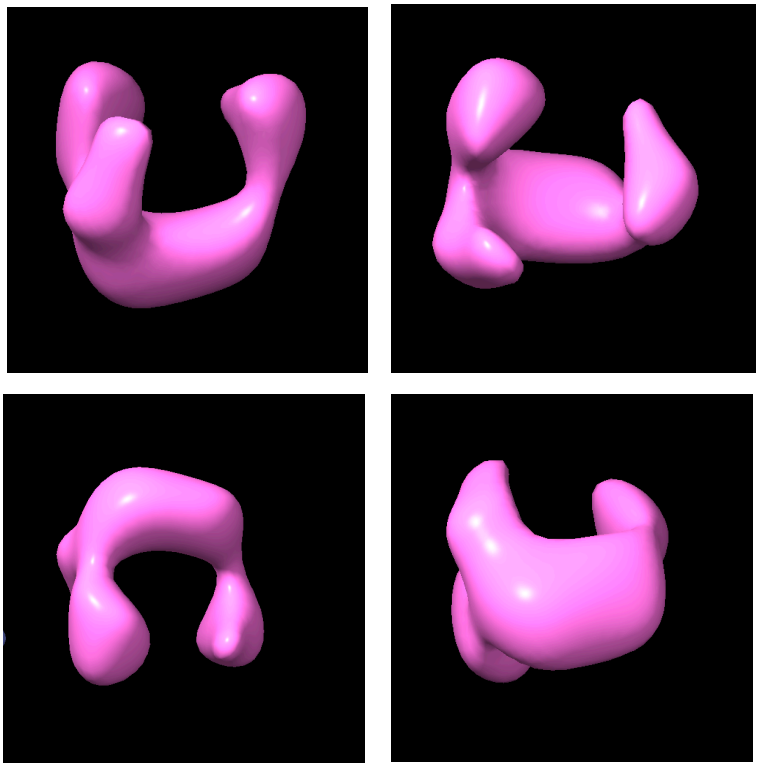
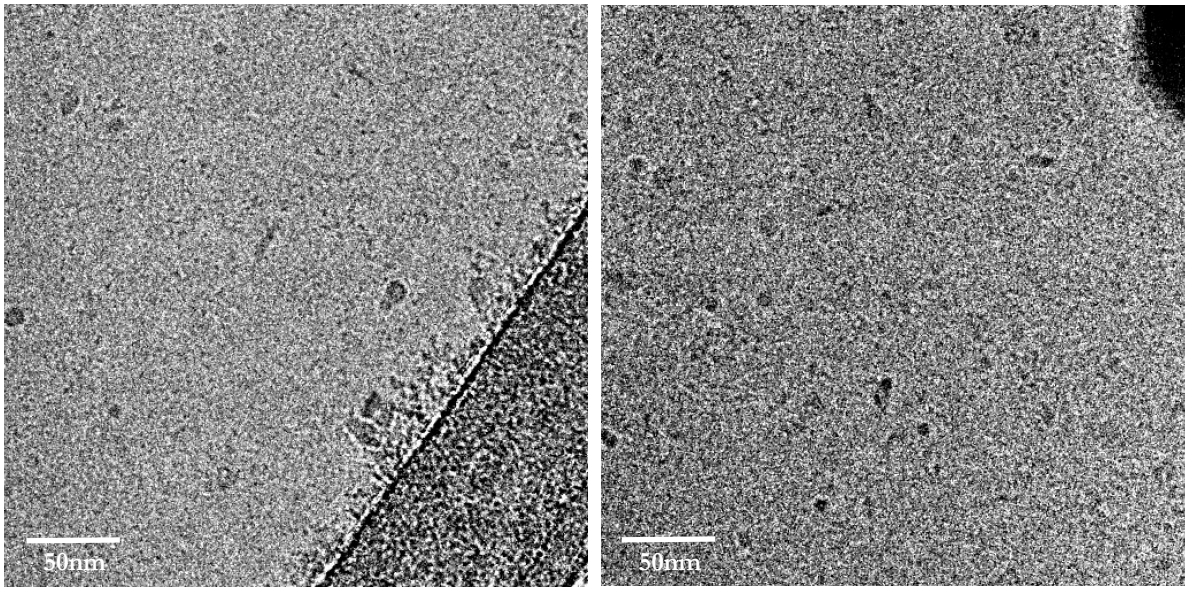


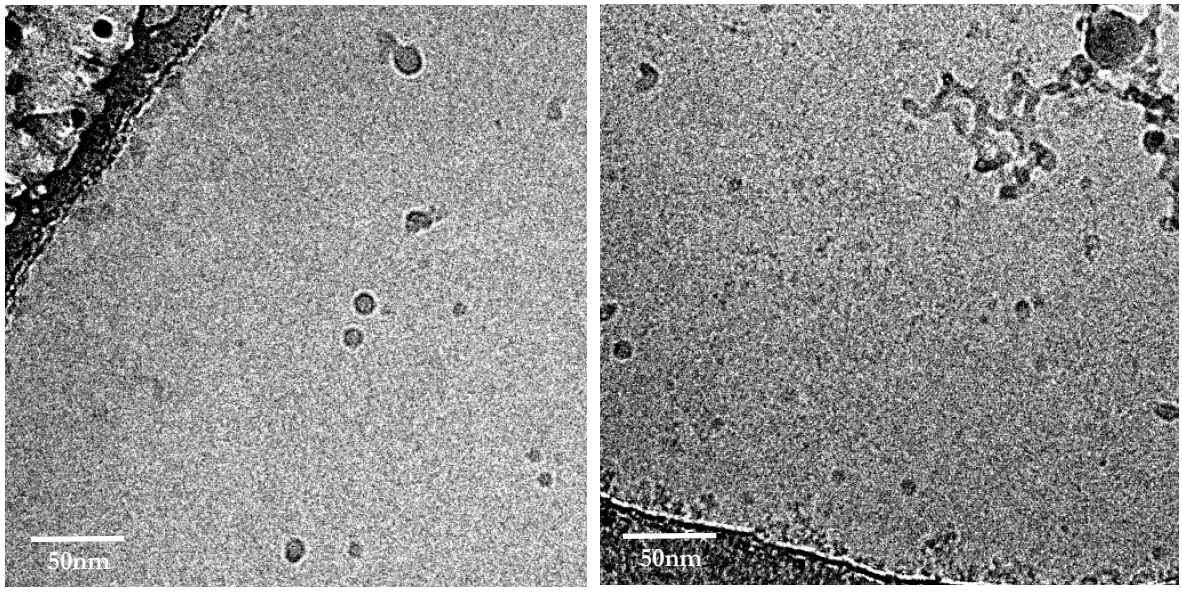
Figure 3-16: Negative staining of monoubiquitinated-FANCD2/FANCI at 0.010 µg/µl. a) Representative micrographs showing the homogeneity of the monoubiquitinated-FANCD2/FANCI sample from the A12 peak fraction from Figure 3-15 at 68 000x magnification, 1.727 Å/pixel. b) 2D classification of monoubiquitinated-FANCD2/FANCI from 2940 particles using EMAN2.2. c) Initial model generated from 2D classification using EMAN2.2.

As the resulting negative staining micrographs show good particle homogeneity and purity, and the 2D classification and 3D model look promising, the sample was prepared for cryo-EM. The sample was prepared as per the negative staining sample above. As with the non-ubiquitinated FANCD2/FANCI sample, a variety of grids were used, including Lacey Carbon copper, Quantifoil copper and graphene oxide grids. For Lacey Carbon grids, the protein sample again seemed to be denatured (Figure 3-17a). For the Quantifoil copper grids, the sample seemed to show significantly less aggregation in comparison to the non-ubiquitinated sample, with only a small amount of protein aggregation at the edges of the holes. The protein particles could be visualised however the particles were very sparse (Figure 3-17b). For the graphene oxide grids, more particles could be seen in comparison to the Quantifoil grids, however there seemed to be signs of aggregation as well as denaturation (Figure 3-17c).

a



b



c

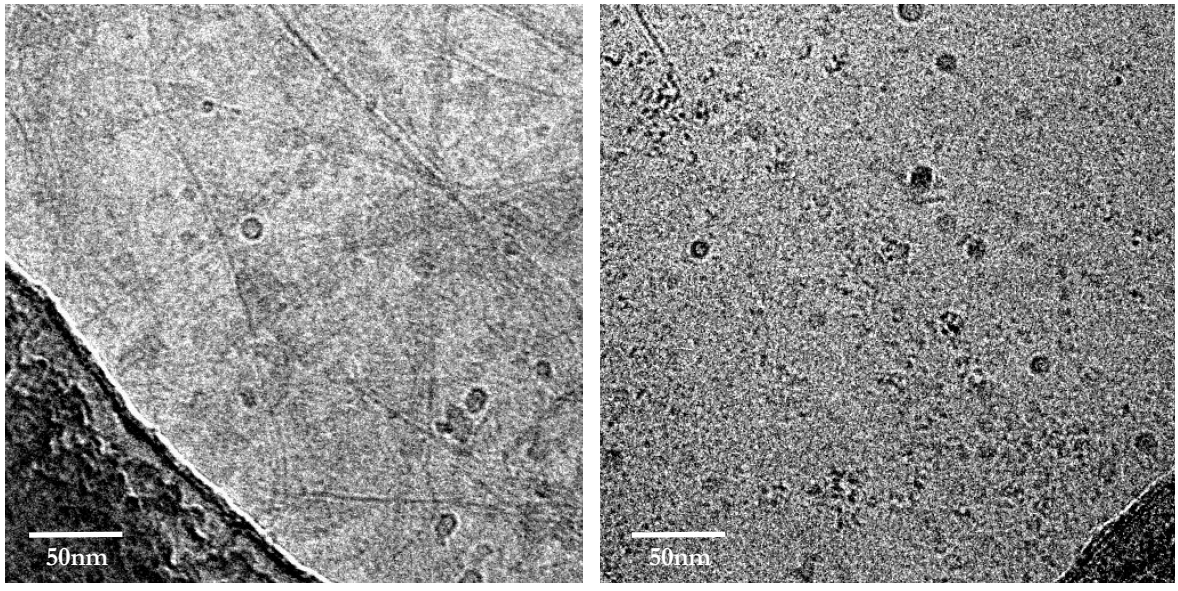


Figure 3-17: Optimisation of cryo-EM conditions for monoubiquitinated-FANCD2/FANCI at 190 000x, 0.78 Å/pixel. a) Representative micrographs of Lacey Carbon Film 300 Mesh Cu grids; protein concentration 0.13 µg/µl. b) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Cu; protein concentration 0.13 µg/µl. c) Representative micrographs of homemade Graphene Oxide Support Film 2/1 300 Mesh Cu; protein concentration 0.13 µg/µl.

3.6 Towards a Cryo-EM Structure of Monoubiquitinated-FANCI/FANCD2

Monoubiquitinated-FANCI was prepared using the ubiquitination assay as described above. The reaction mixture was then dialysed to remove ATP and the other reaction components. FLAG-HA tagged FANCD2 was expressed in Sf9 insect cells and purified via flag affinity purification. The reaction mixture was then incubated with FANCD2 to form the monoubiquitinated-FANCI/FANCD2 complex and applied to a 10kDa concentrator. Following concentration, the monoubiquitinated-FANCI/FANCD2 was treated with benzonase to release the complex from DNA. The complex was then finally purified using size exclusion chromatography (Figure 3-18, Figure 3-19). The resulting pure monoubiquitinated-FANCI/FANCD2 complex was applied to carbon coated grids, which were then negatively stained, and subjected to microscopy to ensure good sample homogeneity and purity. A 2D classification of the complex and a 3D model were generated using EMAN2.2 software (Figure 3-20).

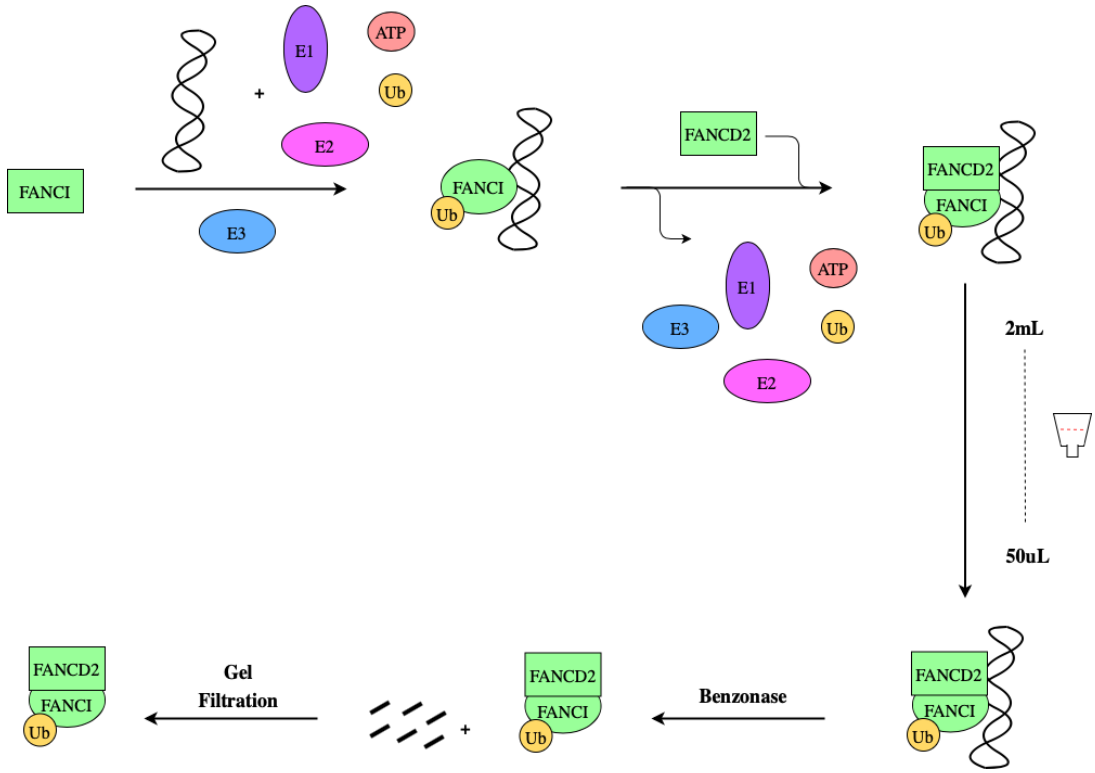


Figure 3-18: Purification schematic of monoubiquitinated-FANCI/FANCD2.

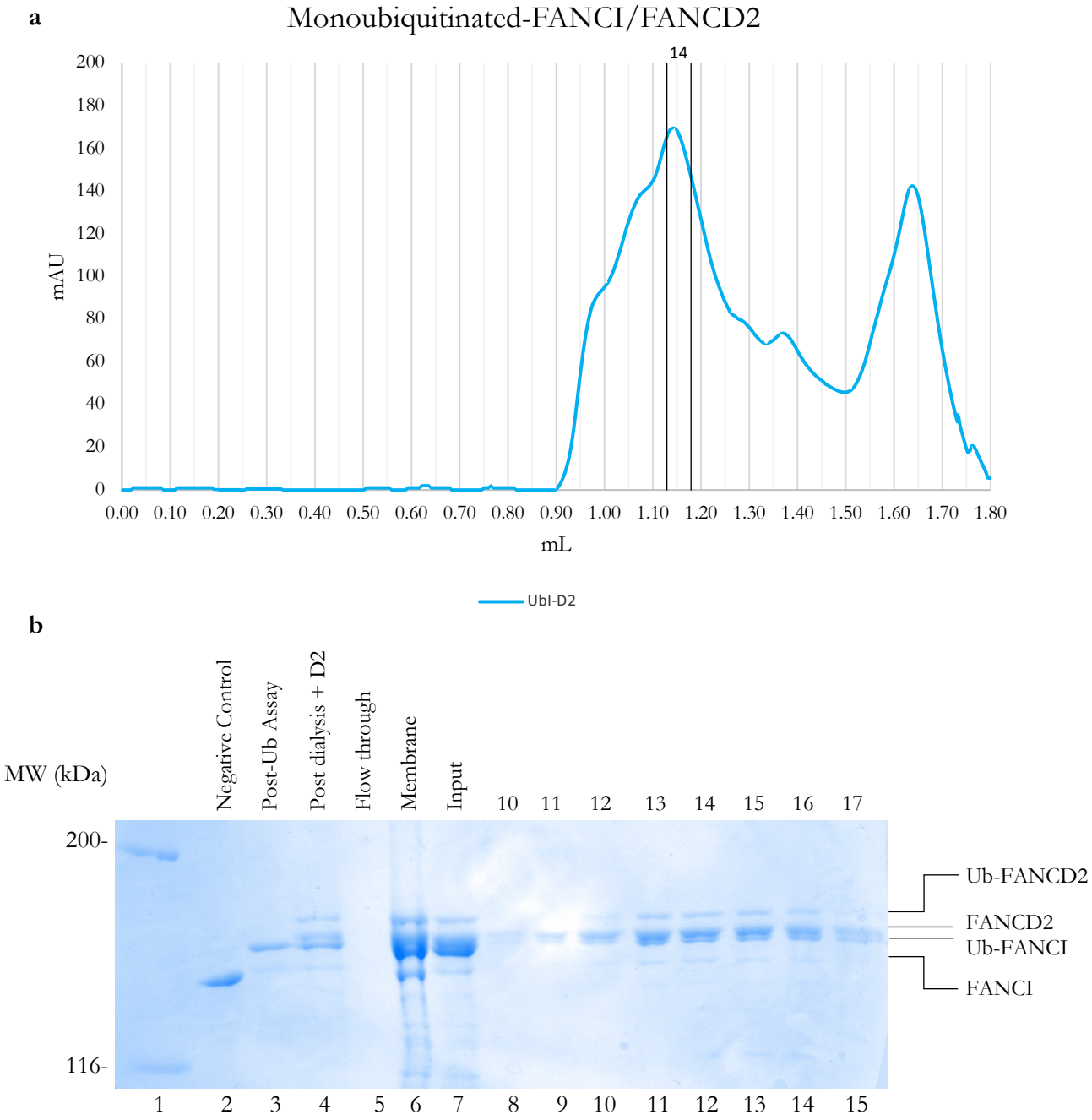
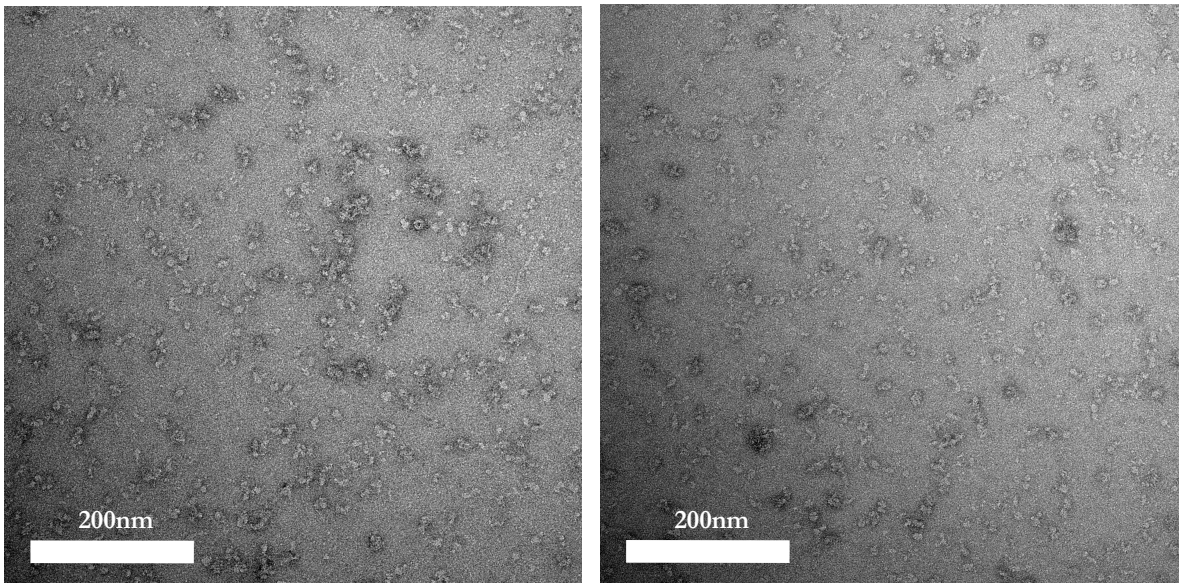
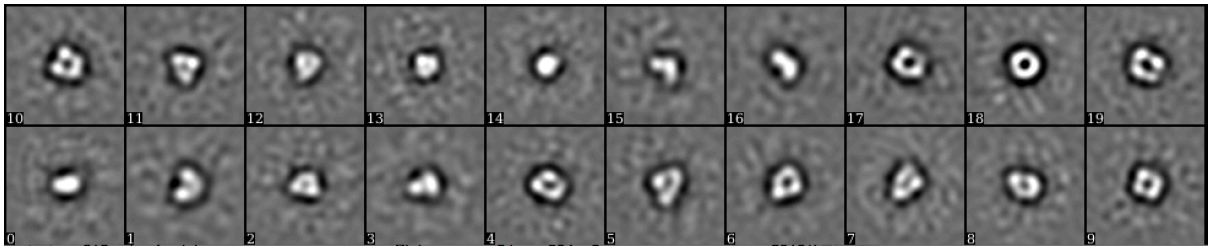


Figure 3-19: Purification of monoubiquitinated-FANCI/FANCD2. a) Chromatogram following purification on a Superdex 200 3.2/300. Monoubiquitinated-FANCI/FANCD2 identified in peak fraction 14. b) Coomassie blue stain of the ubiquitination assay and the purification procedure showing monoubiquitinated-FANCI/FANCD2. Lane 1: marker, lane 2: negative control (no ATP), lane 3: 3 hours post ubiquitination, lane 4: post dialysis with FANCD2 added, lane 5: flow through from 10 kDa cut-off concentrator, lane 6: sample stuck on the membrane, lane 7: sample loaded onto the gel filtration column, lanes 8 – 15: gel filtration peak fractions (lane 8: peak fraction 10 , lane 9: peak fraction 11, lane 10: peak fraction 12, lane 11: peak fraction 13, lane 12: peak fraction 14, lane 13: peak fraction 15, lane 14: peak fraction 16, lane 15: peak fraction 17).

a



b



c

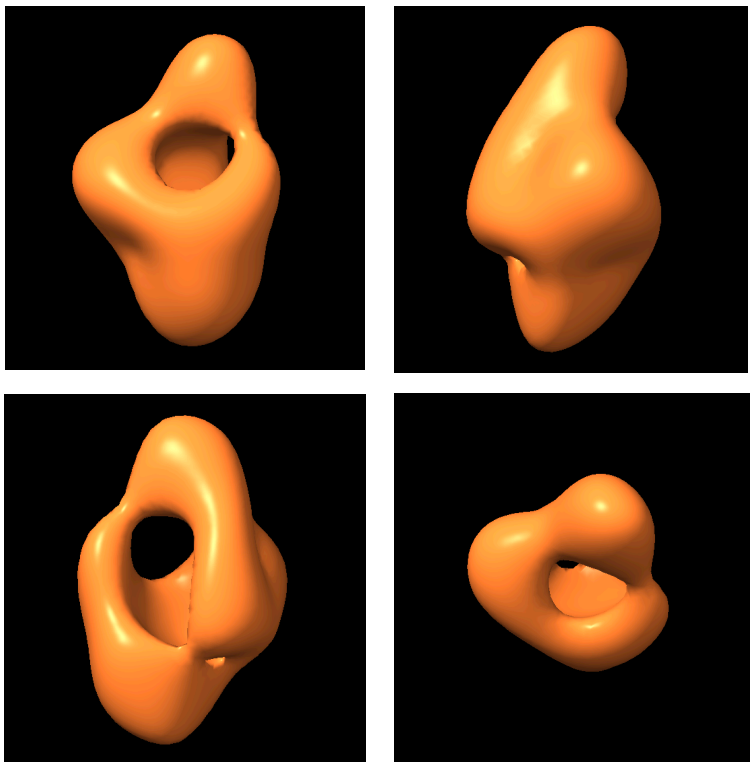
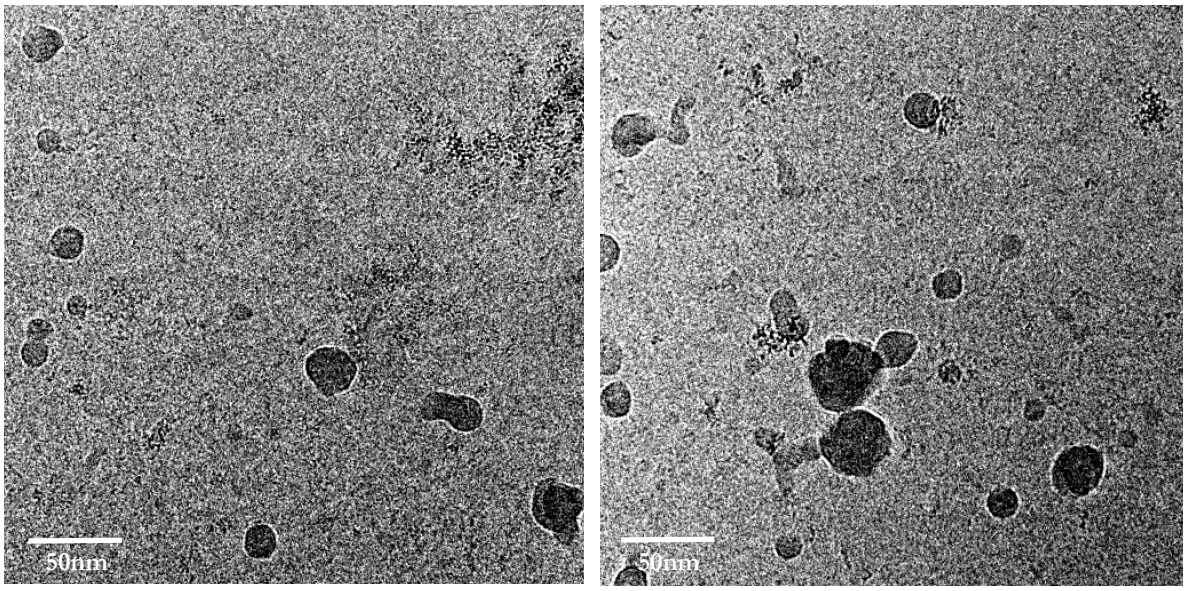


Figure 3-20: Negative staining of monoubiquitinated-FANCI/FANCD2 at 0.010 µg/µl. a) Representative micrographs showing the homogeneity of the monoubiquitinated-FANCI/FANCD2 sample from peak fraction 14 from Figure 3-19 at 68 000x, 1.727 Å/pixel. b) 2D classification of monoubiquitinated-FANCI/FANCD2 from 5266 particles using EMAN2.2. c) Initial model generated from 2D classification using EMAN2.2.

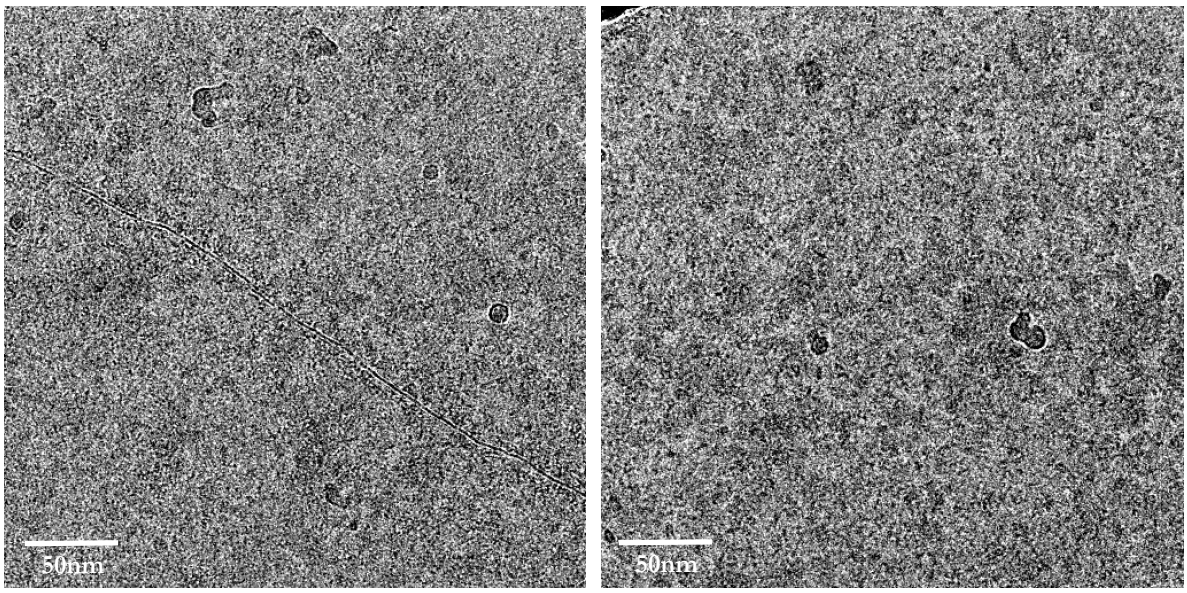
As the negative staining micrographs show good particle homogeneity and purity, and the 2D classification and 3D model look promising, the sample was prepared for cryo-EM. The sample was prepared as per the negative staining sample above. As with the non-ubiquitinated FANCD2/FANCI sample and the monoubiquitinated-FANCD2/FANCI sample, a variety of grids were used, including Quantifoil copper, Quantifoil gold and graphene oxide grids. For the Quantifoil copper grids, the sample seemed to show significant protein denaturation, as well as ethane contamination (Figure 3-21a). For the graphene oxide grids, significantly more protein can be seen, however again, it appears that there is significant degradation (Figure 3-21b). The Quantifoil gold grids looked to be the most promising, with visualisation of individual particles. However, in parts of the grid the contrast between the particles and the background inhibits easy visualisation of the particles, and there also seems to be ethane contamination at a similar size to the protein particles (Figure 3-21c).

Due to the numerous FA complex cryo-EM structures released in late 2019, no further work was continued on the cryo-EM structures of the FANCD2/FANCI, the monoubiquitinated-FANCD2/FANCI and the monoubiquitinated-FANCI/FANCD2 complexes in this DPhil project. The comparisons of the negative staining 3D models discussed in Sections 3.4 – 3.6 and these published structures will be further discussed in Section 3.8.

a



b



c

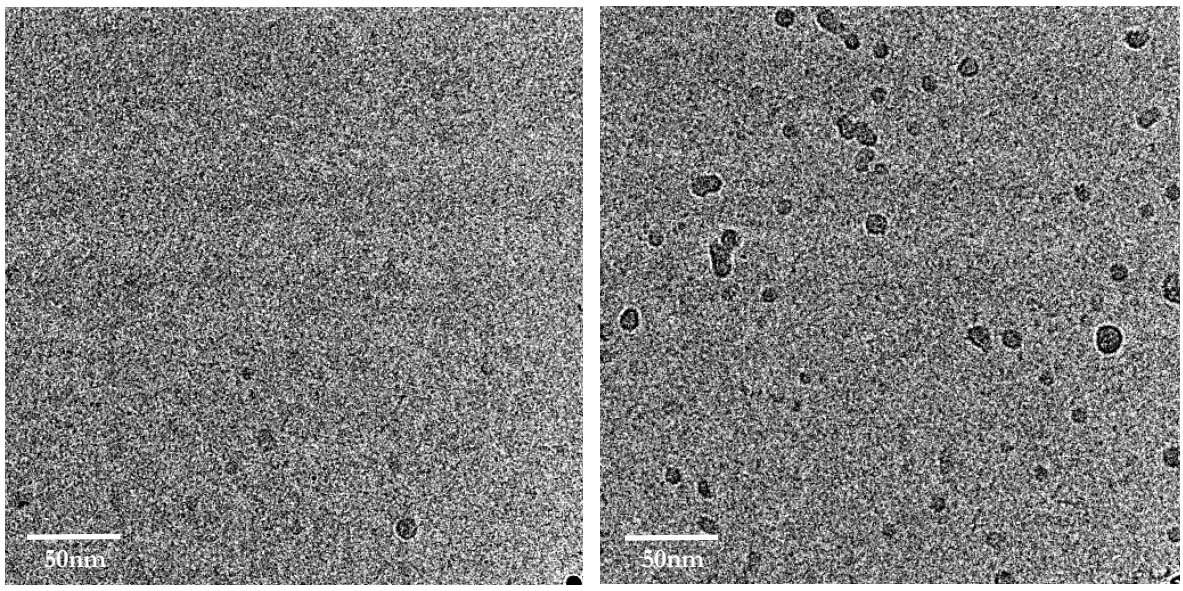


Figure 3-21: Optimisation of cryo-EM conditions for monoubiquitinated-FANCI/FANCD2 at 190 000x, 0.78 Å/pixel. a) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Cu; protein concentration 0.13 µg/µl. b) Representative micrographs of homemade Graphene Oxide Support Film 2/1 300 Mesh Cu; protein concentration 0.13 µg/µl. c) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Au; protein concentration 0.13 µg/µl.

3.7 The Monoubiquitinated-FANCD2/FANCI Intermediate Complex does not Appear to form Filamentous Arrays

During optimisation of purification conditions for the monoubiquitinated-FANCD2/FANCI complex on DNA by other members of the Cohn laboratory, filament formation was observed via negative staining microscopy. In addition, in late 2019 a preprint was released, followed by publication in March 2020, which reported filamentous arrays of the monoubiquitinated-FANCD2/monoubiquitinated-FANCI clamped onto DNA (Tan et al., 2020a). However, due to the purification procedure utilized in the publication, it was not possible to determine whether the filamentous formation was due to the monoubiquitination of FANCD2, the monoubiquitination of FANCI, or both. Therefore, whether the filaments form as a result of formation of the intermediate complex, monoubiquitinated-FANCD2/FANCI, was investigated.

First, to assess whether filament formation occurred as a result of monoubiquitination of FANCD2 in complex with FANCI, the ubiquitination assay was conducted for 0, 1 and 3 hours, as well as a negative control incubated for 3 hours without ATP. The assay was otherwise prepared as described above, however replacing the scDNA with a limiting concentration of the biotinylated DNA substrate, a 2391bp linear strand of DNA (4 μg). The biotinylated DNA substrate was prepared via a large-scale PCR reaction from the scDNA template and purified using the QIAquick PCR Purification Kit. Following the ubiquitination assay using the biotinylated DNA substrate, the reaction mixtures were incubated at 4°C with streptavidin sepharose beads to bind the monoubiquitinated-FANCD2/FANCI coated DNA. The coated DNA was then captured and washed with increasing salt concentration to remove any impurities. The biotinylated DNA was eluted using the restriction enzyme, SmaI, collected, and applied to negative stain EM grids (Figure 3-22). In all four samples, there is clear filament formation (Figure 3-22b, d, f, h). This may indicate that in fact it may not be the monoubiquitination of the FANCD2/FANCI that results in filament formation, but perhaps the recruitment of FANCD2/FANCI to the DNA. This result however, would be in contrast to that published by Tan et al., which claims that the non-ubiquitinated FANCD2/FANCI complex does not form filamentous arrays. However, it is clear that the resulting elution concentrations from the sepharose bead purification strategy is too low to be visualized via SDS-PAGE (Figure 3-22a, c, e, g), and thus the method must be altered to increase the yield of the elution.

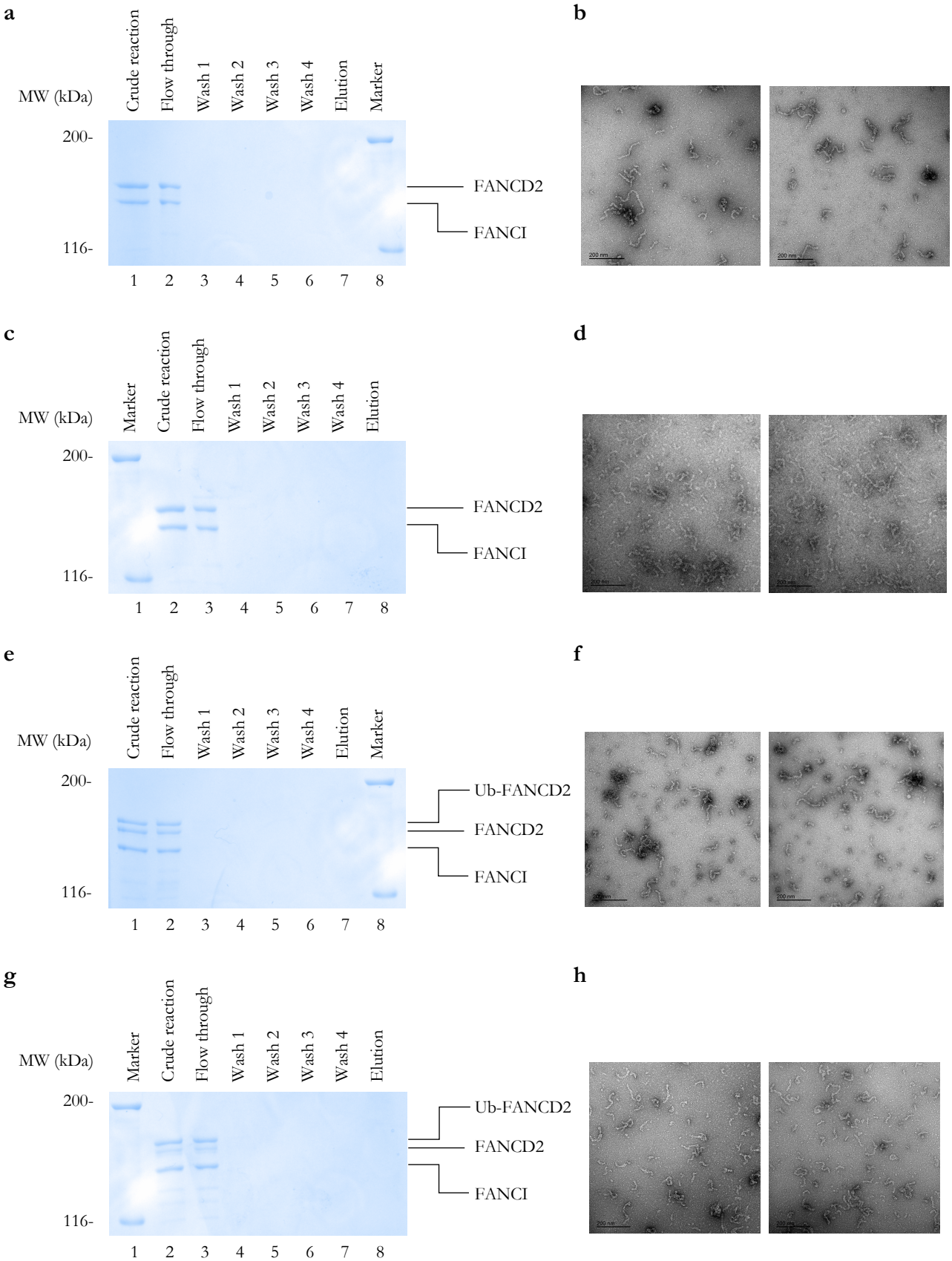
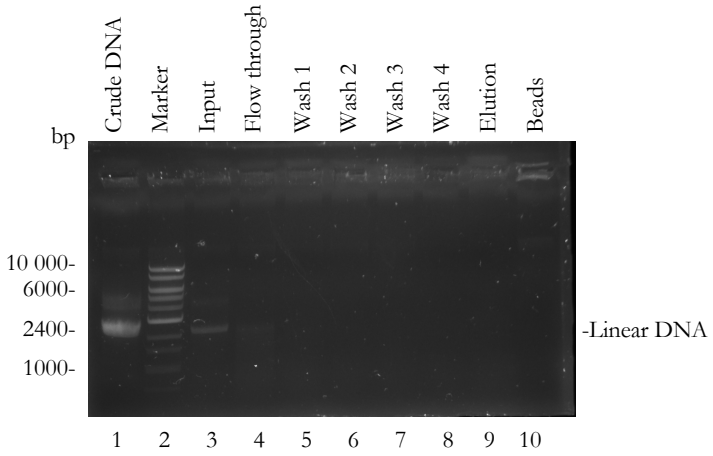

a
MW (kDa)
Crude reaction
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
Marker
200-
116-
FANCD2
FANCI
1 2 3 4 5 6 7 8
b
c
MW (kDa)
Marker
Crude reaction
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
200-
116-
FANCD2
FANCI
1 2 3 4 5 6 7 8
d
e
MW (kDa)
Crude reaction
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
Marker
200-
116-
Ub-FANCD2
FANCD2
FANCI
1 2 3 4 5 6 7 8
f
g
MW (kDa)
Marker
Crude reaction
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
200-
116-
Ub-FANCD2
FANCD2
FANCI
1 2 3 4 5 6 7 8
h

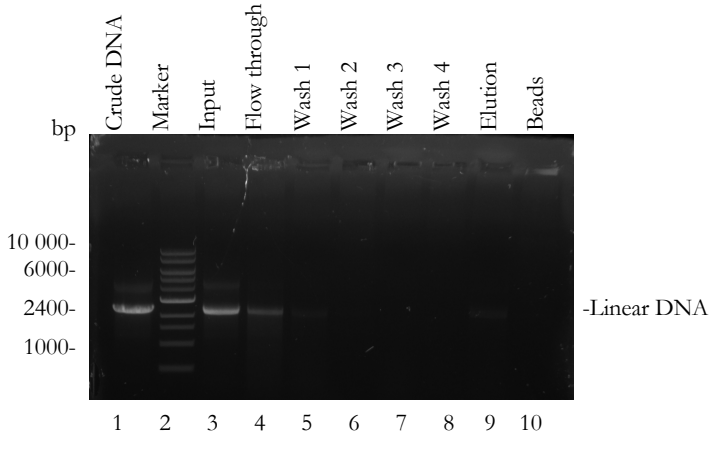
Figure 3-22: Observation of filament formation after monoubiquitination of FANCD2/FANCI via re-purification using sepharose beads using negative stain microscopy. a) Coomassie blue stain of the ubiquitination assay (negative control; no ATP). Lane 1: 3 hours post ubiquitination (no ATP), lane 2: flow through, lane 3: wash one, lane 4: wash two, lane 5: wash three, lane 6: wash four, lane 7: elution, lane 8: marker. b) Representative micrographs of non-ubiquitinated FANCD2/FANCI 3 hour sample from lane 7 in a) at 49 000x magnification. c) Coomassie blue stain of the ubiquitination assay at time 0 hours. Lane 1: marker, lane 2: 0 hours post ubiquitination, lane 3: flow through, lane 4: wash one, lane 5: wash two, lane 6: wash three, lane 7: wash four, lane 8: elution. d) Representative micrographs of monoubiquitinated FANCD2/FANCI 0 hour sample from lane 8 in c) at 49 000x magnification. e) Coomassie blue stain of the ubiquitination assay at time 1 hour. Lane 1: 1 hour post ubiquitination, lane 2: flow through, lane 3: wash one, lane 4: wash two, lane 5: wash three, lane 6: wash four, lane 7: elution, lane 8: marker. f) Representative micrographs of monoubiquitinated FANCD2/FANCI 1 hour sample from lane 7 in e) at 49 000x magnification. g) Coomassie blue stain of the ubiquitination assay at time 3 hours. Lane 1: marker, lane 2: 3 hours post ubiquitination, lane 3: flow through, lane 4: wash one, lane 5: wash two, lane 6: wash three, lane 7: wash four, lane 8: elution. h) Representative micrographs of monoubiquitinated FANCD2/FANCI 3 hour sample from lane 8 in g) at 49 000x magnification.

In order to attempt to increase the yield of the elution, an increase in DNA concentration as well as an alternative bead purification strategy (via streptavidin magnetic beads) were assessed. The streptavidin sepharose beads and streptavidin magnetic beads were directly compared using 4 µg and 10 µg biotinylated DNA substrate. In both cases, the biotinylated DNA substrate was incubated with the respective bead for 2 hours at 4°C to bind the DNA. The beads were then captured and washed with increasing salt concentration and the biotinylated DNA was eluted using SmaI. The eluted DNA was then assessed by agarose gel electrophoresis (Figure 3-23). When comparing these methods, it is clear that the higher concentration of DNA and the streptavidin magnetic beads result in a higher yield. Therefore, the repurification of the monoubiquitinated-FANCD2/FANCI was repeated using the magnetic beads using a higher concentration of DNA.

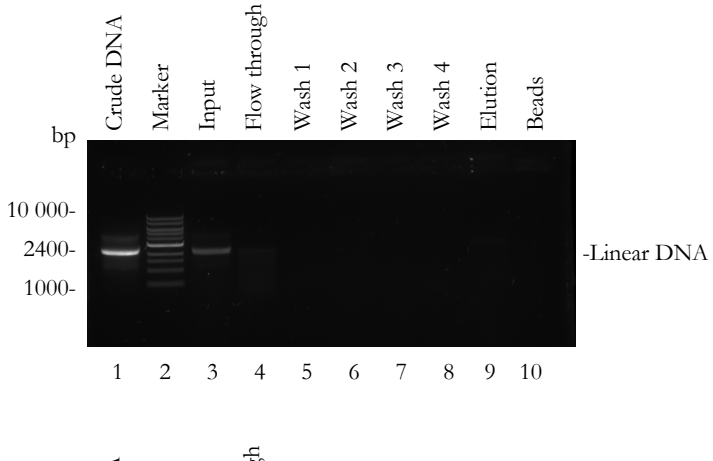
a


Crude DNA
Marker
Input
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
Beads
bp
10 000-
6000-
2400-
1000-
-Linear DNA
1 2 3 4 5 6 7 8 9 10

b


Crude DNA
Marker
Input
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
Beads
bp
10 000-
6000-
2400-
1000-
-Linear DNA
1 2 3 4 5 6 7 8 9 10

c


Crude DNA
Marker
Input
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
Beads
bp
10 000-
2400-
1000-
-Linear DNA
1 2 3 4 5 6 7 8 9 10

d

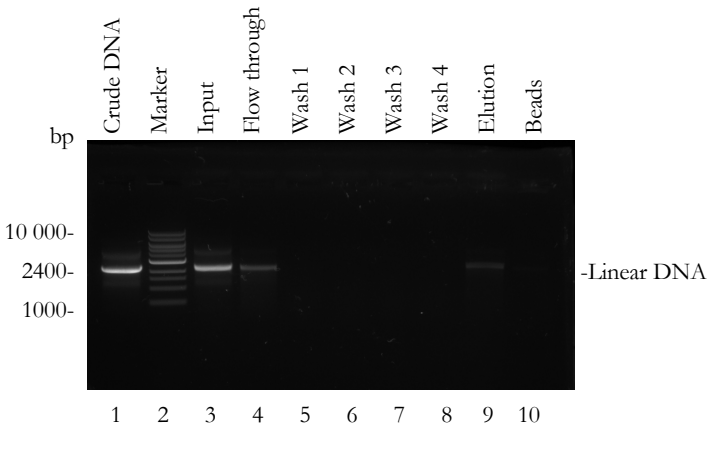
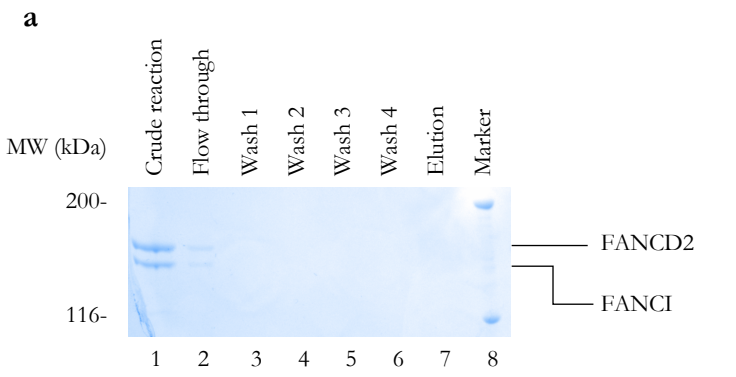
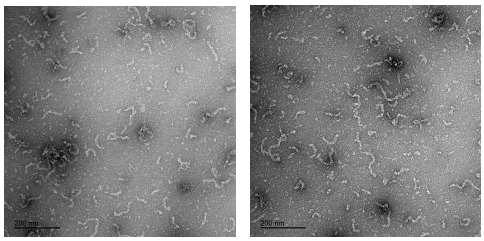
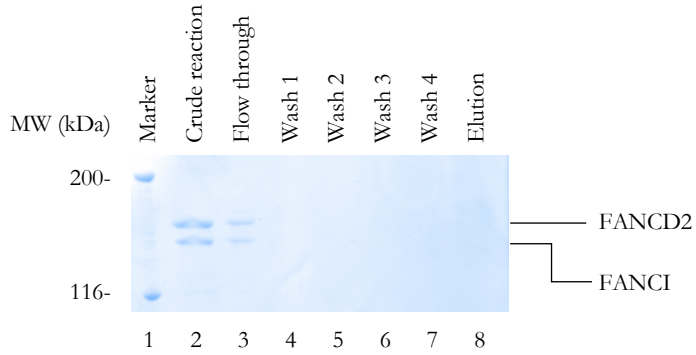

Crude DNA
Marker
Input
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
Beads
bp
10 000-
2400-
1000-
-Linear DNA
1 2 3 4 5 6 7 8 9 10

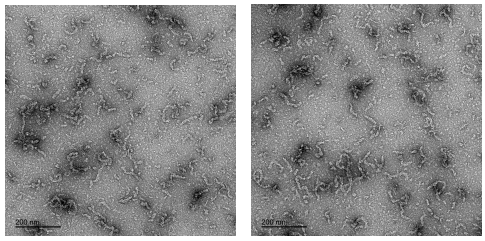
Figure 3-23: Method comparison between re-purification via sepharose beads vs. magnetic beads using varying concentrations showing a higher yield using magnetic beads. a) Repurification strategy for linear DNA using sepharose beads starting from 4 µg DNA. b) Repurification strategy for linear DNA using sepharose beads starting from 10 µg DNA. c) Repurification strategy for linear DNA using magnetic beads starting from 4 µg DNA. d) Repurification strategy for linear DNA using magnetic beads starting from 10 µg DNA. a) – d) Lane 1: crude linear DNA, lane 2: marker, lane 3: input of crude DNA, lane 4: flow through, lane 5: wash one, lane 6: wash two, lane 7: wash three, lane 8: wash four, lane 9: elution, lane 10: beads.

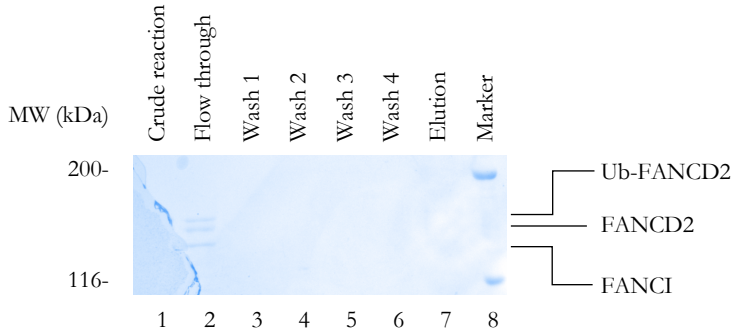
To attempt to assess again whether filament formation occurred as a result of monoubiquitination, the ubiquitination assay was conducted as above, however utilising the magnetic bead repurification strategy. A sample was taken for 0, 1 and 3 hours after ubiquitination, as well as a negative control incubated for 3 hours without ATP using the biotinylated DNA substrate. Following the ubiquitination assay using the biotinylated DNA substrate, the reaction mixtures were incubated at 4°C with streptavidin magnetic beads to bind the monoubiquitinated-FANCD2/FANCI coated DNA. The coated DNA was then captured and washed with increasing salt concentration to remove any impurities. The biotinylated DNA was eluted using SmaI, collected, and applied to negative stain EM grids (Figure 3-24). In all four samples, again, there is clear filament formation (Figure 3-24b, d, f, h). However, despite increasing the yield of the elution, the resulting concentration of FANCD2/FANCI and monoubiquitinated-FANCD2/FANCI is still too low to be visualized by SDS-PAGE. As filament formation is observed in all four samples however, this may again indicate that it may not be the monoubiquitination of the FANCD2/FANCI that results in filament formation, but perhaps is the result of the recruitment of FANCD2/FANCI to the DNA. However, it would be useful to repeat this protocol and use a higher sensitivity visualization method such as western blot or silver stain.

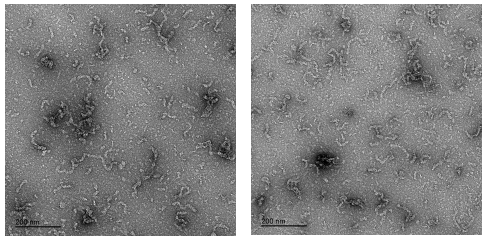

a
MW (kDa)
Crude reaction
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
Marker
200-
116-
FANCD2
FANCI
1 2 3 4 5 6 7 8

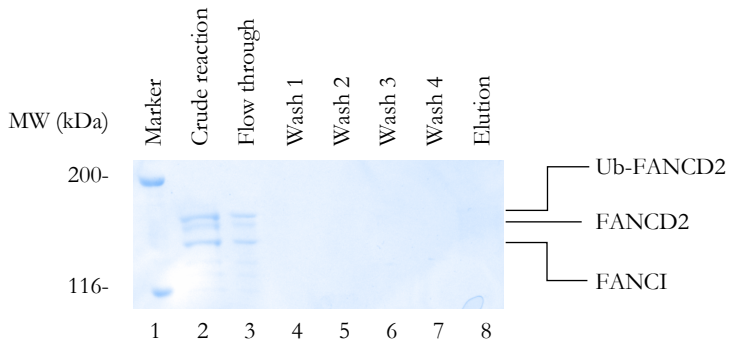

b


c
MW (kDa)
Marker
Crude reaction
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
200-
116-
FANCD2
FANCI
1 2 3 4 5 6 7 8


d


e
MW (kDa)
Crude reaction
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
Marker
200-
116-
Ub-FANCD2
FANCD2
FANCI
1 2 3 4 5 6 7 8


f


g
MW (kDa)
Marker
Crude reaction
Flow through
Wash 1
Wash 2
Wash 3
Wash 4
Elution
200-
116-
Ub-FANCD2
FANCD2
FANCI
1 2 3 4 5 6 7 8

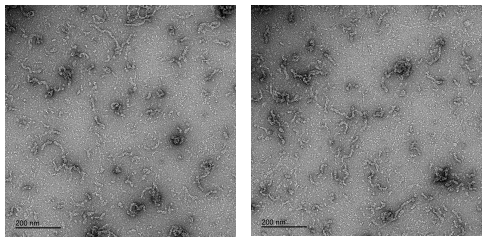

h

Figure 3-24: Observation of filament formation via re-purification using magnetic beads after monoubiquitination of FANCD2/FANCI using negative stain microscopy. a) Coomassie blue stain of the ubiquitination assay (negative control; no ATP). Lane 1: 3 hours post ubiquitination (no ATP), lane 2: flow through, lane 3: wash one, lane 4: wash two, lane 5: wash three, lane 6: wash four, lane 7: elution, lane 8: marker. b) Representative micrographs of non-ubiquitinated FANCD2/FANCI 3 hour sample from lane 7 in a) at 49 000x magnification. c) Coomassie blue stain of the ubiquitination assay at time 0 hours. Lane 1: marker, lane 2: 0 hours post ubiquitination, lane 3: flow through, lane 4: wash one, lane 5: wash two, lane 6: wash three, lane 7: wash four, lane 8: elution. d) Representative micrographs of monoubiquitinated FANCD2/FANCI 0 hour sample from lane 8 in c) at 49 000x magnification. e) Coomassie blue stain of the ubiquitination assay at time 1 hour. Lane 1: 1 hour post ubiquitination, lane 2: flow through, lane 3: wash one, lane 4: wash two, lane 5: wash three, lane 6: wash four, lane 7: elution, lane 8: marker. f) Representative micrographs of monoubiquitinated FANCD2/FANCI 1 hour sample from lane 7 in e) at 49 000x magnification. g) Coomassie blue stain of the ubiquitination assay at time 3 hours. Lane 1: marker, lane 2: 3 hours post ubiquitination, lane 3: flow through, lane 4: wash one, lane 5: wash two, lane 6: wash three, lane 7: wash four, lane 8: elution. h) Representative micrographs of monoubiquitinated FANCD2/FANCI 3 hour sample from lane 8 in g) at 49 000x magnification.

To further investigate the hypothesis that the filaments form as a result of recruitment of the FANCD2/FANCI complex to DNA, rather than due to monoubiquitination of FANCD2/FANCI, a series of control negative staining grids were prepared. First, a set of grids were prepared of the repurified biotinylated substrate alone, without FANCD2/FANCI (Figure 3-25a) as well as a set of grids of the repurified biotinylated substrate incubated with FANCD2/FANCI (Figure 3-25b). Unexpectedly, filament formation was observed in both of these samples. Therefore, a set of grids were prepared of the crude biotinylated DNA substrate and the SmaI enzyme (Figure 3-25c) as well as a set of grids of the SmaI enzyme alone (Figure 3-25d). Again, filaments were observed on both of these grids. Taken together, this indicates that the observed filaments are an artefact of the system, due to the SmaI digestion during the elution step. To further confirm this, the crude biotinylated DNA substrate was applied to negative staining grids (Figure 3-25e), and also the crude biotinylated DNA substrate incubated with FANCD2/FANCI (Figure 3-25f). In both of these samples, no filament formation can be seen, further confirming that the recruitment of FANCD2/FANCI to DNA does not appear to form filaments in this system. To further confirm this and that the filaments are due to the addition of SmaI to the system, negative staining grids were prepared of crude scDNA alone (Figure 3-25g), crude scDNA incubated with SmaI enzyme (Figure 3-25h), crude scDNA incubated with FANCD2/FANCI (Figure 3-25i) and FANCD2/FANCI alone (Figure 3-25j). When analyzing all of these controls, it is clear that filaments are only formed when SmaI is present, irrespective of the presence of DNA or FANCD2/FANCI. However, these controls do not fully rule-out whether there may be filament formation resulting from the monoubiquitination of FANCD2/FANCI, only that they cannot be visualized using this purification method.

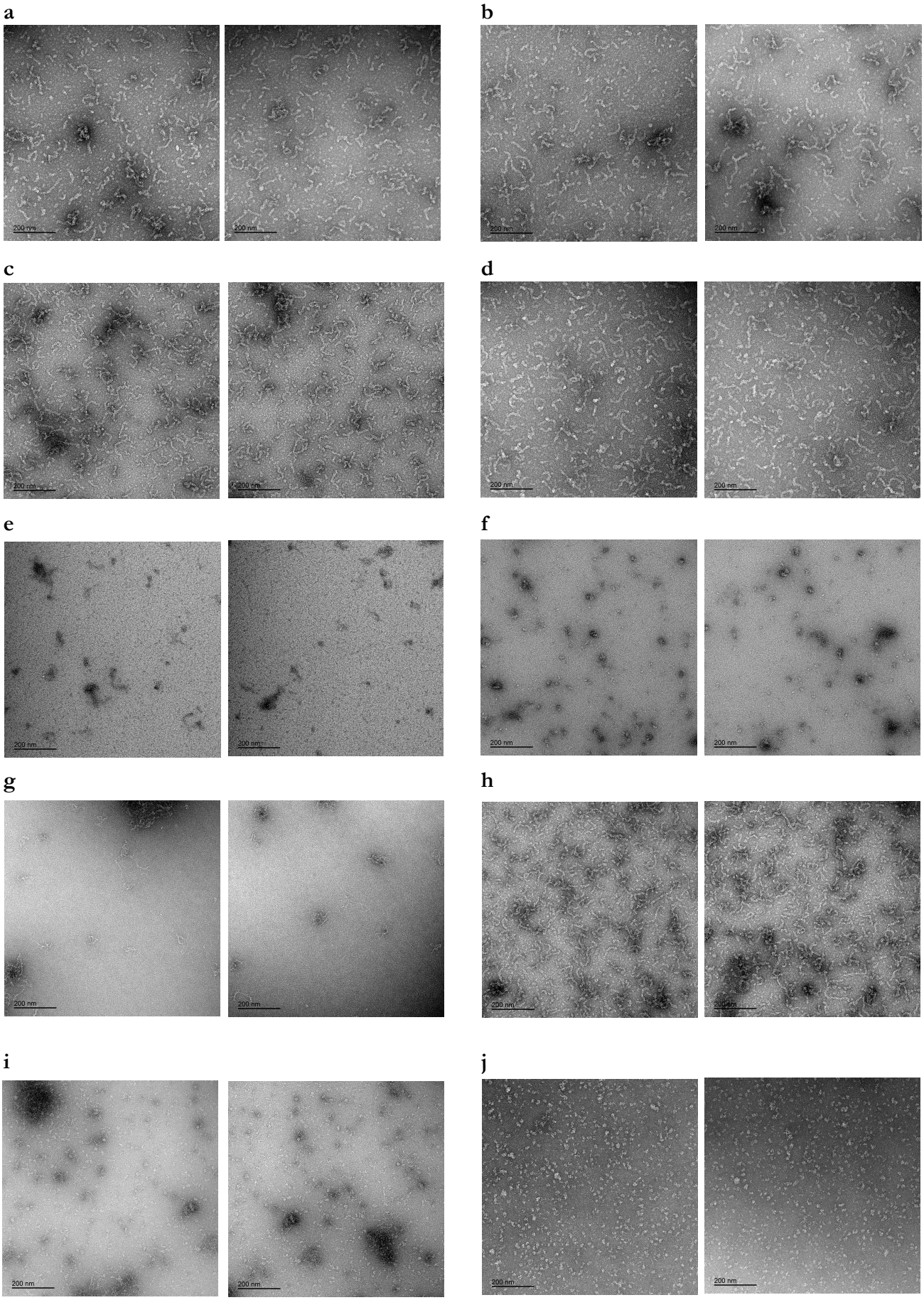
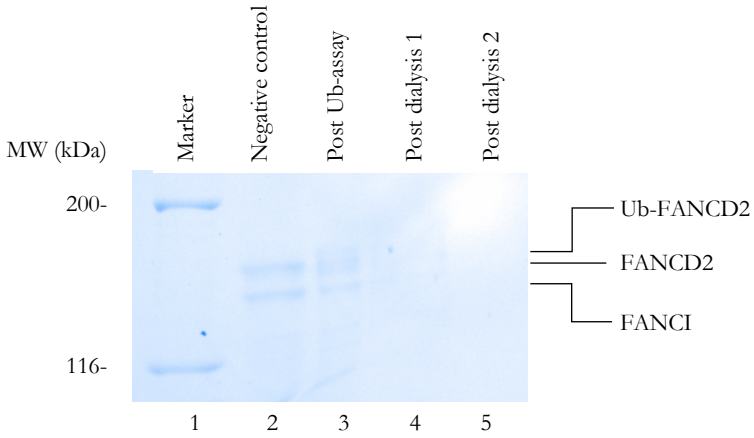

a
b
c
d
e
f
g
h
i
j
200 nm

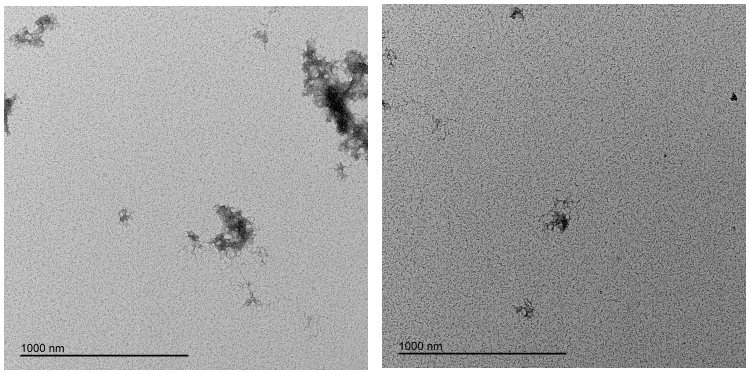
Figure 3-25: Completion of various controls for filament formation using negative stain microscopy at 49 000x magnification. a) Representative micrographs of repurified linear DNA alone. b) Representative micrographs of repurified linear DNA + FANCD2/FANCI. c) Representative micrographs of crude linear DNA + SmaI enzyme. d) Representative micrographs of SmaI alone. e) Representative micrographs of crude linear DNA. f) Representative micrographs of crude linear DNA + FANCD2/FANCI. g) Representative micrographs of crude scDNA. h) Representative micrographs of crude scDNA + SmaI. i) Representative micrographs of crude scDNA + FANCD2/FANCI. j) Representative micrographs of crude FANCD2/FANCI alone.

To further investigate the hypothesis that the filaments form as a result of monoubiquitination of FANCD2/FANCI, the crude monoubiquitination assay of FANCD2/FANCI using scDNA was conducted, dialysed to remove other ubiquitination assay components, and applied to negative staining grids. However, the dialysis treatment seemed to allow the monoubiquitinated-FANCD2/FANCI complex to dialyse out of the membrane, as the sample cannot be visualized following dialysis using SDS-PAGE (Figure 3-26a). Positive staining micrographs were collected for naked scDNA alone (Figure 3-26b), as well as positive (top panel) and negative (bottom panel) staining micrographs for the post-dialysis monoubiquitinated-FANCD2/FANCI sample (Figure 3-26c). These micrographs display that the scDNA can be visualized, as well as the monoubiquitinated-FANCD2/FANCI protein particles, but there is no evidence of filament formation. However, as this sample cannot be visualized using SDS-PAGE, it remains unclear that the corresponding protein particles correspond to monoubiquitinated-FANCD2/FANCI.

a


MW (kDa)
Marker
Negative control
Post Ub-assay
Post dialysis 1
Post dialysis 2
200-
116-
Ub-FANCD2
FANCD2
FANCI
1
2
3
4
5

b


1000 nm
1000 nm

c

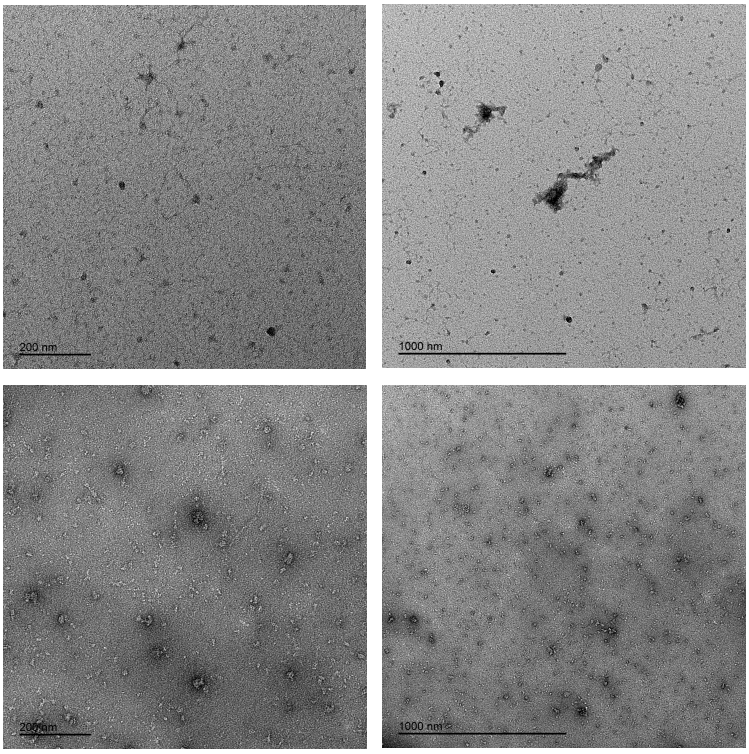
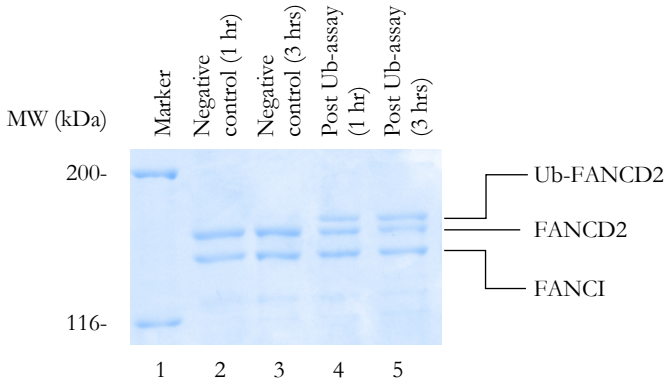

200 nm
1000 hm
200 nm
1000 nm

Figure 3-26: Lack of filament formation after monoubiquitination of FANCD2/FANCI followed by dialysis treatment. a) Coomassie blue stain of the monoubiquitination assay of FANCD2/FANCI and following dialysis. Lane 1: marker, lane 2: negative control (no ATP), lane 3: 3 hours post ubiquitination, lane 4: post dialysis with 300 mM NaCl, lane 5: post dialysis with 100 mM NaCl. b) Representative micrographs of naked scDNA at 0.008 µg/µl at 23 000x magnification (positive stain). c) Representative micrographs of monoubiquitinated FANCD2/FANCI 3 hour sample from lane 5 in a) at 0.01 µg/µl at 49 000x (left) and 23 000x (right) magnification (top panel: positive stain; bottom panel: negative stain).

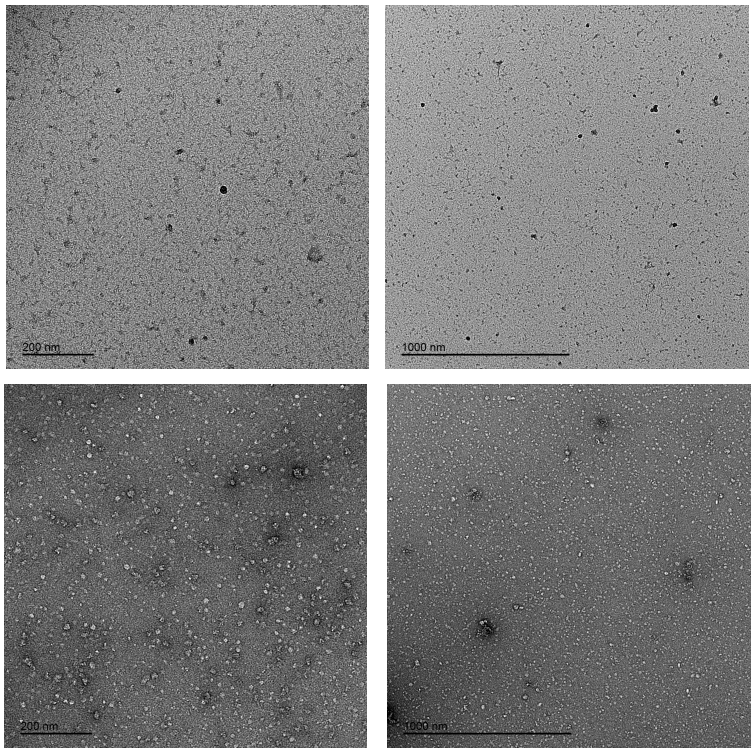
Therefore, to finally determine whether there is filament formation following monoubiquitination of FANCD2/FANCI, the crude monoubiquitination assay of FANCD2/FANCI using scDNA was conducted and applied directly to negative staining grids. The negative control (without ATP) and the 3 hour ubiquitinated sample were visible via SDS-PAGE (Figure 3-27a). The negative control showing all reaction components and scDNA shows no filament formation, consistent with Tan et al. (Figure 3-27b). The top panel displays positive staining data to visualise the scDNA more easily, and the bottom panel displays negative staining data to visualise the FANCD2/FANCI protein complex more easily. The 3 hour monoubiquitinated-FANCD2/FANCI sample also shows no evidence of filament formation (Figure 3-27c). The top panel again displays positive staining data to visualise the scDNA, and the bottom panel displays negative staining data to visualise the monoubiquitined-FANCD2/FANCI complex. It can be seen, particularly in the top panel of Figure 3-27c, that monoubiquitinated-FANCD2/FANCI molecules interact with the scDNA, however this appears to be in a sporadic fashion, and the protein does not form filamentous arrays.

Therefore, through the series of controls conducted, it is clear that the filament formation is due to the addition of the SmaI enzyme and is thus an artefact of the repurification system. In addition, no filament formation can be seen as a result of the recruitment of FANCD2/FANCI to DNA, nor due to formation of the intermediate monoubiquitinated-FANCD2/FANCI when analysing the crude reaction product.

a


MW (kDa)
Marker
Negative control (1 hr)
Negative control (3 hrs)
Post Ub-assay (1 hr)
Post Ub-assay (3 hrs)
200-
116-
Ub-FANCD2
FANCD2
FANCI
1
2
3
4
5

b


200 nm
1000 nm
200 nm
1000 nm

c

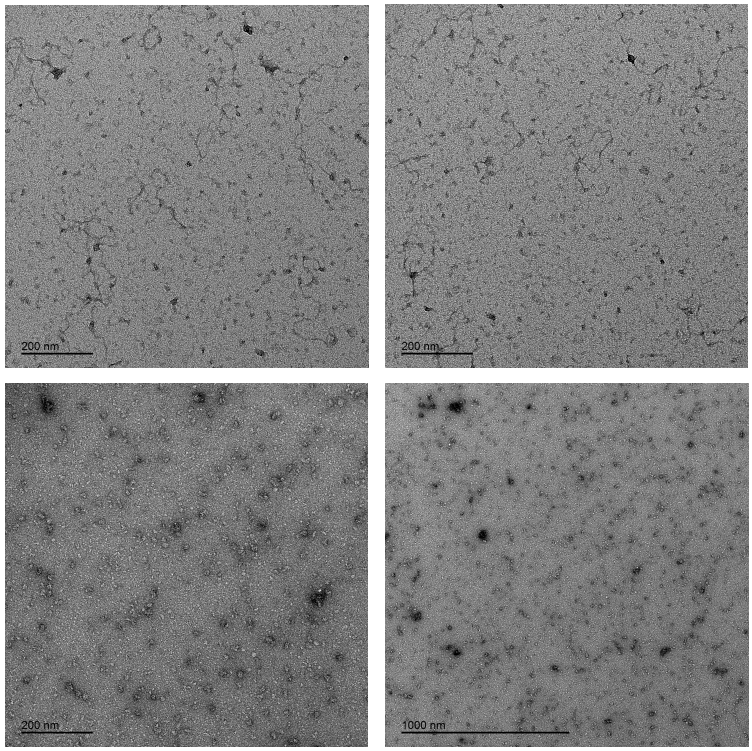

200 nm
200 nm
200 nm
1000 nm

Figure 3-27: Lack of filament formation after crude monoubiquitination of FANCD2/FANCI. a) Coomassie blue stain of the monoubiquitination assay of FANCD2/FANCI. Lane 1: marker, lane 2: 1 hour negative control (no ATP), lane 3: 3 hours negative control (no ATP), lane 4: 1 hour post ubiquitination, lane 5: 3 hours post ubiquitination. b) Representative micrographs of the negative control from lane 3 in a) at 0.01 µg/µl at 49 000x (left) and 23 000x (right) magnification (top panel: positive stain; bottom panel: negative stain). c) Representative micrographs of monoubiquitinated FANCD2/FANCI sample from lane 5 in a) at 0.01 µg/µl at 49 000x (top panel and bottom left) and 23 000x (bottom right) magnification (top panel: positive stain; bottom panel: negative stain).

3.8 Discussion

Monoubiquitination of FANCD2/FANCI is the pivotal event in the FA pathway. This chapter describes the optimisation of the formation of monoubiquitinated-FANCD2/FANCI and monoubiquitinated-FANCI using increasing levels of scDNA. In addition, deubiquitination of both species is described, showing the DNA independence of this deubiquitination reaction. The preliminary 3D models of FANCD2/FANCI, monoubiquitinated-FANCD2/FANCI and monoubiquitinated-FANCI/FANCD2 solved by negative staining EM are described, and progress towards optimizing cryo-EM grid conditions. The possibility of filament formation upon ubiquitination of FANCD2/FANCI is also investigated. Each of these will be further discussed below.

Efficiency of monoubiquitination of FANCD2/FANCI and FANCI is DNA-dependent *in vitro*

There have been varying reports regarding the monoubiquitination of FANCD2/FANCI *in vivo* and whether this occurs prior to or following the complex's recruitment to DNA. For example, it has been reported that the mutant form of FANCD2 that is unable to be monoubiquitinated (FANCD2-K561R) cannot form damage-induced FANCD2 foci (Garcia-Higuera et al., 2001; Smogorzewska et al., 2007). However, it has since been reported that FANCD2-K561R can be recruited to laser-induced ICL damage in cells, albeit at a lower level to the FANCD2 wild type (Liang et al., 2016). It was suggested that monoubiquitination is necessary to stabilize FANCD2 on DNA, and thus there was a lower accumulation of the mutant FANCD2-K561R in comparison to the wild type at ICL damage. In contrast, there has been increasing evidence that the monoubiquitination of FANCD2/FANCI *in vitro* is dependent on DNA. For example, it has been demonstrated that monoubiquitination of FANCD2 in complex with FANCI is stimulated by the presence of circular dsDNA, dsDNA fragments and ssDNA (Sareen et al., 2012). It has also been

demonstrated that ssDNA, dsDNA and branched DNA stimulates the monoubiquitination of FANCD2 in complex with FANCI to a level comparable to *in vivo* monoubiquitination levels, with the maximum ubiquitination achieved with 5' flapped DNA (mimicking an arrested replication fork) (Sato et al., 2012b). In addition, FANCI has been shown to play a role in the binding of the complex to DNA, with evidence demonstrating that FANCI possesses DNA binding activity that prefers branched structures (Longerich et al., 2009; Yuan et al., 2009). Despite this DNA interaction with FANCI, it has also been shown that in the presence of the avian core complex, FANCD2 monoubiquitination is enhanced in complex with FANCI, whilst FANCI ubiquitination in complex with FANCD2 remains unstimulated (Rajendra et al, 2014; Shakeel et al, 2019). FANCD2 has also been shown to contain DNA binding domains, where mutations in these domains result in a reduction of FANCD2 monoubiquitination (Niraj et al., 2017). In addition, a phosphorylation cluster has been identified on FANCD2 which when phosphorylated, reduces ubiquitination *in vivo* as a result of a loss of DNA binding (Lopez-Martinez et al., 2019).

Here, it has been demonstrated that the efficiency of monoubiquitination of FANCD2/FANCI and FANCI alone is dependent on the level of scDNA *in vitro*. For FANCD2/FANCI, the efficiency increases with increasing scDNA concentrations, whilst for FANCI alone, the monoubiquitination efficiency occurs at a peak of 1 μg. Taken together, this data therefore supports the literature that the FANCD2/FANCI complex is recruited to sites of damage prior to the pivotal monoubiquitination event. This in turn supports the hypothesis that there is a conformational change of the FANCD2/FANCI complex upon DNA binding and subsequent monoubiquitination (Walden and Deans, 2014).

Efficiency of deubiquitination of monoubiquitinated-FANCD2/FANCI and monoubiquitinated-FANCI is DNA-independent *in vitro*

The monoubiquitination of FANCD2/FANCI has been shown to be crucial to a functioning FA pathway. However, it has also been shown that the deubiquitination of the monoubiquitinated FANCD2/FANCI complex is also essential for the completion of FA pathway-mediated ICL repair (Arkinson et al., 2018; Jacq et al., 2013). USP1/UAF1 is the complex that reverses the monoubiquitination, and it has been shown that disrupting this deubiquitination results in DNA crosslinker sensitivity (Oestergaard et al., 2007). However, there have been significant discrepancies surrounding whether the deubiquitination of the monoubiquitinated complex occurs on or off DNA.

In 2017 it was demonstrated that the deubiquitination of the monoubiquitinated-FANCD2/monoubiquitinated-FANCI complex by USP1/UAF1 occurs only when the ubiquitinated complex is disengaged from DNA (van Twest et al., 2017). In contrast, in 2019 it was reported that the deubiquitination of the monoubiquitinated-FANCD2/FANCI complex required DNA for effective deubiquitination, and that removal of the DNA decreased the deubiquitination by up to 95% (Liang et al., 2019). This discrepancy was attributed to the use of the human FA core complex proteins to ubiquitinate the *Xenopus laveis* FancD2/FancI, as well as human USP1/UAF1 to deubiquitinate the monoubiquitinated *Xenopus* FancD2/FancI in the van Twest study, in comparison to the use of only human proteins in the Liang report (Liang et al. 2019). In an additional report, it was shown that USP1 can deubiquitinate monoubiquitinated-FANCD2 and monoubiquitinated-FANCI as monomeric substrates in a UAF1-dependent manner (Arkinson et al., 2018). When the deubiquitination of monoubiquitinated-FANCD2/monoubiquitinated-FANCI was examined, it was observed that USP1/UAF1 could deubiquitinate a portion of the monoubiquitinated-FANCD2 and monoubiquitinated-FANCI on DNA, but that a portion remained resistant to deubiquitination. In contrast, when monoubiquitination of FANCI was prevented, the monoubiquitinated-FANCD2/FANCI complex could be fully deubiquitinated on or off DNA. Therefore, the study suggests that it is not in fact the DNA that inhibits the deubiquitination in the van Twest study, but in fact the monoubiquitination of FANCI (Arkinson et al., 2018). In agreement with this study, here it was demonstrated that USP1/UAF1 could effectively deubiquitinate the monoubiquitinated-FANCD2/FANCI, as well as deubiquitinate the monoubiquitinated-FANCI monomeric substrate, in the presence or absence of DNA. In a recent study, the ubiquitination of FANCI has been shown to protect the monoubiquitinated-FANCD2 from deubiquitination by the USP1/UAF1 complex, showing that whilst the monoubiquitinated-FANCD2 monomer and the monoubiquitinated-FANCD2/FANCI dimer can be readily deubiquitinated, the monoubiquitinated-FANCD2/monoubiquitinated-FANCI dimer was resistant to deubiquitination by USP1/UAF1. They therefore conclude that the monoubiquitinated-FANCD2 substrate is less accessible to USP1/UAF1 when in complex with monoubiquitinated-FANCI and DNA (Rennie et al., 2020).

Monoubiquitination of FANCD2/FANCI causes a conformational change resulting in the complex "locking" onto DNA

Despite the monoubiquitination of FANCD2/FANCI being a crucial step in the FA pathway, mechanistic insight into this event has remained elusive. The first structural insight into the

FANCD2/FANCI complex was in 2011, with the murine X-ray crystal structure published to 3.4Å resolution. This structure revealed an open, trough-like structure, with the two monoubiquitination sites at the dimerization interface (Joo. et al., 2011). This structure was followed in 2016, with a low resolution human cryo-EM structure, solved to 22Å resolution. As mentioned above, the structures show good agreement, albeit with a different conformation for the "Tower" domain in the cryo-EM structure (Liang et al., 2016). However, there has been much speculation surrounding the monoubiquitination event of the complex, and whether this event results in a conformational change of the heterodimer. The hypothesis at the conception of this project was that the monoubiquitination of FANCD2/FANCI would result in a conformational change of the complex, resulting in the complex locking onto the DNA molecule. Before the completion of this project, in late 2019 and 2020, several cryo-EM structures were released, including the structure of FANCD2/FANCI, the monoubiquitinated-FANCD2/FANCI and the monoubiquitinated-FANCD2/monoubiquitinated-FANCI (Alcón et al., 2020; Wang et al., 2019; Wang et al., 2020; Rennie et al., 2020). Below, the progress of this project will be discussed in light of these new insights as well as the resulting new mechanistic understanding of the monoubiquitination of FANCD2/FANCI.

As described above, the mechanism of monoubiquitination of the FANCD2/FANCI complex has remained a mystery. To form the monoubiquitinated-FANCD2/monoubiquitinated-FANCI from the FANCD2/FANCI precursor, multiple possible routes could be envisaged. It has been described that FANCD2 needs to be in complex with FANCI to enable efficient ubiquitination *in vitro* (Chaugule et al., 2020; Sareen et al., 2012). However, once this intermediate monoubiquitinated-FANCD2/FANCI has been formed, there have been great difficulties subsequently ubiquitinating FANCI, to form the fully monoubiquitinated-FANCD2/monoubiquitinated-FANCI complex. There have also been reports that FANCI can be efficiently ubiquitinated in its monomeric state, raising the possibility that FANCI is ubiquitinated first to form monoubiquitinated-FANCI, prior to dimeric formation with FANCD2, and subsequent ubiquitination of the FANCD2 protein. In addition to these possible routes, it has also been suggested that monoubiquitination of FANCD2 and FANCI could occur simultaneously, due to the two copies of the E3 ligase, FANCL, present in the core complex (Rajendra et al, 2014; Shakeel et al., 2019; Swuec et al., 2017). Therefore, this project focused on the generation of structural data for the FANCD2/FANCI precursor, as well as the two possible intermediate states, monoubiquitinated-FANCD2/FANCI and monoubiquitinated-FANCI/FANCD2. To be able to determine if there was a conformational change due to the

monoubiquitination of either FANCD2 or FANCI, an improved structure of the human FANCD2/FANCI was also investigated.

To first determine if the purified complexes were homogenous, pure and not aggregated, each of FANCD2/FANCI, monoubiquitinated-FANCD2/FANCI and monoubiquitinated-FANCI/FANCD2 were prepared for screening via negative staining microscopy. These micrographs were then used to generate initial 3D models using EMAN2.2. Figure 3-28 shows these 3D models in a side-by-side comparison. It is evident that the FANCD2/FANCI 3D model (yellow) seems to be most similar to the monoubiquitinated-FANCI/FANCD2 3D model (orange). In comparison, the monoubiquitinated-FANCD2/FANCI 3D model (pink) seems to show a slightly different 3D model, indicating that perhaps there is a greater conformational change with regards to the monoubiquitination of FANCD2. However, there is also significantly less density for this complex in comparison to the FANCD2/FANCI and monoubiquitinated-FANCI/FANCD2, so perhaps the monoubiquitinated-FANCD2/FANCI complex has several preferential orientations on negative staining grids resulting in missing density in the 3D model. In all three negative staining models, a channel running from the top, down and out the side of the structure can be seen. This groove is presumably where the DNA binds to the complex.

Progress towards optimising cryo-EM grids for each of the complexes is also presented, showing across all three complexes, that individual particles could be seen, however in a sparse and hard to visualise manner. There also seems to be evidence of protein aggregation and protein degradation, creating a high amount of background. As suggested in the literature, the graphene oxide grids, as well as the gold mesh Quantifoil grids seemed to herald the most promising grids. Unfortunately, during optimisation of the cryo-EM grid conditions and increasing the concentration of the samples, several papers were released by the Pavletich and Passmore laboratories, detailing the cryo-EM structures of FANCD2/FANCI, monoubiquitinated-FANCD2/FANCI and the monoubiquitinated-FANCD2/monoubiquitinated-FANCI. These papers also indicated that the monoubiquitinated-FANCD2/monoubiquitinated-FANCI complex is formed via the monoubiquitinated-FANCD2/FANCI intermediate. These papers halted further work on this project, however, comparisons of the negative staining models with these published cryo-EM structures and the mechanistic insight gained from these structures is discussed below.

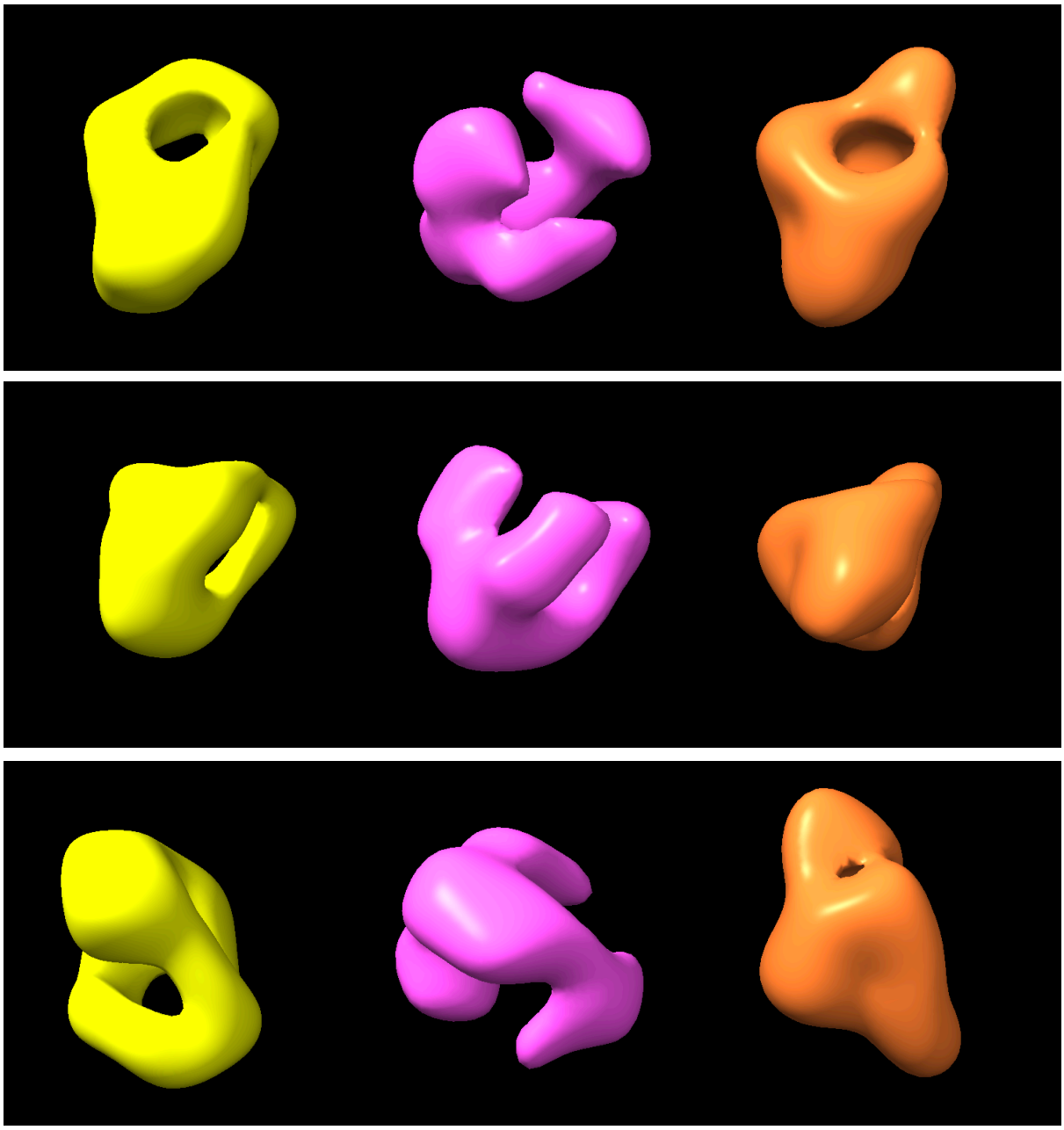
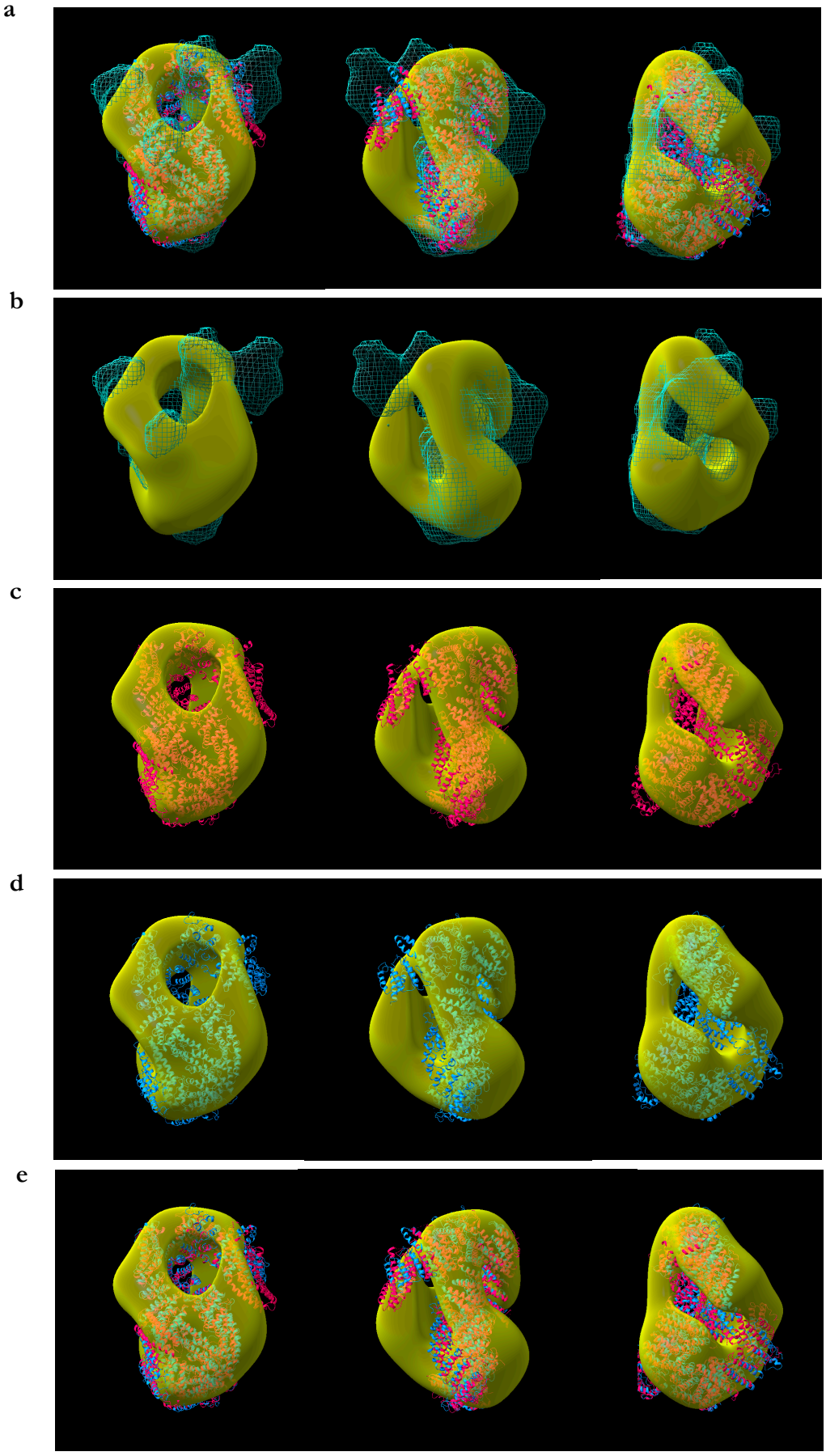


Figure 3-28: Side-by-side comparison of initial models generated from negative staining microscopy of FANCD2/FANCI (yellow), monoubiquitinated-FANCD2/FANCI (pink) and monoubiquitinated-FANCI/FANCD2 (orange).

In late 2019, the Pavletich group released two preprint manuscripts detailing the full-length human FANCD2/FANCI cryo-EM structure to 3.4Å and the monoubiquitinated-FANCD2/monoubiquitinated-FANCI cryo-EM structure to 3.5Å, as well as the FANCD2/FANCI complex poised to be ubiquitinated by the FA core complex and the E2 conjugating enzyme UBE2T cryo-EM structure to 4.2Å (Wang et al., 2019; Wang et al., 2020). Following the release of these manuscripts, the Passmore group released an additional preprint manuscript detailing the *Gallus gallus* FANCD2/FANCI cryo-EM structure to 4.1Å, and the intermediate, monoubiquitinated-FANCD2/FANCI cryo-EM structure to 3.8Å (Alcón et al., 2020). Most of these manuscripts and the resulting structures were published in various journals between March – July 2020, and the structural maps uploaded to the PDB and EMDB.

When comparing the non-ubiquitinated FANCD2/FANCI, the new structures show an open, trough-like architecture, similar to the previously solved X-ray crystal structure. The maps show that each of FANCD2 and FANCI consist of an N-terminal domain (NTD) and a C-terminal domain (CTD) joined together by a helical domain. The NTD domains consist of helical repeats that form an α–α solenoid and the CTD domains consist of a helical repeat domain. It was determined that the CTD of FANCD2 exhibited substantial conformational flexibility. The FANCD2 and FANCI protomers assemble in a top-to-tail like conformation with the ubiquitination sites buried at the dimer interface (as indicated previously in the X-ray crystal structure). Interestingly, Wang et al. reported that upon DNA binding, the conformation of the FANCD2/FANCI structure did not change. When dsDNA is bound, the DNA binds to a basic groove of FANCI that runs along part of the NTD, CTD and joining helical domain. The dsDNA is also bound to part of the NTD and CTD of FANCD2, but not to the acidic helical domain. However, when DNA mimicking replication forks was bound (such as 5'-flap DNA), the DNA was mainly bound via FANCI, and the CTD of FANCD2 did not bind to the DNA. As the FA pathway requires converging replication forks, perhaps the FANCD2/FANCI complex is recruited to dsDNA nearby to a replication fork and then travels along the dsDNA before reaching the replication fork, upon which, due to the change in DNA structure, loses the binding between FANCD2 and DNA. Figure 3-29 displays the overlay of the two newly solved cryo-EM structures (Wang et al.: pink, Alcón et al.: blue) of FANCD2/FANCI, the previously solved cryo-EM structure (Liang et al.: turquoise mesh) of FANCD2/FANCI and the negative staining model of FANCD2/FANCI produced in this project (transparent yellow).


a
b
c
d
e

f

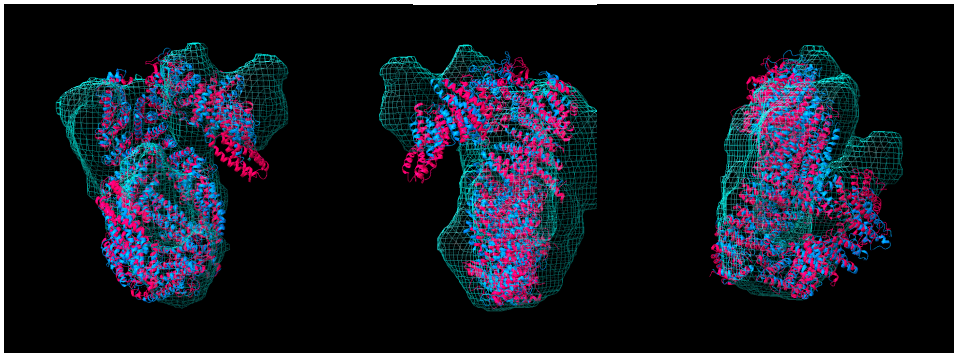


Figure 3-29: Overlay of initial negative staining model of FANCD2/FANCI with previously published cryo-EM structure and recently published cryo-EM structures of FANCD2/FANCI. a) Overlay of negative staining model (transparent yellow) with all cryo-EM published structures: Liang et al., 2016 (turquoise mesh) (EMD-8141), Wang et al., 2020 (pink) (PDB: 6VAD), Alcón et al., 2020 (blue) (PDB: 6TNG). b) Overlay of negative staining model (transparent yellow) with Liang et al., 2016 cryo-EM structure (turquoise mesh). c) Overlay of negative staining model (transparent yellow) with Wang et al., 2020 cryo-EM structure (pink). d) Overlay of negative staining model (transparent yellow) with Alcón et al., 2020 cryo-EM structure (blue). e) Overlay of negative staining model (transparent yellow) with Wang et al., 2020 cryo-EM structure (pink) and Alcón et al., 2020 cryo-EM structure (blue). f) Overlay of cryo-EM published structures: Liang et al., 2016 (turquoise mesh), Wang et al., 2020 (pink) and Alcón et al., 2020 (blue).

Figure 3-29a shows that there is a good agreement of the central portion of the FANCD2/FANCI structure across all four models/maps. In addition, it can be seen that the central pore where the DNA binds is hollow in all four models/maps. Figure 3-29b shows a direct overlay of the previously solved low resolution cryo-EM structure (Liang et al., 2016) and the negative staining model. There is fairly good agreement between the two models, however the Tower domain identified in the cryo-EM model cannot be seen in the negative staining model. Figures 3-29c and 3-29d display a direct overlay of the negative staining model and the Wang et al. and Alcón et al. structures respectively. Figure 3-29e displays the negative staining model overlaid with both of these newly solved cryo-EM FANCD2/FANCI structures. Figures 3-29c-e show that there is quite a good agreement between the negative staining model and the cryo-EM structures, and potentially better agreement than the negative staining model and the previous low resolution cryo-EM structure (Figure 3-29b), as the Tower domain identified in the previous cryo-EM model cannot be seen in the newly solved cryo-EM FANCD2/FANCI structures (in a similar way to the negative staining model). This may be due to the inherent conformational flexibility of the CTD, where the Tower domain is located (Swuec et al., 2017). However, the negative staining model is missing some density and also contains an extra bulge of density when viewed after the first 120°

rotation that is not present in any of the cryo-EM structures. However, as negative staining EM is a lower resolution technique it is unlikely to be in full agreement with the higher resolution cryo-EM structures. When comparing the two new cryo-EM structures with the previously solved low resolution cryo-EM structure, it can be seen that the Tower domain identified in the 2016 structure cannot be identified in the new cryo-EM structures, in a similar way to the X-ray crystal structure. As the Tower domain is part of the FANCD2 CTD, it is possible that the low-resolution structure reflects an alternative conformation, or an amalgamated representation of the flexible conformations of the CTD. As mentioned above, this could be attributed to the conformational flexibility of the CTD of FANCD2 (Swuec et al., 2017). In addition, there is missing density and several extra bulges of density in comparison to the high resolution cryo-EM structures, in a similar manner to the negative staining model.

The second released preprint manuscript from the Pavletich group, which currently remains unpublished, solves a mystery surrounding the sequence of the monoubiquitination events of FANCD2/FANCI. They describe that FANCD2 in complex with FANCI is monoubiquitinated first, followed by monoubiquitination of FANCI. In comparison to previous reports regarding limited FANCI monoubiquitination following FANCD2 monoubiquitination *in vitro*, Wang et al. report complete monoubiquitination of FANCI following monoubiquitination of FANCD2 in the FANCD2/FANCI complex using the complete human FA core complex expressed in HEK 293F cells. Wang et al.'s core complex structure reports that there are two copies of FANCB, FAAP100, FANCA, FANCG and FANCL, but only a single copy of FANCC, FANCE and FANCF. This is in contrast to the previously reported cryo-EM structure of the FA core complex by Shakeel et al., that shows only two copies of FANCB, FAAP100 and FANCL. However, it was also reported that a smaller subcomplex made up of two copies each of FANCB, FAAP100, FANCL and FANCG was also observed. The Shakeel et al., 2019 structural study also could not observe FANCA due to conformational heterogeneity. However, Wang et al. also report the FA core complex in complex with FANCD2/FANCI and UBE2T, which elucidates the reason behind the unusual stoichiometry reported. Wang et al. show that the core of the complex is made up around a dimer of FANCB-FAAP100 heterodimers, consistent with Shakeel et al.. However, they then report that FANCC, FANCE, FANCF, FANCG and FANCL assemble on one side of the dimer, whilst on the other side only FANCG and FANCL assemble. This second copy of FANCL and FANCG display a different conformation to the other copies found in complex with FANCC, FANCE and FANCF. FANCA assembles at each FANCG, however again due to the different conformations of each copy of FANCG, they also adopt different conformations. Wang et al.

show that this unusual stoichiometry is crucial for the binding of the FANCD2/FANCI complex. The second copy of FANCL binds to FANCI, and to a lesser extent to FANCD2. In addition, it binds to the CTD of FANCE, which in turn interacts with FANCD2. This copy of FANCL also recruits UBE2T, placing its active site close to the FANCD2 ubiquitination site, indicating that FANCD2 is the first of the protomers to be monoubiquitinated. The first copy of FANCL cannot make these contacts with the FANCD2/FANCI complex, nor UBE2T, due to being sequestered by the FANCC, FANCE and FANCF subcomplex. Once FANCD2/FANCI bind to the core complex, the NTD-NTD interface of the heterodimer opens by a rotation of FANCI and FANCD2 by ~8°. This is due to the RING hairpin of the second copy of FANCL inserting into the interface. In addition, the FANCD2 NTD rotates ~17° away from FANCI, presumed to be a direct consequence of binding to the CTD of FANCE. These rotations result in a loss of the FANCD2/FANCI interface contacts, showing the complex poised to be monoubiquitinated.

The first preprint manuscript released by the Pavletich group contained the cryo-EM structure of the monoubiquitinated-FANCD2/monoubiquitinated-FANCI complex. The following Passmore group paper described cryo-EM structure of the intermediate, monoubiquitinated-FANCD2/FANCI. The subsequent publication of the original preprint manuscript from the Pavletich group contained both the fully monoubiquitinated-FANCD2/monoubiquitinated-FANCI as well as the intermediate, monoubiquitinated-FANCD2/FANCI. However, Wang et al. report that there are no major structural differences between the intermediate and the fully ubiquitinated complex, except localised around the empty FANCI ubiquitin binding site. However, across both papers, the same conclusion is drawn, that the monoubiquitination of FANCD2 results in closure of the FANCD2/FANCI complex, to form a ring around the DNA, transforming the conformation of the open non-ubiquitinated FANCD2/FANCI to the closed, locked, ubiquitinated state.

Wang et al. report the cryo-EM structure for the monoubiquitinated-FANCD2/monoubiquitinated-FANCI complex bound to four different DNA molecules, ICL DNA, 5'flap DNA, nicked DNA and dsDNA. However, the best resolution structure was obtained with the nicked DNA complex, leading to a 3.5 Å reconstruction. The monoubiquitination of the complex causes a conformational change central to the two ubiquitin molecules, whereby the ubiquitin conjugated to one protomer, binds to the other protomer in a reciprocal fashion. Following ubiquitination, the FANCD2/FANCI interface opens through a rotation of FANCI and FANCD2 by 59°. Alcón et al. report that the intermediate,

monoubiquitinated-FANCD2/FANCI undergoes a rotation of almost 70°. None of the intermolecular contacts from the non-ubiquitinated dimer interface remain after ubiquitination. The rotation of the two protomers results in a repositioning of the CTD domains, which swing towards each other. Previously unstructured C-terminal sections of FANCD2 and FANCI form an intermolecular β-sheet zipper. Once the zipper is formed, this completes the rotation of the complex, to result in locking onto DNA. The importance of this zipper region is highlighted by the position of the FANCI residue R1285 (located within the zipper) whereby an arginine to glutamine mutation is associated with FA (Colnaghi et al., 2011; Dorsman et al., 2007; Sato et al., 2012b; Smogorzewska et al., 2007; Yuan et al., 2009). The conformational change to the ubiquitinated complex is largely attributed to FANCD2, as the FANCI protomer does not undergo a substantial conformational change upon ubiquitination. The FANCD2 CTD rotates by 20° to form the zipper interface, and its NTD rotates 38° towards FANCI which allows FANCD2 to interact with the FANCI ubiquitin moiety. Both FANCD2 and FANCI interact with the other protomer's conjugated ubiquitin via their NTD solenoid residues. These residues cover the entire hydrophobic patch of ubiquitin. Alcón et al. report that as a result of the new CTD conformations, the FANCD2 ubiquitination site becomes accessible, and upon ubiquitination, the ubiquitin docks onto the surface of FANCI, likely stabilising the structure. Both Wang et al. and Alcón et al. report that the closure of the ubiquitinated FANCD2/FANCI complex remodels the bound DNA, converting the non-colinear DNA duplexes into a continuous, but kinked DNA duplex. The DNA is bound at either end via the FANCI groove and the CTD of FANCD2. These new DNA contacts result in a loss of ssDNA binding, and thus the ubiquitination of FANCD2/FANCI is reported to result in a loss of its preference for branched DNA structures, which in turn allows the ubiquitinated complex to slide along the DNA. Perhaps this ability to move away from the ICL whilst remaining clamped on the DNA allows the ubiquitinated complex to mediate interactions with downstream repair factors whilst allowing these factors access to the ICL, or alternatively this sliding ability may protect the enclosed DNA from surrounding nucleases.

Once the FANCD2/FANCI complex has been monoubiquitinated to form its intermediate, monoubiquitinated-FANCD2/FANCI, it is described that the resulting closed conformation cannot revert back to the non-ubiquitinated open, trough-like conformation due to the location of the FANCD2 ubiquitin moiety. This ubiquitin acts as a molecular pin at the now open FANCD2/FANCI dimer interface, wedging the NTD-NTD domains apart. This results in the FANCI ubiquitination site being exposed. However, how FANCI is subsequently ubiquitinated,

and which interactions with the core complex, FANCD2 and DNA are important, remain to be elucidated.

Figure 3-30 displays the overlay of the two newly solved cryo-EM structures (Wang et al.: yellow, Alcón et al.: grey) of monoubiquitinated-FANCD2/FANCI, and the negative staining model of monoubiquitinated-FANCD2/FANCI produced in this project (transparent pink). Figure 3-31 displays the overlay of the newly solved cryo-EM structure of the monoubiquitinated-FANCD2/monoubiquitinated-FANCI (Wang et al.: green) and the negative staining model of monoubiquitinated-FANCI/FANCD2 produced in this project (transparent orange), as well as an overlay of the new cryo-EM structures of the non-ubiquitinated FANCD2/FANCI (Wang et al.: pink, Alcón et al.: blue), the previously solved cryo-EM structure (Liang et al.: turquoise mesh) of non-ubiquitinated FANCD2/FANCI and the negative staining model of monoubiquitinated-FANCI/FANCD2 produced in this project (transparent orange).

Figure 3-30a displays the negative staining model of monoubiquitinated-FANCD2/FANCI overlaid over both newly solved cryo-EM structures. Figures 3-30b and 3-30c show a direct overlay of the negative staining model and the Wang et al. and Alcón et al. structures respectively. It is evident that the negative staining model does not show good agreement with the cryo-EM structures. The model appears to have a more open conformation than the cryo-EM structures however does retain the DNA binding channel seen in the cryo-EM structures. It can be seen that in the second and third 120° rotations, there are large spaces of density in the model that are not filled by the cryo-EM maps, and also large patches of missing density where the cryo-EM structures occupy. This could be due to an orientation preference of the monoubiquitinated-FANCD2/FANCI on the negative staining grids, resulting in a biased 3D reconstruction of certain orientations of the complex leading to extra and missing density. Figure 3-30d compares both Wang et al. and Alcón et al.'s monoubiquitinated-FANCD2/FANCI cryo-EM structures, displaying good agreement showing a more closed structure than the open non-ubiquitinated FANCD2/FANCI.

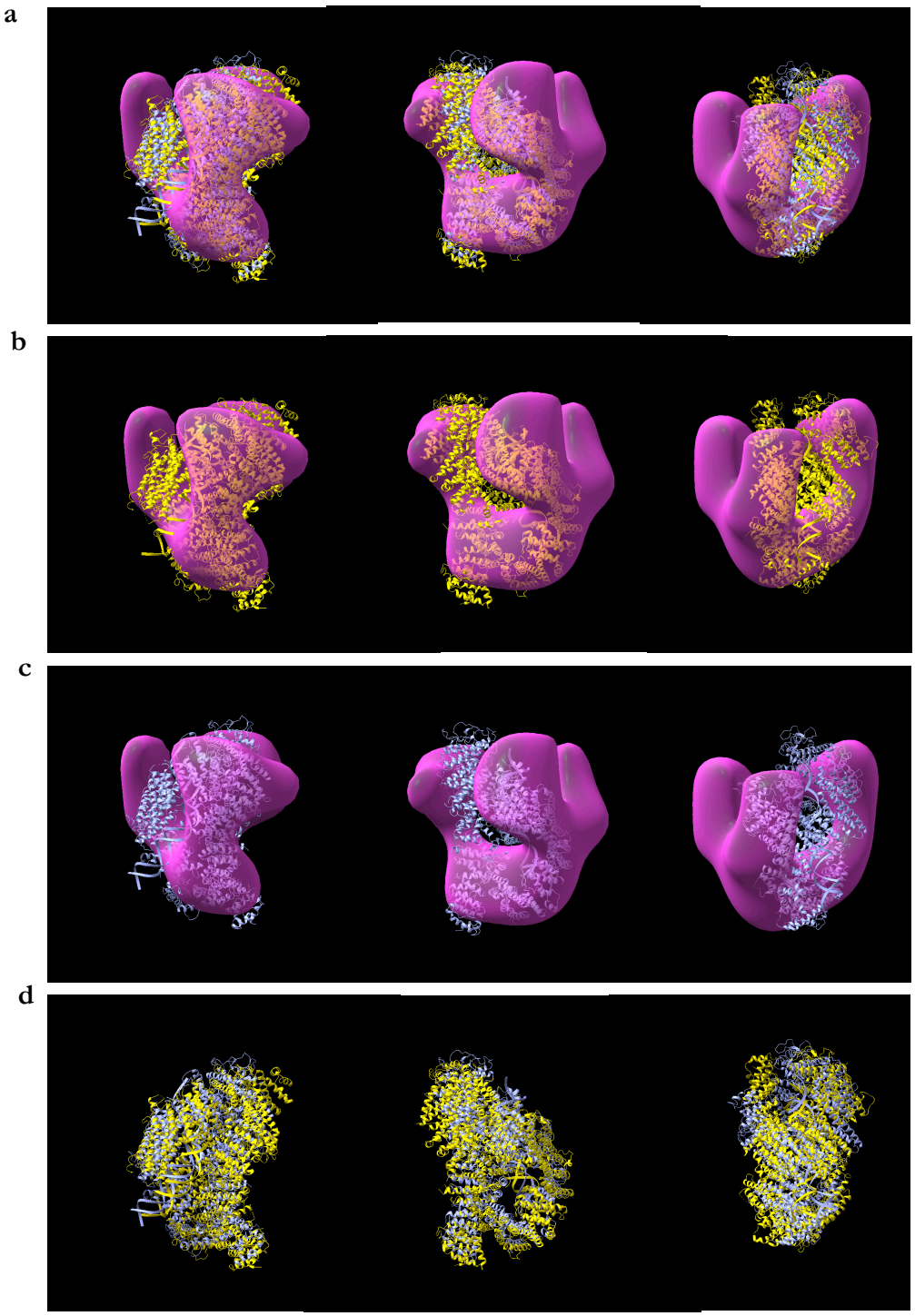


Figure 3-30: Overlay of initial negative staining model of monoubiquitinated-FANCD2/FANCI with recently published cryo-EM structures of monoubiquitinated-FANCD2/FANCI. a) Overlay of negative staining model (transparent pink) with both cryo-EM published structures: Wang et al., 2020 (yellow) (PDB: 6VAF), Alcón et al., 2020 (grey) (PDB: 6TNF). b) Overlay of negative staining model (transparent pink) with Wang et al., 2020 cryo-EM structure (yellow). c) Overlay of negative staining model (transparent pink) with Alcón et al., 2020 cryo-EM structure (grey). d) Overlay of both cryo-EM published structures: Wang et al., 2020 (yellow) and Alcón et al., 2020 (grey).

As the true intermediate of the monoubiquitination of FANCD2/FANCI has been revealed to be the monoubiquitinated-FANCD2/FANCI, the alternative intermediate investigated in this project, the monoubiquitinated-FANCI/FANCD2, has no cryo-EM map for direct comparison. Therefore, the negative staining model of monoubiquitinated-FANCI/FANCD2 was compared to the monoubiquitinated-FANCD2/monoubiquitinated-FANCI cryo-EM structure (Figure 3-31a). However, as the monoubiquitination of FANCD2 results in the major conformational change in the complex, and the monoubiquitination of FANCI does not contribute significantly, the negative staining model was also compared to the non-ubiquitinated FANCD2/FANCI cryo-EM structures (Figure 3-31b – c).

Figure 3-31a shows good agreement between the monoubiquitinated-FANCD2/monoubiquitinated-FANCI cryo-EM structure and the negative staining model. At the top of the model there is a bulge of extra density that the cryo-EM map does not occupy, however, this is where the non-ubiquitinated-FANCD2 would be situated. This is evidenced in Figures 3-31b, c and d, whereby this bulge of density is occupied by the non-ubiquitinated FANCD2/FANCI cryo-EM maps. Figure 3-31a also shows a patch of missing density for the model, presumably due to the rotation caused by the ubiquitination of FANCD2, and thus would not be seen in the model. Figure 3-31b displays the negative staining model overlaid with the previously solved low resolution cryo-EM structure of FANCD2/FANCI (Liang et al., 2016). It can be seen that there is some agreement, however there are large patches of extra density seen for both the model and the cryo-EM structure. Once again, the density of the Tower domain depicted in the cryo-EM structure cannot be seen in the negative staining model, in a similar manner to the newly published cryo-EM structures. Figures 3-31c and 3-31d display the negative staining model overlaid with the cryo-EM structure of FANCD2/FANCI from the Pavletich and Passmore groups respectively. There is good agreement of these non-ubiquitinated maps and the negative staining model, except for some missing density in the model likely owing to the non-ubiquitinated FANCI. This is presumably due to the slight rotation of FANCI upon ubiquitination into a slightly more closed state, resulting in a more closed conformation. This is also consistent with Figure 3-31a, whereby this density is not missing in the monoubiquitinated-FANCD2/monoubiquitinated-FANCI structure. Interestingly, in early 2020, Rennie et al. reported that the monoubiquitination of FANCI may play an important role in protecting monoubiquitinated-FANCD2 from deubiquitination by USP1/UAF1. In addition, they observed that FANCD2 is deubiquitinated faster than FANCI in the monoubiquitinated-FANCD2/monoubiquitinated-FANCI complex, leading to the formation of a

monoubiquitinated-FANCI/FANCD2 intermediate complex. Therefore, the importance of the monoubiquitinated-FANCI/FANCD2 complex remains to be investigated. However, this project has provided low resolution structural information for this potential intermediate.

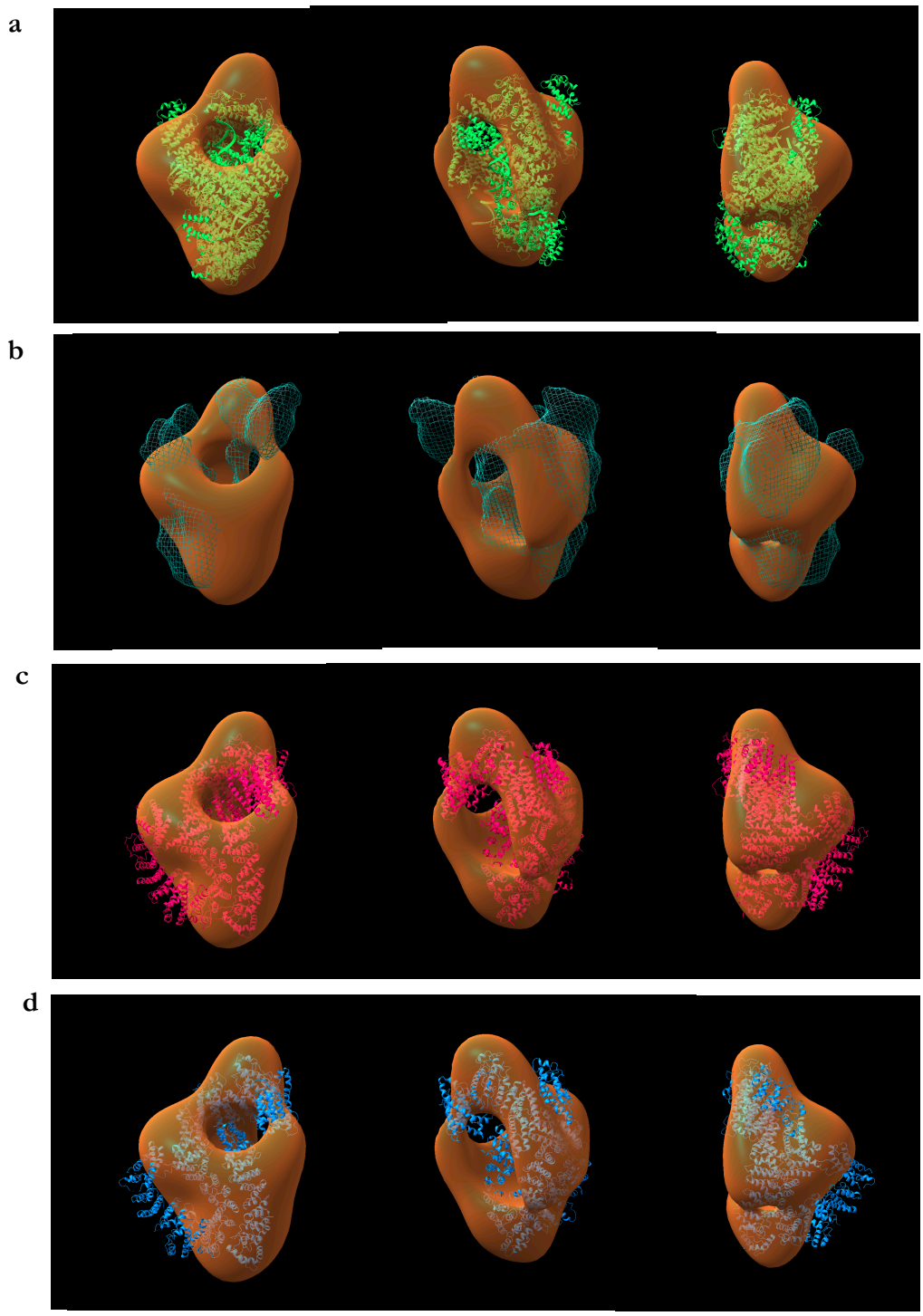


Figure 3-31: Overlay of initial negative staining model of monoubiquitinated-FANCI/FANCD2 with recently published cryo-EM structure of the monoubiquitinated-FANCD2/monoubiquitinated-FANCI, and previously and recently published cryo-EM structures of FANCD2/FANCI. a) Overlay of negative staining model (transparent orange) with cryo-EM monoubiquitinated-FANCD2/monoubiquitinated-FANCI published structure by Wang et al., 2020 (green) (PDB: 6VAE). b) Overlay of negative staining model (transparent orange) with Liang et al., 2016 cryo-EM FANCD2/FANCI structure (turquoise mesh). c) Overlay of negative staining model (transparent orange) with Wang et al., 2020 cryo-EM FANCD2/FANCI structure (pink). d) Overlay

of negative staining model (transparent orange) with Alcón et al., 2020 cryo-EM FANCD2/FANCI structure (blue).

Monoubiquitination of FANCD2 in the FANCD2/FANCI complex does not appear to cause filamentous arrays

Filamentous protein arrays are present in a number of processes in the cell. For example, the bacterial protein, RecA, and its eukaryotic homologue, RAD51, both form nucleoprotein filaments, which protect DNA from degradation by nucleases during HR (Yu et al., 2001). This filamentous formation therefore plays a role in DNA repair and genome stability (Lin et al., 2006). Therefore, the possibility that FANCD2/FANCI forms filamentous arrays upon ubiquitination was an exciting prospect. This filament formation could have perhaps also acted to protect the DNA, in a similar manner to RecA/RAD51, protecting newly synthesized DNA at replication forks. However, through the series of control experiments, it is clear that the filamentous arrays observed here are not formed upon FANCD2/FANCI recruitment to DNA (both linear DNA and scDNA), nor are they formed upon formation of the intermediate monoubiquitinated-FANCD2/FANCI complex when analyzing the crude ubiquitination reaction. It appears that the filament formation observed is due to the addition of the SmaI enzyme.

Interestingly, as mentioned above, in March 2020, Tan et al. reported the observance of filamentous arrays of the monoubiquitinated FANCD2/monoubiquitinated-FANCI clamped on to DNA. However, they note that the filamentous arrays formed do not actually represent fibrous filaments, as they do not show clear fibre diffraction. This observation of filamentous arrays is in contrast to the evidence presented here. However, this discrepancy could be explained by the nature of the purification procedure used as it was only possible for Tan, et al. to purify the fully monoubiquitinated-FANCD2/monoubiquitinated-FANCI complex and to compare this to the crude non-ubiquitinated FANCD2/FANCI. The purification method involved purifying the ubiquitinated complex through an Avi-ubiquitin tag that contained a 3C protease site between the Avidin-tag and the N-terminus of ubiquitin. The ubiquitinated complex was then purified via the Avidin-sepharose interaction. Once the complex was bound to the bead, the Avidin-tag was cleaved using 3C protease to release the ubiquitinated complex from the sepharose beads. However, they found that when using this purification method, they could not obtain the monoubiquitinated-FANCD2/FANCI intermediate complex, which was presumed to be due to the N-terminus of the FANCD2-ubiquitin (and thus the Avidin-tag) being buried within the complex. Therefore, it is possible that only the fully ubiquitinated, monoubiquitinated-

FANCD2/monoubiquitinated-FANCI complex can form filamentous arrays, whilst the intermediate monoubiquitinated-FANCD2/FANCI complex cannot. This would imply that the monoubiquitination of FANCI plays a crucial role in filamentous array formation.

Another possibility is that the due to the purification method, once the complex is locked onto DNA upon ubiquitination, all of the complexes on the DNA slid along the DNA ending up close together to produce what appears to be a filament. This sliding mechanism of the fully monoubiquitinated complex is presented in Wang et al., whereby they found that the monoubiquitinated-FANCD2/monoubiquitinated-FANCI complex dissociated from linear DNA not from the complex opening up, but through the complex sliding off the end of the DNA. As the DNA used during Tan et al.'s purification was circular plasmid DNA, the complex could not fall off, and thus may have slid together to form a filamentous array. Another possibility is that the filamentous arrays formed in this study are an artefact due to the purification method, in a similar way presented here with regard to the addition of SmaI. There is a possibility that through the addition of the 3C protease, filamentous arrays form. Unfortunately, a true negative control of this method is not included in the report whereby just the 3C protease is analysed. In addition, the negative control of the non-ubiquitinated FANCD2/FANCI cannot be purified in the same manner to the ubiquitinated complex, and so is not a direct comparison to the ubiquitinated complex. Therefore, the possibility that these filaments are artefacts of the purification method cannot be ruled out.

4. Structural Studies of UHRF1

UHRF1 has previously been identified as an ICL sensor protein involved in the FA pathway and has been shown to promote repair through the recruitment of FANCD2 to ICL sites (Liang et al., 2015a; Liang et al., 2015b). In addition, UHRF1 has been implicated in the monoubiquitination of FANCD2/FANCI complex *in vivo* (Motnenko et al., 2018). However, how UHRF1 senses the ICL, as well as how UHRF1 interacts with the ICL remains to be elucidated. In addition, how this interaction results in FANCD2 recruitment remains to be explored. Despite the extensive structural data available for UHRF1, no full-length structure is available. However, from the structural data available for the isolated UHRF1 domains, it is clear that there are extensive intramolecular contacts when UHRF1 is in its free state, but that these contacts are disrupted upon DNA and/or histone binding (Fang et al., 2016; Gao et al., 2018; Kori et al., 2019). Therefore, the full-length UHRF1 structure would provide insight not only into the interdomain connectivity of the protein but would also provide information regarding UHRF1's conformation in its free state. This structural data would be able to assist in gaining mechanistic insight into UHRF1's role in the FA pathway, in DSB repair and in DNA maintenance methylation.

This chapter investigates the role of UHRF1 in the monoubiquitination of FANCD2/FANCI *in vitro*. Specifically, the effect of the addition of UHRF1 to the ubiquitination assay is described, as well as the resulting change in monoubiquitination from the addition of excess scDNA. Whether UHRF1 and FANCD2/FANCI form a complex *in vitro* is also investigated, as well as the effect of the different UHRF1 domain deletions on the monoubiquitination of FANCD2/FANCI. Finally, structural studies are presented regarding UHRF1, providing the first structural insight into the full-length human protein, both through negative staining EM, as well as initial cryo-EM structural data.

4.1 Experimental Design

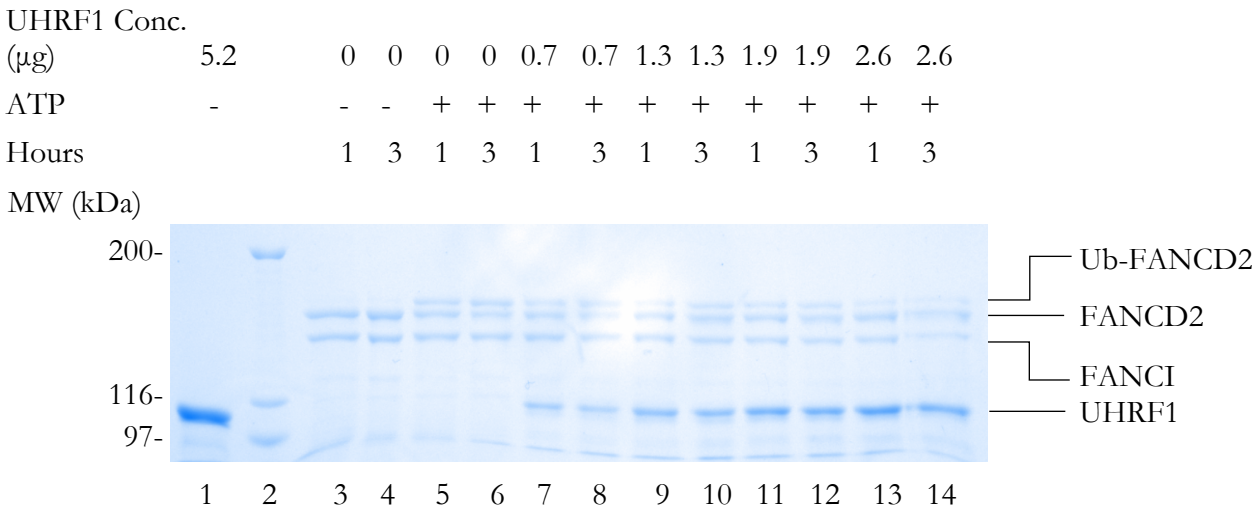
This project involves determining the structure of the full length human UHRF1 protein via cryo-EM. As with the approach for determining the FANCD2/FANCI complexes, this includes purification of the protein, assessment of the sample purity and homogeneity using EM through negative staining analysis and obtaining the structure via cryo-EM.

4.2 UHRF1 hinders the Ubiquitination of FANCD2/FANCI

UHRF1 has been reported to be essential for FANCD2 recruitment to ICLs *in vivo* (Liang et al., 2015). However, whether UHRF1 is essential for the recruitment of FANCD2 alone, or whether it plays a role in the ubiquitination of the FANCD2/FANCI complex has not been elucidated. To gain further mechanistic understanding of the role of UHRF1 in the FA pathway and whether it has an effect on the monoubiquitination of the FANCD2/FANCI complex *in vitro*, the ubiquitination assay was conducted with varying UHRF1 concentrations to determine if there was enhanced ubiquitination efficiency.

The ubiquitination assay was prepared by incubating in reaction buffer the appropriate concentrations of UBA1, UBE2T, FANCL, FANCD2/FANCI, ubiquitin, ATP and 2 μg of scDNA at room temperature. Varying concentrations of FLAG-HA-tagged UHRF1, expressed in Sf9 insect cells and purified via flag affinity purification, were also added. Samples were taken at 1 hour and 3 hours and analysed by SDS-PAGE. The gel bands corresponding to the 3 hour time points were quantified using Image J (Figure 4-1). Surprisingly, it is clear that increasing UHRF1 concentrations decreases the efficiency of monoubiquitination of FANCD2 in complex with FANCI *in vitro*. It is known that UHRF1 binds to DNA, interacts with FANCD2, and is also an E3 RING ligase (Citterio et al., 2004, Liang et al., 2015b; Mori et al., 2004). Therefore, the mechanism of this decreased monoubiquitination efficiency could be due to a number of factors, including the UHRF1 protein sequestering the DNA, FANCD2 or ubiquitin, or a combination of these factors. To address this, first the investigation of whether UHRF1 sequesters the DNA was investigated.

a



b

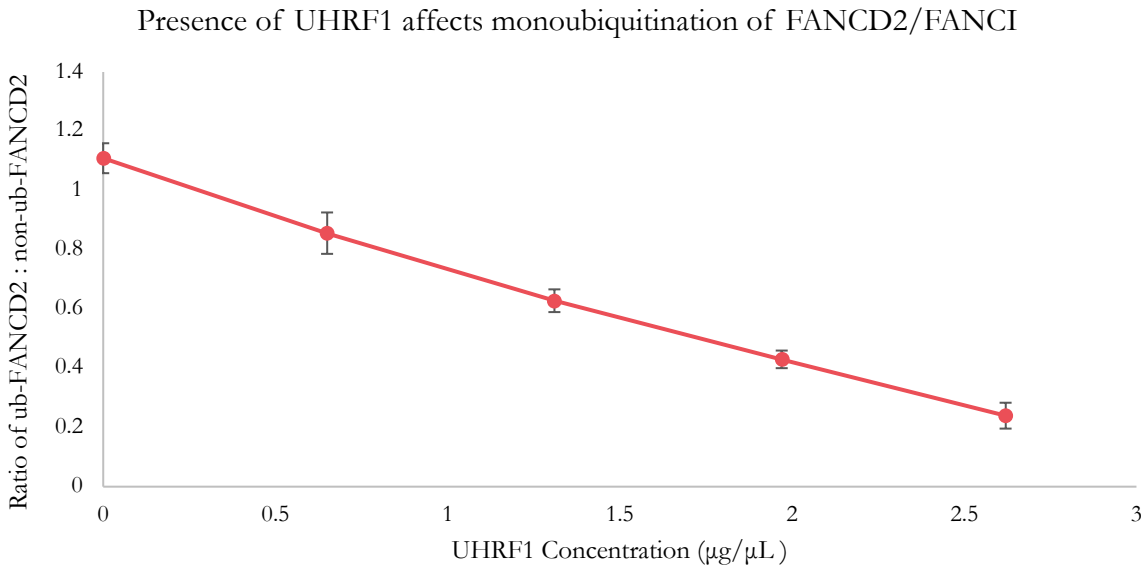
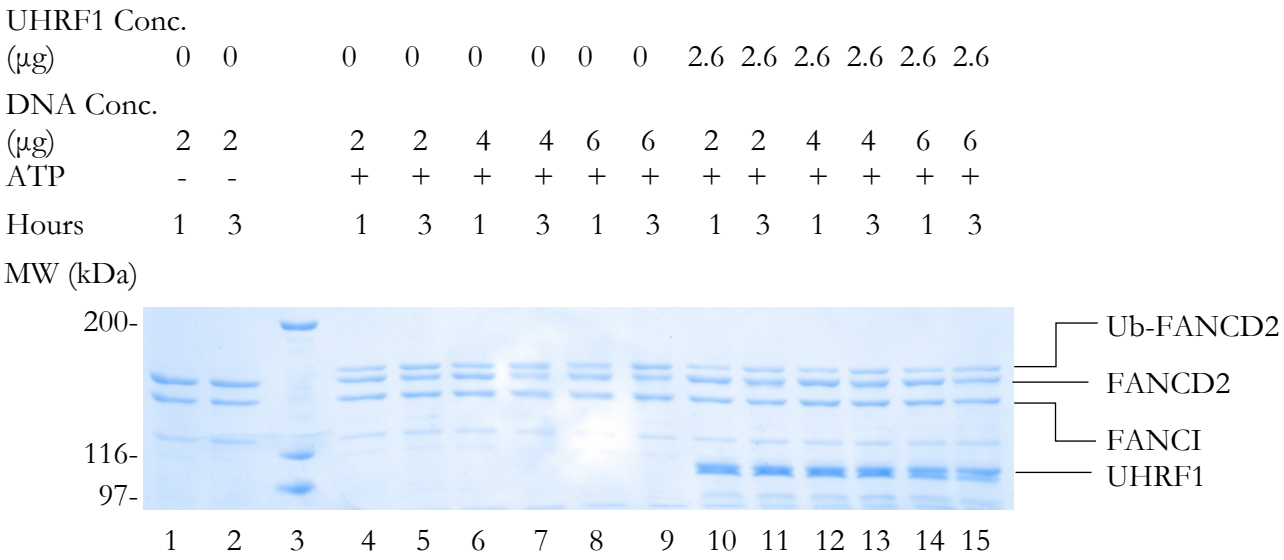


Figure 4-1: UHRF1 hinders monoubiquitination of FANCD2/FANCI in vitro. a) Representative Coomassie blue stain of monoubiquitination of FANCD2/FANCI with increasing UHRF1 concentration. Lane 1: UHRF1, lane 2: marker, lane 3: 1 hour negative control (no ATP), lane 4: 3 hours negative control (no ATP), lane 5: 1 hour post ubiquitination with 0.0 µg UHRF1, lane 6: 3 hours post ubiquitination with 0.0 µg UHRF1, lane 7: 1 hour post ubiquitination with 0.65 µg UHRF1, lane 8: 3 hours post ubiquitination with 0.65 µg UHRF1, lane 9: 1 hour post ubiquitination with 1.31 µg UHRF1, lane 10: 3 hours post ubiquitination with 1.31 µg UHRF1, lane 11: 1 hour post ubiquitination with 1.7 µg UHRF1, lane 12: 3 hours post ubiquitination with 1.7 µg UHRF1, lane 13: 1 hour post ubiquitination with 2.62 µg UHRF1, lane 14: 3 hours post ubiquitination with 2.62 µg UHRF1, b) Quantification of monoubiquitination of FANCD2/FANCI with increasing UHRF1 concentration at time = 3 hours (Mean±SD, n=3).

To determine whether UHRF1 is sequestering the DNA from the ubiquitination reaction, the ubiquitination assay was conducted with increasing levels of scDNA, with and without UHRF1. The ubiquitination assay was prepared by incubating in reaction buffer the appropriate concentrations of UBA1, UBE2T, FANCL, FANCD2/FANCI, ubiquitin, ATP and either 2, 4 or 6 μg of scDNA at room temperature. Equal volumes of either buffer containing no UHRF1 or 2.6 μg of FLAG-HA-tagged UHRF1, was also added. Samples were taken at 1 hour and 3 hours and analysed by SDS-PAGE. The gel bands corresponding to the 3 hour time points were quantified using Image J (Figure 4-2). It is clear that increasing DNA concentration slightly increases the ubiquitination efficiency, irrespective of the presence of UHRF1. In addition, it is clear that by increasing the DNA concentration in the reaction containing UHRF1, it very slightly restores the ubiquitination efficiency to that of the reaction without UHRF1, however not to the full extent. This would imply that UHRF1 does sequester some of the DNA away from the monoubiquitination reaction potentially blocking or preventing important interactions necessary for ubiquitination, however it does not explain the full observation of decreased ubiquitination efficiency. Therefore, whether UHRF1 interacts directly with the FANCD2/FANCI complex was next investigated.

a



b

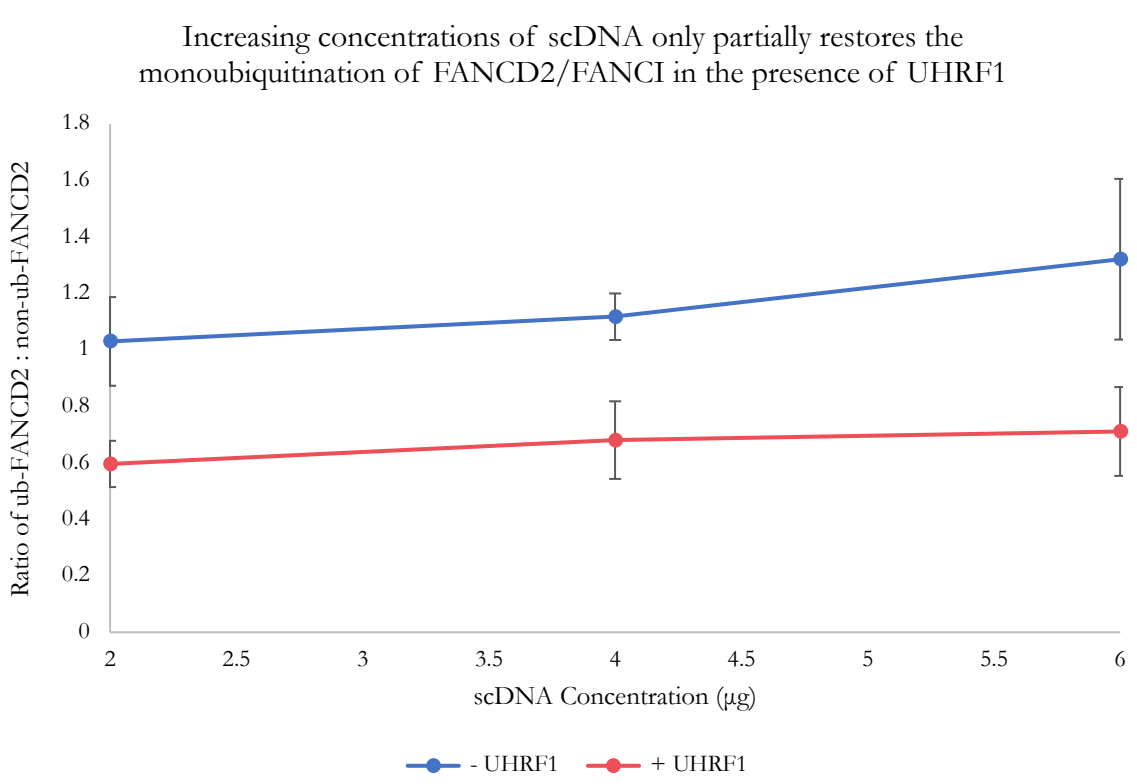
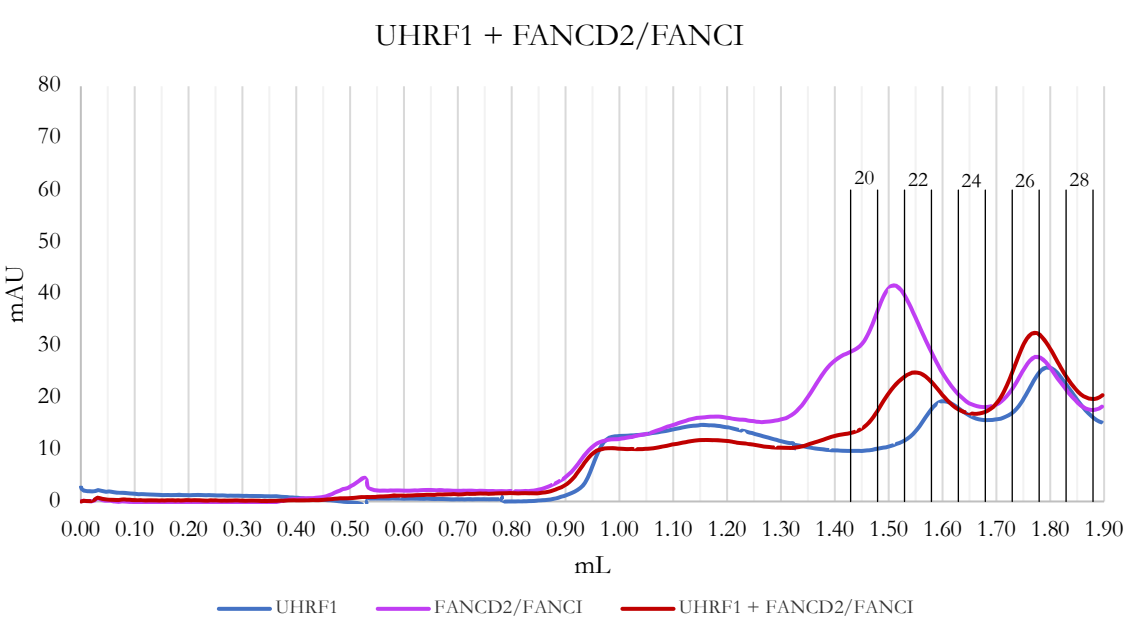


Figure 4-2: Increasing the concentration of scDNA only partially restores the monoubiquitination of FANCD2/FANCI in the presence of UHRF1. a) Representative Coomassie blue stain of monoubiquitination of FANCD2/FANCI with and without 2.6 µg UHRF1 with increasing scDNA concentration. Lane 1: 1 hour negative control (no ATP), lane 2: 3 hours negative control (no ATP), lane 3: marker, lane 4: 1 hour post ubiquitination with 2 µg scDNA (-UHRF1), lane 5: 3 hours post ubiquitination with 2 µg scDNA (-UHRF1), lane 6: 1 hour post ubiquitination with 4 µg scDNA (-UHRF1), lane 7: 3 hours post ubiquitination with 4 µg

scDNA (-UHRF1), lane 8: 1 hour post ubiquitination with 6 µg scDNA (-UHRF1), lane 9: 3 hour post ubiquitination with 6 µg DNA (-UHRF1), lane 10: 1 hour post ubiquitination with 2 µg scDNA (+UHRF1), lane 11: 3 hours post ubiquitination with 2 µg scDNA (+UHRF1), lane 12: 1 hour post ubiquitination with 4 µg scDNA (+UHRF1), lane 13: 3 hours post ubiquitination with 4 µg scDNA (+UHRF1), lane 14: 1 hour post ubiquitination with 6 µg scDNA (+UHRF1), lane 15: 3 hour post ubiquitination with 6 µg DNA (+UHRF1), b) Quantification of monoubiquitination of FANCD2/FANCI with or without UHRF1 with increasing scDNA concentration at time = 3 hours (Mean±SD, n=2).

As it is known that UHRF1 has a direct interaction with FANCD2, there is a possibility that URHF1 is sequestering the FANCD2/FANCI complex through a direct interaction, and potentially blocking ubiquitination. To determine whether there is a direct interaction, UHRF1 and FANCD2/FANCI were incubated at room temperature for three hours and then run on a Superose 6 3.2/300 to determine if a complex of UHRF1-FANCD2/FANCI is formed. In addition, FANCD2/FANCI alone and UHRF1 alone were run on the gel filtration column for comparison (Figure 4-3a). It can be seen that FANCD2/FANCI is eluted in fractions 20-22 (corresponding to a size of approximately 330 kDa) (Figure 4-3c), whilst UHRF1 is eluted in fractions 26-27 (corresponding to a size of approximately 100 kDa) (Figure 4-3b). However, when the combined sample was run on the column, no trimeric complex peak can be seen in the fractions prior to fraction 20, and FANCD2/FANCI can be visualised in the dimer fractions 20-22, and UHRF1 visualised faintly in the monomer fractions 26-27 (Figure 4-3d). This would indicate that UHRF1 does not interact directly with FANCD2/FANCI and would be unlikely to hinder the ubiquitination reaction. However, the protein concentration is very low and difficult to visualise via Coomassie stain, and thus if a low proportion of the UHRF1 and FANCD2/FANCI formed a trimeric complex, it would not be able to be seen. It would be useful to use a higher sensitivity visualization method such as western blot or silver stain to truly determine if there is a direct interaction between FANCD2/FANCI and UHRF1. Additionally, the formation of a complex is dependent on the concentration of the starting components, and thus perhaps if the concentration of UHRF1 and FANCD2/FANCI were increased, the association between the proteins would also increase. Another possibility is that UHRF1 and FANCD2/FANCI only form a trimeric complex when in the presence of DNA. To further investigate how UHRF1 hinders the monoubiquitination of FANCD2/FANCI, whether certain domains of UHRF1 are sequestering elements of the ubiquitination reaction, for example the RING domain interacting with ubiquitin, was investigated.

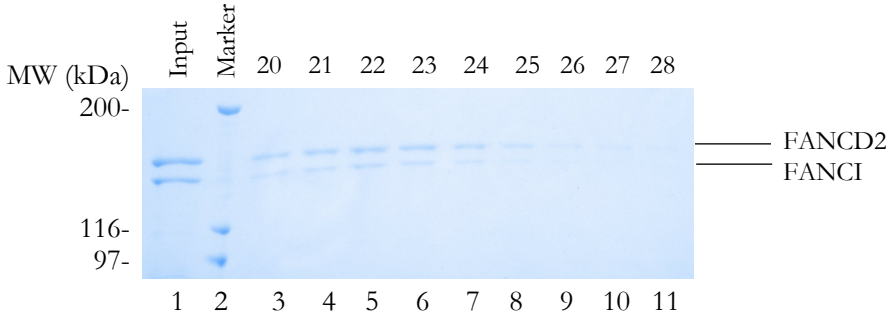
a


UHRF1 + FANCD2/FANCI
mAU
mL
20
22
24
26
28
UHRF1
FANCD2/FANCI
UHRF1 + FANCD2/FANCI

b


MW (kDa)
Input
Marker
UHRF1

c


MW (kDa)
Input
Marker
FANCD2
FANCI

d

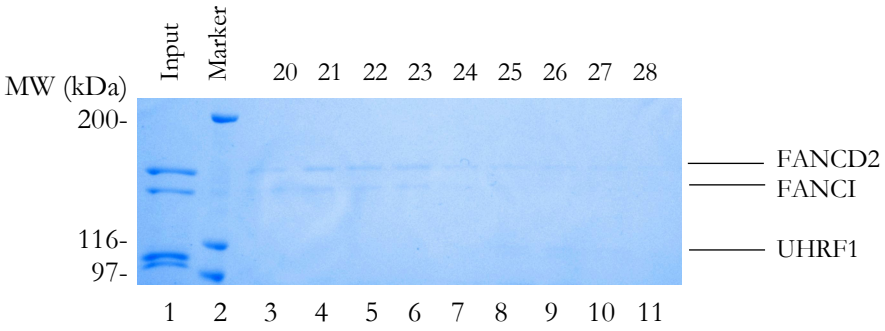

MW (kDa)
Input
Marker
FANCD2
FANCI
UHRF1

Figure 4-3: Purification of UHRF1, FANCD2/FANCI complex and UHRF1 + FANCD2/FANCI complex on a Superose 6 3.2/300. a) Chromatogram following purification on a Superose 6 3.2/300. UHRF1 identified in peak fractions 26-27, FANCD2/FANCI complex identified in peak fractions 20-23. b) Coomassie blue stain of UHRF1. Lane 1: input, lane 2: marker, lanes 3 – 11: gel filtration peak fractions (lane 3: peak fraction 20, lane 4: peak fraction 21, lane 5: peak fraction 22, lane 6: peak fraction 23, lane 7: peak fraction 24, lane 8: peak fraction 25, lane 9: peak fraction 26, lane 10: peak fraction 27, lane 11: peak fraction 28). c) Coomassie blue stain of the FANCD2/FANCI complex. Lane 1: input, lane 2: marker, lanes 3 – 11: gel filtration peak fractions (lane 3: peak fraction 20, lane 4: peak fraction 21, lane 5: peak fraction 22, lane 6: peak fraction 23, lane 7: peak fraction 24, lane 8: peak fraction 25, lane 9: peak fraction 26, lane 10: peak fraction 27, lane 11: peak fraction 28). d) Coomassie blue stain of UHRF1 incubated with the FANCD2/FANCI complex. Lane 1: input, lane 2: marker, lanes 3 – 11: gel filtration peak fractions (lane 3: peak fraction 20, lane 4: peak fraction 21, lane 5: peak fraction 22, lane 6: peak fraction 23, lane 7: peak fraction 24, lane 8: peak fraction 25, lane 9: peak fraction 26, lane 10: peak fraction 27, lane 11: peak fraction 28).

In order to investigate the role of the individual domains of UHRF1 in hindering the ubiquitination assay, the assay was conducted in the presence of UHRF1 domain deletion mutants. First, the mutants were expressed in Sf9 insect cells and purified via flag affinity purification (Figure 4-4). The ubiquitination assay was prepared as above by incubating in reaction buffer UBA1, UBE2T, FANCL, FANCD2/FANCI, ubiquitin, ATP, 2 μg of scDNA and equal volumes of either buffer containing no UHRF1, or 2.6 μg of full-length FLAG-HA-tagged UHRF1, FLAG-HA-tagged UHRF1-ΔUBL, FLAG-HA-tagged UHRF1-ΔTTD, FLAG-HA-tagged UHRF1-ΔPHD, FLAG-HA-tagged UHRF1-ΔSRA or FLAG-HA-tagged UHRF1-ΔRING, at room temperature. Samples were taken at 1 hour and 3 hours and analysed by SDS-PAGE. The gel bands corresponding to the 3 hour time points were quantified using Image J (Figure 4-5). The statistical analysis was completed on GraphPad Prism. It is clear that the addition of UHRF1 and UHRF1-ΔUBL statistically significantly reduces the monoubiquitination efficiency. In addition, the addition of UHRF1-ΔRING shows a statistically significant improvement in the monoubiquitination efficiency in comparison to addition of the full-length UHRF1. This would imply that the RING domain of UHRF1 is interacting with and sequestering components of the ubiquitination assay. Overall, in comparison to the addition of full-length UHRF1, the trend shows that deletion of the UBL, TTD, SRA and RING domains all result in an increased efficiency of ubiquitination. The addition of UHRF1-ΔPHD was also assessed and results in a further decreased efficiency in ubiquitination than full length UHRF1. However, it should be noted that the PHD domain deletion mutant protein contained many contaminants and thus these results must be interpreted

with caution. Therefore, the effect of the PHD domain deletion mutant protein on the monoubiquitination was not quantified and included in Figure 4-5b.

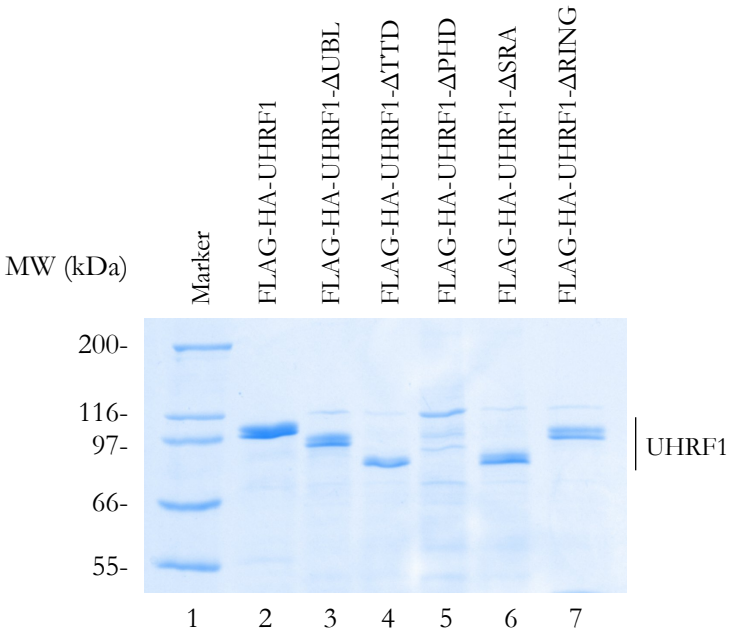
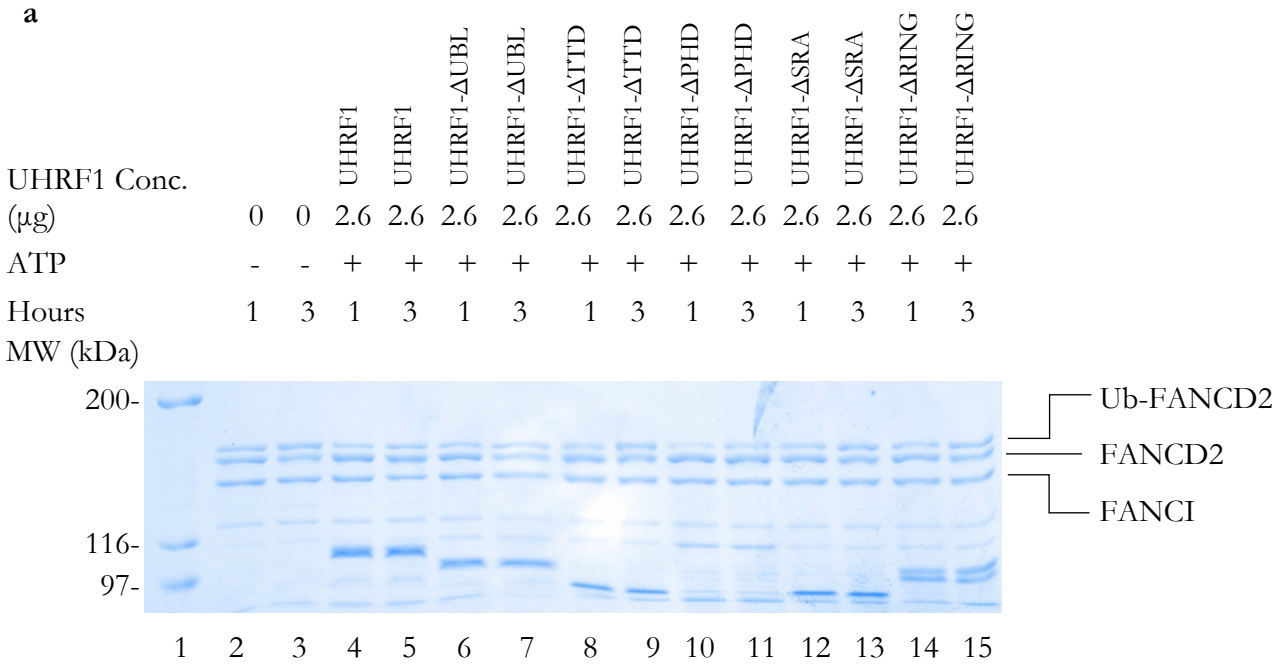


Figure 4-4: Purification of UHRF1 mutants. Coomassie blue stain of the purification of UHRF1 mutants. Lane 1: marker, lane 2: FLAG-HA-UHRF1, lane 3: FLAG-HA-UHRF1-⊿UBL, lane 4: FLAG-HA-UHRF1-⊿TTD, lane 5: FLAG-HA-UHRF1- ⊿PHD, lane 6: FLAG-HA-UHRF1-⊿SRA, lane 7: FLAG-HA-UHRF1-⊿RING.

a



b

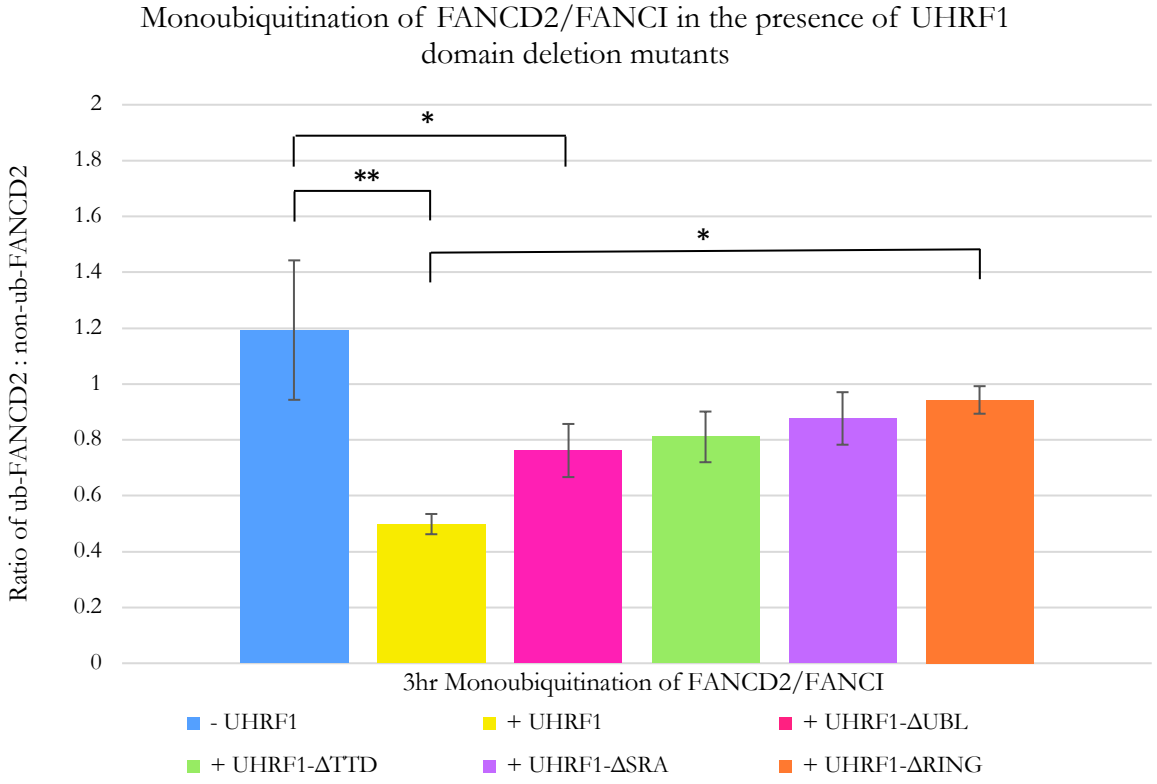


Figure 4-5: Monoubiquitination of FANCD2/FANCI in the presence of UHRF1 domain deletion mutants. a) Representative Coomassie blue stain of monoubiquitination of FANCD2/FANCI with 2.6 µg of UHRF1 domain deletion mutants. Lane 1: marker, lane 2: 1 hour post ubiquitination without UHRF1, lane 3: 3 hours

*post ubiquitination without UHRF1, lane 4: 1 hour post ubiquitination with UHRF1, lane 5: 3 hours post ubiquitination with UHRF1, lane 6: 1 hour post ubiquitination with UHRF1-⊿UBL, lane 7: 3 hours post ubiquitination with UHRF1-⊿UBL, lane 8: 1 hour post ubiquitination with UHRF1-⊿TTD, lane 9: 3 hours post ubiquitination with UHRF1-⊿TTD, lane 10: 1 hour post ubiquitination with UHRF1-⊿PHD, lane 11: 3 hours post ubiquitination with UHRF1-⊿PHD, lane 12: 1 hour post ubiquitination with UHRF1-⊿SRA, lane 13: 3 hours post ubiquitination with UHRF1-⊿SRA, lane 14: 1 hour post ubiquitination with UHRF1-⊿RING, lane 15: 3 hours post ubiquitination with UHRF1-⊿RING, b) Quantification of monoubiquitination of FANCD2/FANCI with UHRF1 domain deletion mutants at time = 3 hours (Mean±SD, n=2). One-way ANOVA against -UHRF1 and +UHRF1 control groups, * p< 0.05, ** p< 0.01.*

4.3 Towards a Cryo-EM Structure of UHRF1

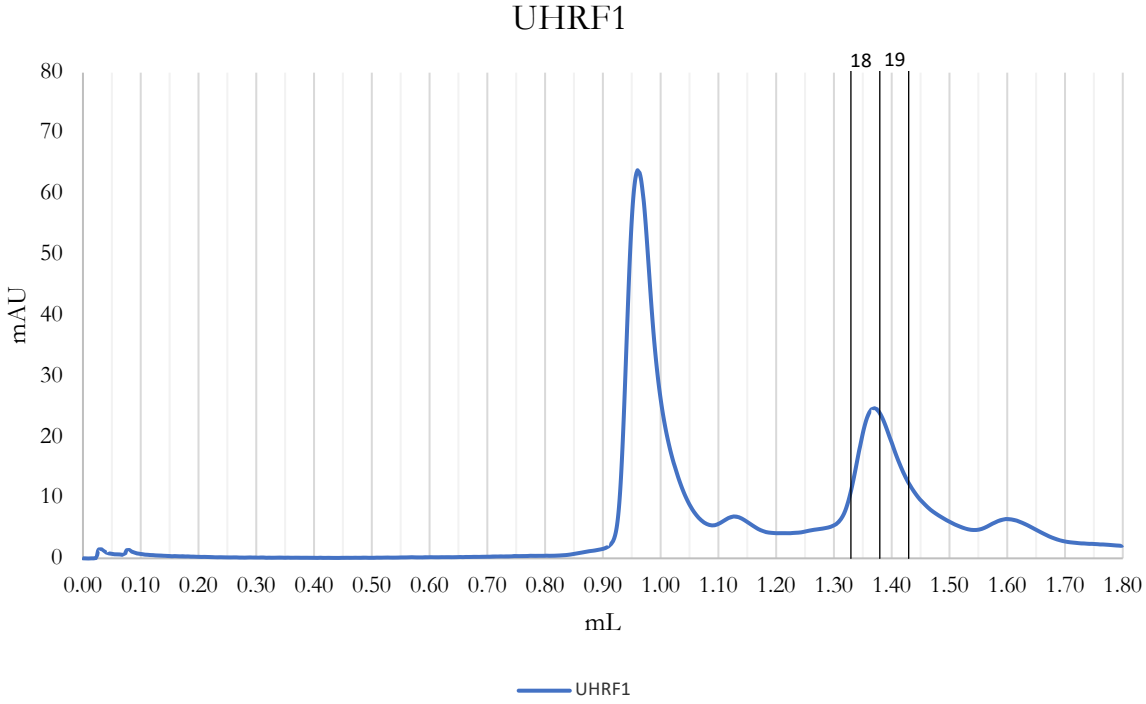
FLAG-HA-tagged UHRF1 was expressed in Sf9 insect cells and purified via flag affinity purification, as described above. The protein was then purified using size exclusion chromatography with two peaks observed, one for the void volume at fractions 10-12 and one for the UHRF1 sample, identified in fractions 18-19 (Figure 4-6). The resulting pure UHRF1 protein was applied to grids and subjected to negative staining microscopy to ensure good sample homogeneity and purity. A 2D classification of the complex was generated using EMAN2.2 software, and a 3D model generated and further refined (Figure 4-7).

As the sample showed good particle homogeneity and purity, the sample was prepared for cryo-EM. The sample was prepared as per the negative staining sample above, however prior to purification using size exclusion chromatography, the sample was applied to a 10 kDa concentrator. Once again, two peaks were observed with one for the void volume at fractions 10-12, and one for the UHRF1 sample, identified in fractions 18-25 (Figure 4-8). A variety of grids were used, including Quantifoil copper, graphene oxide and Quantifoil gold grids. For all grids, it appears that protein sample can be seen, however due to the small size of UHRF1, it is hard to discern whether the protein sample is present in a homogenous manner, or if the sample is denatured. For the Quantifoil copper grids, the sample can be seen across the grid and it appears that individual particles may be able to be seen. However, there is a high level of ethane contamination across this grid type (Figure 4-9a). For the graphene oxide grids, there appears to be a high concentration of sample across the grid, but it is unclear whether the particles are densely populated, or if the sample is denatured (Figure 4-9b). For the Quantifoil gold grids, the sample can be seen across the grid, and once again it appears that individual particles may be able to be seen (Figure 4-9c).

To be able to discern whether UHRF1 particles are truly present, the sample was prepared as above (Figure 4-10), screened using the Talos Arctica 200 kV microscope (Figure 4-11), and then followed by an initial screen using the higher voltage, Titan Krios 300 kV microscope (Figure 4-12a). A variety of grids were trialled, including the most promising Quantifoil copper (Figure 4-11a) and Quantifoil gold (Figure 4-11b) identified in the previous Arctica screen, as well as an additional type of grid, UltraAuFoil gold (Figure 4-11c). All grids looked promising on the Arctica as it seems that the sample can be seen on each grid, and that individual particles may also be able to be discerned. In addition, the ethane contamination seen on the Quantifoil copper grids seem

to be reduced (Figure 4-11a). The Quantifoil gold grid was subjected to the initial screen using the Krios 300 kV microscope (Figure 4-12a).

a



b

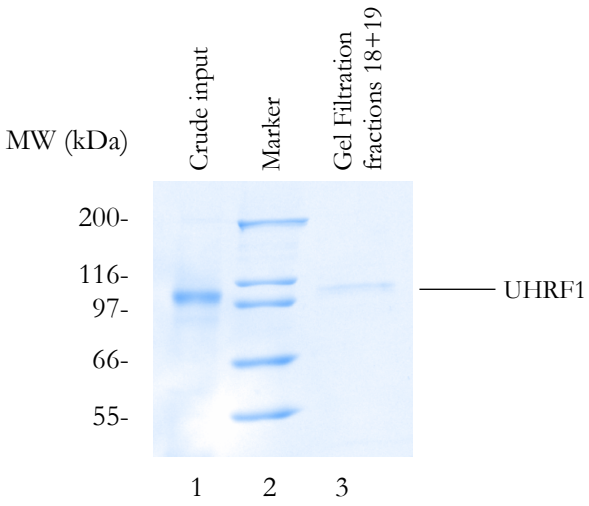
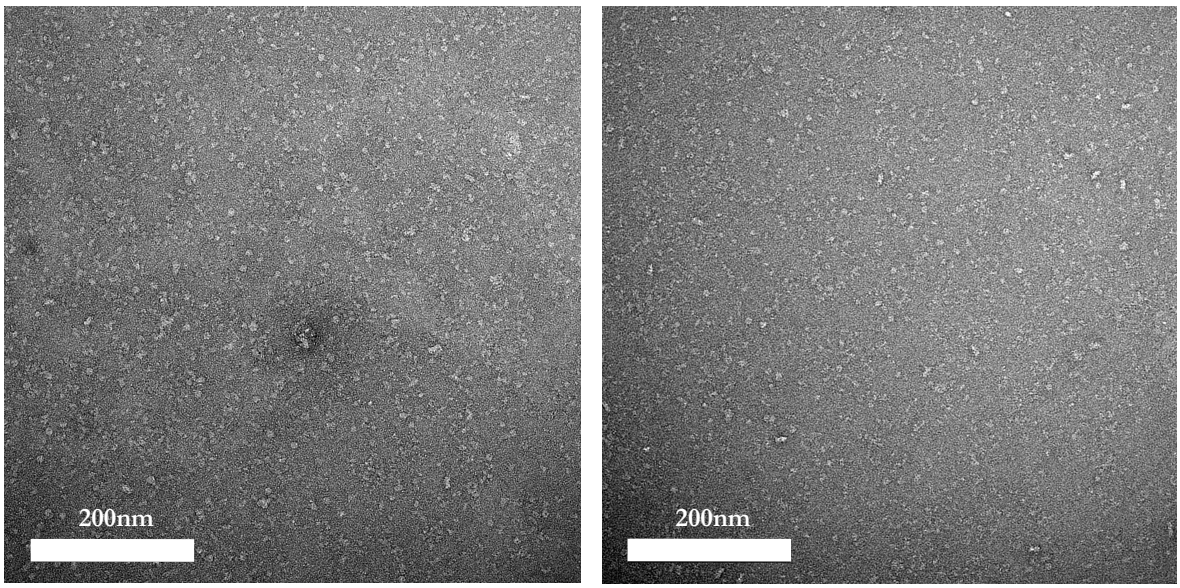
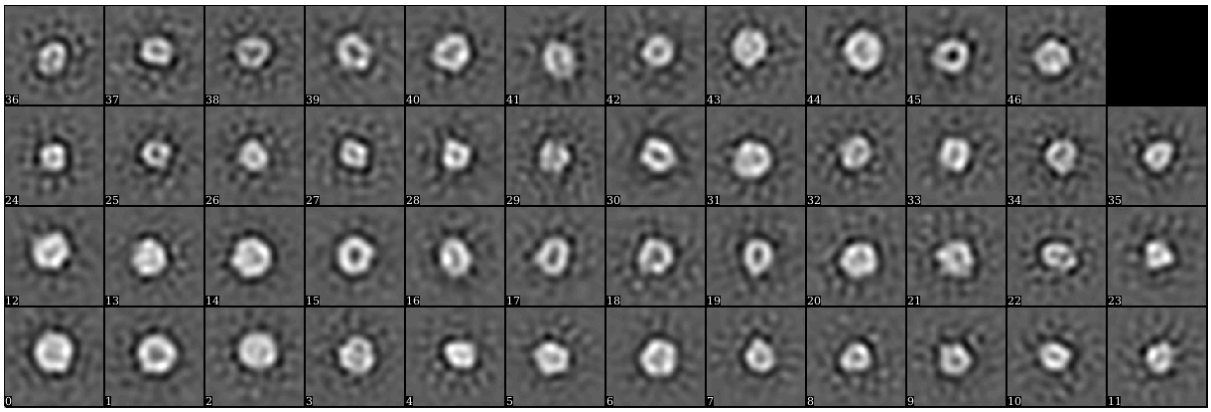


Figure 4-6: Purification of UHRF1 for negative staining microscopy. a) Chromatogram following purification on a Superdex 200 3.2/300. UHRF1 identified in peak fractions 18 and 19. b) Coomassie blue stain of the purification procedure showing UHRF1. Lane 1: crude purification product, lane 2: marker, lane 3: gel filtration peak fractions 18+19.

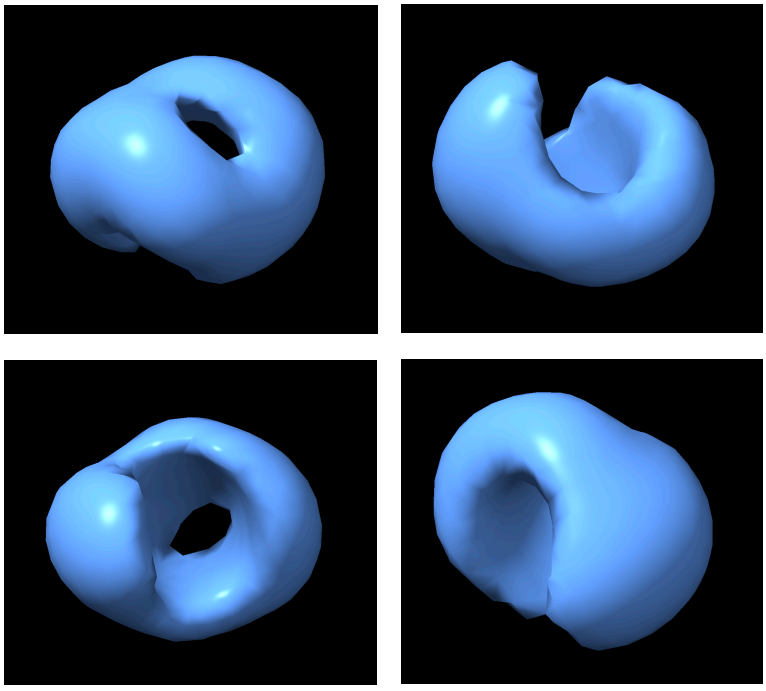
a



b



c



d

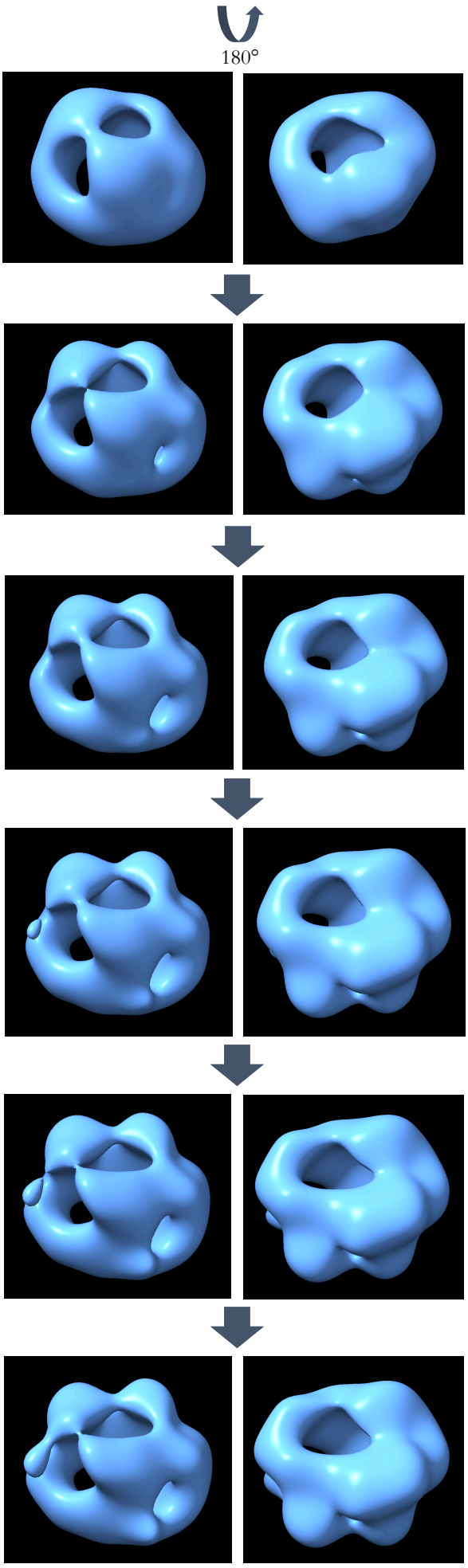
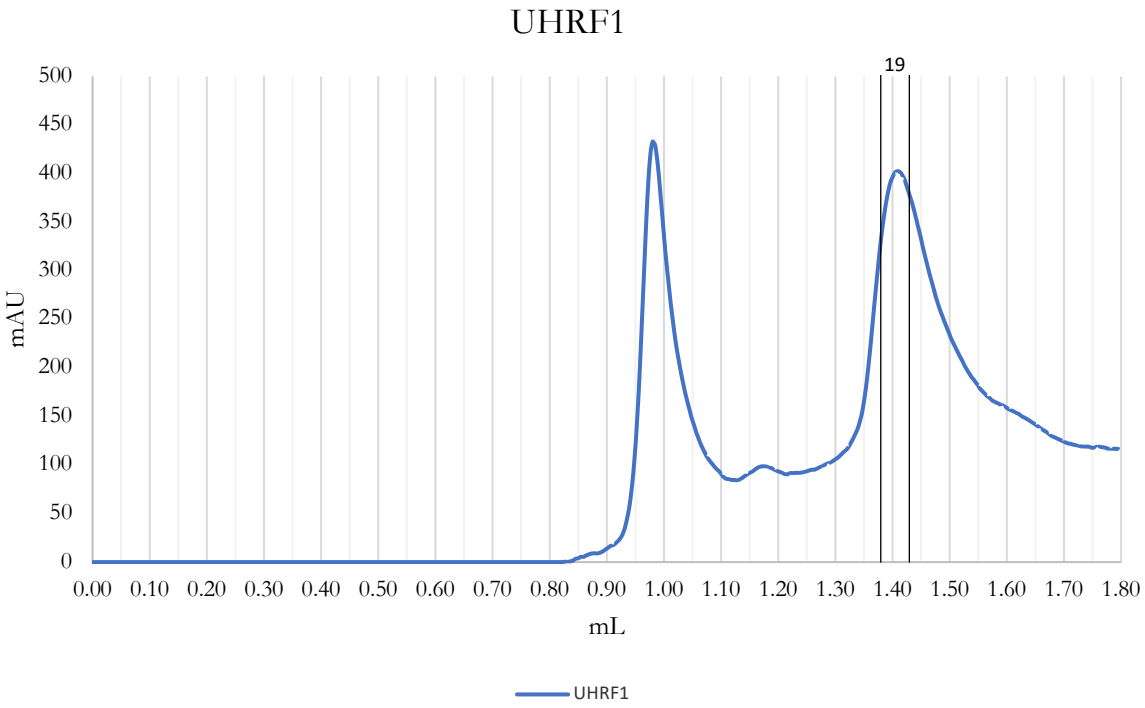


Figure 4-7: Negative staining of UHRF1 at 0.010 µg/µl. a) Representative micrographs showing the homogeneity of the UHRF1 sample from the combined 18 and 19 peak fractions from Figure 4-6 at 68 000x magnification, 1.727 Å/pixel. b) 2D classification of UHRF1 from 6648 particles using EMAN2.2. c) Initial model generated from 2D classification using EMAN2.2. d) Model refined step-wise to 17.7 Å using EMAN2.2.

a



b

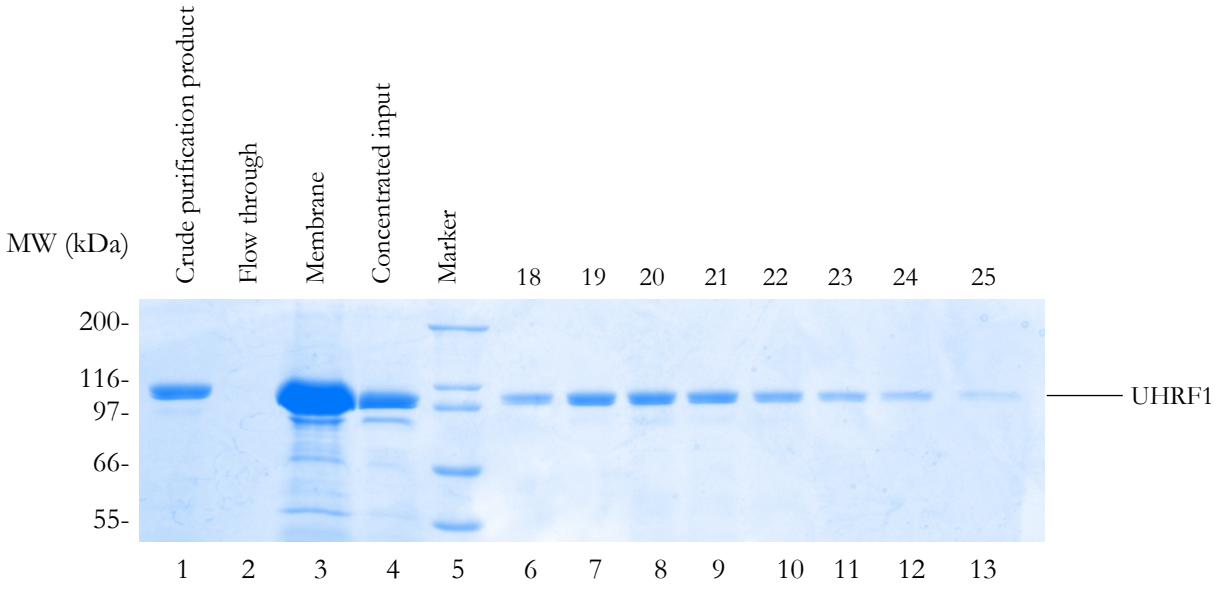
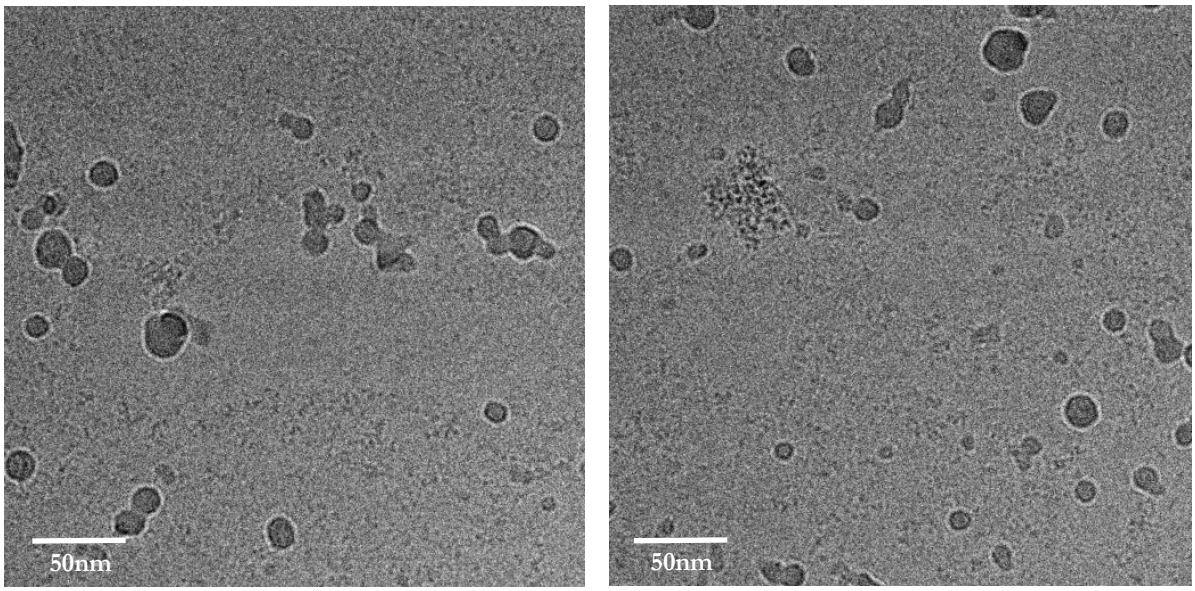
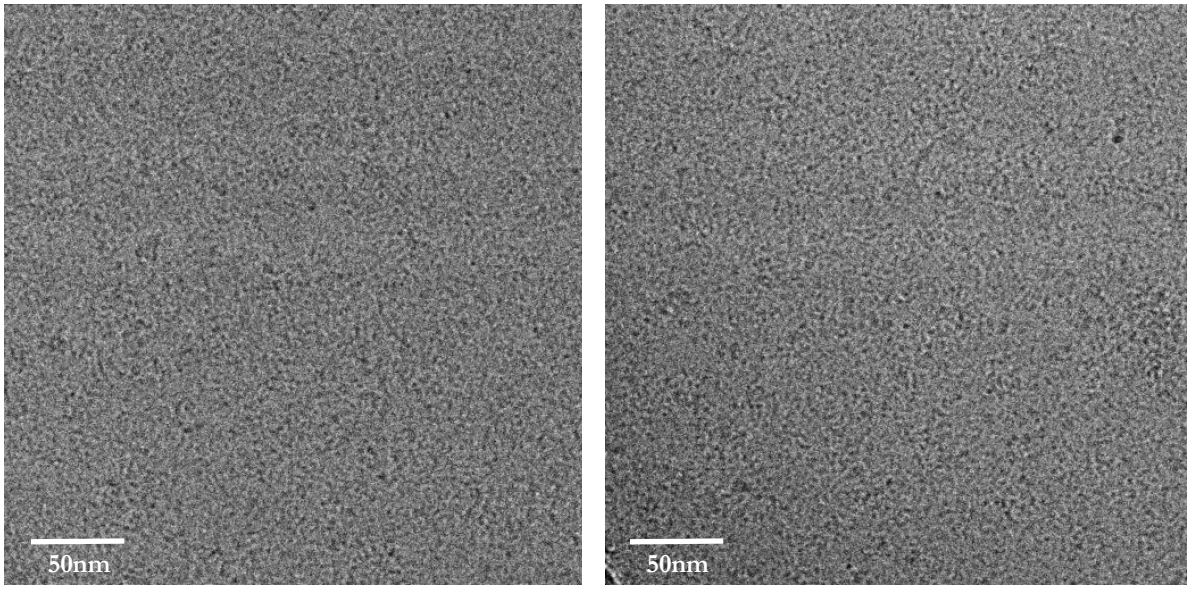


Figure 4-8: Purification of UHRF1 for Cryo-EM. a) Chromatogram following purification on a Superdex 200 3.2/300. UHRF1 identified in peak fractions 18 - 25. b) Coomassie blue stain of the purification procedure showing UHRF1. Lane 1: crude purification product, lane 2: flow through from 10 kDa cut-off concentrator, lane 3: sample stuck on the membrane, lane 4: sample loaded onto the gel filtration column, lane 5: marker, lanes 6 – 13: gel filtration peak fractions (lane 6: peak fraction 18, lane 7: peak fraction 19, lane 8: peak fraction 20, lane 9: peak fraction 21, lane 10: peak fraction 22, lane 11: peak fraction 23, lane 12: peak fraction 24, lane 13: peak fraction 25).

a



b



c

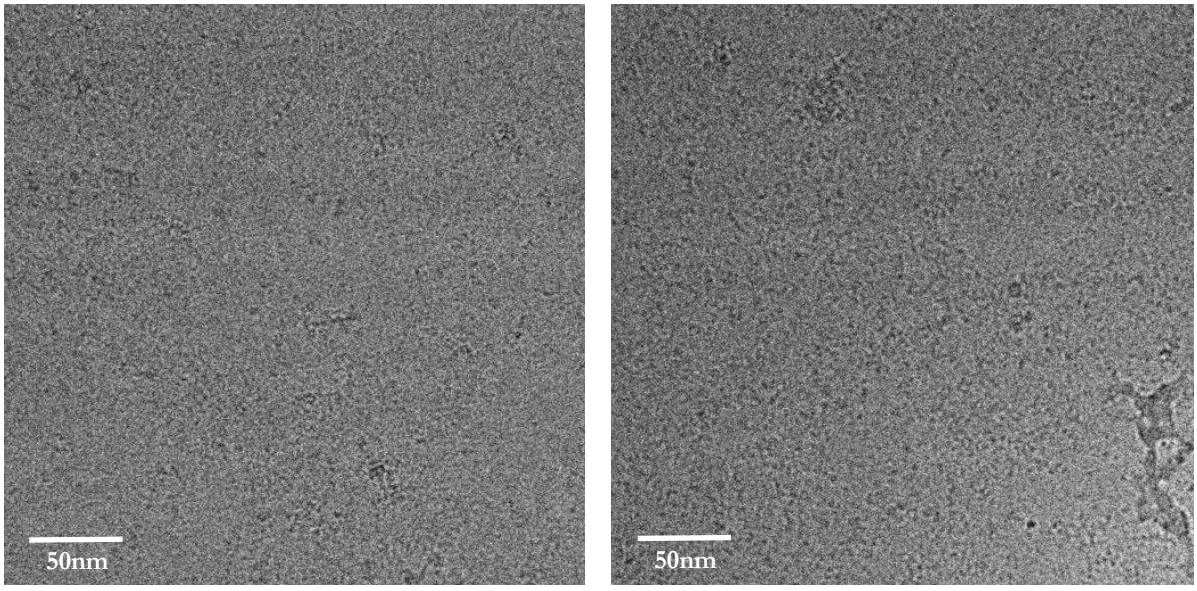
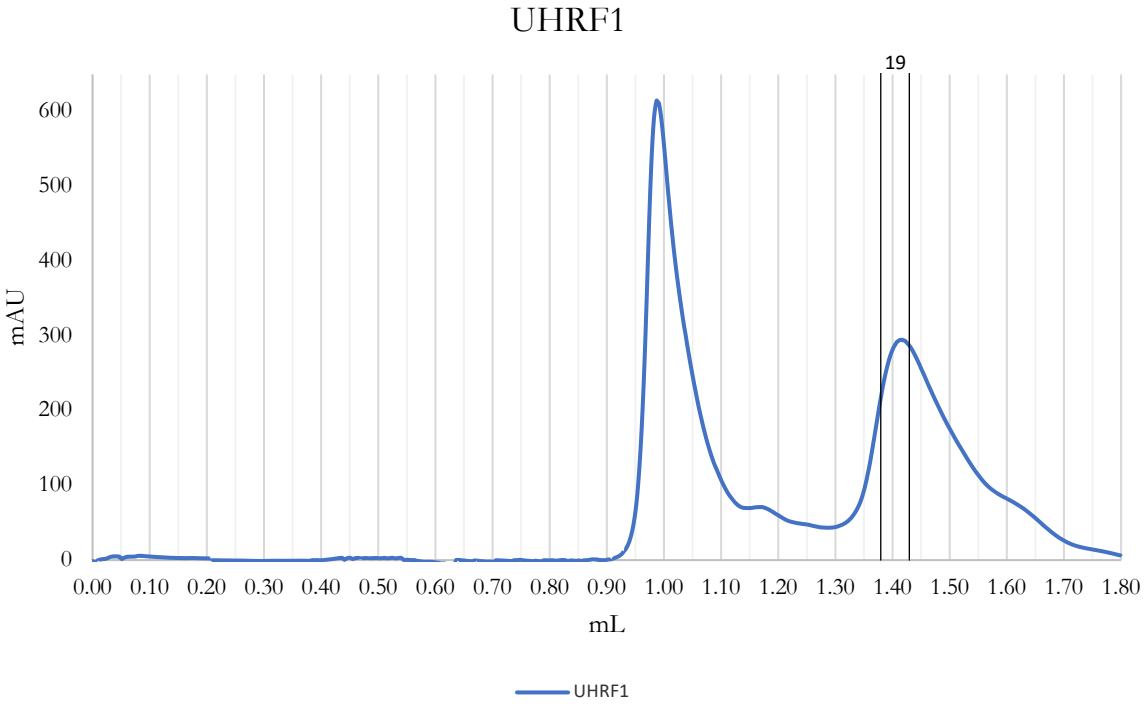


Figure 4-9: Optimisation of cryo-EM conditions for UHRF1 at 190 000x, 0.78 Å/pixel. a) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Cu; protein concentration 0.33 µg/µl. b) Representative micrographs of Quantifoil Graphene Oxide Support Film 2/2 300 Mesh Cu; protein concentration 0.33 µg/µl. c) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Au; protein concentration 0.33 µg/µl.

a



b

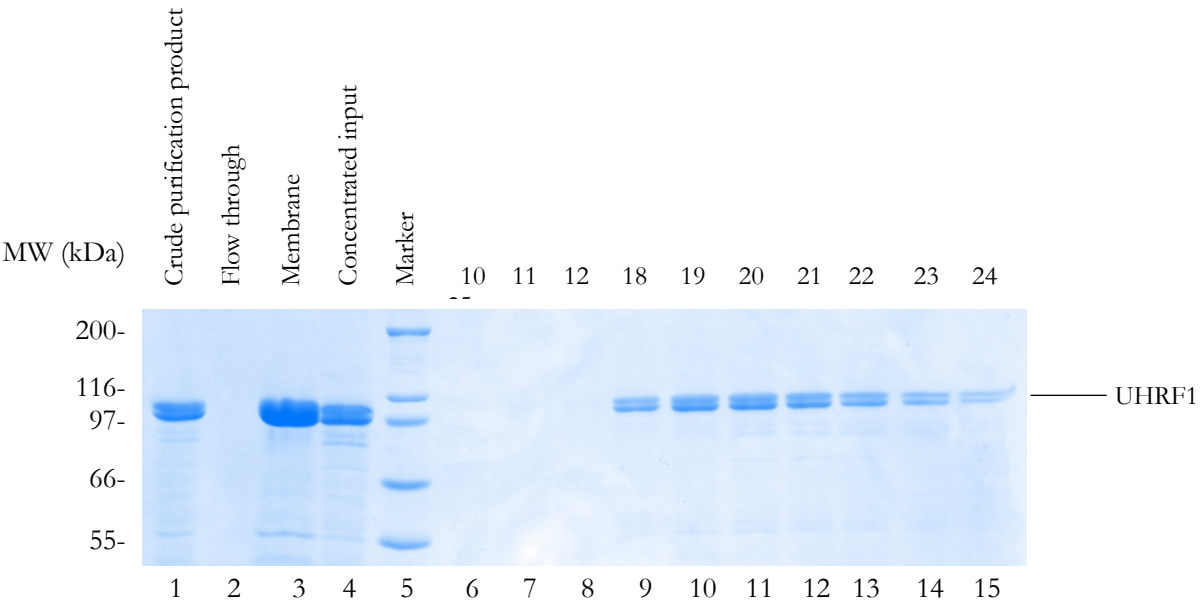
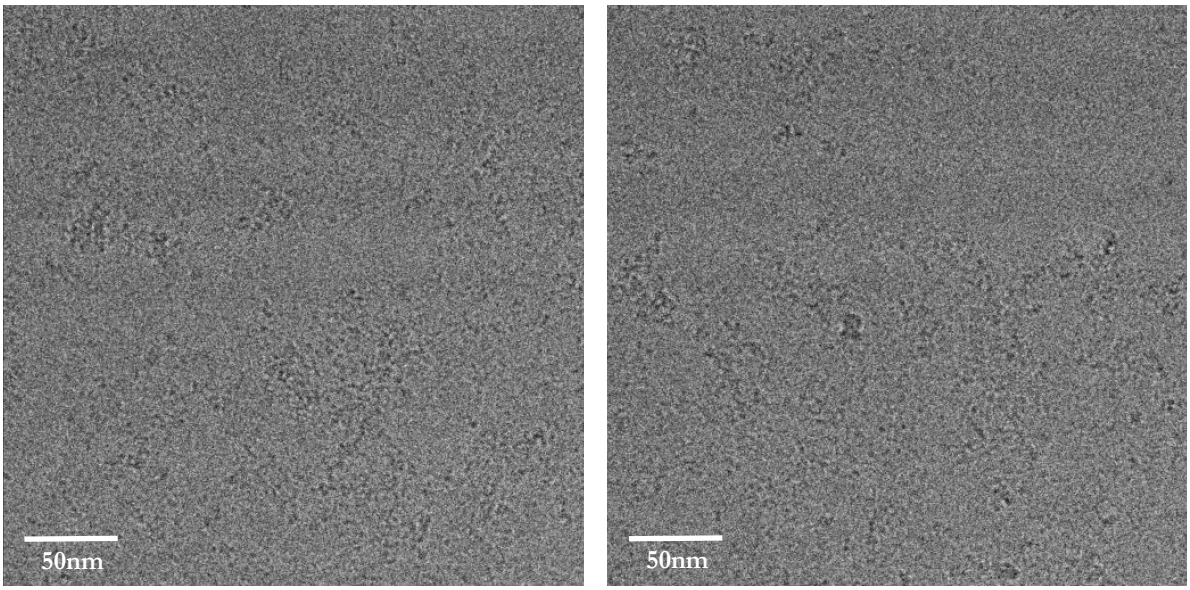
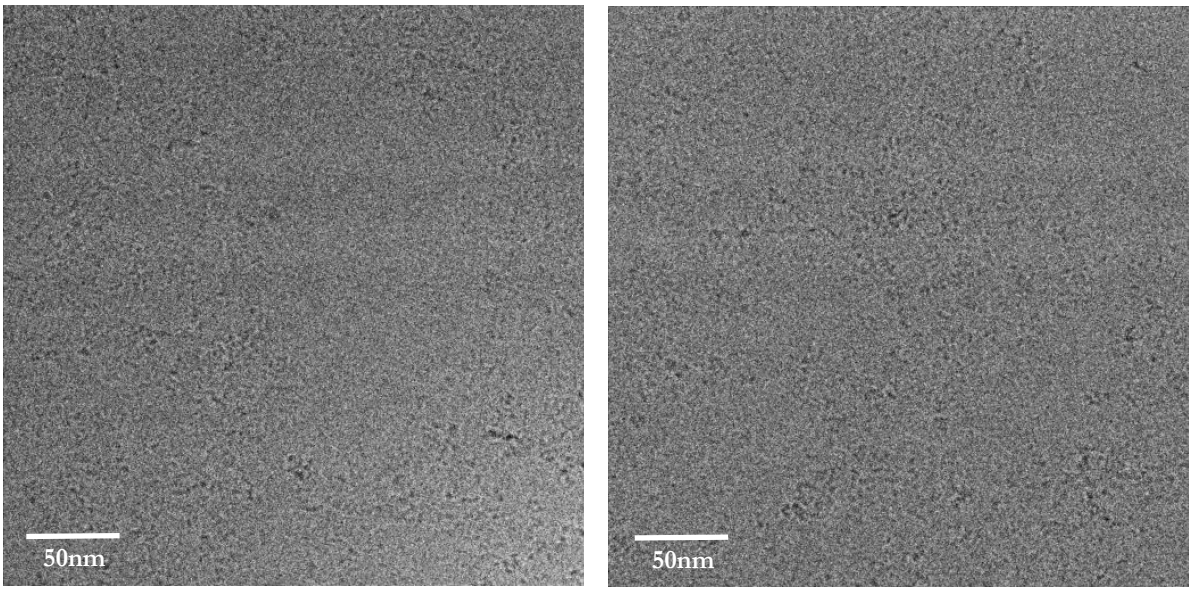


Figure 4-10: Purification of UHRF1 for Cryo-EM. a) Chromatogram following purification on a Superdex 200 3.2/300. UHRF1 identified in peak fractions 18 - 24. b) Coomassie blue stain of the purification procedure showing UHRF1. Lane 1: crude purification product, lane 2: flow through from 10 kDa cut-off concentrator, lane 3: sample stuck on the membrane, lane 4: sample loaded onto the gel filtration column, lane 5: marker, lanes 6 – 15: gel filtration peak fractions (lanes 6 – 8: void volume, lanes 9-15: UHRF1 sample peak) (lane 6: peak fraction 10, lane 7: peak fraction 11, lane 8: peak fraction 12, lane 9: peak fraction 18, lane 10: peak fraction 19, lane 11: peak fraction 20, lane 12: peak fraction 21, lane 13: peak fraction 22, lane 14: peak fraction 23, lane 15: peak fraction 24).

a



b



c

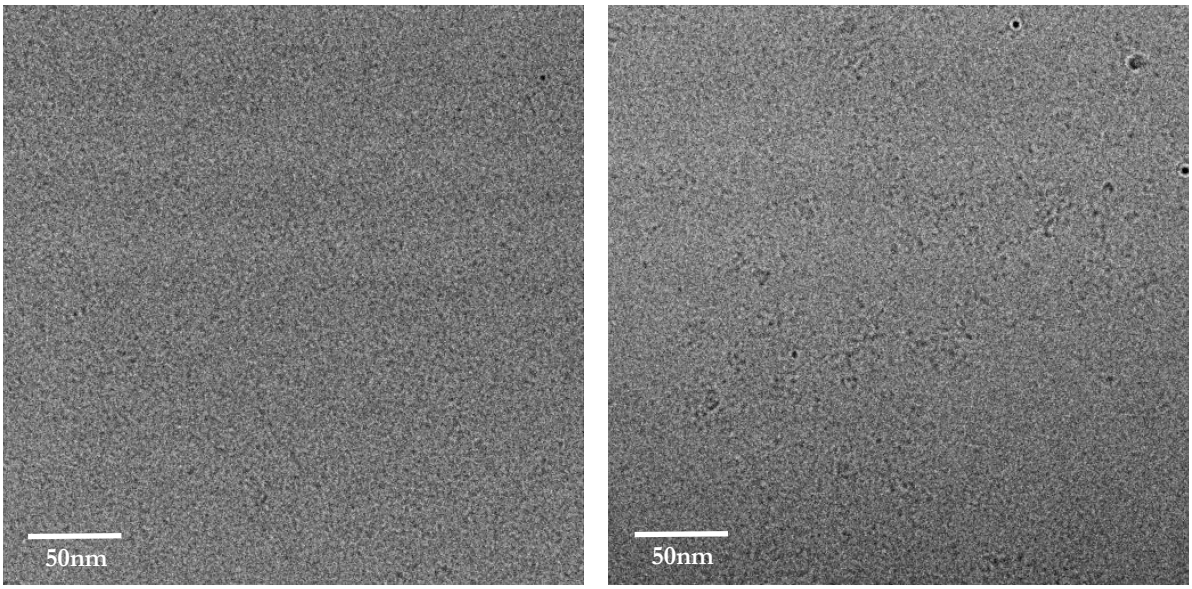
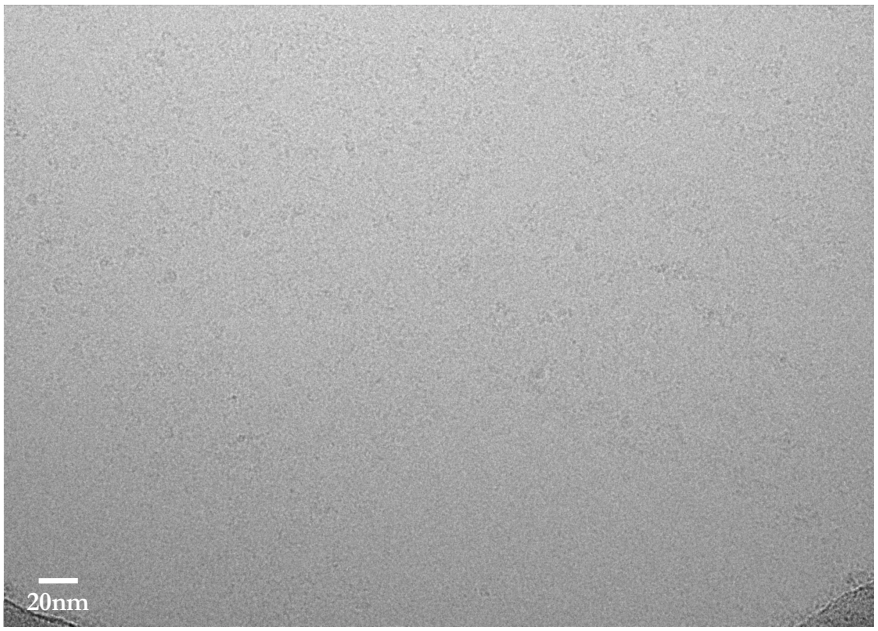


Figure 4-11: Optimisation of cryo-EM conditions for UHRF1 at 190 000x, 0.78 Å/pixel. a) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Cu; protein concentration 0.28 µg/µl. b) Representative micrographs of Quantifoil 1.2/1.3 300 Mesh Au; protein concentration 0.28 µg/µl. c) Representative micrographs of UltraAuFoil 1.2/1.3 300 Mesh Au; protein concentration 0.28 µg/µl.

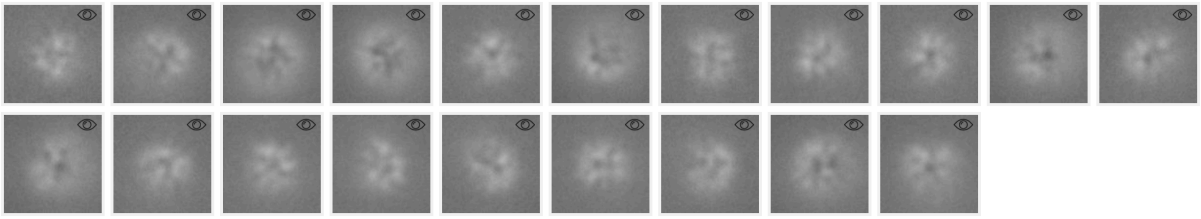
Figure 4-12a displays some representative micrographs from the screen carried out on the Krios 300 kV microscope. Given the small size of UHRF1, it is difficult to easily identify individual particles, however some particles can be discerned. In addition, there is some evidence of protein clumping, which is an indication of some unfolding of the protein during the vitrification process. Nevertheless, an initial 2D classification was produced using SIMPLE from approximately 5000 manually picked particles (Figure 4-12b). It can be seen that the 2D classification generated somewhat resembles the 2D classification generated during negative staining (Figure 4-7b). From this 2D classification, an initial 3D model was generated using SIMPLE (Figure 4-12c). It can be seen that the model displays a globular structure, in a similar manner to the negative staining initial 3D model. Using the initial SIMPLE 2D classification shown in Figure 4-12b as a template for auto-picking, 1079219 particles were auto-picked using SIMPLE from 1328 micrographs. From these auto-picked particles, a low resolution 2D classification was generated (Figure 4-12d). From this low resolution 2D classification, a high resolution 2D classification was generated (Figure 4-12e). However, these 2D classifications seem to be unclear, with no particular secondary structure features that can be identified. This would indicate that the auto-pick was poor, and probably that a large proportion of identified auto-picked particles were background or unfolded protein.

a

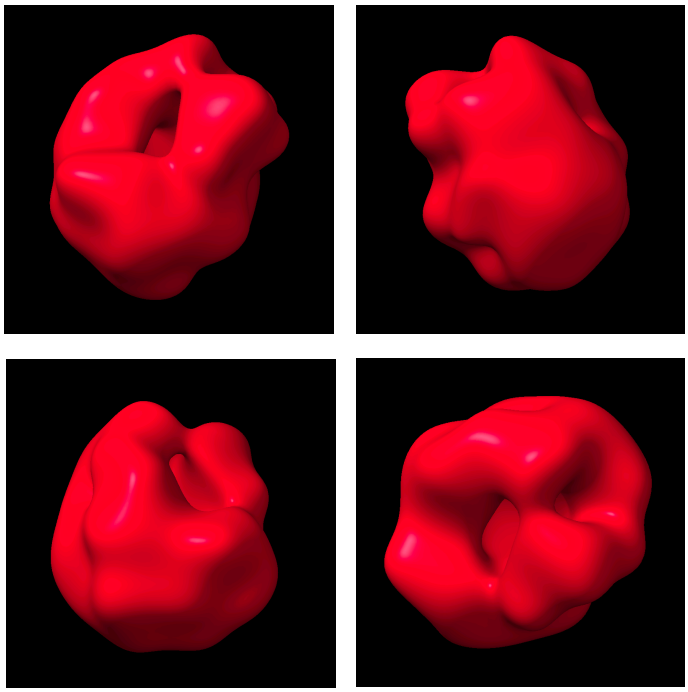




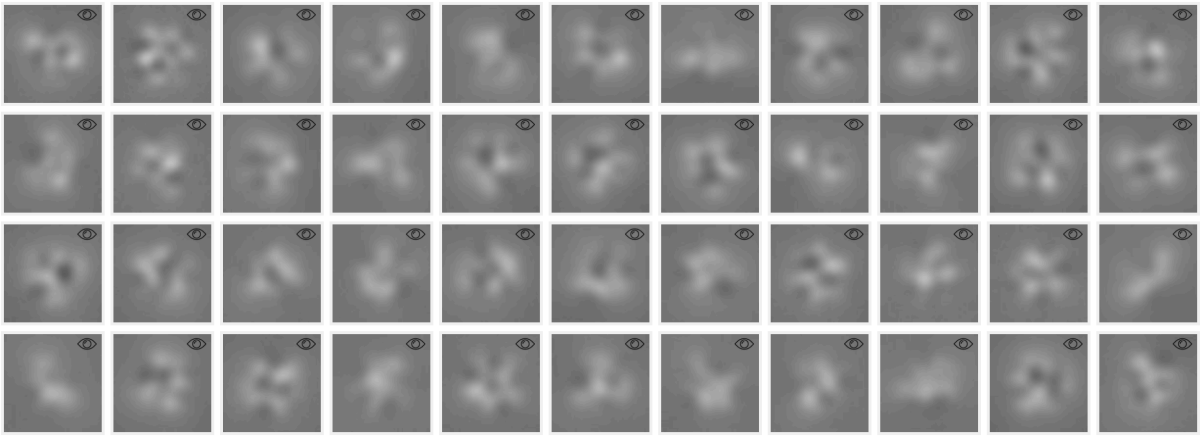
b



c



d



e

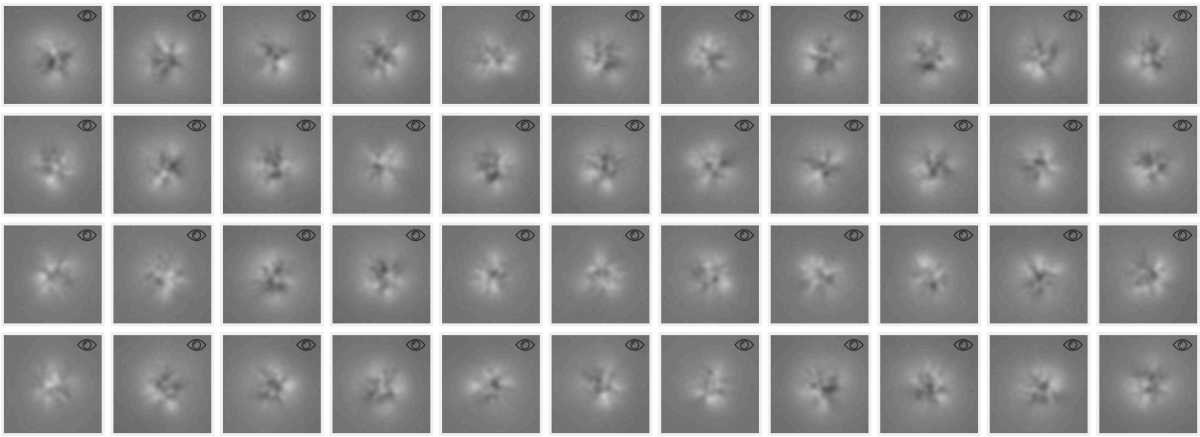


Figure 4-12: Krios 300 kV Initial Screen of UHRF1 on Quantifoil 1.2/1.3 Mesh Au at 105 000x super resolution, 0.416 Å/pixel. a) Representative micrographs from Krios screen; protein concentration 0.28 µg/µl. b) Initial 2D classification of UHRF1 from 5000 manually picked particles using SIMPLE. c) Initial 3D model generated from the 2D classification shown in b). d) Selection of classes from low resolution 2D classification of UHRF1 from 1079219 auto-picked particles from 1328 micrographs using SIMPLE. e) Selection of classes from high resolution 2D classification of UHRF1 from the low resolution 2D classification shown in d) using SIMPLE.

4.4 Discussion

ICL sensing is a crucial first step in the FA pathway. This chapter describes progress towards the structure of full-length human UHRF1, a protein that has been identified as a sensor for ICLs. The effect of UHRF1 on the *in vitro* monoubiquitination of FANCD2/FANCI is described, showing that UHRF1 hinders ubiquitination of FANCD2 in the FANCD2/FANCI complex in this setting. In addition, further work is described looking into the possible mechanistic reasoning behind this reduced monoubiquitination. The preliminary and refined 3D models of UHRF1 solved by negative staining EM are described. In addition, the cryo-EM grid conditions were optimized, and an initial screen on the 300 kV Krios was carried out. The resulting images were analysed and 2D classifications were produced using SIMPLE. From the initial 2D classification, an initial 3D model was solved using SIMPLE. Each of these will be further discussed below.

Efficiency of monoubiquitination of FANCD2/FANCI is hindered by UHRF1 *in vitro*

It has been previously demonstrated that UHRF1 is recruited to ICLs *in vivo*. In addition, *in vitro* it has been demonstrated that the domain that is required for this interaction is the SRA domain, as the interaction between UHRF1-ΔSRA and ICLs is completely abrogated. It has also been shown that the recruitment of UHRF1 to ICLs is required for FANCD2 recruitment *in vivo* as UHRF1 knockdown completely abrogates FANCD2 recruitment to ICLs (Liang et al., 2015b). In 2018, further to the requirement of UHRF1 for the recruitment of FANCD2 to ICLs, it was stated that knockdown of UHRF1 and/or UHRF2 reduced the monoubiquitination of FANCD2 *in vivo*. In addition, it was shown that UHRF1 directly interacts with FANCD2 through its SRA domain, and in a similar manner to the ICL interaction, the interaction was completely abolished between UHRF1-ΔSRA and FANCD2 (Motnenko et al., 2018). Therefore, here it was hypothesized that the addition of UHRF1 to the *in vitro* monoubiquitination assay could promote ubiquitination of the FANCD2/FANCI complex, through promotion of recruitment of the FANCD2/FANCI complex to DNA. Strikingly, rather than promoting the ubiquitination, the addition of UHRF1 decreased the monoubiquitination of FANCD2/FANCI. This could be due to UHRF1 having a higher binding affinity for the DNA than for the FANCD2/FANCI complex, or that UHRF1

interacts with components of the *in vitro* ubiquitination assay, preventing the reaction from occurring in such an efficient manner. However, UHRF1 binds with a higher affinity to ICLs, therefore it would also be informative to assess the effect of the addition of UHRF1 to the ubiquitination assay using ICL-containing DNA. In addition, it should be noted that a true control whereby an alternative protein was added to the ubiquitination assay that did not affect the ubiquitination of FANCD2/FANCI was not analysed. Therefore, the effect of adding UHRF1 in comparison to this negative control could not be assessed, and thus the drastic effects on monoubiquitination may be attributed to adding an additional protein into the *in vitro* system. Nevertheless, how UHRF1 hindered the monoubiquitination of FANCD2/FANCI was investigated.

First, it was hypothesized that the surprising result could be due to UHRF1 binding with a higher affinity to the DNA than the FANCD2/FANCI complex. However, upon testing this hypothesis through increasing the DNA concentration in the ubiquitination assay, the addition of increasing DNA did not fully restore the level of monoubiquitination of FANCD2/FANCI without UHRF1. This would imply that UHRF1 may be sequestering some of the DNA necessary for the ubiquitination assay, however that there is likely more than one factor contributing to the reduction in ubiquitination efficiency.

Next, it was hypothesized that UHRF1 may be directly interacting with FANCD2/FANCI, and thus preventing the complex's ubiquitination. However, upon testing this hypothesis through incubation of UHRF1 and FANCD2/FANCI and subjecting the sample to gel filtration, no trimeric peak was observed, and it could be seen that FANCD2/FANCI was eluted in the dimeric fraction, and UHRF1 was eluted as a monomer. This indicates that UHRF1 and FANCD2/FANCI do not directly interact *in vitro* upon incubation, and thus cannot contribute to the hindrance of ubiquitination. However, it is possible that UHRF1 and FANCD2/FANCI only form a complex whilst interacting with DNA, and this potential interaction would not be observed in the absence of DNA.

Finally, it was hypothesized that the decreased ubiquitination efficiency could be due to UHRF1 binding to one, or multiple of the ubiquitination assay components. For example, the SRA domain has been shown to be crucial to the interaction of UHRF1 and ICLs, as well as for the interaction of UHRF1 and FANCD2. In addition, as UHRF1 is an E3 RING ligase, the RING domain would presumably interact with ubiquitin. Therefore, the ubiquitination assay was conducted with the

addition of UHRF1 domain deletion mutants, UHRF1-ΔUBL, UHRF1-ΔTTD, UHRF1-ΔPHD, UHRF1-ΔSRA and UHRF1-ΔRING. Interestingly, the deletion of the RING domain statistically significantly restored the ubiquitination level in comparison to the full length UHRF1, implying that perhaps the RING domain is interacting with components of the ubiquitination assay. In addition, the deletion of the SRA domain increased the ubiquitination efficiency, albeit not statistically significantly, implying that the interaction of UHRF1 with DNA and potentially with FANCD2 could be hampering the ubiquitination reaction. Overall, the individual deletion of the UBL, TTD, SRA and RING domains increased the ubiquitination efficiency, albeit not to the level of the monoubiquitination efficiency without the addition of UHRF1. Finally, the deletion of the PHD domain seemed to decrease the ubiquitination efficiency further than full length UHRF1. This could be explained as it has been shown that the PHD domain is important for the intramolecular interactions of UHRF1, particularly with the TTD domain. In particular, a linker between the TTD-PHD domains has been shown to be crucial to the intramolecular interactions, and thus the orientation of the TTD and PHD domains with respect to each other. It has been reported that displacement of this intramolecular interaction changes the conformation of UHRF1. Therefore, it has been postulated that isolated UHRF1 adopts a closed structure, but when UHRF1 binds to hemi-methylated DNA and/or histones, UHRF1 changes conformation to adopt an open structure (Fang et al., 2016; Gao et al., 2018; Kori et al., 2019). Perhaps in this assay, the deletion of the PHD domain allows UHRF1 to adopt the open conformation of UHRF1, which may result in an increased sequestering of DNA or other ubiquitination assay components. However, it should be noted that the purification of the UHRF1-ΔPHD contains many contaminants, and thus the experiment should be repeated with a cleaner UHRF1-ΔPHD to confirm these observations.

UHRF1 appears globular in an unbound state

To date there has been much structural data published with respect to the different domains of UHRF1. However, structural data pertaining to the full-length protein has remained elusive. Here, an initial 3D model was produced from negative staining EM displaying a compact globular structure. There also appears to be a channel that runs through the centre of UHRF1. In addition, initial progress towards the cryo-EM structure of UHRF1 was presented. The concentrated UHRF1 protein sample was applied to grids and imaged on an Arctica 200 kV microscope where the protein sample could be visualised. However, due to the small size of the protein, it was difficult to discern if the protein present on the grids was denatured. Therefore, the sample was prepared and further screened on the Krios 300 kV microscope. From the resulting Krios images,

it is still difficult to visualise the individual protein particles. The resulting 2D classifications are also unclear, with no discernible secondary structure features. However, the initial 3D model produced from the manually picked particles displays a similar globular shape to that of the negative staining model.

The structural findings presented here display the first insight into the structure of the full-length UHRF1 protein. UHRF1 appears to adopt a compact globular structure, potentially with a channel passing through the centre of the protein. This closed structure is in keeping with that postulated in the literature for UHRF1 in its unbound state and is presumably a result of intramolecular interactions (Fang et al., 2016; Gao et al., 2018; Kori et al., 2019). It is possible that the channel running through the centre of the UHRF1 protein could accommodate DNA, whether that be the hemi-methylated DNA substrate for UHRF1's role in DNA maintenance methylation, or DNA containing an ICL for UHRF1's role in the FA pathway. However, it is likely that upon binding the DNA that UHRF1 changes conformation to a more open state, altering the intramolecular interactions and forming additional interactions between UHRF1 and the DNA that are not possible in its closed state.

It is clear that further structural work needs to be completed, not only to gain insight into the high resolution structure of UHRF1, but to gain insight into this potential conformational change. As crystallographic studies have proven limited in their ability to gain full-length structural information of UHRF1, here cryo-EM has been presented as an alternative. However, due to the small size of UHRF1, it nears the limits of structural elucidation by cryo-EM. In recent years, the lower size limit of proteins that can be solved by single particle cryo-EM has been reduced, however, despite this only 45 proteins have been solved below 100 kDa, of which 17 are human proteins, and of these only 13 have been solved to better than 8 Å. However, due to the rapid development of cryo-EM, it has opened up the possibility for further full-length structural insight of proteins on the lower end of the cryo-EM resolution limit, to a high structural resolution (Wu and Lander, 2020). The initial progress towards a cryo-EM structure presented here shows that cryo-EM may be a valid technique to provide further structural insight into the full-length UHRF1 protein as the particles can be visualized on the grids. However, despite this, there is clearly further work to be done to reduce sample aggregation and clumping, and thus reduce protein denaturation. In addition, it is clear that the 2D classes presented here are indistinct and that no secondary structure features can be elucidated. Therefore, further work needs to be completed to optimize the sample for structural elucidation, particularly with regard to ensuring the sample is

evenly distributed across the grid. In addition, it would be worth investigating if it is possible to further stabilize the UHRF1 protein with a nanobody or megabody to enable easier visualization of individual particles.

A high resolution structure of UHRF1 obtained via cryo-EM, as well as the structure of UHRF1 bound to an ICL, would provide insight into if DNA binding changes the conformation of UHRF1. This would in turn provide insight into how UHRF1 is able to sense the ICL, and thus allow investigation into its role in the FA pathway. A high resolution full-length UHRF1 structure would also provide insight into its role in DNA maintenance methylation.

5. Discussion

Since the discovery of Fanconi Anaemia in 1927, 22 genes have been identified whereby a defect in these genes causes the FA phenotype (Fang et al., 2020; Lopez-Martinez et al., 2016). Despite significant advances in our understanding of FA and the FA pathway, many intricate mechanisms are yet to be elucidated. For example, it is clear that there may be numerous proteins that detect ICLs, however the interplay between these proteins and whether they have any inter-dependence on each other, or are in fact redundant, remains to be investigated (Liang et al., 2015b; Motnenko et al., 2018, Socha et al., 2020). In addition, whether particular sensor proteins detect certain ICLs based on the different distortions the ICL induces in the helix remains to be explored. It is also unclear how the first identified sensor protein for ICLs, FANCM, is involved, if at all, with these newly identified sensors (Meetei et al., 2005). Following the identification of the ICL, the FANCD2/FANCI complex is recruited. However, how the FANCD2/FANCI complex interacts with the ICL sensor proteins and whether these proteins or the FANCD2/FANCI complex need to undergo any modifications to enable this interaction remains unclear. In addition, it is also unclear whether they play a role in the complex's ubiquitination. To date, it also remains to be elucidated when the sensor proteins have completed their function in the FA pathway and if or when they depart from the ICL.

It is known that the FANCD2/FANCI complex must undergo ubiquitination in order for the FA pathway to be functional (Walden and Deans, 2014). However, what this specific ubiquitination achieves, perhaps through recruitment of different downstream repair factors, remains to be elucidated. In addition, despite the ubiquitination being crucial to the function of the pathway, the deubiquitination of the complex has also been shown to be important for ICL repair (Cohn et al., 2007; Cohn et al., 2009; Nijman et al., 2005). This project has provided insight into these early steps of the FA pathway, and the implications of these findings as well as recent literature will be discussed below.

5.1 UHRF1

In this project, the UHRF1 sensor protein was investigated. It was shown that UHRF1 hinders the ubiquitination of FANCD2/FANCI *in vitro*, and that this decrease in ubiquitination efficiency is significantly restored upon deletion of the RING domain. The ubiquitination efficiency was also increased, albeit not significantly, upon deletion of the UBL, TTD and SRA domains. Also, the addition of excess DNA to the ubiquitination reaction mildly improved the ubiquitination efficiency, but it was not restored to the level of monoubiquitination of the assay without UHRF1.

Finally, it was demonstrated that UHRF1 and FANCD2/FANCI do not appear to have a strong interaction *in vitro*. This finding is extremely interesting, as it has been shown that without UHRF1 recruitment to ICLs *in vivo*, FANCD2 cannot be recruited to sites of damage (Liang et al., 2015b). However, despite the significant reduction of FANCD2 recruitment to ICLs in UHRF1 knockdown cells, the monoubiquitination is only mildly affected (Motnenko et al., 2018). In contrast to these observations, the findings here demonstrate that the addition of UHRF1 to the simplified *in vitro* ubiquitination assay significantly reduces the monoubiquitination efficiency. Based on the evidence presented, it would seem that UHRF1 binds to DNA, perhaps preventing the FANCD2/FANCI complex from binding to DNA, reducing the level of monoubiquitination. In addition, UHRF1 may sequester components of the ubiquitination assay, including ubiquitin and/or FANCD2/FANCI when in the presence of DNA. These findings would indicate that UHRF1 alone is not sufficient to promote the recruitment of FANCD2/FANCI to sites of damage, nor to promote the ubiquitination of FANCD2/FANCI *in vitro*. However, as UHRF1 has been shown to directly interact with FANCD2 *in vitro*, it is likely that there may be a direct interaction in the cell. Therefore, for UHRF1 to promote FANCD2/FANCI recruitment to sites of damage, perhaps either the UHRF1 protein or the FANCD2/FANCI complex must undergo some kind of post-translational modification to assist with this recruitment, and subsequent promotion of ubiquitination. In addition, other factors are present in the cell which may protect the ubiquitination components from sequestration by UHRF1. Furthermore, perhaps there are additional factors or steps in the FA pathway linking UHRF1 to the recruitment and ubiquitination of FANCD2/FANCI. For example, it is known that UHRF1 and BRCA1 form a complex involved in HR, and also that BRCA1 is involved in assisting with CMG unloading at ICLs (Zhang et al., 2016). Perhaps UHRF1 interacts with BRCA1 and this interaction facilitates FANCD2/FANCI recruitment and subsequent ubiquitination. In addition, UHRF1 possesses E3 RING ligase activity, which may ubiquitinate an unknown substrate essential to the recruitment of FANCD2/FANCI. Finally, perhaps UHRF1's role in DNA maintenance methylation plays a role, as UHRF1 is recruited to replicating chromatin which involves the replication fork structure, whilst the FA pathway requires convergence of two replication forks (Avvakumov et al., 2008; Berkyurek et al., 2014; Bostick et al., 2007; Cheng et al., 2013; MacKay et al., 2010; Zhang et al., 2015). Perhaps the mechanism to UHRF1 recruitment is common to both of these processes.

In this project, initial work was completed on the structure of the UHRF1 protein. To date, there is much structural data available for the individual UHRF1 domains, however there have been no insights into the full-length structure. In the field of DNA maintenance methylation, it has been

postulated that UHRF1 forms a closed structure in its free state, forming many intramolecular interactions, but that the domains may lose these interactions and form an open structure upon DNA binding (Fang et al., 2016; Gao et al., 2018; Gelato et al., 2014; Harrison et al., 2016; Vaughan et al., 2018). The findings presented here support this hypothesis and provide the first low resolution structural information for the full-length UHRF1. This structural work shows that UHRF1 appears to adopt a closed, globular structure in its free state. However, whether there is a conformational change upon ICL binding, or perhaps upon binding to FANCD2/FANCI in the presence of DNA, remains to be elucidated. Solving the structure of UHRF1 will therefore not only assist in elucidating its role in the FA pathway, but also its role in the DDR as a whole as well as its role in DNA maintenance methylation.

5.2 FANCD2/FANCI

In this project, the ubiquitination of the FANCD2/FANCI complex, as well as the deubiquitination of the monoubiquitinated FANCD2/FANCI complex was investigated. In addition, low resolution negative staining 3D models of the FANCD2/FANCI complex, the monoubiquitinated-FANCD2/FANCI complex and the monoubiquitinated-FANCI/FANCD2 complex were presented. In keeping with recent literature, increasing levels of DNA in the ubiquitination assay *in vitro* promoted ubiquitination of the FANCD2/FANCI complex. In contrast, the presence of DNA slightly promoted the deubiquitination of the intermediate complex, monoubiquitinated-FANCD2/FANCI, however complete deubiquitination could be seen for the complex both on and off DNA over time. However, this DNA dependence for ubiquitination and DNA independence for deubiquitination is yet to be investigated in a cellular environment.

Recent structural studies have elucidated that FANCD2 is monoubiquitinated first, followed by FANCI, which results in a conformational change of the FANCD2/FANCI complex. The conformational change has been shown to lock the complex onto DNA (Alcón et al., 2020, Wang et al., 2020). It has been demonstrated that the majority of the conformational change is due to the ubiquitination of FANCD2, with only a small secondary conformational change upon the ubiquitination of FANCI (Wang et al., 2020). However, the biological implications of this conformational change, and the resulting locked complex as well as how this ubiquitination event activates the FA pathway, remain to be elucidated. Perhaps FANCD2/FANCI must remain on DNA for downstream repair factors to be recruited, and that this recruitment can only occur with the complex locked onto DNA, either due to interaction with the DNA, or due to the complex's

conformational change. In addition, perhaps through this locking mechanism, the complex protects the enclosed DNA from surrounding nucleases. However, despite mounting evidence *in vitro* and *in vivo,* it remains to be confirmed in a cellular environment that FANCD2 is ubiquitinated first, followed by FANCI, and whether this occurs on DNA or prior to recruitment of the complex to DNA.

This project also presented that the intermediate monoubiquitinated-FANCD2/FANCI complex does not appear to form filamentous arrays following ubiquitination when analyzing the crude ubiquitination assay *in vitro*. This is a novel finding, as it has recently been shown that the monoubiquitinated-FANCD2/monoubiquitinated-FANCI may form filamentous arrays following ubiquitination (Tan et al., 2020a). This would indicate that FANCI plays a key role in forming these filaments. The formation of these filaments could perhaps play a role in DNA fork protection, for example from cellular nucleases including MRE11 and DNA2 (Tan et al., 2020a). Alternatively, perhaps the formation of these filaments explains the accumulation of FANCD2 at fragile sites upon replication stress (Okamoto et al., 2018). In keeping with this protection role whereby filament formation is stimulated following FANCI ubiquitination, a recent publication shows FANCI ubiquitination playing a protective role for the deubiquitination of the ubiquitinated FANCD2/FANCI complex (Rennie et al., 2020). In this study it shows that the ubiquitination of FANCI results in the monoubiquitinated-FANCD2/monoubiquitinated-FANCI complex being resistant to deubiquitination unless the complex has dissociated from DNA. However, Wang et al. show that the ubiquitinated complex did not easily dissociate from DNA, and actually slid along the DNA before falling off the end of a linear DNA fragment. In contrast, for a circular plasmid, the ubiquitinated complex could not easily dissociate from the DNA (Wang et al., 2020). Taking this evidence together, it would seem impossible that the ubiquitinated complex could be deubiquitinated whilst on DNA, and also that the complex is unable to dissociate from the DNA before deubiquitination. A possible explanation for this could be that upon ubiquitination of the FANCD2/FANCI complex, certain nucleases, such as SLX4, are recruited to incise the ICL out of the DNA strand. Upon such incision, it may be possible for the ubiquitinated FANCD2/FANCI complex to slide off the DNA strand, and thus dissociate from DNA and allow its deubiquitination. However, it is clear that despite great advances with regards to the ubiquitination and deubiquitination of the FANCD2/FANCI complex, significant further challenges exist with regards to proving the sequence of events in the cell.

5.3 Future Work and Conclusions

As discussed above, there are many open questions with regards to the first steps of the FA pathway. How UHRF1 recognises an ICL, and whether there are any post translational modifications required that facilitate this interaction needs to be investigated. In addition, is there a conformational change of UHRF1 upon interacting with the ICL, for example from a closed to an open state? Further cryo-EM structural work elucidating the structure of free UHRF1 and also UHRF1 bound to an ICL will be necessary to answer these questions. In addition, discerning UHRF1's role at the ICL is still yet to be investigated. Specifically, why UHRF1 is recruited to the ICL and whether it is essential for any downstream factors remains unclear. Furthermore, it would be interesting to investigate whether the multi-protein complex of UHRF1 and FANCD2/FANCI exists, and if so, how this complex interacts with the ICL. This multi-protein complex would also be larger, and thus easier to visualize via cryo-EM than the UHRF1 protein alone. It would also be interesting to determine how FANCD2/FANCI are recruited, and what role UHRF1 plays in this recruitment, as well as whether the recruitment of this complex results in further conformational changes of UHRF1. In addition, how UHRF1 releases the DNA and allows FANCD2/FANCI to interact with the ICL will be crucial to understanding how this sensor plays its role in the FA pathway.

From a cellular point of view, there are also numerous open questions on the first steps of the FA pathway. It would be interesting to investigate the true interaction of FANCD2/FANCI and UHRF1 in a cellular environment. This could be envisioned through a resonance energy transfer experiment such as Förster resonance energy transfer (FRET), whereby each protein is tagged with a fluorophore and expressed in cells (Scott and Hoppe, 2015). It would then be possible to investigate the interaction between UHRF1 and FANCD2/FANCI, where UHRF1 acts as the FRET donor and FANCD2/FANCI could act as the FRET acceptor. Therefore, UHRF1 would fluoresce in one colour in its free state, however upon interaction with FANCD2/FANCI, would fluoresce in a second colour based on the energy transfer from the donor fluorophore to the acceptor fluorophore. This would truly demonstrate an *in vivo* interaction between UHRF1 and FANCD2/FANCI and would indicate that UHRF1 is required for FANCD2/FANCI recruitment through a direct interaction. However, given that UHRF1 and FANCD2/FANCI interact at ICL sites, it may be possible that the signal seen from this experiment could be due to the proteins being recruited separately to the ICL, rather than through a direct interaction between UHRF1 and FANCD2/FANCI. Perhaps this shortcoming could be overcome through the introduction of DNA-binding mutants on FANCD2/FANCI so that if a FRET signal was observed it would

be due to a direct interaction with UHRF1, as the FANCD2/FANCI complex would be unable to bind to DNA and thus unable to be recruited to ICL sites alone.

In addition to the open questions presented above, there are also many questions with regards to the FANCD2/FANCI complex in a cellular environment. As presented in the recently published literature, *in vitro* FANCD2 is ubiquitinated first, followed by FANCI (Wang et al., 2020). However, not only has this not been proven in a cellular context, whether the FANCD2/FANCI complex even arrives at the ICL as a preformed complex remains unknown. In addition, it has been demonstrated that there is only a weak interaction between the non-ubiquitinated FANCD2 and non-ubiquitinated FANCI *in vivo* (Sareen et al., 2012). In fact, Alcón et al. recently suggested that perhaps FANCD2 exists as a homodimer, and that only upon interaction with FANCI does the heterodimeric FANCD2/FANCI complex form (Alcón et al., 2020). In addition, despite the extensive evidence that monoubiquitination of the FANCD2/FANCI complex is stimulated by DNA *in vitro*, this remains to be seen *in vivo*. This sequence of events *in vivo* could also be investigated using FRET. In the first instance, FANCD2 could be tagged with the fluorophore donor and FANCI could be tagged with the fluorophore acceptor. Therefore it would be possible to determine if FANCD2 travels to the site of damage emitting the donor fluorescence before changing colour to the acceptor fluorescence upon interaction with FANCI, or if this interaction and thus change in colour occurs prior to arrival at the ICL, and thus the complex travels together to the site of damage before becoming stationary. A more complicated, but more informative experiment, would be to complete a three-way FRET experiment, tagging FANCD2 and FANCI and the E2 conjugating enzyme UBE2T. In this way, it would be possible to determine whether FANCD2 or FANCI are ubiquitinated first, and whether this occurs at the site of damage. For example, if FANCD2 and FANCI are tagged with different fluorophore donors, but have a common fluorophore acceptor tagged to UBE2T, it would be possible to track whether UBE2T arrives at FANCD2 or FANCI first, to definitively show the order of ubiquitination in a cell. These cellular experiments would aid in confirming that the *in vitro* findings are also seen in a cellular environment. It would also be interesting to investigate ubiquitination point mutants of FANCD2 and FANCI through FRET, and to determine whether the ubiquitination of one of the substrates is reliant on the other, for example if the UBE2T stalls at one protomer and cannot continue on to recognise the second substrate.

Finally, it is clear that further work needs to be completed surrounding the possibility of filament formation as a consequence of ubiquitination of FANCD2/FANCI. It is clear that the method

utilized here, as well as the method utilized in the recently published paper, does not provide the entire sequence of events (Tan et al., 2020a). The findings presented here demonstrate that the method employed to isolate the DNA resulted in filaments formed through addition of the SmaI enzyme. However, the method employed in the recently published study cannot be used to visualize a true negative control, the non-ubiquitinated FANCD2/FANCI complex. This is due to the fact that the method utilizes a tag on ubiquitin to pull the ubiquitinated complex out of the reaction mixture (Tan et al., 2020a). Therefore, a method would need to be developed whereby it is possible to directly compare FANCD2/FANCI, the intermediate monoubiquitinated-FANCD2/FANCI and the fully monoubiquitinated-FANCD2/monoubiquitinated-FANCI to monitor at what stage, if at all, the filaments occur. In addition, it would need to be controlled that the filaments are not forming simply due to the fact that the ubiquitinated complex has the ability to slide along the DNA and form as a result of the *in vitro* system, rather than being representative of the cellular environment. If indeed through this method it is demonstrated that the filaments only form during the second monoubiquitination step, upon monoubiquitination of FANCI, this would highlight a new and important role for FANCI in the FA pathway. This would then raise numerous questions including why do the filaments form, are they necessary for the FA pathway to function, and indeed are multiple FANCD2/FANCI complexes required at an ICL in order for the damage to be repaired? Recently, a time-resolved cryo-EM method has been described that allows the investigation of the kinetics and intermediate states in biochemical assays (Mäeots et al., 2020). The test sample presented demonstrated the method's ability to visualise the formation of *E. coli* helical RecA filaments crucial in the DDR. Therefore, perhaps it would be possible to utilize this method of time-resolved cryo-EM to visualize if filaments form upon FANCD2/FANCI ubiquitination, and specifically whether they form only upon the second step of ubiquitination, the monoubiquitination of FANCI.

It is clear that despite significant advances in understanding the FA pathway, there are many further questions to be answered. By understanding the intricate mechanisms of the FA pathway it will be possible to utilize this information to better the lives of FA patients, as well as for improvements in cancer therapeutics. Therefore, further understanding of this DNA repair pathway, as well as the interplay of the FA pathway within the DDR is crucial to furthering potential translational applications of exploiting the FA pathway.

6. Materials and Methods

6.1 UBA1 (E1) Protein Expression and Purification

Expression

UBA1 with an engineered N-terminal FLAG-HA tag (provided by members of the Cohn laboratory) expression was completed using the pFastBac1 vector system (Life Technologies) according to the manufacturer's protocol. Sf9 insect cells (provided by members of the Cohn laboratory) were grown in 100 mL cultures in 1 L vented Erlenmeyer flasks in Sf-900^{TM} II SFM insect cell media at 27°C, shaking at 135 rpm. Stock cultures were split to 1 x 10^6 cells/mL.

The cells were split three days before infection, to result in a density of 6 x 10^6 cells/mL on the day of infection, with the volume not exceeding 10% of the total flask volume. The infection protocol was as per the manufacturer's instructions. For a 15 cm dish, UBA1 expressing baculovirus, 35 x 10^6 insect cells, and 20 mL of Sf-900^{TM} II SFM insect cell media was added. The culture was then incubated at 27°C, for 3 days.

Purification

Sf9 cell pellets were re-suspended in lysis buffer (20 mM Tris pH 7.5, 100 mM KCl, 10% Glycerol, 2 mM $MgCl_2$, 0.2 mM PMSF, 2 mM 2-Mercaptoethanol) and sonicated. Lysates were clarified by centrifugation at 17,000 x *g* and the supernatants incubated with M2 anti-FLAG agarose resin (A2220, Sigma) for 2 hours. The resin was washed with lysis buffer, and the protein was eluted in lysis buffer containing 0.5 mg/mL FLAG peptide.

6.2 UBE2T (E2) Protein Expression and Purification

Expression

UBE2T containing an N-terminal 6xHis-Smt3 tag (provided by members of the Cohn laboratory) was expressed in *E. coli* BL21(DE3) cells (Novagen). Cells were grown in Luria Broth (LB) and supplemented with 50 µg/mL Kanamycin at 37°C. Expression of UBE2T was induced at OD600 = 0.6 with 0.5 mM IPTG and cultured overnight at 16°C.

Purification

The cells were harvested by centrifugation at 4 000 x *g* and were lysed by French Press in lysis buffer (100 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM Imidazole, 250 µM TCEP). The lysate was clarified by centrifugation at 48,000 x *g* and the supernatant was incubated with Ni-NTA agarose (QIAGEN) for 2 hours at 4°C. The resin was washed with wash buffer (100 mM Tris-

HCl pH 8.0, 200 mM NaCl, 20 mM Imidazole, 250 μM TCEP), and the protein was eluted in elution buffer (100 mM Tris-HCl pH 8.0, 200 mM NaCl, 500 mM Imidazole, 250 μM TCEP). The eluate was loaded onto a Superdex 200 16/60 column with 100 mM Tris-HCl pH 8.0, 200 mM NaCl, 10% Glycerol, 250 μM TCEP. The 6xHis-Smt3 tag was removed overnight at 4°C by Ulp1 protease at a w/w ratio of 1:30 Ulp1:His-Smt3-UBE2T. The solution was then supplemented with 20 mM imidazole and incubated with Ni-NTA agarose (QIAGEN) for 2 hours at 4°C. UBE2T flow through was collected, concentrated and loaded onto a Superdex 200 16/60 column with 100 mM Tris-HCl pH 8.0, 200 mM NaCl, 10% Glycerol, 250 μM TCEP. The UBE2T peak fractions were then collected.

6.3 FANCL Protein Expression and Purification

Expression

FANCL with an engineered N-terminal FLAG-HA tag (provided by members of the Cohn laboratory) expression was completed using the pFastBac1 vector system (Life Technologies) according to the manufacturer's protocol. Sf9 insect cells (provided by members of the Cohn laboratory) were grown in 100 mL cultures in 1 L vented Erlenmeyer flasks in Sf-900™ II SFM insect cell media at 27°C, shaking at 135 rpm. Stock cultures were split to 1 x 10^6 cells/mL.

The cells were split three days before infection, to result in a density of 6 x 10^6 cells/mL on the day of infection, with the volume not exceeding 10% of the total flask volume. The infection protocol was as per the manufacturer's instructions. For a 15 cm dish, FANCL expressing baculovirus, 35 x 10^6 insect cells, and 20 mL of Sf-900™ II SFM insect cell media was added. The culture was then incubated at 27°C, for 4 days.

Purification

Sf9 cell pellets were re-suspended in lysis buffer (20 mM Tris pH 7.5, 100 mM KCl, 0.2% Tween 20, 10% Glycerol, 0.2 mM PMSF, 2 mM 2-Mercaptoethanol). Lysates were clarified by centrifugation at 17,000 x *g* and the supernatants incubated with M2 anti-FLAG agarose resin (A2220, Sigma) for 2 hours. The resin was washed with lysis buffer and elution buffer (20 mM Tris pH 7.5, 100 mM KCl, 10% Glycerol, 0.2 mM PMSF, 2 mM 2-Mercaptoethanol), and the protein was eluted in elution buffer containing 0.5 mg/mL FLAG peptide.

6.4 FANCD2/FANCI Protein Complex Expression and Purification

Expression

FANCD2 with an engineered N-terminal FLAG-HA tag (provided by members of the Cohn laboratory) and FANCI with an engineered N-terminal 6x-His tag (provided by members of the Cohn laboratory) expression was completed using the pFastBac1 vector system (Life Technologies) according to the manufacturer's protocol. Sf9 insect cells (provided by members of the Cohn laboratory) were grown in 100 mL cultures in 1 L vented Erlenmeyer flasks in Sf-900™ II SFM insect cell media at 27°C, shaking at 135 rpm. Stock cultures were split to 1 x 10^6 cells/mL.

The cells were split three days before infection, to result in a density of 6 x 10^6 cells/mL on the day of infection, with the volume not exceeding 10% of the total flask volume. The infection protocol was as per the manufacturer's instructions. For a 15 cm dish, FANCD2 expressing baculovirus, excess FANCI expressing baculovirus, 35 x 10^6 insect cells, and 20 mL of Sf-900™ II SFM insect cell media was added. The culture was then incubated at 27°C, for 3 days.

Purification

Sf9 cell pellets were re-suspended in lysis buffer (20 mM Tris pH 7.5, 100 mM KCl, 10% Glycerol, 2 mM $MgCl_2$, 0.2 mM PMSF, 2 mM 2-Mercaptoethanol) and sonicated. Lysates were clarified by centrifugation at 17,000 x *g* and the supernatants incubated with M2 anti-FLAG agarose resin (A2220, Sigma) for 2 hours. The resin was washed with lysis buffer, and the protein was eluted in lysis buffer containing 0.5 mg/mL FLAG peptide. For samples used for negative staining, the complex sample was then applied on a Superdex 200 3.2/300 column with 20 mM Hepes pH 7.6, 120 mM NaCl, 2 mM dithiothreitol (DTT). The FANCD2/FANCI fraction A11 was then collected and applied to grids for negative stain electron microscopy. For samples used for cryo-EM, the complex was applied to a 10 kDa concentrator (Amicon Ultracel 10K) at 12,000 x *g* to 50 µL. The complex sample was then applied on a Superdex 200 3.2/300 column with 20 mM Hepes pH 7.6, 120 mM NaCl, 2 mM DTT. The FANCD2/FANCI fraction A11 was then collected and applied to grids for cryo-EM.

6.5 FANCI Protein Expression and Purification

Expression

FANCI with an engineered N-terminal 6x-His tag (provided by members of the Cohn laboratory) expression was completed using the pFastBac1 vector system (Life Technologies) according to the

manufacturer's protocol. Sf9 insect cells (provided by members of the Cohn laboratory) were grown in 100 mL cultures in 1 L vented Erlenmeyer flasks in Sf-900^{TM} II SFM insect cell media at 27°C, shaking at 135 rpm. Stock cultures were split to 1 x 10^6 cells/mL.

The cells were split three days before infection, to result in a density of 6 x 10^6 cells/mL on the day of infection, with the volume not exceeding 10% of the total flask volume. The infection protocol was as per the manufacturer's instructions. For a 15 cm dish FANCI expressing baculovirus, 35 x 10^6 insect cells, and 20 mL of Sf-900^{TM} II SFM insect cell media was added. The culture was then incubated at 27°C, for 3 days.

Purification

Sf9 cell pellets were re-suspended in lysis buffer (20 mM Tris pH 7.5, 100 mM KCl, 10% Glycerol, 2 mM $MgCl_2$, 0.2 mM PMSF, 2 mM 2-Mercaptoethanol) and sonicated. Lysates were clarified by centrifugation at 17,000 x *g* and the supernatants incubated with Ni-NTA agarose resin (Qiagen) for 2 hours. The resin was washed with lysis buffer, and then washed with wash buffer (20 mM Tris pH 7.5, 100 mM KCl, 40 mM imidazole, 10% glycerol, 2 mM $MgCl_2$). The protein was then eluted in elution buffer (20 mM Tris pH 7.5, 100 mM KCl, 250 mM imidazole, 10% glycerol, 2 mM $MgCl_2$).

6.6 FANCD2 Protein Expression and Purification

Expression

FANCD2 with an engineered N-terminal FLAG-HA tag (provided by members of the Cohn laboratory) expression was completed using the pFastBac1 vector system (Life Technologies) according to the manufacturer's protocol. Sf9 insect cells (provided by members of the Cohn laboratory) were grown in 100 mL cultures in 1 L vented Erlenmeyer flasks in Sf-900^{TM} II SFM insect cell media at 27°C, shaking at 135 rpm. Stock cultures were split to 1 x 10^6 cells/mL.

The cells were split three days before infection, to result in a density of 6 x 10^6 cells/mL on the day of infection, with the volume not exceeding 10% of the total flask volume. The infection protocol was as per the manufacturer's instructions. For a 15 cm dish, FANCD2 expressing baculovirus, 35 x 10^6 insect cells, and 20 mL of Sf-900^{TM} II SFM insect cell media was added. The culture was then incubated at 27°C, for 3 days.

Purification

Sf9 cell pellets were re-suspended in lysis buffer (20 mM Tris pH 7.5, 100 mM KCl, 10% Glycerol, 2 mM $MgCl_2$, 0.2 mM PMSF, 2 mM 2-Mercaptoethanol) and sonicated. Lysates were clarified by centrifugation at 17,000 x *g* and the supernatants incubated with M2 anti-FLAG agarose resin (A2220, Sigma) for 2 hours. The resin was washed with lysis buffer, and the protein was eluted in lysis buffer containing 0.5 mg/mL FLAG peptide.

6.7 Ubiquitin Expression and Purification

Expression

Ubiquitin containing an N-terminal 6xHis tag (provided by members of the Cohn laboratory) was expressed in *E. coli* BL21(DE3) cells (Novagen). Cells were grown in LB and supplemented with 50 µg/mL Kanamycin at 37°C. Expression of UBE2T was induced at OD600 = 0.8 with 0.5 mM IPTG and cultured for two hours at 37°C.

Purification

The cells were harvested by centrifugation at 4 000 x *g* and were lysed by French Press in lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM KCl, 20 mM Imidazole, 10% Glycerol). The lysate was clarified by centrifugation at 48,000 x *g* and the supernatant was incubated with Ni-NTA agarose (QIAGEN) for 2 hours at 4°C. The resin was washed with lysis buffer, and the protein was eluted in elution buffer (20 mM Tris-HCl pH 8.0, 300 mM KCl, 250 mM Imidazole, 10% Glycerol). The eluate was loaded onto a Superdex 200 16/60 column with 20 mM Tris-HCl pH 8.0, 100 mM KCl, 5% Glycerol. The ubiquitin peak fractions were then collected.

6.8 USP1/UAF1 Protein Complex Expression and Purification

Expression

USP1 with an engineered N-terminal FLAG-HA tag (provided by members of the Cohn laboratory) and UAF1 with an engineered N-terminal Strep tag (provided by members of the Cohn laboratory) expression was completed using the pFastBac1 vector system (Life Technologies) according to the manufacturer's protocol. Sf9 insect cells (provided by members of the Cohn laboratory) were grown in 100 mL cultures in 1 L vented Erlenmeyer flasks in Sf-900™ II SFM insect cell media at 27°C, shaking at 135 rpm. Stock cultures were split to 1 x 10^6 cells/mL.

The cells were split three days before infection, to result in a density of 6 x 10^6 cells/mL on the day of infection, with the volume not exceeding 10% of the total flask volume. The infection

protocol was as per the manufacturer's instructions. For a 15 cm dish, USP1 expressing baculovirus, excess UAF1 expressing baculovirus, 35 x 10^6 insect cells, and 20 mL of Sf-900™ II SFM insect cell media was added. The culture was then incubated at 27°C, for 3 days.

Purification

Sf9 cell pellets were re-suspended in lysis buffer (20 mM Tris pH 7.5, 100 mM KCl, 10% Glycerol, 2 mM $MgCl_2$, 0.2 mM PMSF, 2 mM 2-Mercaptoethanol) and sonicated. Lysates were clarified by centrifugation at 17,000 x *g* and the supernatants incubated with M2 anti-FLAG agarose resin (A2220, Sigma) for 2 hours. The resin was washed with lysis buffer, and the protein was eluted in lysis buffer containing 0.5 mg/mL FLAG peptide.

6.9 UHRF1 Protein and UHRF1 Mutant Protein Expression and Purification

Expression

UHRF1 with an engineered C-terminal FLAG-HA tag (provided by members of the Cohn laboratory) and UHRF1ΔUBL, UHRF1ΔTTD, UHRF1ΔPHD, UHRF1ΔSRA and UHRF1ΔRING with engineered N-terminal FLAG-HA tags (all provided by members of the Cohn laboratory) protein expression was completed using the pFastBac1 vector system (Life Technologies) according to the manufacturer's protocol. Sf9 insect cells (provided by members of the Cohn laboratory) were grown in 100 mL cultures in 1 L vented Erlenmeyer flasks in Sf-900™ II SFM insect cell media at 27°C, shaking at 135 rpm. Stock cultures were split to 1 x 10^6 cells/mL.

The cells were split three days before infection, to result in a density of 6 x 10^6 cells/mL on the day of infection, with the volume not exceeding 10% of the total flask volume. The infection protocol was as per the manufacturer's instructions. For a 15 cm dish, UHRF1 or UHRF1 domain deletions expressing baculovirus, 35 x 10^6 insect cells, and 20 mL of Sf-900™ II SFM insect cell media was added. The culture was then incubated at 27°C, for 3 days.

Purification

Sf9 cell pellets were re-suspended in lysis buffer (20 mM Tris pH 7.5, 100 mM KCl, 10% Glycerol, 2 mM $MgCl_2$, 0.2 mM PMSF, 2 mM 2-Mercaptoethanol) and sonicated. Lysates were clarified by centrifugation at 17,000 x *g* and the supernatants incubated with M2 anti-FLAG agarose resin (A2220, Sigma) for 2 hours. The resin was washed with lysis buffer, and the protein was eluted in lysis buffer containing 0.5 mg/mL FLAG peptide. For UHRF1 samples used for negative staining,

the sample was then applied on a Superdex 200 3.2/300 column with 20 mM Hepes pH 7.6, 120 mM NaCl, 2 mM DTT. The UHRF1 fractions 18 and 19 were collected and pooled, and applied to grids for negative stain electron microscopy. For UHRF1 samples used for cryo-EM, the sample was applied to a 10 kDa concentrator (Amicon Ultracel 10K) at 12,000 x *g* to 50 µL. The UHRF1 fraction 19 was collected and applied to grids for cryo-EM.

6.10 scDNA Transformation and Purification

Transformation of scDNA: pBlueScript SKII (+)

Transformation of Thermo Fisher Scientific Top 10 *E. coli* cells was completed following the ThermoFisher Top 10 protocol. In brief, the plasmid DNA was added to the *E. coli* Top 10 competent cells and left on ice for 30 minutes. The mixture was then heated at 42°C for exactly 45 seconds before being placed on ice for 2 minutes. LB was added to the cells and the cells were allowed to grow for one hour at 37°C. The resulting cells were plated onto LB agar plates containing 100 µg/mL ampicillin and left to grow overnight.

Purification of scDNA: pBlueScript SKII (+)

A colony per plate was grown for 12 hours in 2 mL LB, 100 µg/mL ampicillin at 37°C on a 200 rpm shaker. Each culture was then grown in 500 mL LB, 100 µg/mL ampicillin overnight at 37°C on a 200 rpm shaker. The scDNA was then isolated using the QIAGEN Plasmid Midiprep Kit.

6.11 Preparation of Biotinylated DNA Substrate for Filament Formation

The pBlueScript SKII (+) plasmid purified previously was used as a template to generate the biotinylated DNA using primers: 5'- BIO-(T55)-SPC3-GTAAAACGACGGCCAGTGAGCGCG-3' and 5'-ATCGGAATTCGTGGCACTTTTCGGGGAAATGTGC -3'. A large-scale PCR reaction was completed, and the PCR product was purified using the QIAquick PCR Purification Kit.

6.12 FANCD2/FANCI Ubiquitination and Deubiquitination Assays

The *in vitro* ubiquitination assay is a derivative of a previously published method (Hodson et al., 2014). 17 nM FLAG-HA-UBA1, 0.78 mM UBE2T, 1.86 µM FLAG-HA-FANCL, 4.2 mM His-Ubiquitin, 0.25 mM FLAG-HA-FANCD2/His-FANCI complex, 50 mM ATP and 40 mM scDNA substrate were incubated in the reaction buffer (50 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 0.5 mM DTT and 2 mM ATP) at room temperature for the indicated time. For DNA concentration tests, 0 – 80 mM scDNA substrate was added as indicated. For deubiquitination

reactions, the ubiquitination reactions were stopped with 5mM EDTA, and treated with either 500 units of benzonase or MilliQ water. Following this 192 nM FLAG-HA-USP1/Strep-UAF1 was added. The ubiquitination assay and deubiquitination assay was subjected to Coomassie blue staining to determine ubiquitination/deubiquitination efficiency.

6.13 FANCI Ubiquitination and Deubiquitination Assays

The *in vitro* ubiquitination assay is a derivative of a previously published method (Hodson et al., 2014). 17 nM FLAG-HA-UBA1, 0.78 mM UBE2T, 1.86 μM FLAG-HA-FANCL, 4.2 mM His-Ubiquitin, 0.25 mM His-FANCI, 50 mM ATP and 40 mM scDNA substrate were incubated in the reaction buffer (50 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 0.5 mM DTT and 2 mM ATP) at room temperature for the indicated time. For DNA concentration tests, 0 – 80 mM scDNA substrate was added as indicated. For deubiquitination reactions, the ubiquitination reactions were stopped with 5mM EDTA, and treated with either 500 units of benzonase or MilliQ water. Following this 192 nM FLAG-HA-USP1/Strep-UAF1 was added. The ubiquitination assay and deubiquitination assay was subjected to Coomassie blue staining to determine ubiquitination/deubiquitination efficiency.

6.14 Purification of the Monoubiquitinated-FANCD2/FANCI complex

The ubiquitination reaction mixture was applied to the 1 MDa concentrator (Satorius, Vivaspin 500), washed with reaction volume of 150 mM NaCl and concentrated at 50 x *g* to 50 μL. The scDNA-bound monoubiquitinated-FANCD2/FANCI complex was then treated with 500 units of benzonase and left on ice for two hours. The sample was then applied on a Superdex 200 3.2/300 column with 20 mM Hepes pH 7.6, 120 mM NaCl, 2 mM DTT. The monoubiquitinated-FANCD2/FANCI A12 fraction was then collected and applied to grids for negative stain electron microscopy or for cryo-EM.

6.15 Purification of the Monoubiquitinated-FANCI/FANCD2 complex

The ubiquitination reaction mixture was dialysed for one hour in 20 mM Tris pH 7.5, 100 mM KCl, 10% glycerol, 2 mM $MgCl_2$ in 6-8 akDa cut-off tubing (SpectroPor) to remove ATP. The remaining reaction mixture was incubated with 0.125 mM FLAG-HA-FANCD2 for 30 minutes, before being applied to a 10 kDa concentrator (Amicon Ultracel 10K) at 12,000 x *g* to 50 μL. The scDNA-bound monoubiquitinated-FANCI/FANCD2 complex was then treated with 500 units of benzonase and left on ice for two hours. The sample was then applied on a Superdex 200 3.2/300 column with 20 mM Hepes pH 7.6, 120 mM NaCl, 2 mM DTT. The monoubiquitinated-

FANCI/FANCD2 fraction 14 was then collected and applied to grids for negative stain electron microscopy or for cryo-EM.

6.16 Purification of Filamentous Arrays

Following the ubiquitination assay described above using either 8 nM or 20 nM of the biotinylated DNA substrate, the reaction mixture was incubated with either streptavidin sepharose beads (GE healthcare) or streptavidin magnetic beads (Invitrogen Dynabeads Streptavidin) for 2 hours. The beads were captured and washed with wash buffers with stepwise increasing salt concentration (20 mM HEPES pH 7.6, 120 – 175 mM NaCl, 2mM DTT). The biotinylated DNA was eluted in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT, containing 200 units SmaI (New England BioLabs). The eluted samples were collected and applied to grids for negative stain electron microscopy.

6.17 Filamentous Array Negative Control Sample Preparation

Repurified linear DNA alone: 20 nM of the biotinylated DNA substrate was incubated with streptavidin magnetic beads (Invitrogen Dynabeads Streptavidin) for 2 hours, followed by being captured and washed with wash buffers as above. The biotinylated DNA was eluted in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT, containing 200 units SmaI (New England BioLabs). The eluted sample was collected and applied to grids for negative stain electron microscopy.

Repurified linear DNA + FANCD2/FANCI: 20 nM of the biotinylated DNA substrate was incubated with 0.25 mM FLAG-HA-FANCD2/His-FANCI complex for 3 hours. The reaction mixture was then incubated with streptavidin magnetic beads (Invitrogen Dynabeads Streptavidin) for 2 hours, followed by being captured and washed with wash buffers as above. The biotinylated DNA was eluted in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT, containing 200 units SmaI (New England BioLabs). The eluted sample was collected and applied to grids for negative stain electron microscopy.

Crude linear DNA + SmaI: 20 nM of the biotinylated DNA substrate in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT, containing 200 units SmaI (New England BioLabs) was applied to grids for negative stain electron microscopy.

SmaI alone: 200 units SmaI (New England BioLabs) in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT was applied to grids for negative stain electron microscopy.

Crude linear DNA: 20 nM of the biotinylated DNA substrate in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT was applied to grids for negative stain electron microscopy.

Crude linear DNA + FANCD2/FANCI: 20 nM of the biotinylated DNA substrate was incubated with 0.25 mM FLAG-HA-FANCD2/His-FANCI complex for 3 hours in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT. The sample was then applied to grids for negative stain electron microscopy.

Crude scDNA: 20 nM of the scDNA substrate in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT was applied to grids for negative stain electron microscopy.

Crude scDNA + SmaI: 20 nM of the scDNA substrate in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT, containing 200 units SmaI (New England BioLabs) was applied to grids for negative stain electron microscopy.

Crude scDNA + FANCD2/FANCI: 20 nM of the scDNA substrate was incubated with 0.25 mM FLAG-HA-FANCD2/His-FANCI complex for 3 hours in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT. The sample was then applied to grids for negative stain electron microscopy.

Crude FANCD2/FANCI: 0.25 mM FLAG-HA-FANCD2/His-FANCI complex was incubated for 3 hours in 20 mM HEPES pH 7.6, 120 mM NaCl, 2mM $MgCl_2$, 2mM DTT. The sample was then applied to grids for negative stain electron microscopy.

6.18 Negative Stain Electron Microscopy

Freshly prepared FANCD2/FANCI, monoubiquitinated-FANCD2/FANCI, monoubiquitinated-FANCI/FANCD2, UHRF1 and filament samples were applied to glow-discharged, carbon coated grids at 0.01 μg/μL and allowed to adsorb for 120 seconds. The grids were then stained with 2% uranyl acetate for 30 seconds. The samples were imaged at a magnification of 30,000 – 68,000 x, 1.727 Å/pixel with a 16 Megapixel Gatan OneView™ CMOS camera using an FEI Tecnai 12 Transmission Electron Microscope, LaB_6 electron source operating at 120 kV.

6.19 Negative Stain Electron Image Processing

Micrographs were processed in EMAN2.2 with box size 300 for FANCD2/FANCI, monoubiquitinated-FANCD2/FANCI, and monoubiquitinated-FANCI/FANCD2 samples. All particles were picked manually in EMAN2.2, with approximately 3000 - 6000 particles selected. 50 2D classifications were generated, and an initial 3D model projection generated from the best manually selected 18-20 2D classes (Tang et al., 2007).

Micrographs were processed in EMAN2.2 with box size 100 for UHRF1 samples. All particles were picked manually in EMAN2.2, with a minimum of 6000 particles selected. 100 2D classifications were generated, and an initial 3D model projection generated from the best manually selected 45 2D classes. The initial 3D model projection was refined step-wise to yield a 17.7 Å 3D model (Tang et al., 2007).

6.20 Cryo-Electron Microscopy

Freshly prepared FANCD2/FANCI, monoubiquitinated-FANCD2/FANCI, monoubiquitinated-FANCI/FANCD2 and UHRF1 samples were applied to glow-discharged Lacey Carbon Film 300 Mesh Cu, Quantifoil 1.2/1.3 300 Mesh Cu, Quantifoil 1.2/1.3 300 Mesh Au or UltraAuFoil 1.2/1.3 300 Mesh Au grids, or to homemade Graphene Oxide Support Film 2/1 300 Mesh Cu or Quantifoil Graphene Oxide Support Film 2/2 300 Film Cu grids. The grids were then blotted and vitrified using an FEI Vitrobot Mk IV Automated plunge freezer. The samples were imaged in counting mode on an FEI Talos Arctica (Thermo Fisher Scientific) operating at 200 kV with a Falcon 3 detector (Thermo Fisher Scientific) using a pixel size of 0.78 Å. For UHRF1 samples, a grid was also imaged in counting mode on an FEI Titan Krios G3 (Thermo Fisher Scientific) operating at 300 kV with a Gatan Imaging Filter (GIF) energy filter (Gatan) and K3 detector (Gatan) using a pixel size of 0.416 Å with 1328 movies collected.

6.21 Cryo-Electron Microscopy Image Processing

For the FANCD2/FANCI sample, micrographs were processed in SIMPLE with a box size 300. 5770 particles were auto-picked using SIMPLE and 100 2D classifications were generated (Reboul et al., 2018).

For the UHRF1 sample, micrographs were processed in SIMPLE with a box size of 100. 5000 particles were manually picked using SIMPLE and 20 2D classifications generated. An initial 3D model projection was generated from this 2D classification. Subsequently the 2D classification was used as a template to auto-pick 1079219 particles from 1328 micrographs using SIMPLE and 300 2D low-resolution classifications generated. Finally, the low-resolution 2D classifications were used to generate 200 high-resolution 2D classifications (Reboul et al., 2018).

7. References

Abbott, S., Iudin, A., Korir, P. K., Somasundharam, S. and Patwardhan, A., 2018, EMDB web resources, Curr. Protoc. Bioinformatics, 61(1): 5.10.1-5.10.12.

Achour, M., Jacq, X., Rondé, P., Alhosin, M., Charlot, C., Chataigneau, T., Jeanblanc, M., Macaluso, M., Giordano, A., Hughes, A. D., Schini-Kerth, V. B. and Bronner, C., 2008, The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression, *Oncogene*, 27: 2187-2197.

Adamo, A., Collis, S. J., Adelman, C. A., Silva, N., Horejsi, Z., Ward, J. D., Martinez-Perez, E., Boulton, S. J. and La Volpe, A., 2010, Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia, *Molecular Cell*, 39: 25-35.

Alcón, P., Shakeel, S., Chen, Z. A., Rappsilber, J., Patel, K. J. and Passmore, L. A., 2020, FANCD2-FANCI is a clamp stabilized on DNA by monoubiquitination of FANCD2 during DNA repair, *Nature Structural & Molecular Biological*, 27: 240-248.

Alnabati, E. and Kihara, D., 2020, Advances in structure modelling methods for cryo-electron microscopy maps, *Molecules*, 25(82): 1-13.

Alpi, A. F., Pace, P. E., Babu, M. M. and Patel, K. J., 2008, Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI, *Molecular Cell*, 32: 767-777.

Altieri, F. Grillo, C., Maceroni, M. and Chichiarelli, S., 2008, DNA damage and depair: from molecular mechanisms to health implications, *Antioxidants & Redox Signaling*, 10(5): 891-937.

Amunugama, R., Willcox, S., Wu, R. A., Abdullah, U. B., El-Sagheer, A. H., Brown, T., McHugh, P. J., Griffin, J. D. and Walter, J. C., 2018, Replication fork reversal during DNA Interstrand crosslink repair requires CMG unloading, *Cell Reports*, 23(12): 3419-3428.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. and Shirakawa, M., 2008, Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism, *Nature*, 455: 818-822.

Arita, K., Isogai, S., Oda, T., Unoki, M., Sugita, K., Sekiyama, N., Kuwata, K., Hamamoto, R., Tochio, H., Sato, M., Ariyoshi, M. and Shirakawa, M., 2012, Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1, *PNAS*, 109(32): 12950-12955.

Arkinson, C., Chaugule, V. K., Toth, R. and Walden, H., 2018, Specificity for deubiquitination of monoubiquitinated FANCD2 is driven by the N-terminus of USP1, *Life Science Alliance*, 1(5): e201800162.

Arnold, S., A., Albiez, S., Bieri, A., Syntychaki, A., Adaixo, R., McLeod, R. A., Goldie, K. N., Stahlberg, H. and Braun , T., 2017, Blotting-free and lossless cryo-electron microscopy grid preparation from nanolitre-size protein samples and single-cell extracts, J. Structural Biology, 197(3): 220-226.

Auerbach, A. D., 2009, Fanconi anemia and its diagnosis, *Mutat. Res.*, 668: 1-15.

Avvakumov, G. V., Walker, J. R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C. H. and Dhe-Paganon, S., 2008, Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1, *Nature*, 455: 822-826.

Baddock, H. T., Yosaatmadja, Y., Newman, J. A., Schofield, C. J., Gileadi, O. and McHugh, P. J., 2020, The SNM1A DNA repair nuclease, *DNA Repair*, 102941.

Bai, X-c., McMullan, G. and Scheres, S. H. W., 2015, How cryo-EM is revolutionizing structural biology, *Trends in Biochemical Sciences*, 40(1): 49-58.

Bendersky, L. A. and Gayle, F. W., 2001, Electron diffraction using transmission electron microscopy, *Journal of Research of the National Institute of Standards and Technology*, 106(6): 997-1012.

Berkyurek, A. C., Suetake, I., Arita, K., Takeshita, K., Nakagawa, A., Shirawa, M. and Tajima, S., 2014, The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA, *J. Biol. Chem.*, 289(1): 379-386.

Bhat, K. P. and Cortez, D., 2018, RPA and RAD51: fork reversal, fork protection and genome stability, *Nature Structural & Molecular Biology*, 25: 446-453.

Bhattacharjee, S. and Nandi, S., 2017, DNA damage response and cancer therapeutics through the lens of the Fanconi anemia DNA repair pathway, *Cell Communication and Signaling*, 15: 41-51.

Bostick, M., Kim, J. K., Estève, P-O., Clark, A, Pradhan, S. and Jacobsen, S. E., 2007, UHRF1 plays a role in maintaining DNA methylation in mammalian cells, *Science*, 317: 1760-1764.

Bouwman, P. and Jonkers, J., 2012, The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance, *Nature Reviews*, 12: 587-598.

Brooks-Bartlett, J. C. and Garman, E. F., 2015, The Nobel science: one hundred years of crystallography, *Interdisciplinary Science Reviews*, 40(3): 244-264.

Buzon, B., Grainger, R., Huang, S., Rzadki, C. and Junop, M. S., 2018, Structure-specific endonuclease activity of SNM1A enables processing of a DNA Interstrand crosslink, *Nucleic Acids Research*, 46(17): 9057-9066.

Campos, A. and Clemente-Blanco, A., 2020, Cell cycle and DNA repair regulation in the damage response: protein phosphatases take over the reins, *Int. J. Mol. Sci.*, 21: 466-472.

Castella, M., Jacquemont, C., Thompson, E. L., Yeo, J. E., Cheung, R. S., Huang, J-W. Sobeck, A., Hendrickson, E. A. and Taniguchi, T., 2015, FANCI regulates recruitment of the FA core complex at sites of DNA damage independently of FANCD2, *PLoS Genetics*, 11(10): e1005563.

Ceccaldi, R., Sarangi, P. and D'Andrea, A. D., 2016, The Fanconi anaemia pathway: new players and new functions, *Nature Reviews: Mol. Cell Biol.*, 17: 337-349.

Chatterjee, N. and Walker, G. C., 2017, Mechanisms of DNA damage, repair, and mutagenesis, *Environmental and Molecular Mutagenesis*, 58: 235-263.

Chaugule, V. K., Arkinson, C., Rennie, M. L., Kämäräinen, O., Toth, R. and Walden, H., 2020, Allosteric mechanism for site-specific ubiquitination of FANCD2, *Nature Chemical Biology*, 16: 291-301.

Cheng, J., Yang, Y., Xiao, J., Zhu, T., Chen, F., Wang, P., Li, Z., Yang, H. and Xu, Y., 2013, Structural insight into coordinated recognition of trimethylated histone H3 lyine 9 (H3K9me3) by the Plant Homeodomain (PHD) and Tandem Tudor Domain (TTD) of UHRF1 (Ubiquitin-like, containing PHD and RING Finger domains, 1) protein, *J. Biol. Chem.*, 288(2): 1329-1339.
Ciccia, A. and Elledge, S. J., 2010, The DNA damage response: making it safe to play with knives, molecular cell, 40: 179-204.

Cohn, M. A., Kowal, P., Yang, K., Haas, W., Huang, T. T., Gygi, S. P. and D'Andrea, A. D., 2007, A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway, *Mol. Cell.*, 28: 786-797.

Cohn, M. A., Kee, Y., Haas, W., Gygi, S. P. and D'Andrea, A. D., 2009, UAF1 is a subunit of multiple deubiquitinating enzyme complexes, *J. Biol. Chem.*, 284 (8): 5343-5381.

Colnaghi, L., Jones, M. J. K., Cotto-Rios, X. M., Schindler, D., Hanenberg, H. and Huang, T. T., 2011, Patient-derived C-terminal mutation of FANCI cause protein mislocalization and reveals a putative EDGE motif function in DNA repair, *Blood*, 117(7): 2247-2256.

Colombo, C. V., Gnugnoli, M., Gobbini, E. and Longhese, M. P., 2020, How do cells sense DNA lesions, *Biochemical Society Transactions*, 48(2): 677-691.

Curtin, N. J., 2012, DNA repair dysregulation from cancer driver to therapeutic target, *Nature Reviews*, 12: 801-817.

Cussiol, J. R. R., Soares, B. L. and de Oliveira, F. M. B., 2020, From yeast to humans: Understanding the biology of DNA damage response (DDR) kinases, *Genetics and Molecular Biology*, 43: 1-10.

Danev, R. Yanagisawa, H. and Kikkawa, M., 2019, Cryo-electron microscopy methodology: current aspects and future directions, *Trends in Biochemical Sciences*, 44(10): 837-848.

Davis, J. D. and Lin, S-Y., 2011, DNA damage and breast cancer, *World J. Clin. Oncol.*, 2(9): 329-338.

Deans, A. J. and West, S. C., 2013, DNA interstrand crosslink repair and cancer, *Nat. Rev. Cancer*, 11(7): 467-480.

Dorsman, J. C., Levitus, M., Rockx, D., Rooimans, M. A., Oostra, A. B., Haitjema, A., Bakker, S. T., Steltenpool, J., Schuler, D., Mohan, S., Schindler, D., Arwert, F., Pals, G., Mathew, C. G., Waisfisz, Q., De Winter, J. P. and Joenje, H., 2007, Identification of the Fanconi anemia complementation group I gene, FANCI, *Cell. Onc.*, 29(3): 211-218.

Duxin, J. P. and Walter, J. C., 2015, What is the DNA repair defect underlying Fanconi anemia?, *Curr. Opin. Cell Biol.*, 37: 49-60.

Fang, C-B., Wu, H-T., Zhang, M-L., Liu, J. and Zhang, G-J., 2020, Fanconi anemia pathway: mechanism of breast cancer predisposition development and potential therapeutic targets, *Front. Cell Dev. Biol.*, 8(160): 1-15.

Fang, J., Cheng, J., Wang, J., Zhang, Q., Liu, M., Gong, R., Wang, P., Zhang, X., Feng, Y., Lan, W., Gong, Z., Tang, C., Wong, J., Yang, H., Cao, C. and Xu, Y., 2016, Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition, *Nature Communications*, 7: 11197.

Felle, M., Joppien, S., Németh, A., Diermeier, S., Thalhammer, V., Dobner, T., Kremmer, E., Kappler, R. and Längst, G., 2011, The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1, *Nucleic Acids Research*, 39(19): 8355-8365.

Friedberg, E. C., 2003, DNA damage and repair, *Nature*, 421: 436-440.

Gao, L., Tan, X-F., Zhang, S., Wu, T., Zhang, Z-M., Ai, H-w. and Song, J., 2018, An intramolecular interaction of UHRF1 reveals dual control for its histone association, *Structure*, 26: 304-311.

Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M. S., Timmers, C., Hejna, J., Grompe, M. and D'Andrea, A. D., 2001, Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway, *Mol Cell.*, 7(2): 249-262.

Gavande, N. S., VanderVere-Carozza, P. S., Hinshaw, H. D., Jalal, S. I., Sears, C. R., Pawelczak, K. S. and Turchi, J. J., 2016, DNA repair targeting therapy: the past or future of cancer treatment, *Pharmacology & Therapeutics*, 160: 65-83.

Gelato, K. A., Tauber, M., Ong, M. S., Winter, S., Hiragami-Hamada, K., Sindlinger, J., Lemak, A., Bultsma, Y., Houliston, S., Schwarzer, D., Divecha, N., Arrowsmith, C. H. and Fischle, W., 2014, Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate, *Molecular Cell*, 54: 905-919.

Gibbs-Seymour, I., Oka, Y., Rajendra, E., Weinert, B. T., Passmore, L. A., Patel, K. J., Olsen, J. V., Choudhary, C., Bekker-Jensen, S. and Mailand, N., 2015, Ubiquitin-SUMO circuitry controls activated Fanconi anemia ID complex dosage in response to DNA damage, *Molecular Cell*, 57: 150-164.

Glaeser, R. M., 2018, Proteins, interfaces, and cryo-EM grids, *Current Opinion in Colloid & Interface Science*, 34: 1-8.

Gudmundsdottir, K. and Ashworth, A., 2006, The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability, *Oncogene*, 25: 5864-5874.

Hahm, J. Y., Kim, J-Y., Park, J. W., Kang, J-Y., Kim, K-B., Kim, S-R., Cho, H. and Seo, S-B., 2018, Methylation of UHRF1 by SET7 is essential for DNA double-strand break repair, *Nucleic Acids Research*, 47(1): 184-196.

Hakem, R., 2008, DNA-damage repair; the good, the bad and the ugly, *The EMBO Journal*, 27(4): 589-605.

Harrison, J. S., Cornett, E. M., Goldfarb, D., DaRosa, P. A., Li, Z. M., Yan, F., Dickson, B. M., Guo, A. H., Cantu, D. V., Kaustov, L., Brown, P. J., Arrowsmith, C. H., Erie, D. A., Major, M. B., Klevit, R. E., Krajewski, K., Kuhlman, B., Strahl, B. D. and Rothbart, S. B., 2016, Hemi-methylated

DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1, *eLife*, 5: e17101.

Hashimoto, H., Horton, J. R., Zhang, X., Bostick M., Jacobsen, S. E. and Cheng, X., 2008, The SRA domain of UHRF1 flips 5-methylatcytosine out of the DNA helix, *Nature*, 455(7214): 826-829.

Hay, R. T., 2005, SUMO: a history of modification, *Molecular Cell*, 18: 1-12.
Hodson, C., Purkiss, A., Miles, J. A. and Walden, H., 2014, Structure of the human FANCL RING-Ube2T complex reveals determinants of cognate E3-E2 selection, *Structure*, 22: 337-344.

Hosoya, N. and Miyagawa, K., 2014, Targeting DNA damage response in cancer therapy, C*ancer Sci.*, 105(4): 370-388.

Houtgraaf, J. H., Vermissen, J. and van der Giessen, W. J., 2006, A concise review of DNA damage checkpoints and repair in mammalian cells, *Cardiovascular Revascularization Medicine*, 7: 165-172.

Howard, S. M., Yanez, D. A. and Stark, J. M., 2015, DNA damage response factors from diverse pathways, including DNA crosslink repair, mediate alternative end joining, *PLoS Genet*, 11(1): e1004943.

Huang, J., Zhang, J., Bellani, M. A., Pokharel, D., Gichimu, J., James, R. C., Gali, H., Ling, C., Yan, Z., Xu, D., Chen, J., Meetei, A. R., Li, L., Wang, W. and Seidman, M. M., 2019, Remodeling of interstrand crosslink proximal replisomes is dependent on ATR, FANCM and FANCD2, *Cell Rep.*, 27(6): 1794-1808.

Huang, Y., Leung, J. W. C., Lowery, M., Matsushita, N., Wang, Y., Shen, X., Huong, D., Takata, M., Chen, J. and Li, L., 2014, Modularized functions of the Fanconi anemia core complex, *Cell Reports*, 7: 1849-1857.

Jackson, S. P. and Durocher, D., 2013, Regulation of DNA damage responses by ubiquitin and SUMO, *Molecular Cell*, 49: 795-807.

Jacq, X., Kemp, M., Martin, N. M. B. and Jackson, S. P., 2013, Deubiquitylating enzymes and DNA damage response pathways, *Cell Biochem. Biophys.*, 67: 25-43.

Joo, W., Xu, G., Persky, N. S., Smogorzewska, A., Rudge, D. G., Buzovetsy, O., Elledge, S. J. and Pavletich, N. P., 2011, Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway, *Science*, 333: 312-316.

Kim, K. and Cole, P. A., 1997, Measurement of a Brønsted nucleophile coefficient and insights into the transition state for a protein tyrosine kinase, *J. Am. Chem. Soc.*, 119: 11096-11097.

Kori, S., Ferry, L., Matano, S., Jimenji, T., Kodera, N., Tsusaka, T., Matsumura, R., Oda, T., Sato, M., Dohmae, N., Ando, T., Shinkai, Y., Defossez, P-A. and Arita, K., 2019, Structure of the UHRF1 tandem tudor domain bound to a methylated non-histone protein, LIG1, reveals rules for binding and regulation, *Structure*, 27: 485-496.

Lallous, N., Legrand, P., McEwen, A. G., Ramón-Maiques, S., Samama, J-P. and Birck, C., 2011, The PHD finger of human UHRF1 reveals a new subgroup of unmethylated histone H3 tail readers, *PLoS ONE*, 6(11): e27599.

Lehmann, A. R., 2006, Translesion synthesis in mammalian cells, *Experimental Cell Research*, 312: 2673-2676.

Li, N., Wang, J., Wallace, S. S., Chen, J., Zhou, J. and D'Andrea, A. D., 2020, Cooperation of the NEIL3 and Fanconi anemia/BRCA pathways in Interstrand crosslink repair, *Nucleic Acids Research*, 48(6): 3014-3028.

Li, Z., Pearlman, A. H. and Hsieh, P., 2016, DNA mismatch repair and the DNA damage response, *DNA Repair*, 38: 94-101.

Liang, C-C. and Cohn, M. A., 2015a, UHRF1 is a sensor for DNA Interstrand crosslinks, *Oncotarget*, 7(1): 3-4.

Liang, C-C., Zhan, B., Yoshikawa, Y., Haas, W., Gygi, S. P. and Cohn, M. A., 2015b, UHRF1 is a sensor for DNA Interstrand crosslinks and recruits FANCD2 to initiate the Fanconi anemia pathway, *Cell Reports*, 10: 1947-1956.

Liang, C-C., Li, Z., Lopez-Martinez, D., Nicholson, W. V., Vénien-Bryan, C. and Cohn, M. A., 2016, The FANCD2-FANCI complex is recruited to DNA Interstrand crosslinks before monoubiquitination of FANCD2, *Nature Communications*, 7: 12124.

Liang, F., Miller, A. S., Longerich, S., Tang, C., Maranon, D., Williamson, E. A., Hromas, R., Wiese, C., Kupler, G. M. and Sung P., 2019, DNA requirement in FANCD2 deubiquitination by USP1-UAF1-RAD51AP1 in the Fanconi anemia DNA damage response, *Nature Communications*, 10: 2849.

Lim, E. T., Wurtz, P., Havulinna, A. S., Palta, P., Tukiainen, T., Rehnström, K., Esko, T., Mägi, R., Inouye, M., Lappalainen, T., Chan, Y., Salem, R. M., Lek, M., Flannick, J., Sim, X., Manning, A., Ladenvall, C., Bumpstead, S., Hämäläinen, E., Aalto, K., Maksimow, M., Salmi, M., Blankenberg, S., Ardissino, D., Shah, S., Horne, B., McPherson, R., Hovingh, G. K., Reilly, M. P., Watkins, H., Goel, A., Farrall, M., Girelli, D., Reiner, A. P., Stitziel, N. O., Kathiresan, S., Gabriel, S., Barrett, J. C., Lethimäki, T., Laakso, M., Groop L., Kaprio, J., Perola, M., McCarthy, M. I., Boehnke, M., Altshuler, D. M., Lindgren, C. M., Hirschhorn, J. N., Metspalu, A., Freimer, N. B., Zeller, T., Jalkanen, S., Koskinen, S., Raitakari, O., Durbin, R., MacArthur, D. G., Salomaa, V., Ripatti, S., Daly, M. J. and Palotie, A., 2014, Distribution and medical impact of loss-of-function variants in the Finnish founder population, *PLoS Genet.*, 10(7): e1004494.

Lin, Z., Kong, H., Nei, M. and Ma, H., 2006, Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer, *PNAS*, 103(27): 10328-10333.

Liu, W., Palovcak, A., Li, F., Zafar, A., Yuan, F. and Zhang, Y., 2020, Fanconi anemia pathway as a prospective target for cancer intervention, *Cell Biosc.*, 10: 39.

Longerich, S., San Filippo, J., Liu, D. and Sung, P., 2009, FANCI binds branched DNA and is monoubiquitinated by UBE2T-FANCL, *J. of Biol. Chem.*, 284(35): 23182-23186.

Lopez-Martinez, D., Kupculak, M., Yang, D., Yoshikawa, Y., Liang, C-C., Wu, R., Gygi, S. P. and Cohn, M. A., 2019, Phosphorylation of FANCD2 inhibits the FANCD2/FANCI complex and suppresses the Fanconi anemia pathway in the absence of DNA damage, *Cell Reports*, 27: 2990-3005.

Lopez-Martinez, D., Liang, C-C. and Cohn, M. A., 2016, Cellular response to DNA interstrand crosslinks: the Fanconi anemia pathway, *Cell. Mol. Life Sci.,* 73: 3097-3114.

Mäeots, M-E., Lee, B., Nans, A., Jeong, S-G., Esfahani, M. M. N., Ding, S., Smith, D. J., Lee, C-S., Lee, S. S., Peter, M. and Enchev, R. I., 2020, Modular microfluidics enables kinetic insight from time-resolved cryo-EM, *Nature Communications*, 11: 3465.

Markwick, P. R. L., Malliavin, T. and Nilges, M., 2008, *PLoS Computational Biology*, 4(9): e1000168.
Martin, S. A., Lord, C. J. and Ashworth, A., 2008, DNA repair deficiency as a therapeutic target in cancer, *Current Opinion in Genetics & Development*, 18: 80-86.

MacKay, C., Déclais, A-C., Lundin, C., Agostinho, A., Deans, A. J., MacArtney, T. J., Hofmann, K., Gartner, A., West, S. C., Helleday, T., Lilley, D. M. J. and Rouse, J., 2010, Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2, *Cell*, 142: 65-76.

Madhusudan, Trafny, E. A., Xuong, N-H., Adams, J. A., Eyck, L. F. T., Taylor, S. S. and Sowadski, J. M., 1994, cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer, *Protein Science*, 3: 176-187.

Madhusudan, S. and Hickson, I. D., 2005, DNA repair inhibition: a selective tumour targeting strategy, *TRENDS in Molecular Medicine*, 11(11): 503-511.

Matsushita, N., Kitao, H., Ishiai, M., Nagashima, N., Hirano, S., Okawa, K., Ohta, T., Yu, D. S., McHugh, P. J., Hickson, I. D., Venkitaraman, A. R., Kurumizaka, H. and Takata, M., 2005, A FancD2-monoubiquitin fusion reveals hidden functions of Fanconi anemia core complex in DNA repair, *Molecular Cell*, 19: 841-847.

McPherson, A. and Gavira, J. A., 2013, Introduction to protein crystallization, *Acta Cryst.*, F70: 2-20.

Medvar, B., Raghuram, V., Pisitkun, T., Sarkar, A. and Knepper, M. A., 2016, Comprehensive database of human E3 ubiquitin ligases: application to aquaporin-2 regulation, *Physiological Genomics*, 48(7): 502-512.

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 orthologue of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M, *Nat. Genet.*, 37(9): 958-963.

Miles, J. A., Frost, M. G., Carroll, E., Rowe, M., Howard, M. J., Sidhu, A. Chaugule, V. K., Alpi, A. F. and Walden, H., 2015, The Fanconi anemia DNA repair pathway is regulated by an interaction between ubiquitin and the E2-like fold domain of FANCL, *J. Biol. Chem.*, 290(34): 20995-21006.

Moraes, M. C. S., Neto, J. B. C. and Menck, C. F. M., 2012, DNA repair mechanisms protect our genome from carcinogenesis, *Front. in Biosci.*, 17: 1362-1388.

Morreale, F. E., Bortoluzzi, A., Chaugule, V. K., Arkinson, C., Walden, H. and Ciulli, A., 2017a, Allosteric targeting of the Fanconi anemia ubiquitin-conjugating enzyme Ube2T by fragment screening, *J. Med. Chem.*, 60: 4093-4098.

Morreale, F. E., Testa, A., Chaugule, V. K., Bortoluzzi, A., Ciulli, A. and Walden, H., 2017b, Mind the metal: a fragment library-derived zinc impurity binds the E2 ubiquitin-conjugating enzyme Ube2T and induces structural rearrangements, *J. Med. Chem.*, 60(19): 8183-8191.

Motnenko, A., Liang, C-C., Yang, D., Lopez-Martinez, D., Yoshikawa, Y., Zhan, B., Ward, K. E., Tian, J., Haas, W., Spingardi, P., Kessler, B. M., Kriaucionis, S., Gygi, S. P. and Cohn, M. A., 2018, *PLoS Genet.*, 14(10): e1007643.

Mouw, K. W. and D'Andrea, A. D., 2014, Crosstalk between the nuecleotide excision repair and Fanconi anemia/BRCA pathways, *DNA Repair*, 19: 130-134.

Muto, M., Kanari, Y., Kubo, E., Takabe, T., Kurihara, T., Fujimori, A. and Tatsumi, K., 2002, Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks, *J. Biol. Chem.*, 277(37): 34549-34555.

Nady, N., Lemak, A., Walker, J. R., Avvakumov, G. V., Kareta, M. S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X., Wang, Y-X., Bronner, C., Chédein, F., Arrowsmith, C. H. and Dhe-Paganon, S., 2011, Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein, *J. Biol. Chem.*, 286(27): 24300-24311.

Nijman, S. M. B., Huang, T. T., Dirac, A. M. G., Brummelkamp, T. R., Kerkhoven, R. K., D'Andrea, A. D. and Bernards, R., 2005, The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway, *Mol. Cell*, 17(3): 331-339.

Niraj, J., Caron, M-C., Drapeau, K., Bérubé, S., Guitton-Sert, L., Coulombe, Y., Couturier, A. M. and Masson, J-Y., 2017, The identification of FANCD2 DNA binding domains reveals nuclear localisation sequences, *Nucleic Acids Research*, 45(14): 8341-8357.

Nishiyama, A., Yamaguchi, L., Sharif, J., Johmura, Y., Kawamura, T., Nakanishi, K., Shimamura, S., Arita, K., Kodama, T., Ishikawa, F., Koseki, H. and Nakanishi, M., 2013, Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication, *Nature*, 502: 249-255.

Nwanochie, E. and Uversky, V. N., 2019, Structure determination by single-particle cryo-electron microscopy: only the sky (and intrinsic disorder) is the limit, *Int. J. Mol. Sci.*, 20: 4186.

Oestergaard, V. H., Langevin, F., Kuiken, H. J., Pace, P., Niedzwiedz, W., Simpson, L. J., Ohzeki, M., Takata, M., Sale, J. E. and Patel, K. J., 2007, Deubiquitination of FANCD2 is required for DNA crosslink repair, *Molecular Cell*, 28: 798-809.

O'Connor, M. J., 2015, Targeting the DNA damage response in cancer, *Molecular Cell*, 60: 547-560.

O'Grady, S., Finn, S. P., Cuffe, S., Richard, D. J., O'Byrne, K. J. and Barr, M. P., 2014, The role of DNA repair pathways in cisplatin resistant lung cancer, *Cancer Treatment Reviews*, 40: 1161-1170.

Okamoto, Y., Iwasaki, W. M., Kugou, K., Takahashi, K. K., Oda, A., Sato, K., Kobayashi, W., Kawai, H., Sakasai, R., Takaori-Kondo, A., Yamamoto, T., Kanemaki, M. T., Taoka, M., Isobe, T., Kurumizaka, H., Innan, H., Ohta, K., Ishiai, M. and Takata, M., 2018, Replication stress induces accumulation of FANCD2 at central region of large fragile genes, *Nucleic Acids Research*, 46(6): 2932-2944.

Pearl, L. H., Schierz, A. C., Ward, S. E., Al-Lazikani, B. and Pearl, F. M. G., 2015, Therapeutic opportunities within the DNA damage response, *Nature Reviews Cancer*, 15: 166-180.

Pichler, A., Knipscheer, P., Saitoh, H., Sixma, T. K. and Melchior, F., 2004, The RanBP2 SUMO E3 ligase is neither HECT- nor RING-type, *Nature Structural & Molecular Biology*, 11(10): 984-991.

Qian, C., Li, S., Jakoncic, J., Zeng, L., Walsh, M. J. and Zhou, M-M., 2008, Structure and hemimethylated CpG binding of the SRA domain from human UHRF1, *J. Biol. Chem.*, 283(50): 34490-34494.

Raghunandan, M., Chaudhury, I., Kelich, S. L., Hanenberg, H. and Sobeck, A., 2015, FANCD2, FANCJ and BRCA2 cooperate to promote replication fork recovery independently of the Fanconi anemia core complex, *Cell Cycle*, 14(3): 342-353.

Rajendra, E., Oestergaard, V. H., Langevin, F., Wang, M., Dornan, G. L., Patel, K. J. and Passmore, L. A., 2014, The genetic and biochemical basis of FANCD2 monoubiquitination, *Molecular Cell*, 54: 858-869.

Räschle, M., Knipsheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J. D., Ellenberger, T. E., Schärer, O. D. and Walter, J. C., 2008, Mechanism of replication-coupled DNA Interstrand cross-link repair, *Cell*, 134(6): 969-980.

Reboul, C. F., Eager, M., Elmlund, D. and Elmlund, H., 2018, Single-particle cryo-EM-Improved ab initio reconstruction with SIMPLE/PRIME, Protein Sci., 27(1): 51-61.

Renaud, E., Barascu, A. and Rosselli, F., 2016, Impaired TIP60-mediated H4K16 acetylation accounts for the aberrant chromatin accumulation of 53BP1 and RAP80 in Fanconi anemia pathway-deficient cells, *Nucleic Acids Research*, 44(2): 648-656.

Rennie, M. L., Lemonidis, K., Arkinson, C., Chaugule, V. K., Clarke, M., Streetley, J., Spagnolo, L. and Walden, H., 2020, Differential functions of FANCI and FANCD2 ubiquitination stabilize ID2 complex on DNA, *EMBO reports*, 21: e50133.

Richardson, C. D., Kazane, K. R., Feng, S. J., Zelin, E., Bray, N. L., Schäfer, A. J., Floor, S. N. and Corn, J. E., 2018, CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway, *Nature Genetics*, 50: 1132-1139.

Russo, C. J. and Passmore, L. A., 2014, Ultrastable gold substrates for electron cryomicroscopy, Science, 346(6215): 1377-1380.

Sareen, A., Chaudhury, I., Adams, N. and Sobeck, A., 2012, Fanconi anemia proteins FANCD2 and FANCI exhibit different DNA damage responses during S-phase, *Nucleic Acids Research*, 40(17): 8425-8439.

Sato, K., Ishiai, M., Toda, K., Furukoshi, S., Osakabe, A., Tachiwana, H., Takizawa, Y., Kagawa, W., Kitao, H., Dohmae, N., Obuse, C., Kimura, H., Takata, M. and Kurumizaka, H., 2012a, Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair, *The EMBO Journal*, 31: 3524-3536.

Sato, K., Toda, K., Ishiai, M., Takata, M. and Kurumizaka, H., 2012b, DNA robustly stimulates FANCD2 monoubiquitylation in the complex with FANCI, *Nucleic Acids Research*, 40(10): 4554-4561.

Sato, K., Shimomuki, M., Katsuki, Y., Takahashi, D., Kobayashi, W., Ishiai, M., Miyoshi, H., Takata, M. and Kurumizaka, H., 2016, FANCI-FANCD2 stabilises the RAD51-DNA complex by binding RAD51 and protects the 5'-DNA end, *Nucleic Acids Research*, 44(22): 10758-10771.

Schärer, O. D., 2005, DNA Interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses, *ChemBioChem*, 6: 27-32.

Schwab, R. A., Nieminuszczy, J., Shah, F., Langton, J., Lopez-Martinez, D., Liang, C-C., Cohn, M. A., Gibbons, R. J., Deans, A. J. and Niedzwiedz, W., 2015, The Fanconi anemia pathway maintains genome stability by coordinating replication and transcription, *Molecular Cell*, 60: 351-361.

Scott, B. L. and Hoppe, A. D., 2015, Optimising fluorescent protein trios for 3-way FRET imaging of protein interactions in living cells, *Scientific Reports*, 5: 10270.

Sgro, G. G. and Costa, T. R. D., 2018, Cryo-EM grid preparation of membrane protein samples for single particle analysis, *Front. in Mol. Biosci.*, 5(74): 1-8.

Shakeel, S., Rajendra, E., Alcón, P., O'Reilly, F., Chorev, D. S., Maslen, S., Degliesposti, G., Russo, C. J., He, S., Hill, C. H., Skehel, J. M., Scheres, S. H. W., Patel, K. J., Rappsilber, J., Robinson, C. V. and Passmore, L. A., 2019, Structure of the Fanconi anemia monoubiquitin ligase complex, *Nature*, 575: 234-243.

Sharif, J., Muto, M., Takebayashi, S-I., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M. and Koseki, H., 2007, The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA, *Nature*, 450: 908-913.

Sims, A. E., Spiteri, E., Sims III, R. J., Arita, A. G., Lach F. P., Landers, T., Wurm, M., Freund, M., Neveling, K., Hanenberg, H., Auerbach, A. D. and Huang, T. T., 2007, FANCI is a second monoubiquitinated member of the Fanconi anemia pathway, *Nature Structural & Molecular Biology*, 4(6): 564-567.

Singh, T. R., Bakker, S. T., Agarwal, S., Jansen, M., Grassman, E., Godthelp, B. C., Ali, A. M., Du, C. H., Rooimans, M. A., Fan, Q., Wahengbam, K., Steltenpool, J., Andreassen, P. R., Williams, D. A., Joenje, H., de Winter, J. P. and Meetei, A. R., 2009, Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterise Fanconi anemia complementation group M, *Blood*, 114: 174-180.

Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald III, E. R., Hurov, K. E., Luo, J., Ballif, B. A., Gygi, S. P., Hofmann, K. and D'Andrea, A. D., 2007, Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair, *Cell*, 129: 289-301.

Smogorzewska, A., Desetty, R., Saito, T. T., Schlabach, M., Lach, F. P., Sowa, M. E., Clark, A. B., Kunkel, T. A., Harper, J. W., Colaiácovo, M. P. and Elledge, S. J., 2010, A genetic screen identifies

FAN1, a Fanconi anemia-associated nuclease necessary for DNA Interstrand crosslink repair, *Molecular Cell*, 39: 36-47.

Socha, A., Yang, D., Bulsiewicz, A., Yaprianto, K., Kupculak, M., Liang, C-C., Hadjicharalambous, A. Wu, R., Gygi, S. P. and Cohn, M. A., 2020, WRNIP1 is recruited to DNA Interstrand crosslinks and promotes repair, *Cell Rep.,* 32(1): 107850.

Swuec, P., Renault, L., Borg, A., Shah, F., Murphy, V. J., van Twest, S., Snijders, A. P., Deans, A. J. and Costa, A., 2017, The FA core complex contains a homo-dimeric catalytic module for the symmetric mono-ubiquitination of FANCI-FANCD2, *Cell Reports*, 18: 611-623.

Takeuchi, K., Baskaran, K. and Arthanari, H., 2019, Structure determination using solution NMR: is it worth the effort?, *Journal of Magnetic Resonance*, 306: 195-201.

Tan, W., van Twest, S., Leis, A., Bythell-Douglas, R., Murphy, V. J., Sharp, M., Parker, M. W., Crismani, W. M. and Deans, A. J., 2020a, Monoubiquitination by the human Fanconi anemia core complex clamps FANCI:FANCD2 on DNA in filamentous arrays, *eLife*, 9: e54128.

Tan, W., van Twest, S., Murphy, V. J. and Deans, A. J., 2020b, ATR-mediated FANCI phosphorylation regulates both ubiquitination and deubiquitination of FANCD2, *Front. Cell Dev. Biol.*, 8(2): 1-11.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. and Ludtke, S. J., 2007, EMAN2: an extensible image processing suite for electron microscopy, *J. Struct. Biol.*, 157:38-46.

Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. and Ranson, N. A., 2016, An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology, *Methods*, 100: 3-15.

Tian, H., Gao, Z., Li, H., Z, B., Wang, G., Zhang, Q., Pei, D. and Zheng, J., 2014, DNA damage response – a double edged sword in cancer prevention and cancer therapy, *Cancer Letters*, 358: 8-16.

Tian, Y., Paramasivam, M., Ghosal, G., Chen, D., Shen, X., Huang, Y., Akhter, S., Legerski, R., Chen, J., Seldman, M. M., Qin, J. and Li, L., 2015, UHRF1 contributes to DNA damage repair as a lesion recognition factor and nuclease scaffold, *Cell Rep.*, 10(12): 1957-1966.

Tien, A. L., SenBanerjee, S., Kulkarni, A., Mudbhary, R., Goudreau, B., Ganesan, S., Sadler, K. C. and Ukomadu, C., 2011, UHRF1 depletion causes a G2/M arrest, activation of DNA damage response and apoptosis, *Biochem. J.*, 435(1): 175-185.

van Twest, S., Murphy, V. J., Hodson, C., Tan, W., Swuec, P., O'Rourke, J. J., Heierhorst, J., Crismani, W. and Deans, A. J., 2017, Mechanism of ubiquitination and deubiquitination in the Fanconi anemia pathway, *Molecular Cell*, 65: 247-259.

Vaughan, R. M., Dickson, B. M., Cornett, E. M., Harrison, J. S., Kuhlman, B. and Rothbart, S. B., 2018, Comparative biochemical analysis of UHRF proteins reveals molecular mechanisms that uncouple UHRF2 from DNA methylation maintenance, *Nucleic Acids Research*, 46(9): 4405-4416.

Velimezi, G., Robinson-Garcia, L., Muñoz-Martínez, F., Wiegant, W. W., da Silva, J. F., Owusu, M., Moder, M., Wiedner, M., Rosenthal, S. B., Fisch, K. M., Moffat, J., Menche, J., van Attikum, H., Jackson, S. P. and Loizou, J. I., 2018, Map of synthetic rescue interactions for the Fanconi anemia DNA repair pathway identifies USP48, *Nature Communications*, 9: 2280.

Walden, H. and Deans, A. J., 2014, The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder, *Annu. Rev. Biophys.*, 43: 257-278.

Walker, J. R., Avvakumov, G. V., Xue, S., Li, Y., Bountra, C., Weigelt, J., Arrowsmith, C. H., Edwards, A. M., Bochkarev, A. and Dhe-Paganon, S., 2009, Structural Genomics Consortium (SGC), Crystal structure of the ring domain of the E3 ubiquitin-protein ligase UHRF1, doi: 10.2210/pdb3FL2/pdb.

Walker, J. R., Wybenga-Groot, L., Doherty, R. S., Finerty Jr., P. J., Newman, E., Mackenzie, F. M., Weigelt, J., Sundstrom, M., Arrowsmith, C., Edwards, A., Bochkarev, A. and Dhe-Paganon, S., 2005, Structural Genomics Consortium (SGC), Ubiquitin-like domain of human nuclear zinc finger protein NP95, doi: 10.2210/pdb2FAZ/pdb.

Wang, H-W. and Wang, J-W., 2016, How cryo-electron microscopy and X-ray crystallography complement each other, *Protein Science*, 26: 32-39.

Wang, R., Persky, N., Yoo, B., Ouerfelli, O., Smogorzewska, A., Elledge, S. J. and Pavletich, N. P., 2014, Mechanism of DNA Interstrand crosslink processing by repair nuclease FAN1, *Science*, 346(6213): 1127-1130.

Wang, R., Wang, S., Dhar, A., Peralta, C. and Pavletich, N. P., 2020, DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex, *Nature*, 580: 278-285.

Wang, S., Wang, R., Peralta, C., Yaseen, A. and Pavletich, N. P., 2019, Structure of the Fanconi anemia core-UBE2T complex poised to ubiquitinate bound FANCI-FANCD2, *bioRxiv*, doi: 10.1101/854158.

Wang, W., Xu, J., Chong, J. and Wang, D., 2018, Structural basis of DNA lesion recognition for eukaryotic transcription-coupled nucleotide excision repair, *DNA Repair (Amst)*, 71: 43-55.

Wang, Y. and Dasso, M., 2009, SUMOylation and deSUMOylation at a glance, *Journal of Cell Science*, 122: 4249-4252.

Woolfson, M. M., 2018, The development of structural X-ray crystallography, *Phys. Scr.*, 93: 032501.

Wu, J., Lu, L-Y. and Yu, X., 2010, The role of BRCA1 in DNA damage response, *Protein Cell*, 1(2): 117-123.

Wu, M. and Lander, G. C., 2020, How low can we go? Structural determination of small biological complexes using single-particle cryo-EM, *Current Opinion in Structural Biology*, 64: 9-16.

Wu, Q., Liang, S., Ochi, T., Chirgadze, D. Y., Huiskonen, J. T. and Blundell, T. L., 2019, Understanding the structure and role of DNA-PK in NHEJ: how X-ray diffraction and cryo-EM contribute in complementary ways, *Progress in Biophysics and Molecular Biology*, 147: 26-32.

Wüthrich, K., 1990, Protein structure determination in solution by NMR spectroscopy, *The J. of Biol. Chem.*, 265(36): 22059-22062.

Wyatt, H. D. M., Laister, R. C., Martin, S. R., Arrowsmith, C. H. and West, S. C., 2017, The SMX DNA repair tri-nuclease, *Mol. Cell*, 65(5): 848-860.

Yin, L., Liu, Y., Peng, Y., Peng, Y., Yu, X., Gao, Y., Yuan, B., Zhu, Q., Cao, T., He L., Gong, Z., Sun, L., Fan, X. and Li, X., 2018, PARP inhibitor veliparib and HDAC inhibitor SAHA synergistically co-target the UHRF1/BRCA1 DNA damage repair complex in prostate cancer cells, *Journal of Experimental & Clinical Cancer Research*, 37: 153-167.

Yu, X., Jacobs, S. A., West, S. C., Ogawa, T. and Egelman, E. H., 2001, Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA, *PNAS*, 98(15): 8419-8424.

Yuan, F., Hokayem, J. E., Zhou, W. and Zhang, Y., 2009, FANCI protein binds to DNA and interacts with FANCD2 to recognize branched structures, *J. Biol. Chem.*, 284(36): 24443-24452.

Zhang, H., Liu, H., Chen, Y., Yang, X., Wang, P., Liu, T., Deng, M., Qin, B., Correia, C., Lee, S., Kim, J., Sparks, M., Nair, A. A., Evans, D. L., Kalari, K. R., Zhang, P., Wang, L., You, Z., Kaufmann, S. H., Lou, Z. and Pei, H., 2016, A cell cycle-dependent BRCA1-UHRF1 cascade regulates DNA double-strand break repair pathway choice, *Nature Communications*, 7: 10201.

Zhang, H., Chen, J., Wu, M., Wu, H., Arokiaraj, A. W., Wang, C., Zhang, W., Tao, Y., Huen, M. S. Y. and Zang, J., 2013, Structural basis for role of ring finger protein RNF168 RING domain, *Cell Cycle*, 12(2): 312-321.

Zhang, J., Dewar, J. M., Budzowska, M., Motnenko, A., Cohn, M. A. and Walter, J. C., 2015, DNA Interstrand cross-link repair requires replication-fork convergence, *Nature Structural & Molecular Biology*, 22(3): 242-247.

Zhao, X., Wei, C., Li, J., Xing, P., Li, J., Zheng, S. and Chen, X., 2017, Cell cycle-dependent control of homologous recombination, *Acta Biochimica et Biophysica Sinica*, 49(8): 655-668.