Studies of The DNA Repair Proteins FANCD2/FANCI and DNA Polymerase η



Chih-Chao Liang

Department of Biochemistry

and

Oriel College

Trinity 2012

Abstract

Our genome is continuously under various sources of genotoxic stresses. Thus, human has developed several DNA repair pathways to deal with different types of DNA damage. This thesis contains two studies, one focusing on the Fanconi Anemia pathway, which is known to coordinate the DNA interstrand crosslink repair, and the other one on the regulation of translesion synthesis, which is responsible for replication at DNA lesion sites, e.g. UV induced pyrimidine dimers.

Fanconi Anemia (FA) is a genetic disorder caused by mutations in any of 15 FA genes, where patients usually have developmental defects and predisposition to cancers in addition to anemia, and all the patients and patient derived cell lines are hypersensitive to DNA crosslinking agents such as Mitomycin C (MMC) and Cisplatin. Of the 15 FA proteins, FANCD2 and FANCI are key players essential for the functionality of the pathway. Upon genotoxic stress, both proteins are phosphorylated, ubiquitinated, and quickly recruited to sites of DNA damage together with other DNA repair proteins. In this part of our studies, we present the structure of the full-length human FANCD2/FANCI heterodimer, and our attempts at generating ubiquitinated FANCD2/FANCI heterodimer.

Polymerase η is one of the Y-family polymerases that are involved in translesion synthesis. It is known that ubiquitination and phosphorylation are together regulating the recruitment of polymerase η. In this part of our studies, we identified several novel polymerase η interacting proteins, as well as two novel phosphorylation sites, which could participate in the process of polymerase η recruitment upon DNA damage.

Acknowledgments

First of all, I would like to thank my mom and sister who have been great supporters both financially and mentally during my time of study. Without their supports, I wouldn't even be able to study in Oxford.

I would like to thank you Martin for being a fantastic supervisor, which build me into an adequate scientist, and on top of that also being a mentor and a friend guiding me when there were some confusions. It has been an educational and productive journey for me in the past two years.

Thanks for Dr. Catherine Vénien-Bryan, Dr. William Nicholson and Dr. Antonio de Riso on the collaboration of the FANCD2/FANCI structure project.

Thanks Prof. Alan Lehmann for initiating the collaboration of the polymerase η project, and also providing precious suggestions and crucial reagents. It has been a challenging task, but meanwhile rewarding to work on.

Thanks Dr. Nickolas Lakin for all the valuable suggestions during lab meetings as a co-supervisor.

Thank you all the current and past members of Cohn's group and other members in the same lab area, especially Jamie Zhan, Jean Tien, Maarten van del Velden, Dr. Duen-Wei Hsu, Dr. Hong-Yu Wang, Dr. Keiko Watanabe and Dr. Regina Teo. It would be a lot harder and less fun, if I did not have your accompanies and supports.

Thanks all my friends for trying to help me having a better-balanced life, in another words keeping me away from lab, especially Yen-Chen Pan and Yasuhiro Mita.

Thanks Tao-Hsin Chang for proofreading the thesis and providing constructive suggests.

Last but not least, thanks all the people who have helped me during my master study. I couldn't have done the study without your help.

Abbreviations

ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related protein
AUC	analytical ultracentrifugation
BER	base excision repair
CD	catalytic domain
CPD	cyclobutane pyrimidine dimer
cryoEM	transmission electron cryomicroscopy
DDR	DNA damage response
EM	transmission electron microscopy
FA	Fanconi Anemia
FT	flow-through fraction
HA tag	Human Influenza Hemaglutinin tag
HR	homologous recombination
HU	hydroxyurea
Lys (K)	lysine
MMC	Mitomycin C
MMR	mismatch repair
MRN	Mre11-Rad50-NBS1 complex
MS/MS	Tandem mass spectrometry
NER	nucleotide excision repair
NHEJ	non-homologous end joining
NLS	nuclear localization sequence
P-Polη	phosphorylated polymerase η
PCNA	Proliferating Cell Nuclear Antigen

PIKK	Phosphatidylinositol 3-kinase-related kinases
PIP	PCNA-binding motif
Polη	polymerase η
RPA	Replication protein A
SEC	size-exclusion chromatography
Ser (S)	serine
Thr (T)	threonine
TLS	translesion synthesis
TopBP1	DNA topoisomerase 2-binding protein 1
TCA	Trichloroacetic acid
UAF1	Usp1-associated factor 1
ub	ubiquitin
UBD	ubiquitin-binding domain
UBZ	ubiquitin-binding zinc-finger motif
USP1	ubiquitin specific peptidase 1
UV	ultraviolet radiation
XPV	xeroderma pigmentosum variant

Index

Abstract ..... **i**

Acknowledgments ..... **ii**

Index ..... **v**

Chapter 1: Introduction **1**

1.1 DNA damage response **1**

1.2 Phosphorylation and ubiquitination in DDR **2**

1.3 Homologous recombination in DNA double-strand break repair **3**

1.4 Translesion synthesis **4**

1.5 The Fanconi Anemia DNA Repair Pathway **5**

Chapter 2: Structural study of full-length human FANCD2/FANCI complex utilizing electron microscopy (EM) **8**

2.1 Introduction **8**

2.2 Specific aims **9**

2.3 Results **9**

2.3.1 Expression and purification of full-length human FANCD2/FANCI complex 9

2.3.2 Determination of the stoichiometry of purified FANCD2/FANCI complex 11

2.3.3 Structure of FANCD2/FANCI heterodimer using EM negative staining 13

2.3.4 Structure of FANCD2/FANCI heterodimer and DNA by cryoEM 15

2.3.5 Comparison of the full-length human FANCD2/FANCI EM structure and the mouse FANCD2/FANCI crystal structure 17

2.3.6 *in vitro* ubiquitination assay of FANCD2/FANCI complex and monomers 17

2.4 Discussion **21**

Chapter 3: Study of the polymerase η complex after hydroxyl urea treatment 23

3.1 Introduction 23

3.2 Specific aims 24

3.3 Results 24

3.3.1 Generation of phosphorylated polymerase η *in vitro* 24

3.3.2 Purification of the polymerase η complex formed *in vitro* 28

3.3.3 Purification of the *in vivo* polymerase η containing complex 30

3.3.4 MS/MS analysis of the purified wildtype and S601A mutant polymerase η complex 32

3.3.5 Identification of novel phosphorylation sites on polymerase η after HU treatment 35

3.4 Discussion 36

Chapter 4: Discussion 39

4.1 Structural study of full-length human FANCD2/FANCI complex utilizing electron microscopy (EM) 39

4.2 Studies of the polymerase η complex after hydroxyl urea treatment 42

Chapter 5: Materials and methods 44

Chapter 6: References 49

Chapter 7: Appendix 57

7.1 Amino acid sequence alignment of Human and mouse FANCD2 57

7.2 Amino acid sequence alignment of Human and mouse FANCI 59

7.3 Introduction of cryoEM 61

Chapter 1: Introduction

1.1 DNA damage response

In daily life, we are constantly encountering various sources of chemical reagents and radiation, which themselves or their metabolic by-products could pose different levels of threats on our genome integrity. In order to cope with these genotoxic stresses within cells, human has developed several different DNA repair pathways to deal with various types of DNA damage, namely homologous recombination (HR) and non-homologous end joining (NHEJ) for DNA double strand break repair, nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR) for single-strand damage, FA pathway for DNA interstrand crosslink repair and translesion synthesis (TLS) for a damage tolerance process during DNA replication. Some of the DNA repair pathways will be discussed in greater details later. All these pathways form a relatively complex DNA damage response (DDR) network in order to allow cells to repair different DNA lesions with the correspondent DNA repair pathway at an appropriate stage of the cell cycle. Either up or down regulation of any of the DNA repair pathways can diminish the integrity of our genome. For instance, increased level of HR activity results in reduced level of chromosome crossover during M phase and thus decreased diversity of our genome, and hyperactive NHEJ causes loss of important genetic information due to its error-prone repair nature. Post-translational modifications, such as phosphorylation, ubiquitination, sumoylation and ADP-ribosylation are main regulators of DDR. Among all the post-translational modifications, phosphorylation and ubiquitination

are some of the most studied modifications, which will be discussed in greater detail later. Failure of initiating the DNA damage response could cause the accumulation of DNA lesions within cells and thus result in genome instability. An unstable genome has been known to be the cause of various diseases, including cancers. Therefore, our interests lie in understanding the mechanisms underlying the DDR, which would allow us to target disease caused by mutations in the DDR during clinical treatments.

1.2 Phosphorylation and ubiquitination in DDR

In the DDR network, ATM (ataxia telangiectasia mutated) and ATR (ATM and Rad3-related) are the two crucial phosphoinositide-3-kinase-like protein kinases (PIKK), which initiate the signal transduction cascade. ATM is mainly thought to initiate the cell cycle checkpoint event through phosphorylation of Chk2 and p53(Canman et al., 1998; Lee and Paull, 2007), which stops the cell cycle progression after activation. This gives cells time to repair the DNA lesion instead of carrying the mutation-causing lesion to the next stage of the cell cycle (Bakkenist and Kastan, 2003). ATR is thought to initiate the DNA repair signal transduction through the recognition of RPA coated single-stranded DNA via its regulating partner ATRIP (ATR interacting protein). The initial signal is then magnified by its autophosphorylation on Thr1989 (Liu et al., 2011) and the interaction with TopBP1 (DNA topoisomerase 2-binding protein 1) (Kumagai et al., 2006). A large-scale proteomic study of ATM and ATR substrates has revealed the complexity of phosphorylation events in the DDR (Matsuoka et al., 2007).

Ubiquitination events are usually coupled with phosphorylation in the DDR (Harper and Elledge, 2007). The most studied phosphorylation and ubiquitination event in

the DDR is perhaps the phosphorylation of histone variant H2AX on serine 139, and the following ubiquitination events. Once DNA damage is detected within cells, ATM and ATR phosphorylate this Histone variant (Fernandez-Capetillo et al., 2004; Ward and Chen, 2001), causing the recruitment of MDC1 to the site of damage through its BRCT domains directly binding to the phosphorylated Ser139 of γH2AX. The phosphorylated MDC1 is then further recruit RNF8, a Ring domain containing E3 Ligase, to the site of DNA damage. RNF8 initiates the ubiquitination of γH2AX(Mailand et al., 2007), and the ubiquitination is further amplified by RNF168, another Ring domain containing E3 ligase. These phosphorylation and ubiquitination events together create repair factories, which are also known as DNA damage foci, where other DNA repair proteins are then recruited to through their phosphorylation specific interacting domains and ubiquitin binding domains (UBD), and carry out the DNA repair process. These phosphorylations and ubiquitinations are crucial in the DDR because any disruption will lead to genome instability. Similar phosphorylation and ubiquitination cross talk events also exist in many other DNA repair pathways, which will be discussed in greater detail later.

1.3 Homologous recombination in DNA double-strand break repair

Homologous recombination is one of the pathways responsible for the DNA double-strand break repair. More specifically it provides an error free repair process during the S and G2 phase of cell cycle, in which sister chromatids are present providing a template for faithful genomic information exchange. The Mre11-Rad50-Nbs1 (MRN) complex acts as a sensor of DNA double strand breaks, and once a double strand

break is recognised, it activates ATM to trigger the signal transduction cascade. Moreover, the MRN complex itself is also a substrate of ATM. The MRN complex first initiates 5'-3' DNA resection, and then further resection is carried out by CtIP, BLM, Exo1 and DNA2 (Nimonkar et al., 2011; Sartori et al., 2007). This leaves 3' overhangs at both sides of the break, and the ssDNA is quickly bound and stabilized by RPA. The next event is the replacement of RPA by recombinase Rad51, which is mediated by BRCA2. The Rad51 coated ssDNA then invades into the sister chromatid to utilize the genetic information for error-free repair (Jensen et al., 2010; Liu et al., 2010).

1.4 Translesion synthesis

Translesion synthesis is the major DNA damage tolerance process responsible for overcoming the replication fork blocking damages, which are not removed by BER and NER (Bienko et al., 2005). Most of the process is carried out by the Y-family DNA polymerases, which includes Polη, Polκ, Polι, and Rev1 (Ohmori et al., 2001). These polymerases have a spacious active site comparing to the replicative DNA polymerases, and this feature allow them to accommodate lesions such as cyclobutane pyrimidine dimers (CPD), which would block the high-fidelity replicative polymerases(Prakash et al., 2005), but this also means that it is an error-prone process. Hence, the process of switching between high fidelity polymerase and the TLS polymerases need to be tightly regulated. The post-translational modification of Proliferating Cell Nuclear Antigen (PCNA) plays a crucial role in regulating this process. Upon replication fork stalling, PCNA is monoubiquitinated on Lysine 164 by E2 ubiquitin-conjugating enzyme, Rad6, and E3 ubiquitin ligase, Rad18 (Hoege et al., 2002). This modification provides a higher affinity for TLS polymerases, which

contain an ubiquitin-binding domain to allow the replacement of replicative polymerases (Bienko et al., 2005; Kannouche et al., 2004). PCNA is also known to be polyubiquitinated by Lysine 63 chains, which further requires MMS2, UBC13 and RAD5 in addition to the Rad6/Rad18 complex. This modification is believed to stimulate the template switching process, which allows the use of genetic information on the sister chromatids. (Hoege et al., 2002).

1.5 The Fanconi Anemia DNA Repair Pathway

Fanconi Anemia is an autosomal recessive syndrome or X-linked inherited disorder where patients usually have congenital abnormalities, e.g. shorter stature, and are prone to develop bone marrow failure and various types of cancer due to genome instability. Another common feature of these patients suffering from Fanconi Anemia is their hypersensitivity to DNA crosslinking agents like Mitomycin C (MMC) and Cisplatin, which are widely used in chemotherapy for treating leukaemia and other types of cancer (D'Andrea, 2010). So far genetic studies have revealed at least 15 genes to be causative for causing Fanconi Anemia. These genes comprise the current Fanconi Anemia DNA repair pathway (FA pathway), and are also known as FA genes (Table 1). These 15 FA genes can be grouped into three categories. Group I contains the effector proteins, FANCD2 and FANCI, whose activation determines the functionality of the FA pathway. Group II comprises the FA core complex, FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCM and FANCL, who are responsible for the ubiquitination of FANCD2 and FANCI. Group III includes the downstream signal proteins FANCD1 (BRCA2), FANCN (PALB2), FANCJ, RAD51c (Somyajit et al., 2010) and FANCP (SLX4) (Crossan et al., 2011; Kim et al., 2011; Stoepker et al., 2011), who

Table 1. List of FA proteins

Group	Gene	Size of Protein Product (kDa)	Activities
I	FANCD2	155	Ubiquitinated after DNA damage
	FANCI	140	
II	FANCA	163	Subunit of core complex
	FANCB	95	Subunit of core complex
	FANCC	63	Subunit of core complex
	FANCE	60	Subunit of core complex
	FANCF	42	Subunit of core complex
	FANCG	68	Subunit of core complex
	FANCL	43	Subunit of core complex; E3 ligase
	FANCM	250	Subunit of core complex; DNA translocase
III	FANCD1	380	BRCA2; regulate RAD51 filament formation
	FANCJ	140	BRIP1/BACH1; helicase
	FANCN	140	PALB2; an essential partner of BRCA2 stability and nuclear localization
	FANCO	42	RAD51c; essential for HR pathway
	FANCP	200	SLX4; a coordinator of structure-specific endonucleases

Table 1 List of Fanconi Anemia proteins. Group I, effector proteins; Group II, FA core complex; Group III, downstream signal proteins.

act downstream of ubiquitinated FANCD2 and FANCI (ub-FANCD2, ub-FANCI) and facilitate the DNA repair.

Upon DNA damage, FANCD2 and FANCI are phosphorylated at various sites by the ATR, ATM and CHK1 kinases (Ho et al., 2006; Ishiai et al., 2008; Taniguchi et al., 2002b; Zhi et al., 2009) (Fig. 1). These phosphorylations are suggested to be involved in promoting the monoubiquitination of FANCD2 and FANCI and the activation of intra-S-phase checkpoint response. After the phosphorylations, FANCD2 and FANCI are then both monoubiquitinated by the FA core complex in the nucleus, and recruited to the site of damage on chromatin (Fig. 1). Monoubiquitination of FANCD2 and FANCI are interdependent on each other (Smogorzewska et al., 2007). Any loss of the components of the FA core complex or mutation in the monoubiquitination site of FANCD2 (K561) will hypersensitize the cells to crosslinking agents, which

suggests that this monoubiquitination process serves as a molecular switch for the FA pathway. Within group III, there are BRCA2, PALB2, SLX4, RAD50c and BRIP/BACH1. BRCA2 has been shown to interact with FANCD2 (Taniguchi et al., 2002a; Wang et al., 2004), and together with PALB2 and BRCA1 link the FA pathway to homologous recombination (Wang, 2007).

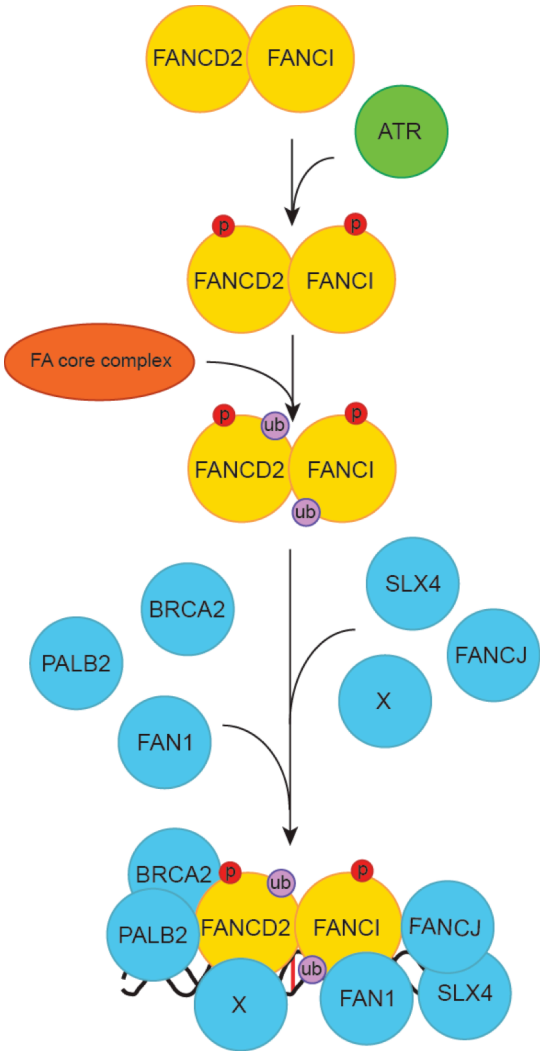


Figure 1 Schematic of the Fanconi Anemia DNA repair pathway.
After DNA damage is detected, ATR will phosphorylate the ID complex at various sites, and the phosphorylated ID complex is then monoubiquitinated by the FA core complex. The monoubiquitinated ID complex is recruited to the sites of DNA damage, and facilitate the repair of interstrand crosslinked DNA together with other proteins, e.g. BRCA2, PALB2, FAN1, SLX4, FANCJ and maybe some unknown proteins.

Chapter 2: Structural study of full-length human FANCD2/FANCI complex utilizing electron microscopy (EM)

2.1 Introduction

Monoubiquitination of FANCD2 and FANCI play a critical role in regulating the activity of the FA pathway. Loss of the monoubiquitination by either disrupting the E3 ligase FANCL-containing FA core complex or mutations on FANCD2 will abolish the DNA interstrand crosslink repair pathway (D'Andrea, 2010). FANCD2 and FANCI have been suggested to be paralogues because both of them have no recognizable domains, both are monoubiquitinated, and have around 20% similarity in amino acid content (Smogorzewska et al., 2007). However, the detailed mechanism of how the FANCD2 and FANCI are functioning in the FA pathway remains elusive. It has been shown previously FANCD2 and FANCI form a heterodimer disregarding the monoubiquitination status (Smogorzewska et al., 2007), and a recent report further suggested that the phosphorylation of FANCD2 and FANCI trigger the dissociation of the FANCD2 and FANCI heterodimer (Sareen et al., 2012). Yet, it is unclear how FANCD2 and FANCI are recruited to the site of damage, and how they coordinate the DNA interstrand crosslink repair in their monoubiquitinated forms (ub-FANCD2 and ub-FANCI). In this project, we carried out structural studies utilizing transmission electronic microscopy (TEM) hoping to answer some of the questions from a structural point of view.

2.2 Specific aims

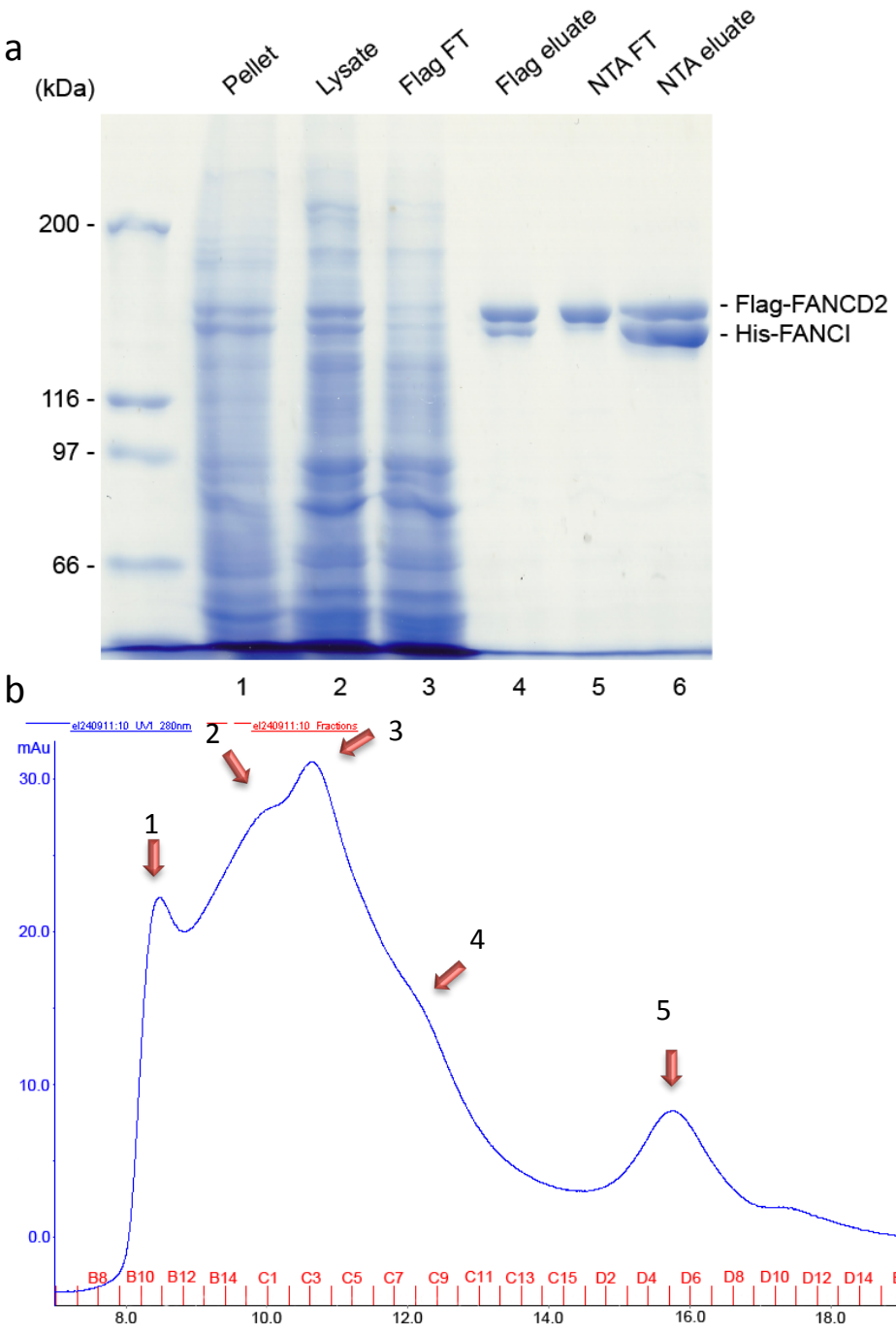
1. Obtain the recombinant human full-length FANCD2/FANCI complex.
2. Study the structure of FANCD2/FANCI heterodimer by EM and cryoEM.
3. Reconstitute an *in vitro* ubiquitination reaction of FANCD2/FANCI and study the structure of ub-FANCD2/FANCI heterodimer.

2.3 Results

2.3.1 Expression and purification of full-length human FANCD2/FANCI complex

We co-expressed full-length human FANCD2 (Flag-HA tagged) and FANCI (6xHis tagged) in Sf9 cells, and under the assumption that they form a heterodimer *in vivo*, we carried out a sequential affinity purification using anti-Flag agarose beads and Ni-NTA agarose beads. Through this two-step purification system, we obtained the purified FANCD2/FANCI complex (Fig. 2a). We analysed our purified FANCD2/FANCI complex by electron microscopy (EM), but our first attempt to study the structure of the complex was not successful due to the heterogeneity of our purified sample (data not shown). We therefore introduced an extra step of size-exclusion chromatography (SEC) in our purification scheme in order to separate different populations of the complex. After passing our crude purified sample through the size-exclusion column, Superdex 200, the chromatogram showed that there were at least 5 different populations in our crude purified sample (Fig. 2b). The 1^{st} peak is the void volume of the column, and analysis of the 2^{nd}, 3^{rd} and 4^{th} peak showed that the ratio between FANCD2 and FANCI varied. We concluded that the 2^{nd} and 3^{rd} peaks

contain the heterodimer, and there are mixed populations of FANCD2 and FANCI monomers in the 4th peak (Fig. 2c).



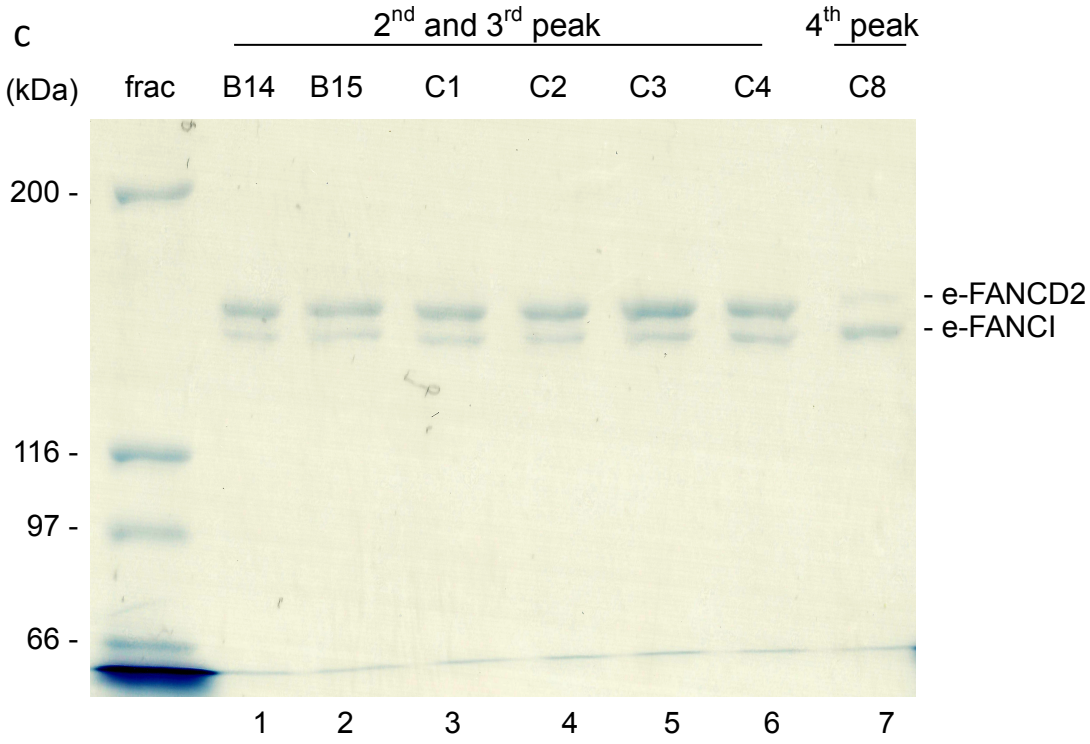


Figure 2 Purification of the FANCD2/FANCI heterodimer.

(a) The intermediates and final eluate were analysed by SDS-PAGE and Coomassie staining. (FT stands for flow-through fraction)
(b) Chromatogram of the FANCD2/FANCI heterodimer on Superdex 200. Arrows and numbers indicate the peaks.
(c) Peak fractions B14-C4 and C8 were analysed by SDS-PAGE and Coomassie staining. The FANCD2 and FANCI ratio in fractions B14-C4 and C8 is different which indicates that we separate heterodimer from monomers through the Superdex 200.

2.3.2 Determination of the stoichiometry of purified FANCD2/FANCI complex

We noticed that sometimes our crude purified sample and fractions from size-exclusion chromatography displayed various ratios between FANCD2 and FANCI according to the Coomassie staining (Fig. 2c), which suggested that perhaps the complex we purified had a 2:1 rather than 1:1 ratio between FANCD2 and FANCI. To determine the stoichiometry of our predicted FANCD2/FANCI heterodimer, we

incubated the sample with anti-HA antibody, which recognises the epitope tag of FANCD2, followed by size-exclusion chromatography. The change in elution profile during gel filtration should then allow us to determine the amount of antibody bound to each complex by calculating the theoretical molecular weight difference between sample with and without antibody bound, and we would then determine the stoichiometry of the purified complex. Unfortunately, the molecular weight of FANCD2/FANCI complex together with anti-HA antibody is close to the limit that Superdex 200 can resolve, thus the calculated theoretical molecular weight was not sufficiently accurate. We therefore decided to analyse our purified FANCD2/FANCI complex using analytical ultracentrifugation (AUC). The preliminary data fitting shows that there are several populations within the crude purified FANCD2/FANCI complex, and that the dominant complex population is the one with a molecular weight of 286kDa (Fig. 3). The theoretical molecular weight of the FANCD2/FANCI heterodimer is 315kDa. Thus, the most abundant species in our sample with a molecular weight above that of a monomer is likely a heterodimer rather than a trimer. We are aware that the calculated molecular weight does not fit perfectly into the theoretical molecular weight, yet the crude purified complex we used for the AUC analysis contained some other larger complex populations, which could affect the accuracy of molecular fitting.

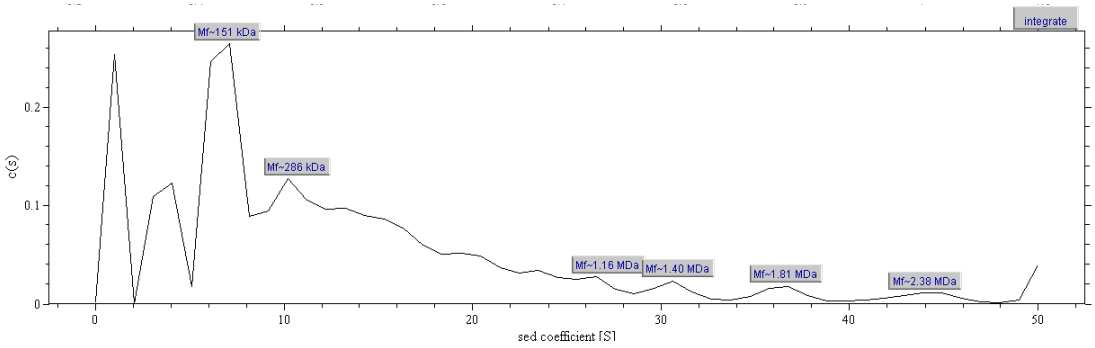


Figure 3 Preliminary fitting of the transmission profile of AUC showed that the majority of our purified complex is heterodimer.

2.3.3 Structure of FANCD2/FANCI heterodimer using EM negative staining

With the refined purification scheme, we were now able to purify FANCD2/FANCI heterodimer to homogeneity, and we proceeded to analyse our FANCD2/FANCI heterodimer using EM negative staining. The micrographs demonstrated that our purified FANCD2/FANCI heterodimer is homogeneous (Fig. 4a), and the classification analysis showed no major difference among 10 different classes (Fig. 4c). In order to obtain a more comprehensive view of the protein complex for the 3D structure reconstruction, the sample grids were tilted in order to be able to look at each particle from different angles (Fig. 4b). After reconstruction using all the particles in the 10 classes, the first structure of full-length human FANCD2/FANCI heterodimer was generated. Our full-length human FANCD2/FANCI heterodimer has some distinctive characteristics. The overall structure is quite compact, and it has a pocket-like main body. There is a thumb-like structure at one side of the main body (1 in Fig. 4d), and there is a fork-like tower protruding out from the other side of the main body (2 in Fig.4d). There are also two spikes at the back of the main body (3 in Fig. 4d).

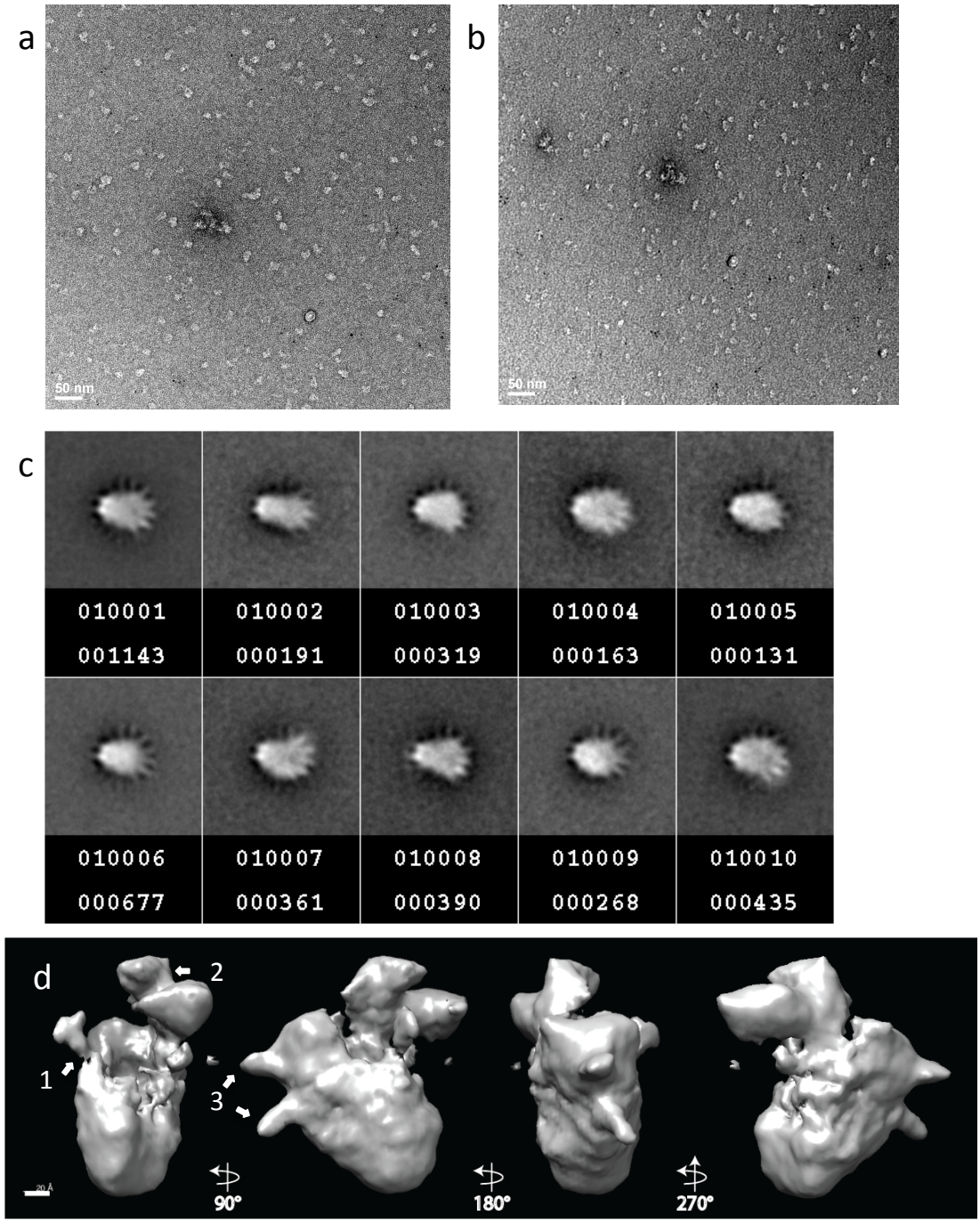


Figure 4 Structure of FANCD2/FANCI heterodimer using EM negative staining.

(a) Micrograph of negative stained FANCD2/FANCI protein particles. (Scale bar: 50nM)

(b) Micrograph of the tilted grid at the same field as in (a). (Scale bar: 50nM)

(c) Classification of the 4078 FANCD2/FANCI heterodimer particles. The upper number of each figure represents the numbering of the classes, and the lower number represents the number of the particles in each class.

(d) The FANCD2/FANCI structure at 26 Å reconstituted from negative staining. (Scale bar: 20 Å) 1 indicates the thumb-like structure, 2 indicates the fork-like tower, and 3 indicates the two spikes at the back.

2.3.4 Structure of FANCD2/FANCI heterodimer and DNA by cryoEM

Sometimes the heavy metal used in negative staining to enhance the contrast between protein particles and background can generate artificial density in addition to the real density originating from the protein. To confirm the structure we obtained from the negative staining, we therefore proceeded to analyse FANCD2/FANCI complex utilizing cryoEM. The overall structure of heterodimeric FANCD2/FANCI from cryoEM still remained the same as the one obtained from negative staining, yet the general noise was greatly reduced, and we correlated the volume with the exact molecular weight of the FANCD2/FANCI heterodimer (Fig. 5a). In the cryoEM structure, the definition of the main body and the other two parts became much clearer due to the lower noise signal. The thumb-like structure (1 in Fig. 5a) remained attached to the main body, while the fork-like tower (2 in Fig.5a) was more standing out from the main body. It has been known that FANCD2/FANCI heterodimer possesses DNA binding activity (Longerich et al., 2009; Yuan et al., 2009). We have tested the affinity of our purified FANCD2/FANCI heterodimer towards different DNA substrates, and found that our unmonoubiquitinated FANCD2/FANCI heterodimer does possess some DNA binding activity, but that it has no preferential binding towards any kind of DNA substrate including interstrand cross-linked DNA substrate (data not shown). With the preliminary FANCD2/FANCI structure by cryoEM, we next studied the structure of FANCD2/FANCI in complex with a DNA structure containing 40 base pairs of double stranded DNA plus 40 nucleotides of single stranded 3’ overhang DNA. The structure of the FANCD2/FANCI/DNA complex (Fig. 5b) revealed a conformational change upon the DNA binding as comparing with FANCD2/FANCI structure (Fig.5a). The characteristic

thumb-like and fork-like tower structures (1 in Fig. 5b) still exist, but the base of the fork-like tower moved slightly away from the back. This difference is more obvious when we overlay the two structures (2 in Fig. 5c). There is also an extra bridge-like density appeared in the middle of the dimer and a small spike at the back of the heterodimer (1 in Fig. 5c), which could be contributed by the additional DNA.

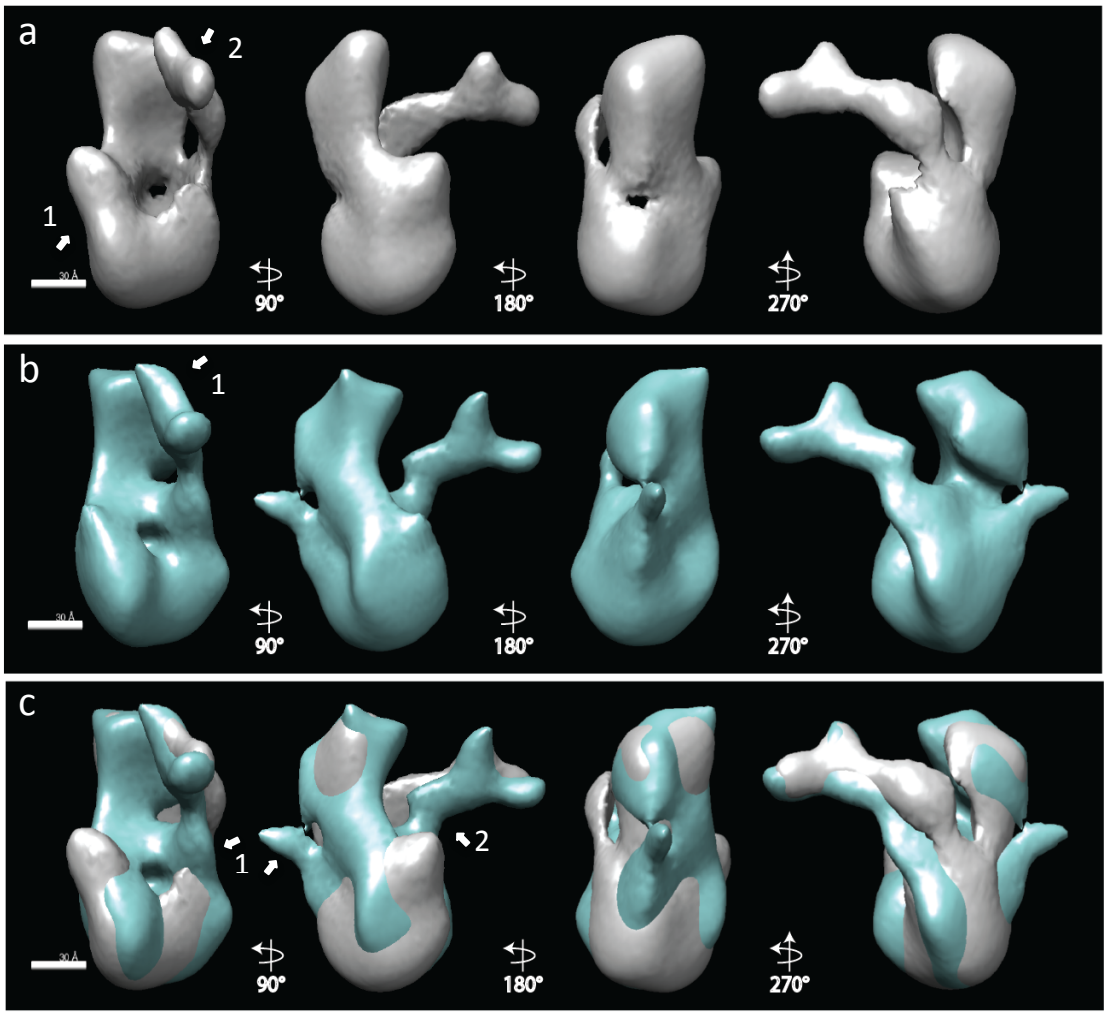


Figure 5 CryoEM structures of the FANCD2/FANCI heterodimer and the heterodimer/DNA complex (Scale bar: 30 Å). 1 indicates the fork-like tower, and 2 indicates the thumb-like structure.

(a) The 44 Å FANCD2/FANCI heterodimer structure reconstituted from cryoEM. 1 indicates the thumb-like structure, and 2 indicates the fork-like tower.

(b) The 49 Å FANCD2/FANCI/DNA complex structure reconstituted from cryoEM. 1 indicates the fork-like structure.

(c) Overlay of the FANCD2/FANCI heterodimer and FANCD2/FANCI/DNA complex. 1 indicates the extra bridge density and spike, which may be formed by DNA, and 2 indicates the base shifting.

2.3.5 Comparison of the full-length human FANCD2/FANCI EM structure and the mouse FANCD2/FANCI crystal structure

During our study, the crystal structure of the mouse FANCD2/FANCI heterodimer was published (Joo et al., 2011). The human and mouse FANCD2/FANCI heterodimer have 75% identities in the amino acid sequences. Furthermore, the mouse FANCD2/FANCI complex contained several deletions across both FANCD2 and FANCI, while we are studying the full-length human FANCD2/FANCI complex (Fig. 6a). We could find some common features and some differences (Fig. 6b). Overall, our human FANCD2/FANCI structure is more compact, while the mouse FANCD2/FANCI crystal structure appears relatively flat. Both of our full-length human FANCD2/FANCI and the mouse FANCD2/FANCI complex have a similar main body. The mouse FANCD2/FANCI complex has two thumb-like structures protruding from each side, which the authors described as the bell of a saxophone, yet we only have one thumb-like structure protruding from one side of the main body, and a much taller fork-like tower replacing the other thumb-like structure.

2.3.6 *in vitro* ubiquitination assay of FANCD2/FANCI complex and monomers

As described in the introduction, both FANCD2 and FANCI are monoubiquitinated after DNA damage, and this monoubiquitination event is important for the activity of the FA pathway. It has recently been shown that the phosphorylation-mimicking FANCI mutant failed to interact with FANCD2 suggesting the phosphorylation status of FANCI may interfere the FANCD2/FANCI complex formation. However, there was

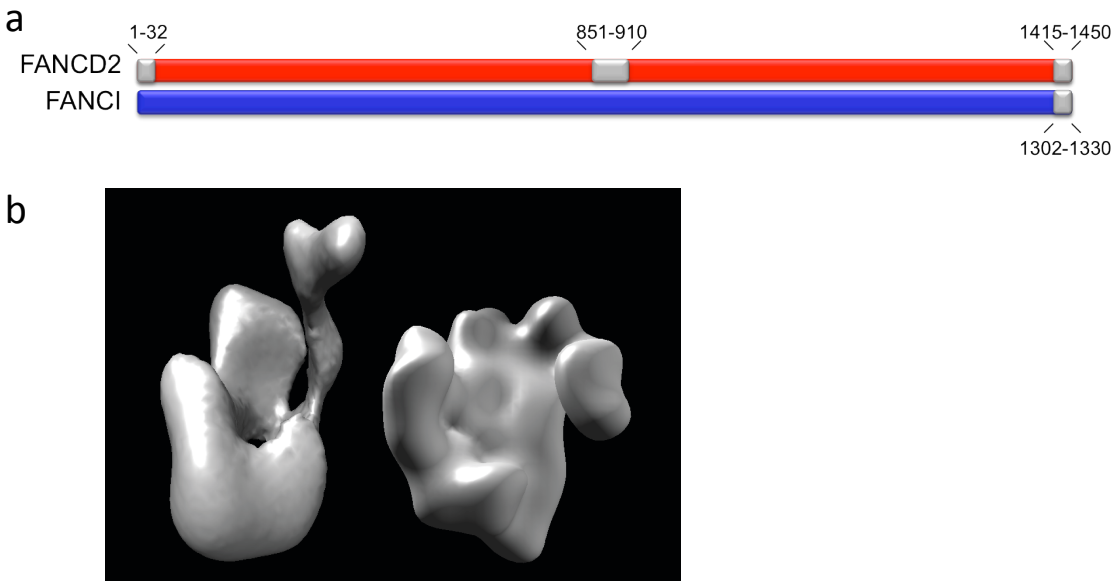


Figure 6 Comparison of the full-length human FANCD2/FANCI EM structure and the mouse FANCD2/FANCI crystal structure

(a) Schematic of the fragment deletion applied in the mouse FANCD2/FANCI crystal structure study.

(b) Comparison of the human FANCD2/FANCI EM structure (left) and the mouse FANCD2/FANCI crystal structure (right). The crystal structure was converted into the space-filling model in order to compare to the EM structure.

still a minor population of ub-FANCD2 interacting with FANCI being detected (Sareen et al., 2012). We have obtained structural insight into the native full-length human FANCD2/FANCI complex. However, it would also be interesting to gain insight into the structure of the ubiquitinated FANCD2/FANCI complex, which could possibly help us to understand how ubiquitination contribute to the activation of the FA pathway.

We first attempted to reconstitute an *in vitro* ubiquitination assay using the E1 protein (UBA1), E2 protein (UBE2T), E3 ligase (FANCL) and ubiquitin. We expressed and purified the recombinant Flag-HA tagged UBA1, UBE2T and FANCL from Sf9 insect cells, and the 6xHis tagged ubiquitin from *E. coli* (Fig. 7a). This work was initiated by two previous lab members, Jean Tian and Maarten van der Velden, but unfortunately we could not detect any ubiquitination under our preliminary setting

(data not shown). It has been suggested by the Kurumizaka group in early 2012 that DNA can stimulate the *in vitro* ubiquitination reaction of FANCD2(Sato et al., 2012b). We carried out the *in vitro* ubiquitination reaction with our purified recombinant proteins under the same reaction condition, but we still could not detect any ubiquitination of our FANCD2/FANCI complex or monomers (Fig. 7b). However, we also could not detect any ubiquitin charging to UBA1 and Ube2t or FANCL auto-ubiquitination with non-reducing SDS-PAGE (Fig. 7c), which is suggesting that the reason why the *in vitro* ubiquitination reaction failed could be in part because the recombinant proteins we generated were not sufficiently active.

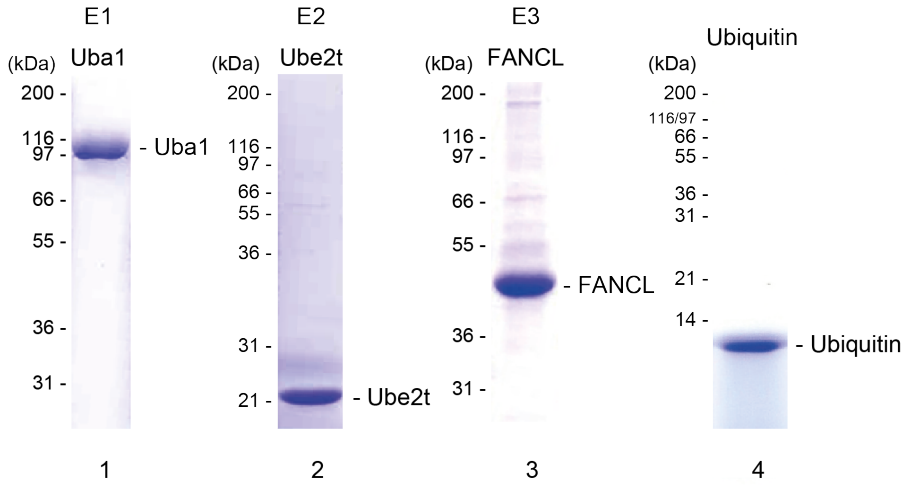
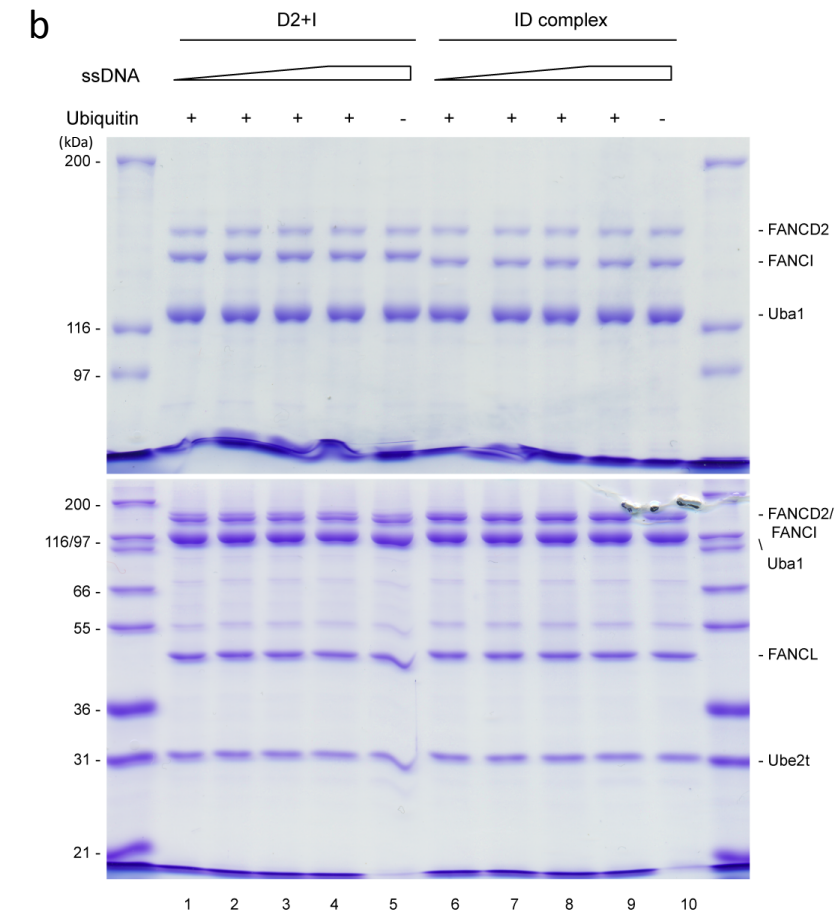


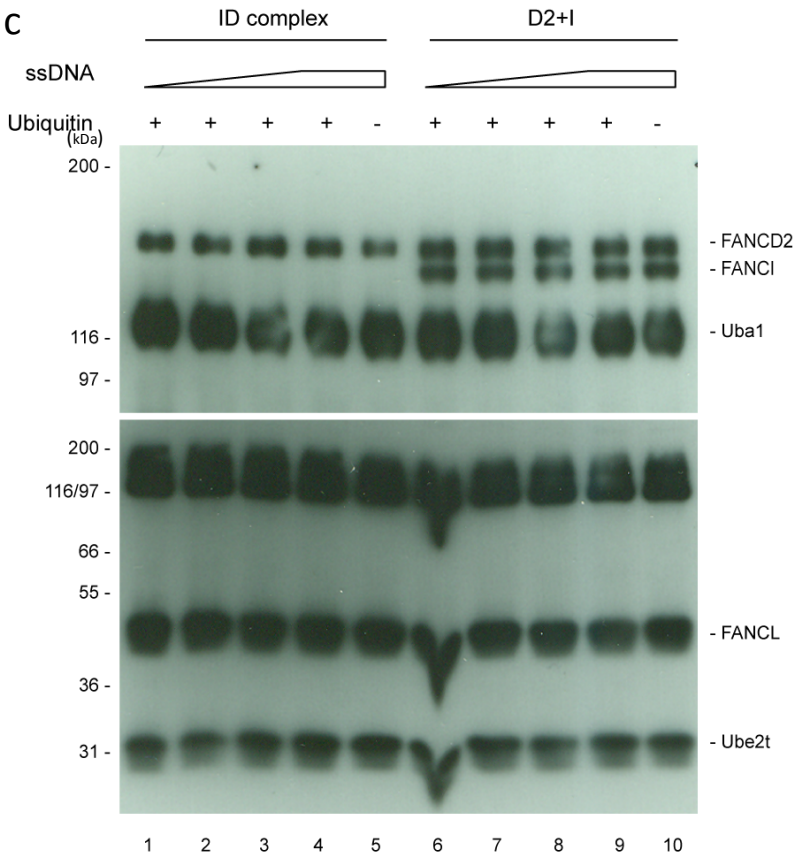
Figure 7 *In vitro* ubiquitination of FANCD2 and FANCI

(a) SDS-PAGE and Coomassie staining analysis of purified E1 (Uba1), E2 (Ube2t), E3 (FANCL) and ubiquitin.

(b) SDS-PAGE and Coomassie staining analysis of *in vitro* ubiquitination assay. The first 5 lanes are the assay with the FANCD2 and FANCI monomers, and the following 5 lanes are the assay with FANCD2/FANCI complex.

(c) Non-reducing SDS-PAGE with anti-HA antibody. The first 5 lanes are the assay with FANCD2/FANCI complex, and the following 5 lanes are the assay with the FANCD2 and FANCI monomers.


b
D2+I
ID complex
ssDNA
Ubiquitin
+ + + + - + + + + -
(kDa)
200 -
116 -
97 -
- FANCD2
- FANCI
- Uba1
200 -
116/97 -
66 -
55 -
36 -
31 -
21 -
- FANCD2/
FANCI
Uba1
- FANCL
- Ube2t
1 2 3 4 5 6 7 8 9 10


c
ID complex
D2+I
ssDNA
Ubiquitin
+ + + + - + + + + -
(kDa)
200 -
116 -
97 -
- FANCD2
- FANCI
- Uba1
200 -
116/97 -
66 -
55 -
36 -
31 -
- FANCL
- Ube2t
1 2 3 4 5 6 7 8 9 10

2.4 Discussion

Here, we show the EM structure of full-length human FANCD2/FANCI heterodimer and the structure of human FANCD2/FANCI in complex with DNA, which both have several distinct features. Interestingly, the position of DNA we observed in our heterodimer/DNA complex fits into the model suggested by Pavletich Group, which lies in the inner grove of the FANCD2/FANCI complex (Joo et al., 2011). From the comparison between our human FANCD2/FANCI and the mouse FANCD2/FANCI heterodimer (Joo et al., 2011), we speculate that the fork-like tower structure could be formed by the C-terminal of FANCD2, as the thumb-like structure of our FANCD2/FANCI complex has more similarity to the FANCI in the mouse FANCD2/FANCI complex. To confirm this, we are currently generating a C-terminal deletion of FANCD2 and we are going to study whether this truncated protein still forms a complex with FANCI. If a complex is formed, we will subject it to EM studying its structure. If the fork-like tower structure disappears in the new structure, we could therefore draw a conclusion on the distribution of FANCD2 and FANCI in our structure. Also, we are going to test whether the C-terminal deletion affects the DNA binding activity of the complex, as we observed a conformational change of the fork-like tower after our FANCD2/FANCI complex was bound to DNA. The C-terminal of both FANCD2 and FANCI have been suggested to be important to the functionalities of the proteins in different aspects (Pace et al., 2010; Sato et al., 2012a; Smogorzewska et al., 2007). We would like to explore more about the significance of the fork-like tower structure on the activity of FANCD2/FANCI complex.

Our attempts to generate ub-FANCD2/FANCI complex have not been successful so far. We are currently investigating whether the present of EDTA in the buffer our recombinant proteins are in could be the cause of the lack of activities. More control experiments are needed to demonstrate that our recombinant proteins are functionally active after we rule out the effect from the composition of our buffers. Ubiquitination of FANCD2 and FANCI *in vivo* requires the whole FA core complex in human (Moldovan and D'Andrea, 2009), yet we only included the responsible E3 ligase, FANCL, in our *in vitro* reaction. Therefore, we do not exclude the possibilities that more factors are needed for a robust *in vitro* reaction.

In addition to optimizing our *in vitro* ubiquitination reaction condition, we are also exploring other possible methods to obtain sufficient amount of ub-FANCD2/ub-FANCI heterodimer for structural studies. It would be exciting and informative to be able to study the structure and DNA binding affinity of the ub-FANCD2/ub-FANCI complex, which is the active form in the FA DNA repair pathway.

Chapter 3: Study of the polymerase η complex after hydroxyl urea treatment

3.1 Introduction

Among the Y-family polymerases, polymerase η is the one that has been best characterized. Mutation of this gene causes the variant form of the skin cancer-prone syndrome Xeroderma Pigmentosum (XPV) (Masutani et al., 1999). Like other TLS polymerases, it has a conserved catalytic domain (CD), an ubiquitin-binding domain (UBZ), a nuclear localization sequence (NLS) and a PCNA interacting motif (PIP box) (Fig. 8). As mentioned earlier, PCNA is monoubiquitinated at Lys 164 by the Rad6/Rad18 complex upon replication fork arrest, which creates a interaction surface for TLS polymerases through their UBZ domain. It has also been shown that monoubiquitination of polymerase η at the NLS inhibits the interaction between PCNA and polymerase η, which serves as a negative regulation to prevent polymerase η loading onto the active replicative site in the absent of damage(Bienko et al., 2010). ATR mediated phosphorylation event on polymerase η also takes part in the regulation of the recruitment of polymerase η. Alan Lehmann's group previously demonstrated that polymerase η is phosphorylated at the Ser 601 residue upon UV damage, and that this phosphorylation event is important for the cellular survival towards UV (Gohler et al., 2011). Furthermore, this phosphorylation requires the physical interaction between polymerase η and Rad18, but both PIP box and monoubiquitination of PCNA are not required (Gohler et al., 2011). However, there are some connections missing in the current model. Do other factors recognise the

phosphorylated polymerase η first, and then recruit polymerase η to monoubiquitinated PCNA? In this study, we carried out the purifications of the polymerase η containing complex in the hope of identifying novel polymerase η interacting proteins, which could be the crucial recruitment factor.



Figure 8 Schematic of polymerase η. CD, catalytic domain; UBZ, ubiquitin-binding zinc-finger motif; N, nuclear localization sequence; P, PIP box PCNA-binding motif.

3.2 Specific aims

1. Purify the polymerase η-containing complex formed *in vitro.*
2. Purify the *in vivo* polymerase η-containing complex.
3. Identify the novel polymerase η interacting proteins by MS/MS.
4. Identify novel phosphorylation sites on polymerase η by MS/MS.
5. Study the roles of novel polymerase η interacting proteins and phosphorylation events in the regulation of polymerase η.

3.3 Results

3.3.1 Generation of phosphorylated polymerase η *in vitro*

Our strategy here was to generate the phosphorylated polymerase η via an *in vitro* kinase assay, and use the phosphorylated protein as bait to purify the interacting proteins from the nuclear extract (Fig. 9).

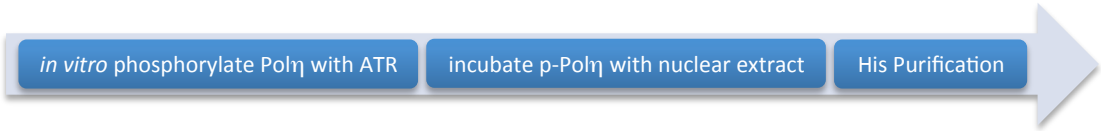


Figure 9 Schematic of the purification strategy of the polymerase η complex formed *in vitro*

We expressed and purified 6xHis tagged polymerase η from Sf9 insect cells and ATR from HeLa S3 cells stably expressing Flag-HA tagged ATR (Fig. 10a). We first carried out a simple kinase assay to confirm the activity of our recombinant ATR. The phosphorylation event was monitored using an antibody specifically recognizing polymerase η phosphorylated on the Serine 601. The polymerase η purified from Sf9 cells had some endogenous level of phosphorylation on serine 601 (Fig. 10b lower panel, lane1). However after adding the recombinant ATR, the phosphorylation level increased dramatically (Fig. 10b lower panel, lane 2). It has been reported that TopBP1 can activate the ATR signalling pathway both *in vivo* and *in vitro* (Kumagai et al., 2006), so we also wanted to test if TopBP1 could further stimulate the kinase activity in the *in vitro* assay. We expressed and purified human TopBP1 from Sf9 insect cells, and tested the kinetics of the phosphorylation on our recombinant polymerase η. Surprisingly, we could not detect any further increase of the phosphorylation upon addition of the recombinant TopBP1 compared to the reaction with only recombinant ATR and polymerase η (Fig. 10b lower panel, lanes 3 and 4). In fact, we even observed a decrease of phosphorylation level. We rationalise what we observed might be because of the lack of activity of our recombinant TopBP1. Or there is an inhibitory effect resulting from the buffer in which TopBP1 is present, since we added in a significant volume of TopBP1 into the reactions, which changed the reaction condition. However, it has been reported that the autophosphorylation of ATR could stimulate the kinase activities, and this process is prior to the stimulation from the interaction between ATR and TopBP1 (Liu et al., 2011). To further optimize the reaction condition aiming to obtain the maximum amount of phosphorylated polymerase η after the kinase assay, we titrated in

increasing amount of ATR in the assay (Fig. 10c). With an increasing amount of ATR added to the reaction (compare lanes 2-6), we could detect a increase of the phosphorylation level, and longer reaction time allowed the polymerase η to be phosphorylated even more (compare lanes 2-6 with 7-11).

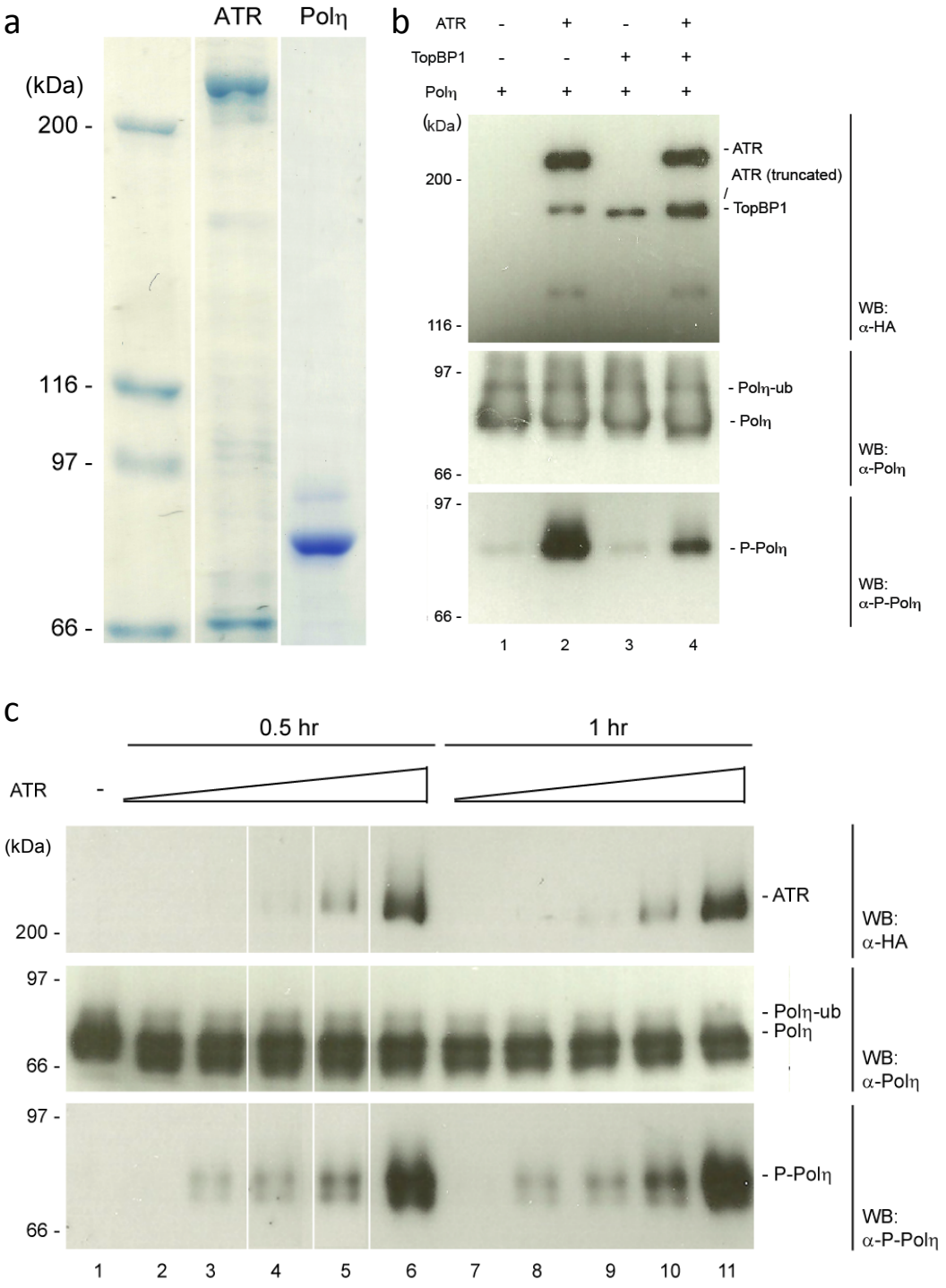


Figure 10 *In vitro* kinase assay

(a) Coomassie stain SDS-PAGE of purified ATR and Polymerase η.

(b) Western blot of the *in vitro* kinase assay determining the activity of recombinant ATR and TopBP1.

(c) Western blot of the *in vitro* kinase assay titration the amount of ATR. Anti-HA, anti-pol η and anti-phospho-pol η antibodies were used in the blots which is labelled at the right hand side.

3.3.2 Purification of the polymerase η complex formed *in vitro*

With the established *in vitro* kinase assay, we were able to obtain sufficient amounts of phosphorylated polymerase η. We proceeded to purify the phospho-S601 polymerase η together with interacting proteins following the schematic shown in Fig. 9. We included a mock purification with only the Ni-NTA beads and the nuclear extract, and a reaction with or without recombinant ATR. Unfortunately, the background binding from the Ni-NTA beads was so strong that we could not distinguish the proteins specifically interacting with polymerase η or phospho-S601 polymerase η from other non-specific interacting proteins (data not shown).

The dilemma here was that if we wanted to lower the non-specific binding from the Ni-NTA beads, we would need to increase the stringency during purification, yet increasing the stringency would also risk dissociating specifically interacting proteins. In order to overcome this issue, we introduced formaldehyde crosslinking to preserve protein-protein interaction, and performed the purification under denaturing condition as shown in the schematic (Fig. 11a). First the phosphorylation status of polymerase η was confirmed by western blot (Fig. 11b). We then analysed our purification products on a silver stained NuPage gradient gel, which showed that the new purification scheme allowed us to purify the polymerase η or phospho-S601 polymerase η interacting complexes with lower levels of non-specific contaminations (Fig. 11c lane 1). When comparing the three lanes, we could observe some common bands present in all three samples suggesting that they are purification contaminants resulting from direct Ni-NTA interaction. Some bands only existed in lanes 2 and 3, which indicated that they are potential polymerase η interacting proteins regardless of the phosphorylation status. The differences between lanes 2

and 3 are relatively minor suggesting that only few proteins interact with phosphorylated polymerase η specifically under these conditions. We decided not to pursue this strategy further because later we failed to purify the *in vitro* complex with similar quality after scaling up the purification.

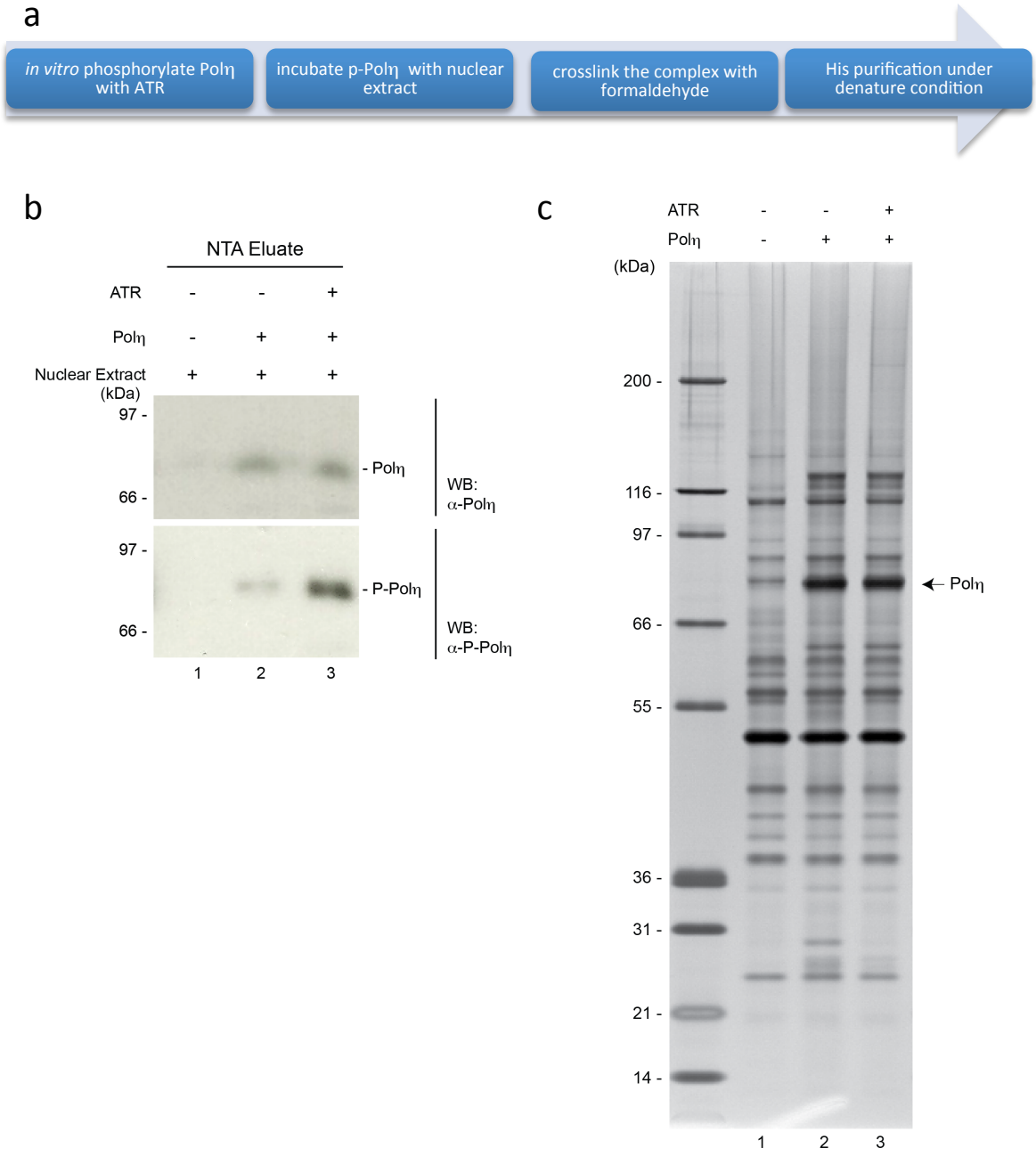


Figure 11 Purification of polymerase η complex formed *in vitro*

(a) Schematic of purification strategy with modifications

(b) Western blot of the purified polymerase η complex formed *in vitro* confirming the phosphorylation of polymerase η in the reaction with ATR.

(c) Silver stain of the purified polymerase η complex formed *in vitro*.

3.3.3 Purification of the *in vivo* polymerase η containing complex

In parallel with the *in vitro* polymerase η complex purification, we also tried to purify the polymerase η complex from HeLa cells treated with hydroxyurea. We generated two HeLa S3 cell-lines stably expressing either Flag-HA tagged human full-length wildtype polymerase η, or the Serine 601 point mutation to alanine (S601A) polymerase η (Fig. 12a). The expression level of exogenous polymerase η is higher than the endogenous protein, yet in this experiment, lane 1 was slightly underloaded, which makes it hard to compare the exact fold difference between the endogenous and exogenous protein. Our main goal is to obtain the active polymerase η complex, which is primarily located in the nucleus, so we first fractionated the cells into cytoplasmic and nuclear fractions. The nuclear fraction was used for sequential anti-Flag and anti-HA purification. The purified polymerase η complexes were then analysed by Western blot to confirm the phosphorylation status of the purified polymerase η (Fig. 12b), and silver stained SDS-PAGE for visualizing the composition of the purified polymerase η containing complexes (Fig. 12c).

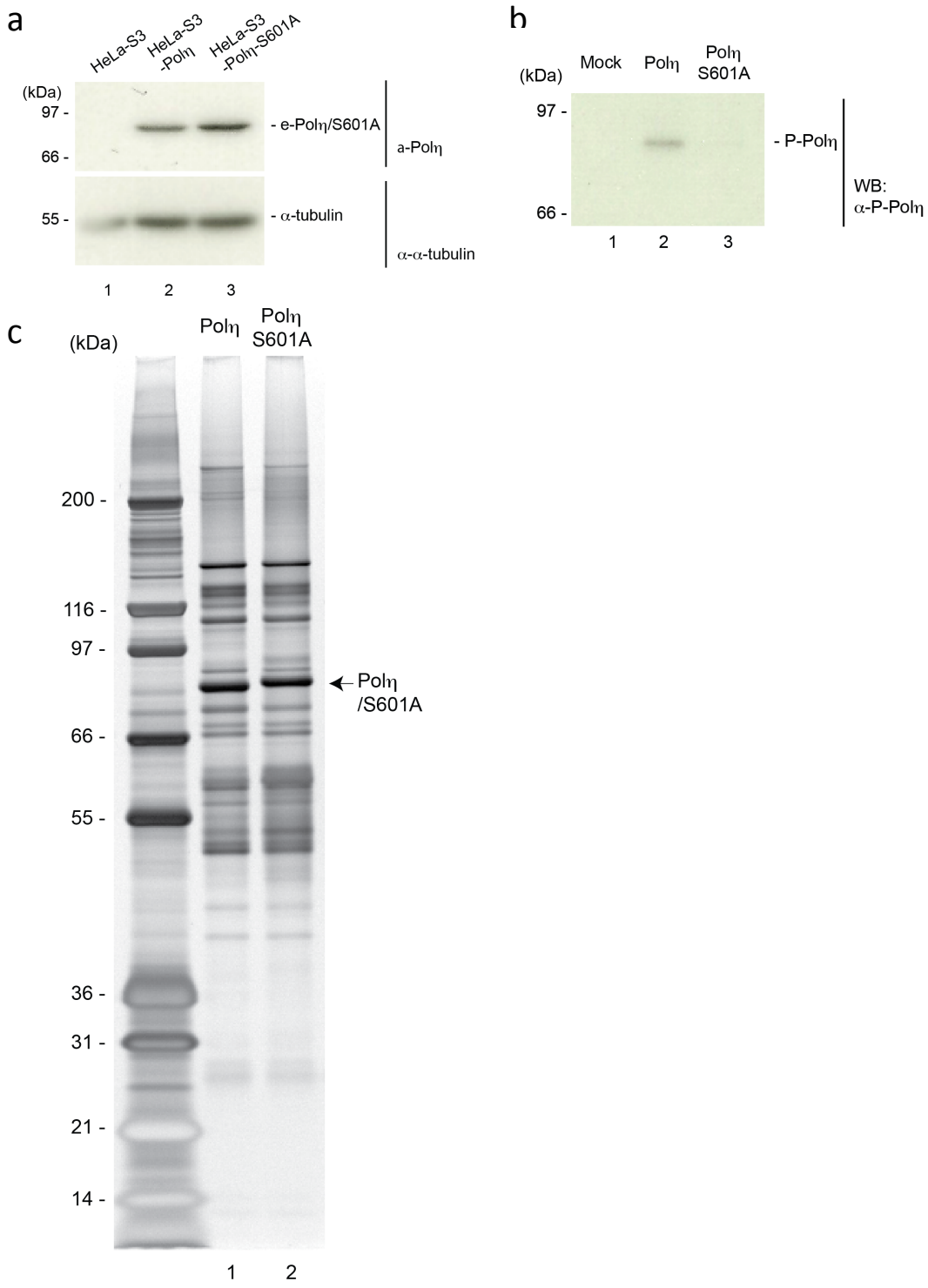


Figure 12 Purification of the *in vivo* polymerase η interacting complex.

(a) Western blot of the HeLa S3, Pol η and Pol η S601A mutant confirming the expression of the exogenous polymerase η.

(b) Western blot of the purified polymerase η complex confirming the phosphorylation of polymerase η on serine 601 residue.

(c) Silver stain of the purified polymerase η complex from wild type and mutant.

3.3.4 MS/MS analysis of the purified wildtype and S601A mutant polymerase η complex

After preliminary analysis of the silver stained gel, we precipitated our purified *in vivo* polymerase η-containing complex with TCA. The protein complexes were then trypsinized and subjected to MS/MS to identify the interacting proteins. Once we obtained the raw data from the mass spectrometry, we first calculated the fold difference of the total peptides between wildtype and S601A polymerase η, and normalized the total peptide number of all peptides against this factor in order to compare the interacting proteins between the wildtype and S601A mutant (Table 2). Rad18, which has be previously reported to interact with polymerase η, existed in both wildtype and S601A polymerase η-containing complex with similar amount of peptides (Yuasa et al., 2006). This gave us confidence that after the correlation, the list of the potential interacting proteins we obtained is still reflecting the real composition of the polymerase η-containing complex. Most of the proteins exist in both purified samples and have similar number of total peptides after normalization with some exceptions, such as AKAP8 and COPS4, for which there are 3- and 2-fold more peptides in the wildtype than S601A mutant, respectively, and UBR5, for which there are 6-fold more peptides in the S601A mutant.

Table 2 List of polymerase η interacting protein candidates.

WT		Mut	
Protein	Total peptides	Total peptides (correlated)	Protein
POLH_IPI:IPI00008011.1	145	145	POLH_IPI:IPI00008011.1
SF3B1_IPI:IPI00026089.3	67	27	SF3B1_IPI:IPI00026089.3
TUBB2C_IPI:IPI00007752.1	51	23	TUBB2C_IPI:IPI00007752.1
EIF4B_IPI:IPI00012079.3	47	21	EIF4B_IPI:IPI00012079.3
SF3B2_IPI:IPI00221106.5	41	22	SF3B2_IPI:IPI00221106.5
SF3B3_IPI:IPI00300371.5	36	16	SF3B3_IPI:IPI00300371.5
RAD18_IPI:IPI00024579.1	29	27	RAD18_IPI:IPI00024579.1
PRMT5_IPI:IPI00064328.3	29	17	PRMT5_IPI:IPI00064328.3
HSPA8_IPI:IPI00003865.1	28	24	HSPA8_IPI:IPI00003865.1
FRYL_IPI:IPI00159652.1	25	32	FRYL_IPI:IPI00159652.1
HSPA5_IPI:IPI00003362.2	23	16	HSPA5_IPI:IPI00003362.2
SR140_IPI:IPI00143753.6	22	12	SR140_IPI:IPI00143753.6
TUBB_IPI:IPI00011654.2	20	11	TUBB_IPI:IPI00011654.2
DHX15_IPI:IPI00396435.3	19	16	DHX15_IPI:IPI00396435.3
TUBA4A_IPI:IPI00007750.1	19	21	TUBA4A_IPI:IPI00007750.1
RBM10_IPI:IPI00375731.1	18	6	RBM10_IPI:IPI00375731.1
SF3A1_IPI:IPI00017451.1	17	12	SF3A1_IPI:IPI00017451.1
HSPB1_IPI:IPI00025512.2	16	9	HSPB1_IPI:IPI00025512.2
CCBL2_IPI:IPI00061178.1	14	6	CCBL2_IPI:IPI00061178.1
SF3A3_IPI:IPI00029764.1	13	6	SF3A3_IPI:IPI00029764.1
SF3B4_IPI:IPI00017339.1	9	5	SF3B4_IPI:IPI00017339.1
RBM17_IPI:IPI00176706.1	9	5	RBM17_IPI:IPI00176706.1
HNRNPM_IPI:IPI00171903.2	8	10	HNRNPM_IPI:IPI00171903.2
RUVBL1_IPI:IPI00021187.4	8	7	RUVBL1_IPI:IPI00021187.4
SF3B3_IPI:IPI00179138.3	8	6	SF3B3_IPI:IPI00179138.3
TUBA1C_IPI:IPI00166768.3	8	6	TUBA1C_IPI:IPI00166768.3
SF3B14_IPI:IPI00032827.1	8	2	SF3B14_IPI:IPI00032827.1
MATR3_IPI:IPI00017297.1	7	8	MATR3_IPI:IPI00017297.1
HNRNPK_IPI:IPI00216049.1	7	5	HNRNPK_IPI:IPI00216049.1
SLC4A1AP_IPI:IPI00306749.8	7	9	SLC4A1AP_IPI:IPI00306749.8
TUBB1_IPI:IPI00006510.1	7	8	TUBB1_IPI:IPI00006510.1
PRDX1_IPI:IPI00000874.1	7	5	PRDX1_IPI:IPI00000874.1
HSPA1A_IPI:IPI00304925.5	6	3	HSPA1A_IPI:IPI00304925.5
RUVBL2_IPI:IPI00009104.7	6	8	RUVBL2_IPI:IPI00009104.7
AKAP8_IPI:IPI00014474.1	6	2	AKAP8_IPI:IPI00014474.1
WDR77_IPI:IPI00012202.1	6	6	WDR77_IPI:IPI00012202.1
THRAP3_IPI:IPI00104050.3	5	2	THRAP3_IPI:IPI00104050.3
HNRNPF_IPI:IPI00003881.5	5	2	HNRNPF_IPI:IPI00003881.5
LOC652595_IPI:IPI00183920.2	4	2	LOC652595_IPI:IPI00183920.2
DNAJA2_IPI:IPI00032406.1	4	4	DNAJA2_IPI:IPI00032406.1
RIF1_IPI:IPI00293845.5	4	3	RIF1_IPI:IPI00293845.5
CHERP_IPI:IPI00333010.8	4	4	CHERP_IPI:IPI00333010.8
COPS4_IPI:IPI00171844.3	4	2	COPS4_IPI:IPI00171844.3

RCN1_IPI:IPI00015842.1	4	2	RCN1_IPI:IPI00015842.1
DNAJA1_IPI:IPI00012535.1	4	4	DNAJA1_IPI:IPI00012535.1
RBMXL2_IPI:IPI00004450.2	4	1	RBMXL2_IPI:IPI00004450.2
SF3A2_IPI:IPI00017341.3	3	3	SF3A2_IPI:IPI00017341.3
UBR5_IPI:IPI00026320.1	3	17	UBR5_IPI:IPI00026320.1
PSMD14_IPI:IPI00024821.1	3	1	PSMD14_IPI:IPI00024821.1
PSMC2_IPI:IPI00021435.3	3	2	PSMC2_IPI:IPI00021435.3
ACTA2_IPI:IPI00008603.1	3	3	ACTA2_IPI:IPI00008603.1
PSMD2_IPI:IPI00012268.3	3	5	PSMD2_IPI:IPI00012268.3
PSMD3_IPI:IPI00011603.2	3	3	PSMD3_IPI:IPI00011603.2
HNRNPH1_IPI:IPI00013881.6	3	1	HNRNPH1_IPI:IPI00013881.6
UBE2A_IPI:IPI00746451.2	3	2	UBE2A_IPI:IPI00746451.2
LDLR_IPI:IPI00000070.1	3	3	LDLR_IPI:IPI00000070.1
ACTB_IPI:IPI00021439.1	3	2	ACTB_IPI:IPI00021439.1
RIOK1_IPI:IPI00171336.3	3	1	RIOK1_IPI:IPI00171336.3
SNRPG_IPI:IPI00016572.1	3	0	SNRPG_IPI:IPI00016572.1
RBMX_IPI:IPI00304692.1	3	2	RBMX_IPI:IPI00304692.1
PSMC6_IPI:IPI00021926.4	2	2	PSMC6_IPI:IPI00021926.4
EEF1A1_IPI:IPI00025447.8	2	0	EEF1A1_IPI:IPI00025447.8
KHDRBS1_IPI:IPI00008575.3	2	1	KHDRBS1_IPI:IPI00008575.3
HSPA1L_IPI:IPI00301277.1	2	5	HSPA1L_IPI:IPI00301277.1
PSMC3_IPI:IPI00018398.4	2	2	PSMC3_IPI:IPI00018398.4
PSMD1_IPI:IPI00299608.3	2	2	PSMD1_IPI:IPI00299608.3
HNRNPR_IPI:IPI00012074.3	2	3	HNRNPR_IPI:IPI00012074.3
RPSAP12_IPI:IPI00398958.3	2	0	RPSAP12_IPI:IPI00398958.3
SNRPD3_IPI:IPI00017964.1	2	1	SNRPD3_IPI:IPI00017964.1
USP11_IPI:IPI00184533.1	2	7	USP11_IPI:IPI00184533.1
PSMD12_IPI:IPI00185374.4	2	1	PSMD12_IPI:IPI00185374.4
DDX5_IPI:IPI00017617.1	2	4	DDX5_IPI:IPI00017617.1
EEF1A2_IPI:IPI00014424.1	2	2	EEF1A2_IPI:IPI00014424.1
PHF5A_IPI:IPI00005511.1	2	1	PHF5A_IPI:IPI00005511.1
SNRPD2_IPI:IPI00017963.1	2		
SNRPB2_IPI:IPI00029267.1	2	1	SNRPB2_IPI:IPI00029267.1
GNB2L1_IPI:IPI00641950.4	2	1	GNB2L1_IPI:IPI00641950.4
SNRPF_IPI:IPI00220528.6	1	0	SNRPF_IPI:IPI00220528.6
AKAP8L_IPI:IPI00297455.5	1	2	AKAP8L_IPI:IPI00297455.5
CLNS1A_IPI:IPI00004795.1	1	2	CLNS1A_IPI:IPI00004795.1
UBE2A_IPI:IPI00375549.1	1	2	UBE2A_IPI:IPI00375549.1
HSPA7_IPI:IPI00011134.2	1	1	HSPA7_IPI:IPI00011134.2
E2F7_IPI:IPI00414604.4	1	2	E2F7_IPI:IPI00414604.4
ACTBL2_IPI:IPI00003269.1	1	1	ACTBL2_IPI:IPI00003269.1
LOC646057_IPI:IPI00887164.2	1	0	LOC646057_IPI:IPI00887164.2
PSMD8_IPI:IPI00010201.4	1	0	PSMD8_IPI:IPI00010201.4
HSPA9_IPI:IPI00007765.5	1	1	HSPA9_IPI:IPI00007765.5
RPS14_IPI:IPI00026271.5	1	0	RPS14_IPI:IPI00026271.5
TUBA4B_IPI:IPI00017454.4	1		
IGKV4-1_IPI:IPI00026197.7	1	0	IGKV4-1_IPI:IPI00026197.7

IPI:IPI00556589.3	1		
DHX8_IPI:IPI00031508.1	1	0	DHX8_IPI:IPI00031508.1
RBM39_IPI:IPI00163505.2	1	4	RBM39_IPI:IPI00163505.2
PSMD4_IPI:IPI00022694.3	1		
RPS18_IPI:IPI00013296.3	1	2	RPS18_IPI:IPI00013296.3
CALU_IPI:IPI00045396.1	1	0	CALU_IPI:IPI00045396.1
NPM1_IPI:IPI00220740.1	1	2	NPM1_IPI:IPI00220740.1
HNRNPCL1_IPI:IPI00027569.1	1	1	HNRNPCL1_IPI:IPI00027569.1
PRC1_IPI:IPI00022629.6	1		
IPI:IPI00176696.2	1		
BCLAF1_IPI:IPI00006079.1	1		

Table 2 The complete peptide list of the wildtype and S601A mutant polymerase η-containing complexes. The first and fourth column contains the gene name, e.g. POLH, and the EMBL-EBI number, e.g. IPI00008011.

3.3.5 Identification of novel phosphorylation sites on polymerase η after HU treatment

As mentioned in the introduction, phosphorylation events play an important role in the regulation of the DDR. We would like to explore if other phosphorylation events also take part in the regulation of polymerase η after replication fork arrest in addition to the S601 site. During the purification of the polymerase η containing complex, we also noticed that our purified wildtype polymerase η and S601A mutant themselves had different mobility on the NuPage gradient gel. We were curious whether the point mutation on Serine 601 residue could cause any further dominant modification to change the mobility of the whole polymerase η population, or the change of the residue from serine to alanine is affecting the overall protein charges resulting in changing the mobility on NuPage gradient gel. The Coomassie stained polymerase η bands were cut out, and subjected to MS/MS analysis. The MS/MS data shows no difference in the phosphorylation profile between the wildtype and S601A mutant, which suggests that the change of mobility is mainly caused by the

change of the residue, or that a differential phosphorylation site was not identified in the mass spectrometry analysis, potentially due to the nature of the correspondent peptides. However, we identified two novel phosphorylation sites on Serine 512 and Serine 687. Multiple species protein sequence alignment shows that S512 is conserved in higher order mammals, and S687 shows conservation in vertebrates (Fig. 13).

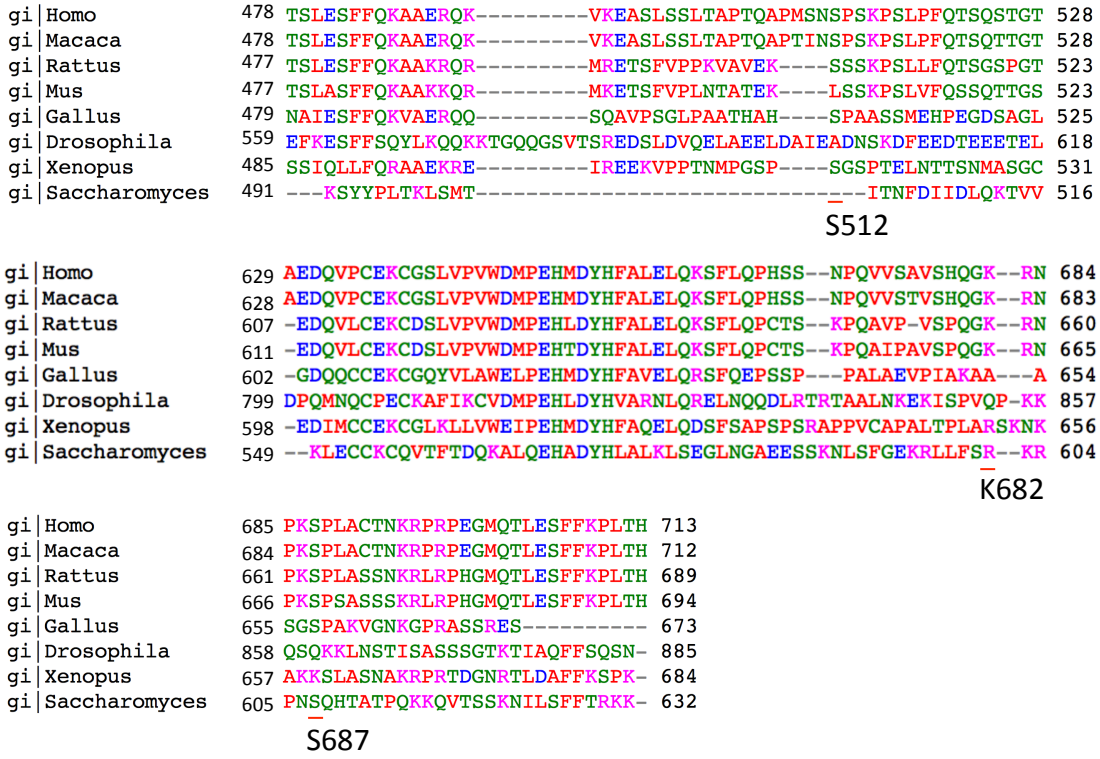


Figure 13 Amino sequence alignment of multiple species polymerase η. Red underlines indicate correspondent serine and lysine residue.

3.4 Discussion

In this study, we report the purification of the polymerase η-containing complex formed *in vitro* or *in vivo.* We analysed the *in vivo* polymerase η complex by MS/MS, while we did not analyse the complex formed *in vitro* because we failed to purify the complex with similar quality after scaling up the purification. Our initial aim was to identify proteins preferentially interacting with the polymerase η, which is

phosphorylated on the serine 601 residue, with the aim to understand better how the polymerase is regulated after the phosphorylation event. However, we were not able to identify any protein interacting only with the wildtype polymerase η while comparing the list of interacting candidates of the wild type and S601A mutant, suggesting that the phosphorylation on S601 does not influence on composition of the polymerase η-containing complex, or the difference of the composition between the two forms of polymerase η is below the sensitivity of MS/MS. Regardless of this, our study is still informative in terms of identifying novel potential polymerase η interacting proteins as well as novel phosphorylation sites. Some interacting proteins identified in this study are barely studied, but have been implied to be involved in the DNA damage response, such as SLC4A1AP, which has an alternative name as Human lung cancer oncogene 3 protein. This is a nuclear protein and is phosphorylated upon DNA damage (Matsuoka et al., 2007), yet its role in the DNA damage response has not been studied. Another example is RIF1, which is also phosphorylated upon DNA damage, and it has been reported to be important for the intra-S-phase checkpoint, but the mechanism still remains elusive. In both cases, no connection has been made between the proteins and translesion synthesis. It would be interesting to investigate their relationships with polymerase η and whether it is involved in the regulation of translesion synthesis.

Looking closer at the position of two phosphorylation sites, we found that S512 locates slightly ahead of the UBZ domain, while the S687 locates in the NLS domain. It has been reported that the monoubiquitination on K682 of polymerase η will prevent the interaction between polymerase η and PCNA, and it is possible that this phosphorylation could affect the monoubiquitination process for instance by

disrupting the interaction between the responsible E3 ligase and polymerase η, or perhaps stimulate the deubiquitination process. It would also be interesting to study which kinase is responsible for these two newly discovered phosphorylation sites, as it would shed light on how the polymerase η is regulated in response to replication fork arrest.

Chapter 4: Discussion

4.1 Structural study of full-length human FANCD2/FANCI complex utilizing electron microscopy (EM)

FANCD2 and FANCI stand an important position in the FA pathway, where its ubiquitination status affect the functionality of the pathway profoundly, and phosphorylation events also take part in the regulation of the activity of the FA pathway. Multiple FA proteins have been shown be the substrates of the ATM, ATR or other kinases after DNA damage, including FANCD2 and FANCI(Wang, 2007). The current model of the removal of DNA interstrand crosslink involves NER to unhook the crosslink nucleotide, TLS synthesize through the lesion and HR to repair the double strand break, and the FANCD2/FANCI complex is believed to participate in coordinating these events to happen at a timewise manner. Thus it has been the interests of the scientists in the DNA repair field to understand how FANCD2 and FANCI perform their functions. However, it is a difficult task because of the nature of FANCD2 and FANCI. Both FANCD2 and FANCI have no clearly defined domains and no enzymatic activities. One possibility is that they serve as a scaffold protein complex to recruit different interacting partners according to different stages of its modification.

The aim of this study was to characterise the two proteins from a structural point of view. Together with a previous study carried out by the Pavletich Group, we answered some questions, but more questions remain. They have suggested that the monoubiquitination sites of FANCD2 and FANCI are embedded in the interface of the

heterodimer, and the phosphorylation sites, which have been reported on FANCD2 and FANCI, are also present in the dimer interface (Joo et al., 2011). This raises a question how can the FA core complex access the monoubiquitination site, and carry out the ubiquitination process. Furthermore, if ubiquitin is embedded in the interface, how can it serve as a signaling platform as it is generally believed to be? Our EM study provided a different FANCD2/FANCI structure with some similarity to the crystal structure published by the Pavletich Group. Because of the limitation of the techniques, we could not identify the protein contribution to the structure we obtained. We predicted that the C terminal of FANCD2 may contribute to the fork-like tower in our EM structure by comparing with the crystal structure. It would be informative if we could confirm this contribution by studying the structure of the C terminal deletion FANCD2, and characterise the DNA binding activity of the deletion mutant.

The ultimate aim of this project is to obtain the structure of the ub-FANCD2/ub-FANCI heterodimer, which really functions at the site of DNA damage. We have already observed a slight conformational change after the native heterodimer bound to DNA, and it would be interesting to compare the structure before and after monoubiquitination, and also before and after DNA binding. Also, studying the DNA binding activity of the ub-FANCD2/ub-FANCI will give us additional information on how the heterodimer functions. As a native complex, FANCD2 and FANCI only possess general DNA binding activities. It is still unclear if monoubiquitination contribute to the recruitment of the heterodimer to the site of damage through the increase affinity to crosslinked DNA or through other regulating partners. It will be

exciting if we can address the remaining questions properly, and shed light on the detailed mechanism of the FA pathway.

4.2 Studies of the polymerase η complex after hydroxyl urea treatment

As other DNA repair pathways in the DDR, translesion synthesis is also heavily regulated by phosphorylation and ubiquitination events. Monoubiquitination of PCNA moves the equilibrium towards the switching of polymerases, while the monoubiquitination of polymerase η inhibits this effect. However, monoubiquitination is not the only determining factor of the recruitment of polymerase η. It has been shown that the UBZ domain of polymerase η is important for its nuclear accumulation (Bienko et al., 2005), but the polymerase η localization is not affected when the monoubiquitination of PCNA is inhibited (Sabbioneda et al., 2008), which suggests that other factors are also involved in the recruitment of polymerase η. In support of this conclusion, Alan Lehmann's group demonstrated that ATR-mediated phosphorylation of polymerase η on S601 is dependent on its UBZ domain, while the PCNA monoubiquitination is negligible, and this phosphorylation is important for the check point response and cellular survival towards UV irradiation (Gohler et al., 2011).

We initiated this study trying to discover novel factors, which may be important for the regulation of polymerase η. We identified several interesting potential binding partners that have not been implied in TLS. The next step will be establishing their roles in the pathway by confirming the interaction and mapping the interacting domain on both sides. Furthermore, if the cells become hypersensitive to UV radiation with reduced expression level of target genes, it would strongly suggest

that these genes play an important role in the TLS pathway. Another role of the potential interacting partners can be to regulate the polymerase activities.

It will be very interesting to study the impact of the two novel phosphorylation sites on the regulation of polymerase η, especially the S687, which is located next to the previously reported Lysine residue in the NLS domain that could be monoubiquitinated. Do these phosphorylation sites crosstalk with the ubiquitination process of polymerase η, or do they play a role in the recruitment of polymerase η to the damage site by creating a landing platform for other factors? Similar to the studies of interacting partners, survival assay could help us confirm the importance of the phosphorylation sites, and finding out the responsible kinases would allow us to fit these newly discovered phosphorylation events into the existing DDR network.

Chapter 5: Materials and methods

Cell culture

Sf9 insect cells were grown in Sf-90 II SFM (Invitrogen) at 27 °C. HeLa cells were grown in Dulbecco's or Joklik's MEM (Sigma) supplemented with 2.5% fetal bovine serum (Sigma) at 37°C. Stable HeLa cells expressing N-terminal Flag-HA tagged ATR, C-terminal Flag-HA tagged Polη and C-terminal Flag-HA tagged Polη S601A were generated as described in (Cohn et al., 2007).

Protein purification from *Sf9* insect cells

Proteins purified from *Sf9* cells were expressed using either the pFastBac-HTa vector (Invitrogen) containing an N-terminal His tag (full-length human FANCI and polymerase η) or the pFastBac vector (Invitrogen) with an engineered Flag tag and purified as described in (Cohn et al., 2007) (full-length human FANCD2 and TopBP1).

For polymerase η purification, the cells infected with baculovirus were lysed with base buffer containing 20mM Tris (pH 8.0), 0.2mM EDTA*, 1% Tween 20, 5mM $MgCl_2$, 0.1M KCl, 20mM Imidazole, 2mM β-me, 0.2mM PMSF and 10% glycerol. The lysate was clarified by centrifuging at 13,300 rpm for 20 minutes. The soluble fraction was collected and incubated with pre-equilibrated Ni^{2+}-NTA beads (Qiagen) for 2 hours at 4 degree. The beads were washed extensively by the lysis buffer, and protein was eluted with the same base buffer containing 250mM Imidazole.

* EDTA is not recommended for Ni^{2+}-NTA purification. In the base buffer there was excess amount of $MgCl_2$, which quenched EDTA and reduced the interference to the Ni^{2+}-NTA purification.

For TopBP1 purification, the cells infected with baculovirus were lysed with base buffer containing 20mM Tris (pH 8.0), 0.2mM EDTA, 1% Tween 20, 5mM $MgCl_2$, 0.1M KCl, 2mM β-me, 0.2mM PMSF and 10% glycerol. The lysate was clarified by centrifuging at 13,300 rpm for 20 minutes. The soluble fraction was collected and incubated with pre-equilibrated M2 beads (Sigma) for 2 hours at 4 degree. The beads were washed extensively by the lysis buffer, and protein was eluted with the same base buffer containing 1mg/ml flag peptide.

For FANCD2/FANCI purification, the cells infected with baculovirus were lysed with base buffer containing 20mM Tris (pH 8.0), 0.2mM EDTA, 1% Tween 20, 5mM $MgCl_2$, 0.1M KCl, 2mM β-me, 0.2mM PMSF and 10% glycerol. The lysate was then centrifuged and purified against Flag tag described as above. The flag eluate was supplemented with 20mM Imidazole and underwent the same Ni^{2+}-NTA purification described as above. The Ni^{2+}-NTA eluate was injected into pre-equilibrated size exclusion chromatography column, Superdex 200, and eluted with the base buffer containing 20mM Tris (pH 8.0), 0.2mM EDTA, 1% Tween 20, 5mM $MgCl_2$, 0.1M KCl, 2mM β-me, 0.2mM PMSF and 5% glycerol.

Electron Microscopy

For negative staining, samples were applied to electron microscope hydrophilic grids coated with carbon film and stained with 2% uranyl acetate. The imaging and analysis process are as previously described previously in (Janganan et al., 2011; Ohi et al., 2004).

In vitro ubiquitination assay

The reaction contained 75nM Uba1, 4uM Ube2T, 2uM FANCL, 10uM ubiquitin, and either 1uM ID complex or 1uM FANCD2 and FANCI monomers. The reaction was incubated at 30 degree for 90 minutes under buffer contained 0.5nM DTT, 2nM ATP, 2nM $MgCl_2$, 50mM Tricl-HCl (pH 7.5), 3.6% Glycerol, 64mM NaCl and ssDNA (0uM, 0.5uM, 20uM and 50uM).

Site directed mutagenesis

To generate the Serine to Alanine mutation in Polη, C-terminal Flag-HA tagged full-length Polη was used as a template, with the following primers: for Flag-HA tagged Polη S601A, 5'-GATTTGGCCCACAACGCCCAAAGCATGCACGCCT-3' as forward and 5'-AGGCGTGCATGCTTTGGGCGTTGTGGGCCAAATC -3' as reverse primer.

Antibodies

Antibodies used were as follows: anti-HA (mouse monoclonal antibody clone 12CA5); mouse anti-α-tubulin (CP06, Calbiochem); goat polyclonal anti-ATR (N-19, Santa Cruz Biotechnology, Inc.). The rabbit polyclonal α-Polη phosphospecific antibody and α-Polη were kindly provided by Prof. Alan Lehmann, which were described in (Gohler et al., 2011).

ATR Kinase assay

For ATR *in vitro* kinase assays, Flag-HA tagged human ATR was purified from HeLa cells stably expression Flag-HA tagged ATR with anti-Flag M2 agarose beads (Sigma-Aldrich) in base buffer containing 20mM Tris (pH 8.0), 0.2mM EDTA, 0.1% Tween 20, 5mM $MgCl_2$, 0.1M KCl and 10% glycerol. The purified recombinant ATR was then incubated with His-tagged Polη purified from baculovirus-infected insect cells in the reaction buffer contains 25mM Hepes (pH 7.4), 50mM NaCl, 10mM $MnCl_2$, 10mM

MgCl2, 1mM DTT and 1mM ATP at 30°C for 30 or 60 min. Phosphorylated proteins were separated by SDS-PAGE, and detected with a phospho-specific antibody.

Purification of the polymerase η-containing complex formed *in vitro*

First the *in vitro* ATR kinase assay on polymerase η was carried out, and the phosphorylated polymerase η was then incubated with nuclear extract prepared from hydroxyurea treated HeLa cells for 30 min at 4°C. Formaldehyde was then added to the mixture to 0.5%, and incubated on ice for 30 min to preserve all the protein-protein interactions. After incubation, 125mM Glycin-Tris (pH 7.8) was added to the mixture to quench the reaction, and the mixture was then dialyzed against 6M guanidine hydrochloride for 1 hour. Ni-NTA beads were incubated with the mixture for 2 hours, and washed extensively with 6M guanidine hydrochloride. The beads were equilibrated with base buffer described above, and eluted in base buffer containing 250mM imidazole. Before analysis with SDS-PAGE, the purified complex was mixed with SDS loading buffer, and heated at 95°C for 20 min to reverse the crosslink.

Cell fractionation and purification of *in vivo* polymerase η containing complex

HeLa stably expressing wildtype and S601A mutant were grown in spinning flasks, and treated with 4mM hydroxyurea for 24 hours before harvesting. Fractionation was carried out, and the nuclear fraction was used for anti-Flag and anti-HA purification as described in (Yuasa et al., 2006).

Mass spectrometry analysis

Proteins were reduced with dithiothreitol, cysteine residues were derivatized with iodoacetamide, and precipitated with 10% Trichloroacetic acid (TCA). The protein pellet was then resuspend in water, and digested with trypsin. Proteins from Coomassie-stained gel bands were in-gel digested with trypsin. These samples were summited to the Taplin Biological Mass Spectrometry Facility at the Harvard Medical School, and the processing procedure is as previously described in (Cohn et al., 2007).

Amino acid sequence alignment

The amino acid sequences of polymerase η from 5 different species were obtained from the NCBI database, and subjected to EMBL-EBI ClaustalW2 database for multiple sequence alignment (Goujon et al., 2010; Larkin et al., 2007). Manual alignment was also carried out, but no significant improvement was observed comparing to the original output from ClaustalW2.

Chapter 6: References

Bakkenist, C.J., and Kastan, M.B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature *421*, 499-506.

Bienko, M., Green, C.M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannouche, P., Wider, G., Peter, M., Lehmann, A.R.*, et al.* (2005). Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. Science *310*, 1821-1824.

Bienko, M., Green, C.M., Sabbioneda, S., Crosetto, N., Matic, I., Hibbert, R.G., Begovic, T., Niimi, A., Mann, M., Lehmann, A.R.*, et al.* (2010). Regulation of translesion synthesis DNA polymerase eta by monoubiquitination. Mol Cell *37*, 396-407.

Canman, C.E., Lim, D.S., Cimprich, K.A., Taya, Y., Tamai, K., Sakaguchi, K., Appella, E., Kastan, M.B., and Siliciano, J.D. (1998). Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. Science *281*, 1677-1679.

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.

Crossan, G.P., van der Weyden, L., Rosado, I.V., Langevin, F., Gaillard, P.H., McIntyre, R.E., Gallagher, F., Kettunen, M.I., Lewis, D.Y., Brindle, K.*, et al.* (2011). Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. Nat Genet *43*, 147-152.

D'Andrea, A.D. (2010). Susceptibility pathways in Fanconi's anemia and breast cancer. N Engl J Med *362*, 1909-1919.

Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., and Nussenzweig, A. (2004). H2AX: the histone guardian of the genome. DNA Repair (Amst) *3*, 959-967.

Gohler, T., Sabbioneda, S., Green, C.M., and Lehmann, A.R. (2011). ATR-mediated phosphorylation of DNA polymerase eta is needed for efficient recovery from UV damage. J Cell Biol *192*, 219-227.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res *38*, W695-699.

Harper, J.W., and Elledge, S.J. (2007). The DNA damage response: ten years after. Mol Cell *28*, 739-745.

Ho, G.P., Margossian, S., Taniguchi, T., and D'Andrea, A.D. (2006). Phosphorylation of FANCD2 on two novel sites is required for mitomycin C resistance. Mol Cell Biol *26*, 7005-7015.

Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G., and Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature *419*, 135-141.

Ishiai, M., Kitao, H., Smogorzewska, A., Tomida, J., Kinomura, A., Uchida, E., Saberi, A., Kinoshita, E., Kinoshita-Kikuta, E., Koike, T.*, et al.* (2008). FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. Nat Struct Mol Biol *15*, 1138-1146.

Janganan, T.K., Bavro, V.N., Zhang, L., Matak-Vinkovic, D., Barrera, N.P., Venien-Bryan, C., Robinson, C.V., Borges-Walmsley, M.I., and Walmsley, A.R. (2011). Evidence for the assembly of a bacterial tripartite multidrug pump with a stoichiometry of 3:6:3. J Biol Chem *286*, 26900-26912.

Jensen, R.B., Carreira, A., and Kowalczykowski, S.C. (2010). Purified human BRCA2 stimulates RAD51-mediated recombination. Nature *467*, 678-683.

Joo, W., Xu, G., Persky, N.S., Smogorzewska, A., Rudge, D.G., Buzovetsky, 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.

Kannouche, P.L., Wing, J., and Lehmann, A.R. (2004). Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. Mol Cell *14*, 491-500.

Kim, Y., Lach, F.P., Desetty, R., Hanenberg, H., Auerbach, A.D., and Smogorzewska, A. (2011). Mutations of the SLX4 gene in Fanconi anemia. Nat Genet *43*, 142-146.

Kumagai, A., Lee, J., Yoo, H.Y., and Dunphy, W.G. (2006). TopBP1 activates the ATR-ATRIP complex. Cell *124*, 943-955.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Lee, J.H., and Paull, T.T. (2007). Activation and regulation of ATM kinase activity in response to DNA double-strand breaks. Oncogene *26*, 7741-7748.

Liu, J., Doty, T., Gibson, B., and Heyer, W.D. (2010). Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. Nat Struct Mol Biol *17*, 1260-1262.

Liu, S., Shiotani, B., Lahiri, M., Marechal, A., Tse, A., Leung, C.C., Glover, J.N., Yang, X.H., and Zou, L. (2011). ATR autophosphorylation as a molecular switch for checkpoint activation. Mol Cell *43*, 192-202.

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

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., and Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell *131*, 887-900.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. Nature *399*, 700-704.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., 3rd, Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y.*, et al.* (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science *316*, 1160-1166.

Moldovan, G.L., and D'Andrea, A.D. (2009). How the fanconi anemia pathway guards the genome. Annu Rev Genet *43*, 223-249.

Nimonkar, A.V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J.L., Wyman, C., Modrich, P., and Kowalczykowski, S.C. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. Genes Dev *25*, 350-362.

Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. Biol Proced Online *6*, 23-34.

Ohmori, H., Friedberg, E.C., Fuchs, R.P., Goodman, M.F., Hanaoka, F., Hinkle, D., Kunkel, T.A., Lawrence, C.W., Livneh, Z., Nohmi, T.*, et al.* (2001). The Y-family of DNA polymerases. Mol Cell *8*, 7-8.

Pace, P., Mosedale, G., Hodskinson, M.R., Rosado, I.V., Sivasubramaniam, M., and Patel, K.J. (2010). Ku70 corrupts DNA repair in the absence of the Fanconi anemia pathway. Science *329*, 219-223.

Prakash, S., Johnson, R.E., and Prakash, L. (2005). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. Annu Rev Biochem *74*, 317-353.

Sabbioneda, S., Gourdin, A.M., Green, C.M., Zotter, A., Giglia-Mari, G., Houtsmuller, A., Vermeulen, W., and Lehmann, A.R. (2008). Effect of proliferating cell nuclear antigen ubiquitination and chromatin structure on the dynamic properties of the Y-family DNA polymerases. Mol Biol Cell *19*, 5193-5202.

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 Res *40*, 8425-8439.

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S.P. (2007). Human CtIP promotes DNA end resection. Nature *450*, 509-514.

Sato, K., Ishiai, M., Toda, K., Furukoshi, S., Osakabe, A., Tachiwana, H., Takizawa, Y., Kagawa, W., Kitao, H., Dohmae, N.*, et al.* (2012a). Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair. EMBO J *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 Res *40*, 4553-4561.

Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald, E.R., 3rd, Hurov, K.E., Luo, J., Ballif, B.A., Gygi, S.P., Hofmann, K., D'Andrea, A.D.*, et al.* (2007). Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. Cell *129*, 289-301.

Somyajit, K., Subramanya, S., and Nagaraju, G. (2010). RAD51C: a novel cancer susceptibility gene is linked to Fanconi anemia and breast cancer. Carcinogenesis *31*, 2031-2038.

Stoepker, C., Hain, K., Schuster, B., Hilhorst-Hofstee, Y., Rooimans, M.A., Steltenpool, J., Oostra, A.B., Eirich, K., Korthof, E.T., Nieuwint, A.W.*, et al.* (2011). SLX4, a

coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. Nat Genet *43*, 138-141.

Taniguchi, T., Garcia-Higuera, I., Andreassen, P.R., Gregory, R.C., Grompe, M., and D'Andrea, A.D. (2002a). S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. Blood *100*, 2414-2420.

Taniguchi, T., Garcia-Higuera, I., Xu, B., Andreassen, P.R., Gregory, R.C., Kim, S.T., Lane, W.S., Kastan, M.B., and D'Andrea, A.D. (2002b). Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. Cell *109*, 459-472.

Wang, W. (2007). Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. Nat Rev Genet *8*, 735-748.

Wang, X., Andreassen, P.R., and D'Andrea, A.D. (2004). Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. Mol Cell Biol *24*, 5850-5862.

Ward, I.M., and Chen, J. (2001). Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress. J Biol Chem *276*, 47759-47762.

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

Yuasa, M.S., Masutani, C., Hirano, A., Cohn, M.A., Yamaizumi, M., Nakatani, Y., and Hanaoka, F. (2006). A human DNA polymerase eta complex containing Rad18, Rad6

and Rev1; proteomic analysis and targeting of the complex to the chromatin-bound fraction of cells undergoing replication fork arrest. Genes Cells *11*, 731-744.

Zhi, G., Wilson, J.B., Chen, X., Krause, D.S., Xiao, Y., Jones, N.J., and Kupfer, G.M. (2009). Fanconi anemia complementation group FANCD2 protein serine 331 phosphorylation is important for fanconi anemia pathway function and BRCA2 interaction. Cancer Res *69*, 8775-8783.

Chapter 7: Appendix

7.1 Amino acid sequence alignment of Human and mouse FANCD2

```
gi|Homo     1 MVSKRRLSKSEDKESLTEDASKTRKQPLSKKTKKSHIANEVEENDSIFVKLLKISGIILK 60
gi|Mus      1 MISKRRRLDSEDKENLTEDASKT--MPLSKLAKKSHNSHEVEENGSVFVKLLKASGLTLK 58
              *:****  .*****.********   **** :**** ::******.*:****** **: **

gi|Homo    61 TGESQNQLAVDQIAFQKKLFQTLRRHPSYPKIIEEFVSGLESYIEDEDSFRNCLLSCERL 120
gi|Mus     59 TGENQNQLGVDQVIFQRKLFQALRKHPAYPKVIEEFVNGLESYTEDSESLRNCLLSCERL 118
              ***.****.***: **:****:**:**:***:*****.***** **.:*:**********

gi|Homo   121 QDEEASMGASYSKSLIKLLLGIDILQPAIIKTLFEKLPEYFFENKNSDEINIPRLIVSQL 180
gi|Mus    119 QDEEASMGTFYSKSLIKLLLGIDILQPAIIKMLFEKVPQFLFESENRDGINMARLIINQL 178
              ********: ******************** ****:*:::**.:* * **:.***:.**

gi|Homo   181 KWLDRVVDGKDLTTKIMQLISIAPENLQHDIITSLPEILGDSQHADVGKELSDLLIENTS 240
gi|Mus    179 KWLDRIVDGKDLTAQMMQLISVAPVNLQHDFITSLPEILGDSQHANVGKELGELLVQNTS 238
              *****:*******:::******:** *****:***************:*****.:**::***

gi|Homo   241 LTVPILDVLSSLRLDPNFLLKVRQLVMDKLSSIRLEDLPVIIKFILHSVTAMDTLEVISE 300
gi|Mus    239 LTVPILDVFSSLRLDPNFLSKIRQLVMGKLSSVRLEDFPVIVKFLLHSVTDTTSLEVIAE 298
              ********:********** *:*****.****:****:***:**:*****   :*****:*

gi|Homo   301 LREKLDLQHCVLPSRLQASQVKLKSKGRASSSGNQESSGQSCIILLFDVIKSAIRYEKTI 360
gi|Mus    299 LRENLNVQQFILPSRIQASQSKLKSKGLASSSGNQENSDKDCIVLVFDVIKSAIRYEKTI 358
              ***:*::*: :****:**** ****** ********.*.:.**:*:**************

gi|Homo   360 SEAWIKAIENTASVSEHKVFDLVMLFIIYSTNTQTKKYIDRVLRNKIRSGCIQEQLLQST 420
gi|Mus    359 SEAWFKAIERIESAAEHKALDVVMLLIIYSTSTQTKKGVEKLLRNKIQSDCIQEQLLDSA 418
              ****:****.  *.:***.:*:***:*****.***** :::*:****:*.*******:*:

gi|Homo   421 FSVHYLVLKDMCSSILSLAQSLLHSLDQSIISFGSLLYKYAFKFFDTYCQQEVVGALVTH 480
gi|Mus    419 FSTHYLVLKDICPSILLLAQTLFHSQDQRIILFGSLLYKYAFKFFDTYCQQEVVGALVTH 478
              **.*******:*.*** ***:*:** ** ** ****************************

gi|Homo   481 ICSGNEAEVDTALDVLLELVVLNPSAMMMNAVFVKGILDYLDNISPQQIRKLFYVLSTLA 540
gi|Mus    479 VCSGTEAEVDTALDVLLELIVLNASAMRLNAAFVKGILDYLENMSPQQIRKIFCILSTLA 538
              :***.***************:***.*** :**.*********:*:********:* :*****

gi|Homo   541 FSKQNEASSHIQDDMHLVIRKQLSSTVFKYKLIGIIGAVTMAGIMAADRSESPSLTQERA 600
gi|Mus    539 FSQQPGTSNHIQDDMHLVIRKQLSSTVFKYKLIGIIGAVTMAGIMAEDRSVPSNSSQRSA 598
              **:*  :*.************************************ *** ... :*. *

gi|Homo   601 NLSDEQCTQVTSLLQLVHSCSEQSPQASALYYDEFANLIQHEKLDPKALEWVGHTICNDF 660
gi|Mus    599 NVSSEQRTQVTSLLQLVHSCTEHSPWASSLYYDEFANLIQERKLAPKTLEWVGQTIFNDF 658
              *:*.** ************:*:** **:***********..** **:*****:** ***

gi|Homo   661 QDAFVVDSCVVPEGDFPFPVKALYGLEEYDTQDGIAINLLPLLFSQDFAKDGGPVTSQES 720
gi|Mus    659 QDAFVVDFCAAPEGDFPFPVKALYGLEEYSTQDGIVINLLPLFY-QECAKDASRATSQES 717
              ******* *..*****************.*****.******:: *: ***.. .*****
```

```
gi|Homo   721 GQKLVSPLCLAPYFRLLRLCVERQHNGNLEEIDGLLDCPIFLTDLEPGEKLESMSAKERS 780
gi|Mus    718 SQRSMSSLCLASHFRLLRLCVARQHDGNLDEIDGLLDCPLFLPDLEPGEKLESMSAKDRS 777
              .*: :*.****.:******** ***:***:*********:**.**************:**

gi|Homo   781 FMCSLIFLTLNWFREIVNAFCQETSPEMKGKVLTRLKHIVELQIILEKYLAVTPDYVPPL 840
gi|Mus    778 LMCSLTFLTFNWFREVVNAFCQQTSPEMKGKVLSRLKDLVELQGILEKYLAVIPDYVPPF 837
              :**** ***:*****:******:**********:***.:**** ******** ******:

gi|Homo   841 GNFDVETLDITPHTVTAISAKIRKKGKIE-RKQKTDGSKTSSSDTLSEEKNSECDPTPSH 899
gi|Mus    838 ASVDLDTLDMMPRSSSAVAAKNRNKGKTGGKKQKADSNKASCSDTLLTEDTSECDMAPSG 897
              ...*::***: *:: :*::** *:***   :***:*..*:*.****  *..**** :**

gi|Homo   900 RGQLNKEFTGKEEKTSLLLHNSHAFFRELDIEVFSILHCGLVTKFILDTEMHTEATEVVQ 959
gi|Mus    898 RSHVDKESTGKEGKTFVSLQNYRAFFRELDIEVFSILHSGLVTKFILDTEMHTEATEVVQ 957
              *.:::** **** ** : *:* :***************.**********************

gi|Homo   960 LGPPELLFLLEDLSQKLESMLTPPIARRVPFLKNKGSRNIGFSHLQQRSAQEIVHCVFQL 1019
gi|Mus    958 LGPAELLFLLEDLSQKLENMLTAPFAKRICCFKNKGRQNIGFSHLHQRSVQDIVHCVVQL 1017
              ***.**************.***.*:*:*:  :**** :*******:***.*:*****.**

gi|Homo  1020 LTPMCNHLENIHNYFQCLAAENHGVVDGPGVKVQEYHIMSSCYQRLLQIFHGLFAWSGFS 1079
gi|Mus   1018 LTPMCNHLENIHNFFQCLGAEHLSADDKARATAQEQHTMACCYQKLLQVLHALFAWKGFT 1077
              *************:****.**: .. * . ...** * *:.***:***::*.****.**:

gi|Homo  1080 QPENQNLLYSALHVLSSRLKQGEHSQPLEELLSQSVHYLQNFHQSIPSFQCALYLIRLLM 1139
gi|Mus   1078 HQSKHRLLHSALEVLSNRLKQMEQDQPLEELVSQSFSYLQNFHHSVPSFQCGLYLLRLLM 1137
              : .::.**:***.***.**** *:.******:***. ******:*:*****.***:****

gi|Homo  1140 VILEKSTASAQNKEKIASLARQFLCRVWPSGDKEKSNISNDQLHALLCIYLEHTESILKA 1199
gi|Mus   1138 ALLEKSAVPNQKKEKLASLAKQLLCRAWPHGEKEKNPTFNDHLHDVLYIYLEHTDNVLKA 1197
              .:****:.. *:***:****:*:***.** *:***.   **:** :* ******:.:***

gi|Homo  1200 IEEIAGVGVPELINSPKDASSSTFPTLTRHTFVVFFRVMMAELEKTVKKIEPGTAADSQQ 1259
gi|Mus   1198 IEEITGVGVPELVSAPKDAASSTFPTLTRHTFVIFFRVMMAELEKTVKGLQAGTAADSQQ 1257
              ****:*******:.:****:*************:************** ::.********

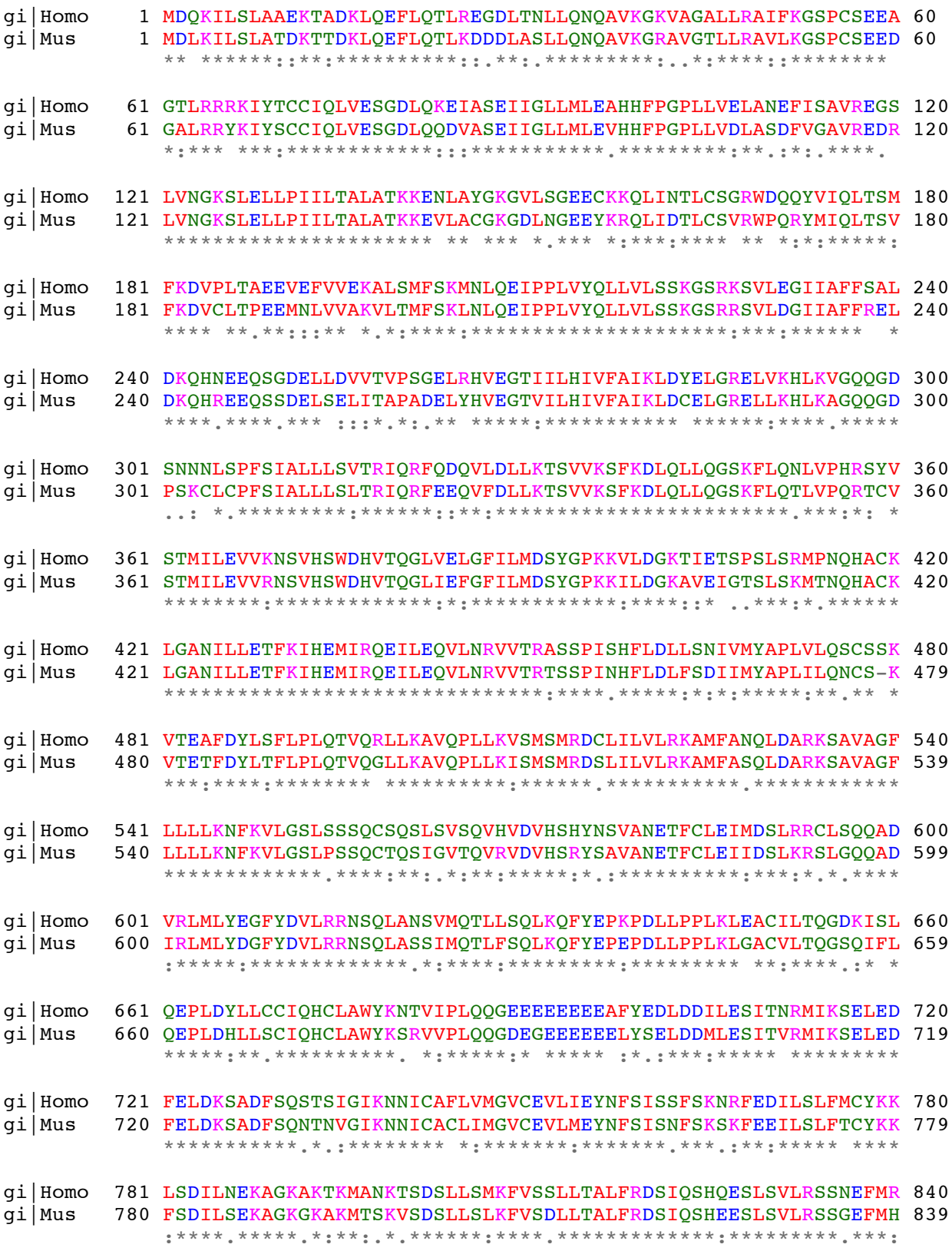
gi|Homo  1260 IHEEKLLYWNMAVRDFSILINLIKVFDSHPVLHVCLKYGRLFVEAFLKQCMPLLDFSFRK 1319
gi|Mus   1258 VHEEKLLYWNMAVRDFSILLNLMKVFDSYPVLHVCLKYGRRFVEAFLKQCMPLLDFSFRK 1317
              :******************:**:*****:********** ******************

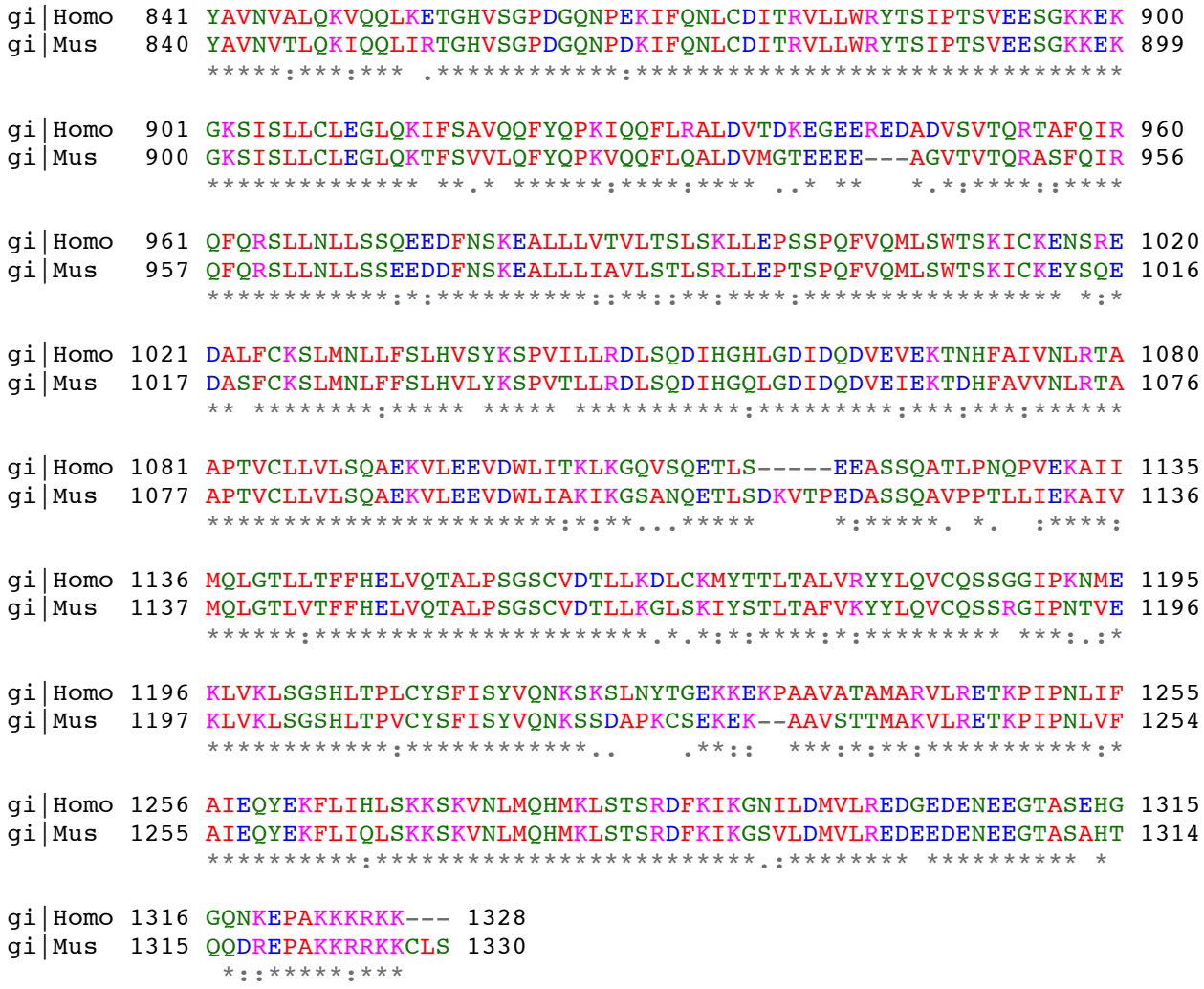
gi|Homo  1320 HREDVLSLLETFQLDTRLLHHLCGHSKIHQDTRLTQHVPLLKKTLELLVCRVKAMLTLNN 1379
gi|Mus   1318 HREDVLSLLQTLQLNTRLLHHLCGHSKIRQDTRLTKHVPLLKKSLELLVCRVKAMLVLNN 1377
              *********:*:**:*************:******:*******:***********.***

gi|Homo  1380 CREAFWLGNLKNRDLQGEEIKSQN--SQESTADESEDDMSSQASKSKATEVSLQNPPESG 1437
gi|Mus   1378 CREAFWLGTLKNRDLQGEEIISQDPSSSESNAEDSEDGVTSHVSRNRATEDGEDEASDEQ 1437
              ********.********** **:  *.**.*::***.::*:.*:.:*** . ::..:.

gi|Homo  1438 TDGCILLIVLSWWSRTLPTYVYCQMLLCPFPFPP 1471
gi|Mus   1438 KD------QDSDESDDSSS--------------- 1450
              .*        *
```

7.2 Amino acid sequence alignment of Human and mouse FANCI




gi|Homo 841 YAVNVALQKVQQLKETGHVSGPDGQNPEKIFQNLCDITRVLLWRYTSIPTSVEESGKKEK 900
gi|Mus 840 YAVNVTLQKIQQLIRTGHVSGPDGQNPDKIFQNLCDITRVLLWRYTSIPTSVEESGKKEK 899
gi|Homo 901 GKSISLLCLEGLQKIFSAVQQFYQPKIQQFLRALDVTDKEGEEREDADVSVTQRTAFQIR 960
gi|Mus 900 GKSISLLCLEGLQKTFSVVLQFYQPKVQQFLQALDVMGTEEEE---AGVTVTQRASFQIR 956
gi|Homo 961 QFQRSLLNLLSSQEEDFNSKEALLLVTVLTSLSKLLEPSSPQFVQMLSWTSKICKENSRE 1020
gi|Mus 957 QFQRSLLNLLSSEEDDFNSKEALLLIAVLSTLSRLLEPTSPQFVQMLSWTSKICKEYSQE 1016
gi|Homo 1021 DALFCKSLMNLLFSLHVSYKSPVILLRDLSQDIHGHLGDIDQDVEVEKTNHFAIVNLRTA 1080
gi|Mus 1017 DASFCKSLMNLFFSLHVLYKSPVTLLRDLSQDIHGQLGDIDQDVEIEKTDHFAVVNLRTA 1076
gi|Homo 1081 APTVCLLVLSQAEKVLEEVDWLITKLKGQVSQETLS-----EEASSQATLPNQPVEKAII 1135
gi|Mus 1077 APTVCLLVLSQAEKVLEEVDWLIAKIKGSANQETLSDKVTPEDASSQAVPPTLLIEKAIV 1136
gi|Homo 1136 MQLGTLLTFFHELVQTALPSGSCVDTLLKDLCKMYTTLTALVRYYLQVCQSSGGIPKNME 1195
gi|Mus 1137 MQLGTLVTFFHELVQTALPSGSCVDTLLKGLSKIYSTLTAFVKYYLQVCQSSRGIPNTVE 1196
gi|Homo 1196 KLVKLSGSHLTPLCYSFISYVQNKSKSLNYTGEKKEKPAAVATAMARVLRETKPIPNLIF 1255
gi|Mus 1197 KLVKLSGSHLTPVCYSFISYVQNKSSDAPKCSEKEK--AAVSTTMAKVLRETKPIPNLVF 1254
gi|Homo 1256 AIEQYEKFLIHLSKKSKVNLMQHMKLSTSRDFKIKGNILDMVLREDGEDENEEGTASEHG 1315
gi|Mus 1255 AIEQYEKFLIQLSKKSKVNLMQHMKLSTSRDFKIKGSVLDMVLREDEEDENEEGTASAHT 1314
gi|Homo 1316 GQNKEPAKKKRKK--- 1328
gi|Mus 1315 QQDREPAKKRRKKCLS 1330

7.3 Introduction of cryoEM

Cryo-electron microscopy (cryoEM), or electron cryomicroscopy, is a form of transmission electron microscopy (EM) where the sample is studied at cryogenic temperatures (generally liquid nitrogen temperatures). The advantage of cyoEM is that the samples are studied in their native environment, which could avoid the functionally irrelevant conformational changes resulted from placing the sample into a non-physiological environment. The protein sample was spread out onto an electron microscopy grid and preserved in a frozen-hydrated state by rapid freezing. Protein sample are imaged by low-dose technique because of the radiation sensitive nature. The image was then analysed by single particle alignment, where all the particles were selected, centered and grouped into different classes by comparing the intensity of each pixel. This analysis could reinforce the common features existing in most particles by averaging out the random distributed noise signal. In this study, these images were then used to refine the 3D model we obtained from tilted negative stained EM.