Characterisation of a novel gene with a potential role in the Fanconi anaemia tumour suppressor pathway

By
Durva Sevak

UNIVERSITY OF BIRMINGHAM

A thesis submitted to the University of Birmingham for the degree of MRes in Cancer Sciences

School of Cancer Research
College of Medical and Dental Sciences
University of Birmingham
August 2014

Abstract

The large nuclear protein PFAP termed as Putative Fanconi Anaemia protein is a novel uncharacterised protein, approximately 550 KDa in size, with few identifiable domains. However, till date its function remains elusive. According to unpublished results from our laboratory, cells lacking PFAP exhibit loss of phenotypes such as hypersensitivity to cross-linking agents, accumulation in G2 phase and increased chromosomal breakage after exposure to MMC (mitomycin C), as previously observed in cells lacking known Fanconi anaemia proteins. Mitomycin C is an anti-tumour alkylating drug, with an potential of forming cross-links within the DNA. The Fanconi anaemia pathway is important in the repair of inter-strand crosslinks (ICLs) and the maintenance of genome instability. Based on two reports, there is a possibility of PFAP playing a key role in regulating histone lysine methylation, via its interaction with the SET/COMPASS like histone methyltransferases SET1A/SET1B. It was previously shown that PFAP interacts with SET1B. However, unpublished studies carried out by our laboratory have suggested that PFAP interacts with SET1A. Therefore, we wished to investigate these interactions in further detail. In this study, we demonstrated that PFAP interacts with SET1A histone methyltransferase, but not SET1B. In addition, we studied localisation of SET1A to the sites of ICL repair sites, as proteins involved in DNA repair often relocalises to form sub-nuclear foci in response to DNA- damaging agents such as MMC. Immunofluorescence staining reveals that FLAG-tagged SET1A and RAD51 each localises to a largely non-overlapping regions of nucleus, suggesting that both FLAG-SET1A and RAD51 each targets specific genomic sites and therefore, has distinct functions in ICL repair. Finally, we demonstrated that the Fanconi anaemia proteins: FANCD2 and FANCJ interacts with PFAP, thus indicating that PFAP is involved in FA pathway.

Acknowledgement

First and foremost, I would like to thank, Dr. Grant Stewart and Dr. Martin Higgs for providing me this opportunity to work on this project; my parents and my brother, for being my strength and support throughout my MRes. My deepest and heartfelt go gratitude to Dr. Martin Higgs for supervising and tolerating me throughout the project. I would also like to thanks members of the Stewart lab, in particularly Dr. Edward Miller for providing constructs of GST-tagged PFAP fragments, and other colleagues in neighbouring group studying DNA damaging. Special thanks to my two lovely friends and office desk neighbour Thea Constantinou and Harpreet Dibra for daily coffee sessions and being my support during difficult phases of the project. Finally, I would also like to thank Dr. Roger Watson for guiding me as a friend. Without the support of these people and other friends of mine, I would not have been able to come this far.

Abbreviations

ASH2L:	Absent Small Homeotic Discs-2 Like
ATM:	Ataxia telangiectasia mutated
ATR:	ATM Rad3- related
BER:	Base Excision Repair
BOD1:	Bi-orientation of chromosomes in cell division 1
BOD1L:	Bi-orientation of chromosomes in cell division 1-like
BRCA1:	Breast cancer 1, early onset
BRCA2:	Breast cancer 1, early onset
BRCT:	BRCA1 C Terminus
BRIP1:	BRCA1 interacting protein C terminal helicase 1
BSA:	Bovine Serum Albumin
Cfp1:	CXXC Finger Protein 1
COMPASS:	Complex Proteins Associated with Set1
COSMIC:	Catalogue of Somatic Mutations In Cancer
DEB:	Dipeoxybutane
DSBs:	Double strand breaks
ECL:	Enhanced Chemiluminescence
EDTA:	Ethylene diamine tetra acetic acid
EME1:	Essential meiotic structure specific endonuclease 1
ERCC1:	Excision Repair Cross Complementation group1
FA:	Fanconi anaemia
FAAP24:	Fanconi anaemia associated protein 24
FANCA:	Fanconi Anaemia Complementation group A

FANCB:	Fanconi Anaemia Complementation group B
FANCC:	Fanconi Anaemia Complementation group C
FANCD1 (BRCA2):	Fanconi Anaemia Complementation group D1
FANCD2:	Fanconi Anaemia Complementation group D2
FANCE:	Fanconi Anaemia Complementation group E
FANCF:	Fanconi Anaemia Complementation group F
FANCG:	Fanconi Anaemia Complementation group G
FANCI:	Fanconi Anaemia Complementation group I
FANCJ (BRIP1):	Fanconi Anaemia Complementation group J
FANCL:	Fanconi Anaemia Complementation group L
FANCM:	Fanconi Anaemia Complementation group M
FANCN (PALB2):	Fanconi Anaemia Complementation group N
FANCO (RAD51C)	Fanconi Anaemia Complementation group O
FANCP (SLX4):	Fanconi Anaemia Complementation group P
FCS:	Fetal calf serum
FLAG-SET1A:	FLAG-tagged SET domain containing 1A
g:	Relative centrifugal force in units of gravity
HCF-1:	Host Cell factor-1
HEPES:	(4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid)
HMTs:	Histone Methyltransferases
HR:	Homologous Recombination
ICL:	Interstrand crosslinks
IF:	Immunofluorescence staining
IgG:	Immunoglobulin G

IGEPAL:	OctyIphenoxypolyethoxyethanol
IPTG:	Isopropyl-β D-thiogalactoside
mA:	milliAmpere
MMC:	Mitomycin C
NER:	Nucleotide excision repair
NHEJ:	Non-homologous end joining
PBS:	Phosphate buffered saline
PFA:	Paraformalaldehyde
PFAP:	Putative Fanconi Anaemia Protein
PALB2:	Partner and localizer of BRCA2
PHD:	Plant Homeo domain
PTMs:	Posttranslational modifications
RAD51:	RAD51 recombinase
RPA:	Replication Protein A
RPM:	Revolution per minute
RRM:	RNA recognition motif
SET:	Su (var) 3-9, Enhancer of Zeste, Trithorax
SET1A:	SET domain containing 1A
SET1B:	SET domain containing 1B
SLX4:	Structure specific endonuclease subunit SLX4
TLS:	Translesion synthesis
WDR5:	WD repeat- containing protein 5
WDR82:	WD repeat- containing protein 82
XP F / C:	Xeroderma Pigmentosum F / C

Table of Content

1.0 Introduction

Fanconi Anaemia (FA) pathway functions to protect the cell against DNA highly toxic lesion termed interstrand crosslinks (ICLs). This project involves characterization of a novel gene called PFAP, whose loss would phenotypically resembles loss of Fanconi anaemia proteins involved in the Fanconi anaemia pathway. This project include investigating in vitro interactions of PFAP with components of histone methyltransferase complexes and other FA proteins, to elucidate the role of PFAP in histone methylation and with other fanconi anaemia proteins involved in FA pathway.

1.1 Introduction to DNA damage and repair

Our cellular genomes are under constant attack from variety of DNA damaging agents that could pose a threat to genomic integrity (Bernstein *et al.,* 2013). DNA damage is a change in the structure of DNA; this change can be a chemical addition or disruption to a base possibly, giving rise to bulky adducts or a break within the phosphate backbone of DNA (See Hoeijmakers, 2001 for review). Such breaks may affect single strand or both strands of DNA (See Hoeijmakers, 2001; Ciccia and Elledge, 2010 for reviews). DNA damaging lesions are caused by environmental mutagens (UV, ionizing radiation and various genotoxic causes), by product(s) of normal cellular metabolism (reactive oxygen species, superoxide dismutase and glutathione peroxidase) and spontaneous disintegration of chemical bonds in DNA caused by physiological conditions such as hydrolysis, deamination. Consequences of lesions caused by all of the above sources could trigger cell cycle arrest, induction of apoptosis, inhibition of transcription and replication and also cause mutations resulting in oncogenesis (Hoeijmakers, 2001).

Double strand breaks (DSBs) are the most potent lesions, and one unrepaired DSB is sufficient to cause apoptosis due to its ability to inhibit transcription, prevent DNA replication and cause problems during chromosome segregation (Hoeijmakers, 2001; Chapman *et al*., 2012). Due to the above reasons, DSBs are referred as one of the primary causes of cancer (Jackson and Bartek, 2009; McKinnon, 2009). Therefore, DNA breaks need to be repaired efficiently by specific DNA repair mechainsms. DNA damage recognition repair is carried out by activities of many different enzymes including nucleases, kinases, demethylases, polymerases and phosphatases which act in various repair mechanisms such as nucleotide excision repair, base excision repair, homologous recombination, non-homologous end joining, mismatch repair (Ciccia and Elledge, 2010) (Figure 1.1).

1.1.2 DNA repair mechanisms

Eukaryotic organisms have evolved sophisticated, interwoven DNA repair systems to cope with the injuries "imposed on a cell's vital genetic information" (Hoeijmakers, 2001). The five main types of repair pathways in mammals include nucleotide-excision repair (NER), base-excision repair (BER), homologous recombination (HR) and non-homologous end joining (NHEJ) (Reviewed by Lindahl and Wood, 1999; Hoeijmakers, 2001) (Figure 1.1). While these pathways deal with different types of lesions, however, there is a considerable overlap between these mechanisms (Martin *et al*., 2008). For example, NER deals with range of helix-distorting lesions arising mostly by exogenous agents, and some caused by oxidative lesions. Most of the oxidative lesions are mostly repaired by BER, the repair pathway responsible dealing small chemical alterations of bases (Pecorino, 2012) (Figure 1.1). Furthermore, if these single strand breaks (ssDNA) are not repaired by BER, then these ssDNA breaks may form into DSBs, which are then repaired by homologous recombination

repair or non-homologous end joining (Martin *et al*., 2008). Homologous recombination dominates in S and G2 when the DNA is replicated, providing a second copy of the sister chromatid for aligning the breaks. In contrast, non-homologous end joining is most prevalent in the G1 phase of the cell cycle, where a second copy of sister chromatid as template is not available (Pecorino, 2012; Takata *et al*., 1998). Each of these repair mechanism mentioned above forms part of a series of orchestrated pathways called the DNA damage response (DDR) (Spycher *et al*., 2008; Ciccia and Elledge, 2010).

1.2 DNA Damage Response (DDR)

The DDR is a highly orchestrated signal transduction pathway that acts to sense and repair DNA damage; and responds in a well orchestrated way. The activation of DDR pathway is dependent on the type of DNA damage that occurs and the activation of apical kinases called ATM (ataxia telangiectasia mutated), activated in response to DSBs, ATR (ataxia telangiectasia- Rad3 related) activated in response to exposed single stranded DNA and DNA-PKcs (DNA-dependent protein kinase catalytic subunits). These apical kinases activate various downstream kinases and proteins that coordinate the arrest of cell cycle progression, cellular senescence and in some cases apoptosis to maintain the genomic integrity (Spycher *et al*., 2008; Sulli *et al*., 2012).

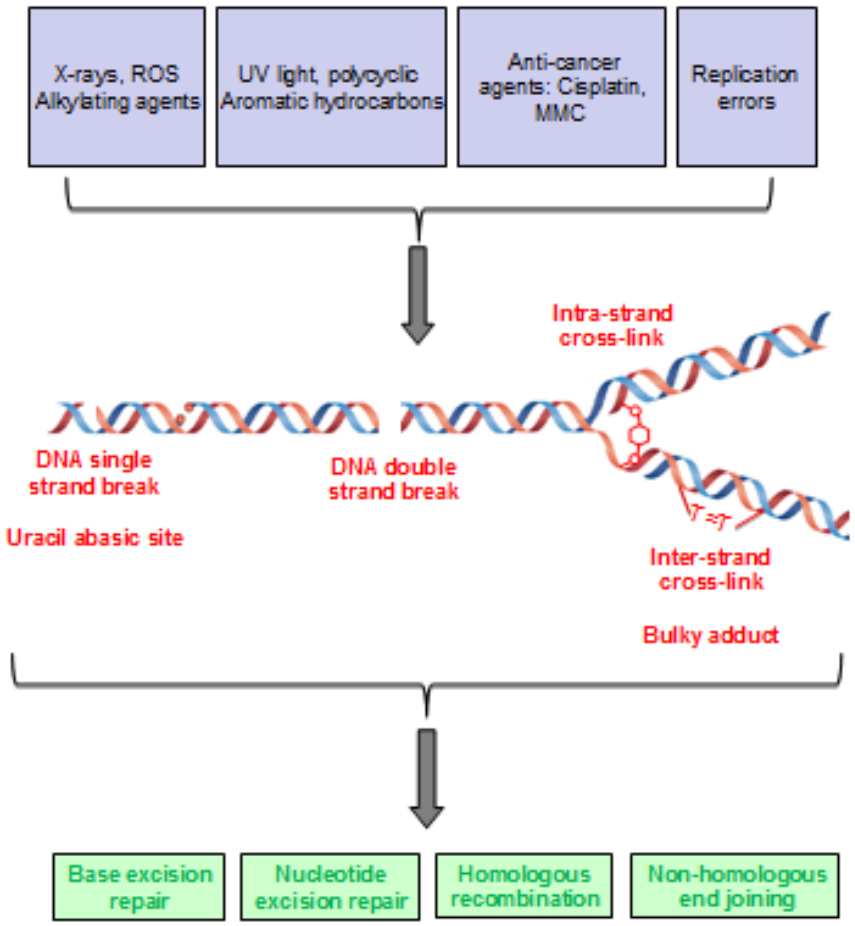

Figure 1.1: Schematic diagram of common DNA damage-inducing agents and repair mechanisms. Examples of DNA damage inducing agents (blue), types of lesions induced by DNA damaging agents (red) and mechanisms involved for the removal of the lesions (green). **(Adapted from Hoeijmakers, 2001).**

1.3 Fanconi Anaemia (FA)

1.3.1 Clinical phenotypes of FA

Fanconi anaemia (FA) is an autosomal recessive disorder affecting almost 1 in 100,000 births (Moldovan and D'Andrea, 2009; Kottemann and Smogorzewska, 2013). Due to the diversity of clinical phenotypes presented by patients diagnosed with FA, it is termed as "clinically heterogenous"(Moldovan and D'Andrea, 2009). Fanconi anaemia was initially described by Swiss paediatrician Guido Fanconi in 1927; and characterized as aplastic anaemia with short stature, hypogonadism and skin pigmentation (Fanconi, 1927). Several clinical phenotypes are observed in FA patients including skeletal abnormalities (primordial dwarfism, congenital defects), progressive bone marrow failure, pancytopenia, microcephaly (Garcia-Higuera *et al*., 2001; Tischkowitz and Hodgson, 2003). Importantly, patients with FA also exhibits an increased risk of developing acute myeloid leukaemia (AML) and squamous cell carcinoma of head and neck, although the underlying mechanisms remain unclear (Moldovan and D'Andrea, 2009).

1.3.2 Cellular phenotypes of FA

In addition to the clinical manifestations described above, the most striking feature of cells defective in the FA pathway is their tendency to exhibit hypersensitivity to DNA cross-linking agents such as mitomycin C (MMC), cisplatin, dipeoxybutane etc. This crucial phenotype underlines the importance of proteins in the FA pathway in repairing inter-strand crosslinks (described in section 1.3.3-1.3.4) induced by these cross-linking agents (Moldovan and D'Andrea, 2009). Accordingly, exposure to these crosslinking agents can also increase the frequency of chromosomal breakage and radial chromosomes in FA cells, (Figure 1.2). Thus, Fanconi anaemia is referred as genomic instability disorder.

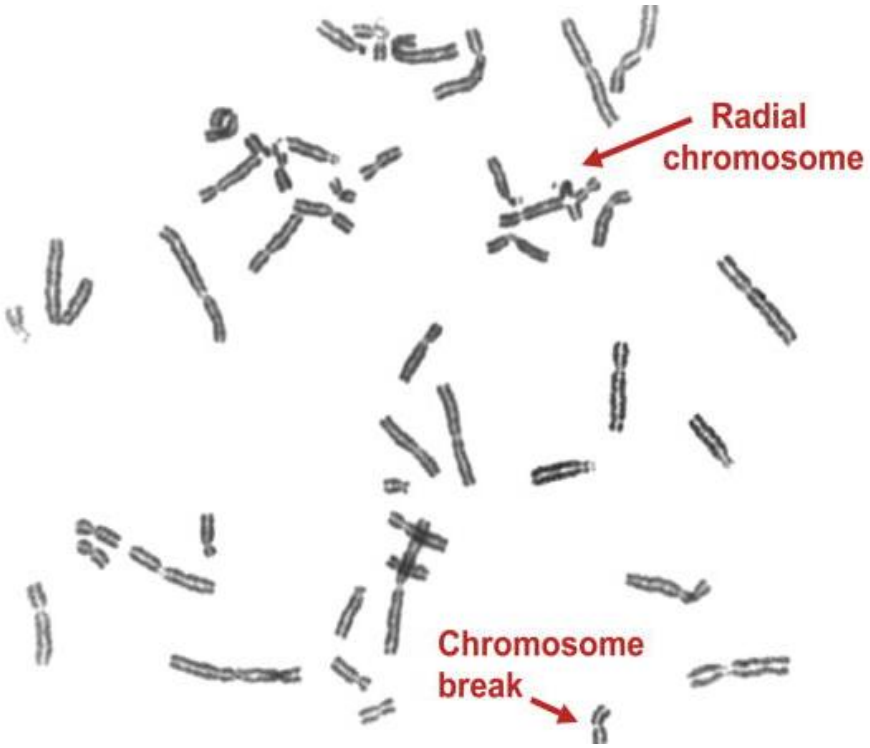

Figure 1.2: Representative image of a metaphase spread in cells defective for the FA pathway after MMC exposure. Chromosome breakage and radial chromosomes are indicated in red. **(Adapted from Moldovan and D'Andrea, 2009)**

Therefore, cells defective in FA are becoming a highly valued model system to study ICL repair.

1.3.3 Fanconi anaemia proteins

Thus far, sixteen Fanconi anaemia gene products have been identified (also known as complementation groups: FANCA, B, C, D1, D2 E, F, G, I, J, L, M, N, O, P, Q) (Refer to table 1.0 for more details). Eight of these FA proteins (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL) forms FA core complex of the fanconi anaemia pathway, which gets recruited to the chromatin for the mono-ubiquitination of FANCD2-FANCI, once ICL lesion is detected by FANCM and accessory protein called FAAP24. Other FA proteins such as FANCD1 (BRCA2), FANCN (PALB2), FANCO (paralog of RAD51) and FANCJ are involved in the homologous recombination repair (downstream pathway of FA pathway). FANCP and newly identified FA protein called FANCQ (Walden, 2014) are involved in the nucleolytic activity downstream of mono-ubiquitination of FANCD2. FANCP (SLX4) forms the scaffold or co-factor for the other nucleases: XPF-ERCCI and MUS81-EME1 for nucleolytic incision stages of FA pathway (table 1.0) (See reviews by Moldovan and D'Andrea, 2009).

1.3.4 The FA pathway and ICL repair

As a result of numerous studies, it is now clear that components of the FA pathway are involved in the both of repair of interstrand crosslinks. In this section, I will introduce the role of FA proteins in ICL repair.

ICLs are highly toxic lesions that covalently link two strands of DNA, when exposed to DNA-cross-linking agents such as MMC, cisplatin and DEB etc (Ciccia and Elledge, 2010). The presence of these crosslinks inhibits DNA replication, as ICL blocks replication

forks and also interfere with gene transcription. Regardless at which stage of the cell cycle ICL has formed, cells with ICLs arrest in the S phase of the cell cycle, allowing time for the ICL to be repaired before replication can continue (Akkari *et al*., 2000; Ward *et al*., 2012). Due to their ability to inhibit vital cellular processes, ICLs needs to be repaired to prevent fatal cellular consequences.

Table 1.0: List of Fanconi anaemia proteins and their functions. **(Adapted from Kottemann and Smogorzewska, 2013; Walden and Dean, 2014).**

FA genes	**Function**	**Link to Fanconi anaemia**
FANCA FANCB FANCC FANCD2 FANCE FANCF FANCI FANCG FANCL FANCM	Core complex member and required FANCD2-FANCI ubiquitination; FANCL is an E3 ubiquitin ligase	Genes mutated in Fanconi anaemia
$FANCD_1$ (BRCA2)	HR mediator; downstream of FANCD2-FANCI ubiquitination	Genes mutated in Fanconi anaemia
FANCJ (BRIP1)	Fanconi anaemia, homologous recombination, 3' to 5' DNA helicase	Genes mutated in Fanconi anaemia
FANCN (PALB2)	Fanconi anaemia, homologous recombination and promotes BRCA2	Genes mutated in Fanconi anaemia
FANCO (RAD51C)	Homologous recombination	Genes mutated in Fanconi anaemia
FANCP (SLX4)	Fanconi anaemia, coordinates XPF-ERCC1, MUS81-EME1 and SLX1 nucleases	Genes mutated in Fanconi anaemia
FAN1	Interstrand crosslink repair	Interacts with I-D2 complex
FANCQ (ERCC4)	Interstrand crosslink and nucleotide excision repair	Gene mutated in Fanconi anaemia

ICL lesions are repaired through the orchestrated actions of various nucleases, translesion polymerases and many more DNA repair factors, which are co-ordinated by members of the

FA pathway. Although the precise mechanisms are not yet understood, current evidence suggests that ICL maybe repaired in two ways - replication dependent and replication independent (Figure 1.3) (Kottemann and Smogorzewska, 2013; Huang and Li, 2013). In the absence of active replication forks, ICL lesions are repaired independent of replication and recombination (RIR). This repair process mainly occurs in cells during G0/G1 phase of the cell cycle. RIR is a "critical process to maintain genomic integrity and allowing transcription in non-dividing or slowly dividing cells as suggested by Williams and co (2012). NER factor termed XPC recognises ICL lesion. This result in the recruitment of NER nucleases: XPF-ERCC1 and XPG to initiates unhooking step on one of the damaged strand by introducing dual incisions on both sides of the ICL lesion, followed by translesion synthesis in which a translesion polymerase zeta (TLS pol ζ) fills the gap on one of the damaged strand and exonucleases excises the unhooked ICL lesion to complete the repair process (Figure 1.3) (Huang and Li, 2013).

On the other hand, replication and recombination dependent pathway is activated when an ICL lesion is encountered during S phase and the converging forks from different directions stalls a few nucleotides from a particular ICL lesion site. Subsequently one fork moves forwards and stops just before lesion (Moldovan and D'Andrea, 2009). The recruitment of endonucleases allows excision of the region surrounding the ICL, and forming a DSB. This results in the uncoupling of one sister chromatid from the other, followed by the unhooking of the cross-links by endonucleases, bypassing of the unhooked lesion by translesion polymerases which are able to replicate through DNA lesions (Neidzweidz *et al.,* 2004). The DSB formed by endouclease activity is repaired by homologous recombination and invasion

of the repaired chromatid by the sister (Figure 1.4). Fanconi anaemia pathway is the classic example of replication-dependent repair of ICLs.

1.3.4.1 Replication dependent ICL repair - FA pathway

Recruitment of the FA core complex and mono-ubiquitination of FANCI-FANCD2

One of the crucial stages in the activation of the FA pathway during replication-dependent ICL repair is the mono-ubiquitination of FANCI- FANCD2 complex by the FA core complex. The FA proteins form the FA core complex (refer to section 1.3.3) and acts as a multi-subunit ubiquitin ligase. The FA core complex is recruited to the chromatin during S phase of the cell cycle for FANCI-FANCD2 monoubiquitination complex, by FANCM and FAAP24, upon ICL encounter (Reviewed by Patel and Joenje, 2007; Ciccia *et* al., 2007; Moldovan and D'Andrea, 2009; Kottemann and Smogorzewska, 2013) (Figure 1.4).

Prior to mono-ubiquitination, phosphorylation of FANCI at SQ and TQ sites is required for the efficient ubiquitination of I-D2 complex (Ishiai *et al*., 2008; Knipscheer *et al*., 2009). Mono-ubiquitination of FANCD2 and FANCI occurs at lysines lys561 and lys523 respectively through the action of the PHD domain of FANCL and UBET E2 ubiquitin enzyme (Figure 1.4) (Taniguchi *et al*., 2002; Smogorzewska *et al.,* 2007).

Upon mono-ubiquitination on lys561, FANCD2 relocalizes to chromatin as DNA damage foci, where it binds and co-localises with BRCA2 (Garcia-Higuera *et al*., 2001; Taniguchi *et al* 2002; Wang *et al*., 2004) and thus enable FANCD2 to orchestrate the actions of downstream repair pathway including translesion bypass, nuclease recruitment and homologous recombination repair (Figure 1.4) (Moldovan and D'Andrea, 2009).

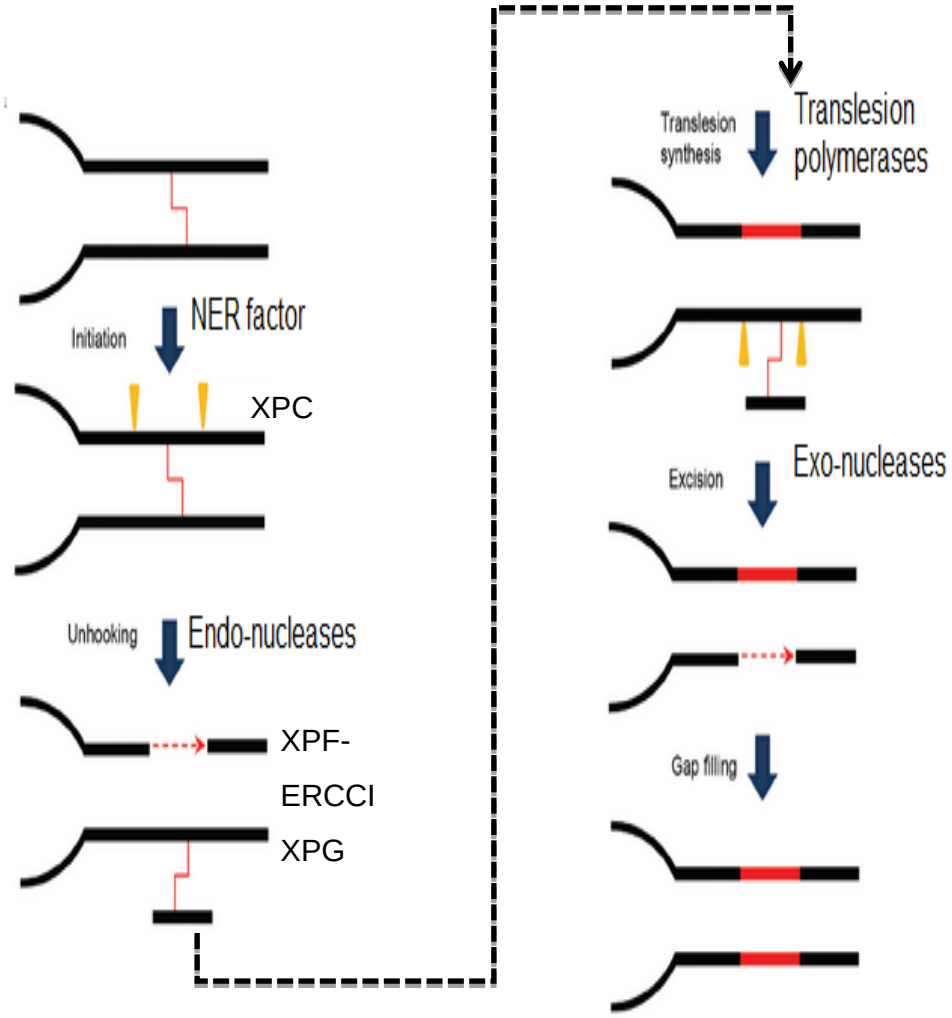

Figure 1.3: Replication and recombination independent ICL repair. A. Replication independent interstrand cross link repair. Pathway comprises by recognition of ICL lesion by NER factor XPC. XPF-ERCC1 and XPG unhook the lesion, translesion polymerase zeta fills the gap one of damaged strands and the unhooked lesion is removed or excised by exonucleases. **(Adapted from Huang and Li, 2013)**.

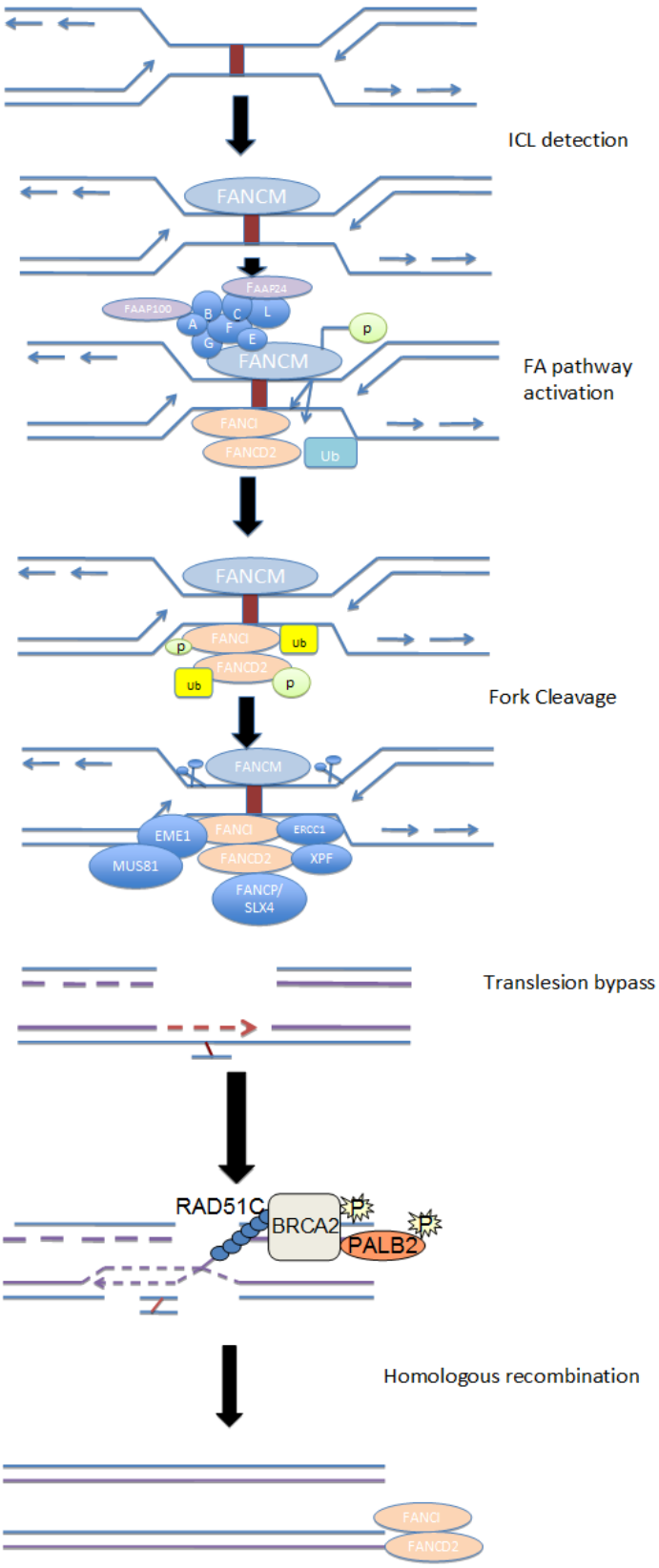

Figure 1.4: Replication-dependent ICL repair.
The core complex is made up of Fanconi anaemia proteins and accessory proteins (such as FAAP24 and FAAP20). Upon detection of the crosslink, FANCM is recruited and activated by ATR-mediated phosphorylation, recruiting the core complex to chromatin. This complex ubiquitinates the I-D2 complex. Upon monoubiquitination, ubiquitinated I-D2 complex localizes to the chromatin and orchestrates the action of downstream repair factors including SLX4, XPF, and MUS81/EME1 that functions at the site of DNA damage to make incisions on either side of the lesion strand. On the incised strand, TLS polymerases are recruited to by-pass the unhooked crosslink. This break is then repaired by BRCA2, PALB2, and RAD51C via homologous recombination. **(Adapted from Ciccia and Elledge, 2010)**

Regulation of D2-I mono-ubiquitination

The regulation of the FA pathway is poorly studied. However, USP1 (ubiquitin specific peptidase 1) protein is involved in the regulation of FA pathway via deubiquitination of FANCD2 (Patel and Joenje, 2007; Moldovan and D'Andrea, 2009; Kottemann and Smogorzewska, 2013).

Under normal circumstances, i.e. in the absence of DNA damage, mono- ubiquitination of FANCD2 is kept under control by USP1 deubiquitylating enzymes, with the help of UAF1 (WD repeat containing protein) by forming USP1-UAF1 complex. Upon DNA damage, the transcription of USP1 is turned off and the USP1 protein present in the cells are degraded by the proteasome and thus, resulting in excessive accumulation of mono-ubiquitinated FANCD2 (Moldovan and D'Andrea, 2009). However, even in the absence of USP1, excess FANCD2 are still unable to localise to foci, even though they are able to accumulate on the chromatin. This suggests that USP1 is required for the completion of DNA repair and / or re-localisation of FANCD2 to the sites of ICL damage.

Nucleolytic removal of the ICL and TLS

The initial incisions that unhook the cross-links are performed by NER endonculeases: XPF- ERCC1 and Mus81-EME1, with help of SLX4 (FANCP) acting as a scaffold or co-factor. ICL nucleolytic cleavage is performed in highly coordinated manner, with Mus81-EME1 performing the initial incision at the 3' end of the lesion and ERCC1-XPF performs secondary incision at 5' end cleavage (de Winter and Joenje, 2008; Ciccia *et al*., 2008).

Translesion DNA polymerases are also important components of ICL repair. Studies in *Xenopus laevis* egg extracts demonstrated that the translesion polymerases including polymerase Rev1 and polymerase polζ (comprised of Rev3 and Rev7) have essential roles in the removal of ICL. Upon ICL unhooking, Rev1 inserts a base into the position across the

ICL lesion on the complementary strand, followed by the polζ to extend the unrepaired strand (Huang and Li, 2013).

Homologous Recombination Repair (HR)

FANCD1 (BRCA2), FANCN (PALB2) and FANCJ do not form the part of the core complex required for the mono-ubiquitination of FANCD2-FANCI, but are required for homologous recombination repair of the ICL. Chromatin bound ubiquitinated FANCD2-FANCI, recruits BRCA2, PALB2 (interactor of BRCA2) and Rad51 to initiate ICL repair via homologous recombination (HR). BRCA2 relocalises with mono-ubiquitinated FANCD2 and thus, initiates homologous recombination by promoting assembly of RAD51 onto ssDNA. This RAD51-ssDNA complex then invades homologous duplex DNA to generate a displacement (D-loop) loop. At the D-loop, DNA synthesis occurs to extend the invading strand and finally, recombination process occurs via double holiday junction (Pecorino, 2012). FANCJ (BRIP1 or BACH1) is the second helicase in the FA pathway. FANCJ interacts with BRCT domains of BRCA1 and localizes to DNA repair structures containing BRCA2 and RPA (Replication Protein A) (Moldovan and D'Andrea, 2009). "Although, the function of FANCJ in ICL repair is unknown, its 5' to 3' helicase activity is likely used to remodel DNA structures, thus facilitating or activating repair"(Moldovan and D'Andrea, 2009).

The roles of FA proteins are not limited to ICL repair, but are also involved in the stabilisation of replication fork and tumour suppression (de Winter and Joenje, 2009; Kottemann and Smogorzewska, 2013).

1.3.5 FA and stabilisation of replication fork

Apart from its role in interstrand cross-link repair, FA pathway is also important for the protection and stabilisation of DNA replication forks under replicative stress (Moldovan and D'Andrea, 2012; Schlacher *et al*., 2012). In addition to ICLs, FA pathway can also be activated by replicative stress agents mainly hydroxyurea (HU) and ultraviolet light (UV). Schlacher *et al* has demonstrated that components of the FA pathway are required for the stabilisation and protection of replication forks from nucleolytic degradation, thus preventing genomic instability and tumorigenesis (Schlacher *et al*., 2012). Schlacher and co-workers demonstrated that mono-ubiquitinated FANCD2 (FANCD2-ub) is functionally epistatic with BRCA2 and Rad51, thus providing a complete understanding of Fanconi anaemia proteins involved in maintaining replication fork stability from ICL and other DNA stresses (Schlacher *et al*., 2012). FANCD2-ub interacts with proliferating cell nuclear antigen (PCNA) (Howlett *et al*., 2009), BRCA2, RAD51. In the absence of FANCD2, BRCA2, Rad51, are insufficient to stabilise replication fork, therefore FANCD2, acts as mediator for the association of BRCA2/Rad51 with PCNA, for stabilising replication forks and preventing it from collapsing (Schlacher *et al*., 2012). However, in the absence of BRCA2 and FANCD2, Rad51 filaments are sufficient to for fork protection. This is because of gain of function by Rad51 mutant or overexpressed Ras51, results in the formation of hyperfilaments of Rad51 which stabilises stalled replication fork in the absence of BRCA2 and FANCD2 (Morrison *et al*., 1999).

1.4 PFAP (Putative Fanconi Anaemia Protein)

Our laboratory has discovered a novel uncharacterized protein, that we have termed PFAP (Putative Fanconi Anaemia Protein). PFAP is a large nuclear protein of 3051 amino acid and approximately 550 KDa in size. PFAP has few identifiable protein domains, but interestingly contains several ATM/ATR phosphorylation sites scattered throughout PFAP (Matsuoka *et al*., 2007), mainly located within the central region of PFAP (Figure 1.4). In addition, the N-terminus of PFAP contains a region of sequence homologous to the mitotic protein BOD1 (Biorientation Of Cell Division 1) (Porter *et al*., 2007) and the yeast COMPASS protein Shg1, that we have named the Shg1-homology (Shg1H) domain (Figure 1.4).

According to unpublished data from our laboratory, cells lacking PFAP-defective cells exhibit similar cellular phenotypes to cells lacking Fanconi anaemia proteins (as mentioned in earlier sections 1.3.2), including hypersensitivity to DNA crosslinking agents, accumulation in the G2 phase after MMC and increased chromosomal breakages after exposure to MMC.

Thus far, there is no literature describing PFAP and its role in maintaining genomic stability. However, there are two reports which raises the possibility that PFAP may play role in the regulating histone methylation, via interactions with the methylated forms of Lys4 of histone H3 (Eberl, 2013), and through a possible interaction with SET/COMPASS like histone methyltransferases SET1A/SET1B (Nuland *et al*., 2013). Recent studies carried out by Nuland and co-workers (2013), demonstrated that PFAP and BOD1 interacted with SET1B (Nuland *et al*., 2013). However, data from our laboratory suggests that PFAP interacts with SET1A, but not SET1B. Thus, this project is aimed to investigate this interaction in further detail.

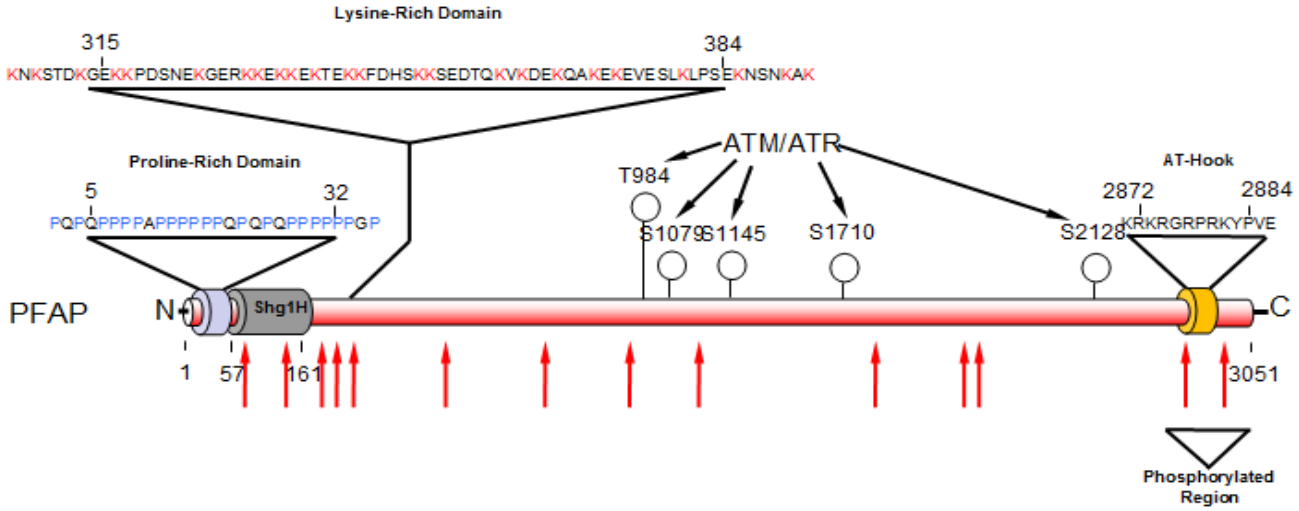

Figure 1.5: Schematic representation of PFAP and its domains. The Shg1H domain, which is homologous to both the yeast COMPASS member Shg1, and the human protein BOD1, is indicated in grey. The N-terminal 'Pro-rich domain' is in blue, and the C-terminal A-T hook/NLS domain is in orange. Known ATM/ATR phosphorylation sites are encircled (Matsuoka *et al.,*2007), and putative ATM/ATR phosphorylation sites are indicated with arrows.

1.5 The SET1A and SET1B histone methyltransferase

SET1A and SET1B comprise two of the six SET/COMPASS like-histone methyltransferases (HMTs) (Figure 1.6), responsible for regulating global levels of methylation of lysine four on histone H3 (H3K4). SET1A and SET1B, along with the other four histone methyltransferase complexes, shares non-catalytic subunits such as WDR5, Rbbp5, ASH2L and DPY30 also known as (WRAD) Cfp1, WDR82 (Figure 1.6). The WRAD forms a sub-complex and associates with catalytic subunit of HMT complexes (Dehe *et al.*, 2006; Nuland *et al*., 2013).

Both SET1A and SET1B contains SET-domain, post-SET domain at their carboxyl terminal and RNA recognition motif at the amino terminal end (Figure 1.7). SET1A and SET1B also exhibit 39% identity and 56% similarity to their counterpart yeast Set1 protein (Figure 1.7) (Lee *et al.*, 2007). Although both SET1A and SET1B are similar, confocal microscopy reveals that both SET1A and SET1B exhibit non-overlapping sub-nuclear localization. This shows both target distinct set of genomic sites and therefore, has distinct functions in regulation of chromatin structure and gene expression (Lee *et al*., 2007).

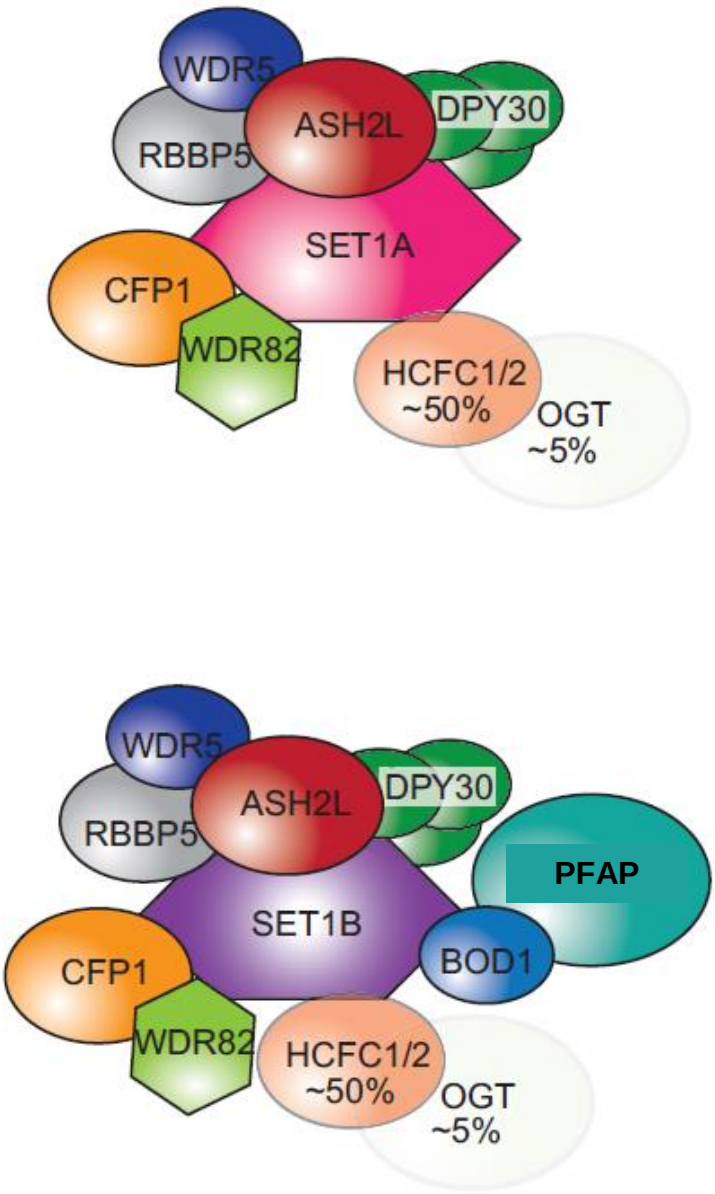

Figure 1.6: SET1A/SET1B histone methyltransferase complexes.
The SET1A (upper) and SET1B (lower) histone methyltransferase complexes consists of common non-catalytic subunits and the unique enzymatic subunits SET1A and SET1B. NB: PFAP and BOD1 are illustrated to be part of the SET1B-containing complex in this diagram, although our data indicates that PFAP is a part of the SET1A-containg complex **(Adapted from Nuland *et al.,* 2013).**

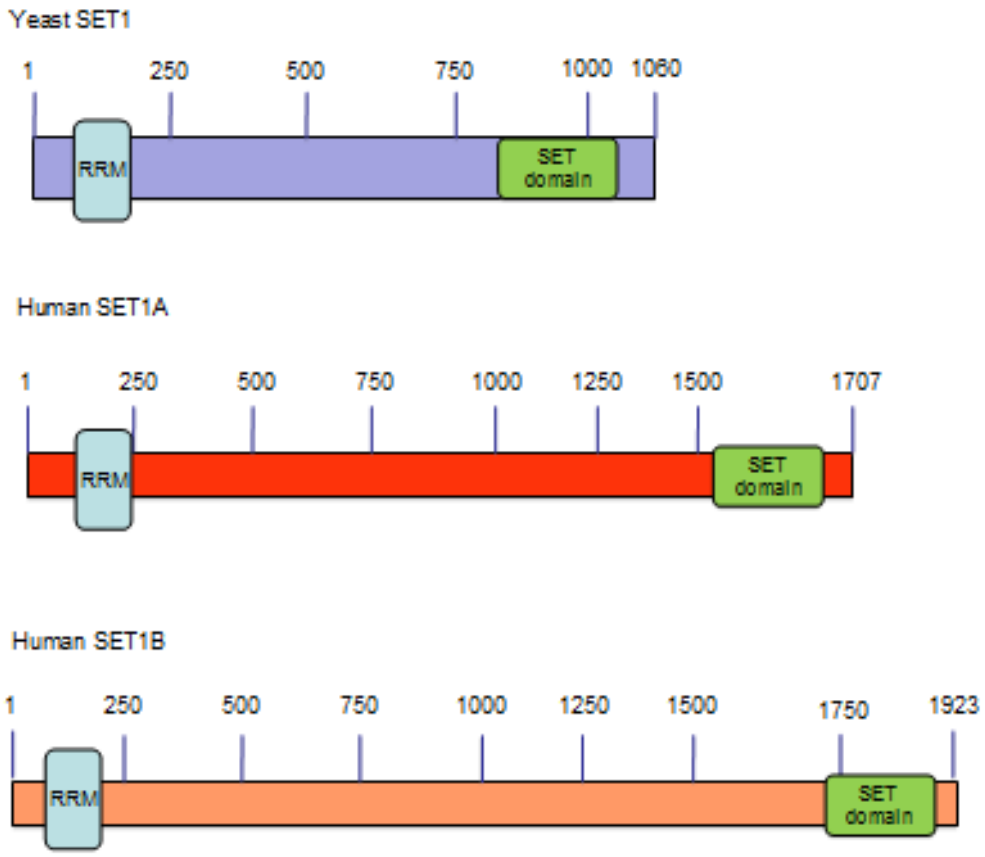

Figure 1.7: Schematic representations of human SET1A/SET1B and yeast SET. The enzymatic domains are indicated, with the RRMs (RNA recognition motif) in blue, and the catalytic SET domains in green. **(Adapted from Lee *et al*., 2013).**

1.5.1 Histone methylation by SET1A/SET1B

"Chromatin" is the structure in which DNA is embedded, and the importance of chromatin structure is increasingly understood to be crucial to the DDR and how lesions are recognised and repaired (Jorgensen *et al.*, 2013).

Chromatin is composed of nucleosomes, which in turn is composed of an octamer of four core histone proteins namely H2A, H2B, H3, H4 (Review by Martin and Zhang, 2005). Histone proteins are subject to various posttranslational modifications (PTMs) such as phosphorylation, methylation, acetylation, ubiquitination, sumoylation that are crucial in maintaining the chromatin structure and regulating gene expression but which are important in DNA repair, DNA replication and cell division (Review by Martin and Zhang, 2005).

Histone methylation is the process where methyl groups are transferred to amino acids of histone proteins. Histone methylation occurs on lysine and arginine residues on histone three and histone four and are catalysed by enzymes belonging to three distinct families of proteins: Arg residues are methylated by PRMT (Protein Arginine Methyltransferases), and lysine residues are methylated by SET-domain methyltransferases (SET1A) and non-SET domain proteins (DOT1/DOT1L) (Zhang and Reinberg, 2001; Bannister and Kouzarides, 2005). "Methylated lysines and arginine residues act to recruit effector proteins to specific genomic loci to impose their specific regulatory function upon the underlying DNA" (Jenuwein and Allis, 2001).

A large body of evidence has demonstrated that histone lysine methylation regulates gene expression by acting as a transcriptional activator or transcriptional repressor, depending on the site of the methylation (Zhang and Reinberg, 2001; Martin and Zhang, 2005). To date,

there is no published evidence that methylation of H3K4 by SET1A or SET1B regulates the DDR, or that they contribute to maintaining genomic instability. This is currently under investigation by our laboratory.

Aims of the Project

As part of our ultimate goal, which is to characterise the function and role of PFAP in the DDR, and in particular its role in the FA pathway, we wished to explore further interaction between SET1A and PFAP. Given our knowledge outlined above, this project therefore had three main objectives:

1. To examine the interaction of PFAP and SET1A in vitro, using two approaches:
    - To confirm previous data obtained in our laboratory, we will first examine the interaction of PFAP with the histone methyltransferases SET1A, SET1B *in vitro* by immunoprecipitation analyses

    - To extend and complement (a), we aim to determine the area(s) of PFAP necessary for interaction with SET1A using GST-tagged fragments of PFAP in GST-pull down assays

2. To determine whether PFAP interacts with other known Fanconi anaemia proteins by co-immunoprecipitation

3. To study whether SET1A localises to Rad51-containing sub-nuclear foci induced by mitomycin c exposure by immunofluorescence

2.0 Materials and Methods

2.1 Plasmids

The following plasmids were used in this study

Plasmid	**Origin/Supplier**	**Brief description**
pGEX-3X	GE Healthcare	Bacterial vector for expressing GST fusions under the control of tac promoter. Encodes Amp^R
pGEX-PFAP-F1	Dr. M.Higgs/Dr. E. Miller University of Birmingham	Amino acids 1-600 of PFAP cloned into the SalI-NotI restriction sites of pGEX-3X
pGEX-PFAP-F2	Dr. M.Higgs/Dr. E. Miller University of Birmingham	Amino acids 500-1000 of PFAP cloned into the SalI-NotI restriction sites of pGEX-3X
pGEX-PFAP-F3	Dr. M.Higgs/Dr. E. Miller University of Birmingham	Amino acids 900-1500 of PFAP cloned into the SalI-NotI restriction sites of pGEX-3X
pGEX-PFAP-F4	Dr. M.Higgs/Dr. E. Miller University of Birmingham	Amino acids 1400-2000 of PFAP cloned into the SalI-NotI restriction sites of pGEX-3X
pGEX-PFAP-F5	Dr. M.Higgs/Dr. E. Miller University of Birmingham	Amino acids 1900-2500 of PFAP cloned into the SalI-NotI restriction sites of pGEX-3X

pGEX-PFAP-F6	Dr. M.Higgs/Dr. E. Miller University of Birmingham	Amino acids 2400- 3051 of PFAP cloned into the SalI-NotI restriction sites of pGEX-3X

2.2 Cell culture

HeLa-FLAG-SET1A cells were from Prof.D.Skalnik (IUPUI, Indianapolis, USA), and were cultured in Dulbecco's Modified Eagle's Media supplemented with 10 % tetracycline-free fetal calf serum (Gibco, UK). U2OS-FLAG-SET1A cells were created by Dr. G. Stewart (University of Birmingham), and cultures in McCoys' media supplemented with 10 % tetracycline-free fetal calf-serum. Where appropriate, doxycycline was added to cell culture medium at 1 μg/ml. Mitomycin c was from Sigma Aldrich (UK), and was used at a final concentration of 50 ng/ml.

2.3 Antibodies

The following antibodies were used in this study

Table 2.1: List of primary antibodies used

Primary antibodies	**Host**	**Dilution for immunoblotting**	**Application**	**Source**
SET1A	Rabbit	1 in 500	IP, pull down	Home-made by David Skalnik, USA
SET1B	Rabbit	1in 250	IP, pull down	Bethyl laboratories (USA)
FLAG-SET	Mouse	1in 500	IF	Bethyl laboratories (USA)
RAD51	Rabbit	1in 50	IF	Calbiochem (UK)

PFAP	Rabbit	1in 500	IP	Santa cruz Biotechnology (UK)
FANCM	Rabbit	1in 500	IP	Fanconi anaemia Research funds (USA)
FANCJ	Rabbit	1in 500	IP	Fanconi anaemia Research funds (USA)
FANCD2	Mouse	1in 500	IP	Santa cruz Biotechnology UK

Table 2.2: List of secondary antibodies used

Secondary antibodies	**Host**	**Dilutions**	**Application**	**Source**
HRP Conjugate Reliablot	Rabbit	1in 3000	IP, pull downs	Bethyl laboratories (USA)
Fluorescein FITC	Goat	1in 500	IF	Life technologies (UK)
Alexa Fluor® 595	Mouse	1 in 500	IF	Life technologies (UK)
Polyclonal anti-goat IgG/ HRP	Rabbit	1in 3000	IP	Dako (UK)
Polyclonal swine IgG/HRP	Mouse	1in 1000	IP	Dako (UK)

2.4 Chemicals

Chemicals were purchased from Sigma Aldrich (UK) except the exceptions listed below. Glycine was from VWR (UK), glutathione was from Roche Biochemicals (USA) and acrylamide was purchased from BioRad (UK).

2.5 Bacterial transformations

One hundred microlitres of BL21 *Escherichia coli* competent cells (Agilent Technologies) were aliquoted into prechilled tubes. To each set of competent cells, 2 µl of diluted β-mercaptoethanol (Agilent Technologies, UK) and approximately 10-50 ng of pGEX-3X or pGEX-3X encoding PFAP fragments (refer to section 2.1) were added. Cells were placed on ice for 30 minutes, heat-shocked for 25 seconds at 42 °C and incubated on ice for another 2-3 minutes. Eight hundred microlitres of SOC medium was added to the transformation reactions and incubated for 1 h at 37 °C in a shaking incubator at 200 rpm. After incubation, 100 µl of transformation reactions were placed onto agar plates containing 500 µl of ampicillin and left to form colonies overnight at 37 °C.

2.6 Small-scale protein expression

Single ampicillin-resistant colonies were picked from each agar plates and grown in 15 ml of Luria Broth (LB) supplemented with 100 µg/ml ampicillin at 30 °C, until they reached an optical density (OD_{600nm}) of 0.4-0.6. Once the expected OD was reached, IPTG was added to a final concentration of 1 mM and incubated for 3 h at 30 °C.

2.7 Large scale protein expression

Single bacterial colonies were picked and grown as described in section 2.6. Fifteen mililitre starter cultures were grown for 6-7 hours at 30 °C and 2 ml of this culture was aliquoted into 500 ml of LB supplemented with 100 µg/ml ampicillin. Bacterial cultures were incubated at 30 °C, until they reached an optical density (OD_{600nm}) between 0.4-0.6. Once the expected OD was reached, IPTG was added to a final concentration of 1 mM and incubated for 3 hours at 30°C. Bacterial pellets were collected by centrifugation x g for 20 minutes at 4 °C and bacterial pellets frozen down at -80 °C until purification.

2.8 Purification of recombinant proteins

Prior to purification, Glutathione Sepharose beads (GE Healthcare, UK) were centrifuged at x g for 30 seconds to remove excess ethanol and washed three times with 10 minute washes of 10 ml of PBS + 1 % Triton X-100 and twice in 10 ml of PBS +1 mM EDTA for ten minutes. Beads were left in PBS + 1 mM EDTA until further use.

Bacterial pellets were resuspended in lysis buffer (PBS +1 % Triton X- 100, plus protease inhibitor tablets) and lysed by sonicating the samples six cycles at max power for 20 seconds on ice. To each cell lysates 4 ml pre-prepared glutathione sepharose beads were added and the mixture was incubated for 2 h on a rotator at 4 °C. After 2 h, lysates were spun and second set of GST beads were added and incubated for further 2 h. After centrifugation the recovered beads were washed three times with 40 ml of PBS+ 1 % Triton X-100 + 1 mM EDTA for ten minutes each. Beads were further washed two times with 10 ml of PBS+1 mM

EDTA. After the last wash, the beads were resuspended in 4 ml of elution buffer (25 mM Glutathione pH 8, 2.5 ml of 50 mM Tris pH 8) and incubated for another 1 hour. The eluted protein was transferred into pre-chilled 15 ml tube and protein re-eluted with a two further incubations of 4 ml of elution buffer. All 3 eluted proteins fractions were transferred into dialysis tubing (pre-boiled in 2 mM EDTA and 3 % sodium bicarbonate; and washed three times in distilled water before use) and dialysed overnight in dialysis buffer (50 mM Tris pH 7.5, 200 ml of 10 % glycerol, 2 ml of 1 mM DTT and 1698 ml of distilled water). The concentration of eluted dialysed proteins was measured by Bradford assay (Bio-rad, UK) and estimated by Sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) in conjunction with pre-diluted BSA protein standards.

2.9 GST pull down assays

For GST pull-down assays, 5 µg of purified GST or GST fusion proteins were incubated with 1 ml of pre-cleared HeLa nuclear extracts (Cilbiotech, Belgium) at 4 °C for 3 h. Glutathione sepharose beads (pre-washed in NTN buffer (100 mM NaCl, 20 mM Tris pH 7.5, 0.05 % NP40)) were then added and extracts were further incubated at 4 °C for another 4 hours. The beads were then washed three times with 1 ml of NTN buffer and resuspended in 2X SDS loading buffer (Bio-Rad) and boiled for ten minutes at 95 °C before analysing by SDS-PAGE and immunoblotting.

2.10 Immunoprecipitation

Five micrograms of immunoprecipitating antibody or IgG was added to 200 µl of pre-cleared HeLa nuclear cell extracts and incubated at 4 °C for 3 h. Twenty five microlitres of

protein A sepharose beads (GE Healthcare, UK) were added and the lysates incubated at 4 °C for further 2 h. The immune complexes were washed with three times with 1 ml NTN buffer and resuspended in 2X SDS loading buffer, boiled at 95 °C for ten minutes, before analysing by SDS-PAGE and western blotting.

2.11 Whole cell extract preparation

Whole cell extracts (WCE) were prepares from cultured cells using UTB buffer (8 M urea, 50 mM tris, 150 mM β-mercaptoethanol). Briefly, approximately 0.5 x 10^7 were trypsinsed, washed in ice cold PBS, and resuspended in 150 µl UTB. Lysates were sonicated for 2 cycles at max power for 5 seconds on ice, and protein concentrations determined by Bradford assay.

2.12 Sodium-dodecyl sulphate polyacrylamide (SDS-PAGE) gel electrophoresis

GST pull downs, immunoprecipitates or whole cell extracts were analysed by SDS-PAGE. Briefly 50 µg of WCE, or the entirety of GST pull downs or immunoprecipitates were analysed on 6% Tris-bicine gels in 1 X Tris-bicine buffer (1 M of Tris/bicine, 10 % SDS) for 4 hours at 25 mA. Separated proteins were visualized by staining Coomassie stain (0.1 % (w/v) Brilliant Blue G, 25 % (v/v) methanol and 5 % (v/v) acetic acid) (Sigma Aldrich, UK) followed by destaining (10 % of acetic acid and 10 % methanol) or transformed to nitrocellulose membrane for subjected to immunoblotting.

2.13 Western Blotting/ Immunoblotting

For immunoblotting, separated polypeptides were transferred onto nitrocellulose membrane (GE Healthcare). Transferred proteins were visualized by Ponceau S stain for five minutes to determine equal loading or the presence of GST fusion proteins. Membranes were blocked with 5 % milk in with 1% TBST [(10 % TBS- 24 g Tris base, 88 g of NaCl, 900 ml of distilled water) 9 L of distilled water, 0.1% tween] then incubated with appropriate primary antibodies at 4 °C overnight (see section 2.3, table 2.1). Membranes washed with three times with TBST, exposed to HRP-coupled secondary antibodies for 1 h at room temperature (see section 2.3, table 2.2) and further incubated with secondary antibodies (table 2.2) and finally washed three times with TBST. Proteins were detected by chemiluminescence ECL (GE healthcare) or hypersensitive ECL- Immobilon (Millipore).

2.14 Immunofluorescence Staining (IF)

For immunofluorescence staining, cells were grown on glass coverslips ($0.5x10^5$ cells per coverslip), Cells were fixed in 4 % paraformaldehyde for ten minutes at room temperature, permeabilised with extraction buffer ((20 mM HEPES pH 7.5, 20 mM NaCl, 5 mM MgCl2, 1 mM DTT, 0.5% IGEPAL (A-630)) and blocked in 10 % of FCS in PBS for at least for one hour at 4 °C. Alternatively for foci analysis, cells were permeabilised with extraction buffer for five minutes on ice, then fixed in paraformalaldehyde and blocked as above. Coverslips were incubated with primary antibodies for 1 h at room temperature (mouse anti-FLAG SET1A and rabbit anti-Rad51; diluted in PBS), followed by three PBS washes and incubated with appropriate secondary antibodies (anti-rabbit FITC in 1:500 and anti-mouse 595 in 1:500) for a further hour in dark at room temperature. Coverslips were washed as above, rinsed with distilled water and mounted onto glass slides with Vectashield (Dako)

containing 4, 6-diamidino-2-phenylindole (DAPI). Cells were visualized by fluorescence microscopy using a Nikon Eclipse E600 (Nikon) with a 60 x or 100 x oil immersion objectives in conjunction with a Hamamatsu C474295 camera. Images were captured using Velocity software (Perkin Elmer).

2.15 Bioinformatics analyses

COSMIC (Catalogue Of Somatic Mutations in Cancer) (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/) and cbioportal (http://www.cbioportal.org/public-portal/) databases were interrogated to identify substitution mutations in PFAP associated with different types of cancer. The impact of these mutations was predicted using Polyphen (http://genetics.bwh.harvard.edu/pph2/) and Mutation Taster (www.mutationtaster.org/). Sequence alignments of human PFAP amino acid sequences were performed using Clustal W (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Table 2.3: Accession numbers of sequences used in this study

Sequences	**Accession numbers**
Biorientation of chromosomes in cell division protein 1-like 1 [*Homo sapiens*]	NP_683692
Biorientation of chromosomes in cell division 1-like [*Mus musculus*]	NP_001074891
Biorientation of chromosomes in cell division 1-like [*Gallus gallus*]	XP_420784
Biorientation of chromosomes in cell division 1-like *Macacca mulatta*	XP_001098928
Biorientation of chromosomes in cell division 1-like *Gorilla gorilla*	XP_004038496
Biorientation of chromosomes in cell division 1-like *Xenopus (Silurana) tropicalis*	XP_002940788

3.0 Results

3.1 PFAP interacts with SET1A in vitro

Previous publications have suggested that PFAP interacts with SET1B, but not with SET1A (Nuland *et al.,* 2013). However, data from our laboratory suggests that PFAP and SET1A and not SET1B, act together to protect cells from DNA damage (data not shown). Therefore, we hypothesize that PFAP (BOD1L) interacts with SET1A histone methyltransferase complex. To investigate this hypothesis, we examined whether PFAP interacted with SET1A or SET1B by immunoprecipitation (Figure 3.1).

SET1A or SET1B were immunoprecipitated from HeLa nuclear cell extracts using specific commercially available antibodies and protein A beads. Protein A beads will bind to the immune complexes and pulled down specific antibodies/ antigen complex. As a negative control, an IgG immunoprecipitation was included. The resulting immune complexes were washed, proteins were separated by SDS-PAGE and analysed by western blotting using SET1A, SET1B and PFAP antibodies. In parallel, a fraction of HeLa nuclear cell extract was also analysed as input, to show the presence of protein targets that are not immunoprecipitated (Figure 3.1).

Endogenous SET1A and SET1B were successfully immunoprecipitated by their respective antibodies but not by IgG, and migrated similar distances to SET1A and SET1B as visible in the input. However, cross-reactivities of immunoblotting SET1A and SET1B

antibody resulted in the detection of both endogenous SET1A and SET1B proteins in SET1A and SET1B immunoblots (Figure 3.1 panel A, B).

In addition, immunoblotting of PFAP revealed that, whilst PFAP was present in the input sample and immune complexes using the SET1A antibody, it was absent from immunoprecipitates of SET1B (Figure 3.1 panel C). Reciprocal experiments performed in parallel in our laboratory confirmed that immunoprecipitates isolated with a PFAP-specific antibody contained SET1A but not SET1B (Figure 3.2). These immunoprecipitation analyses establish that PFAP interacts with histone methyltransferase SET1A *in vitro* but not with SET1B.

3.2 Identification of the SET1A binding region of PFAP

Since our initial experiments strongly suggested that PFAP interaction with SET1A *in vitro*, we therefore, wished to identify the region of PFAP necessary for this interaction. Since PFAP is a large protein of over 3000 amino acids, therefore, we first created several fragments of PFAP, each encoding roughly 500-600 amino acids, with the aim of using these to identify regions of PFAP important for interactions.

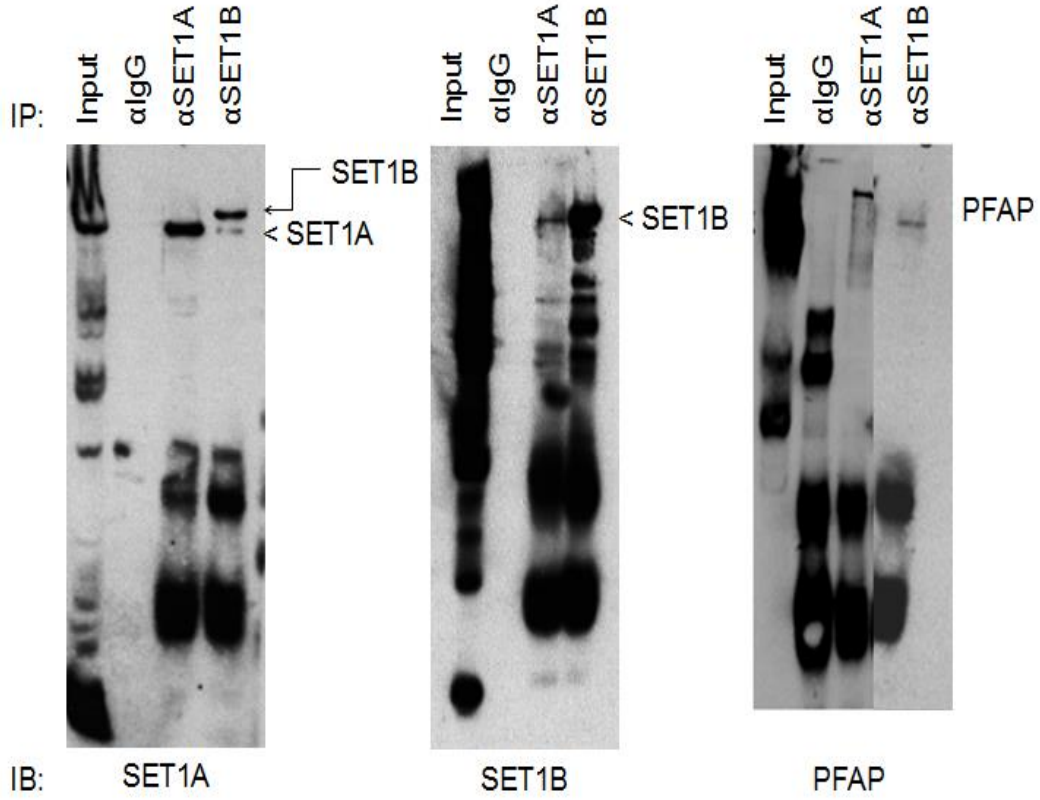

Figure 3.1: Analyses of *in vitro* interactions between PFAP and SET1A/SET1B by co-immunoprecipitation. HeLa nuclear extracts were immunoprecipitated with either anti-SET1A or anti-SET1B antibodies. IgG was used as a negative control. Antibody-bound proteins were collected with Protein A beads, immune complexes washed and separated by SDS-PAGE, and then analyzed by Western blotting. Blots were probed with anti-SET1A (panel A), anti-SET1B (panel B) and anti-PFAP (panel C) antibodies.

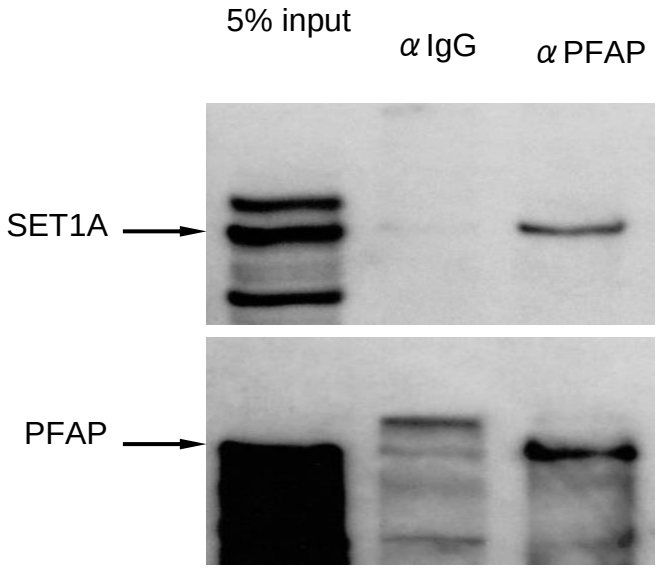

Figure 3.2: Analysis of *in vitro* interactions between PFAP and SET1A by co-immunoprecipitation. HeLa nuclear extracts were immunoprecipitated with anti-PFAP antibody or IgG. Immune complexes were analysed as in Figure 3.1 using anti-SET1A or anti-PFAP antibodies. Data courtesy of Dr. M. Higgs.

3.2.1 Constructions of GST-tagged PFAP fragments

As mentioned above, six overlapping of PFAP (previously constructed in our laboratory), termed fragments 1-6 (F1-F6) were available (see figure 3.3 for a schematic). These fragments also encoded NotI and SalI restriction sites at their termini to aid cloning (Figure 3.4). These six fragments were cloned into the pGEX-3X vector, downstream of glutathione sepharose transferase (GST) as shown in figure 3.4. To insert these fragments, pGEX 3X vector was digested using NotI and SalI (Figure 3.4). This enabled the IPTG-inducible expression of these fragments as GST-tagged fusion proteins (Figure 3.5).

3.2.2 Small-scale expression of GST-tagged PFAP fragments

To express the fusion proteins described above, the plasmids encoding the PFAP fragments, plus an empty (pGEX-3X) vector used as a negative control, were transformed into BL21 *Escherichia coli.* Single colonies of transformed cells were picked and grown into starter cultures using Luria broth (LB) media containing ampicillin in the absence (uninduced) or in the presence of IPTG. GST-fusion expressions were induced by adding isopropyl-β-D-thiogalactoside (IPTG). Expression of these GST-fusions were monitored by SDS-PAGE, subjected to Brilliant coomassie blue stain and visualized by autoradiography. In parallel, a fraction of uninduced culture was used as an uninduced control (Figure 3.6).

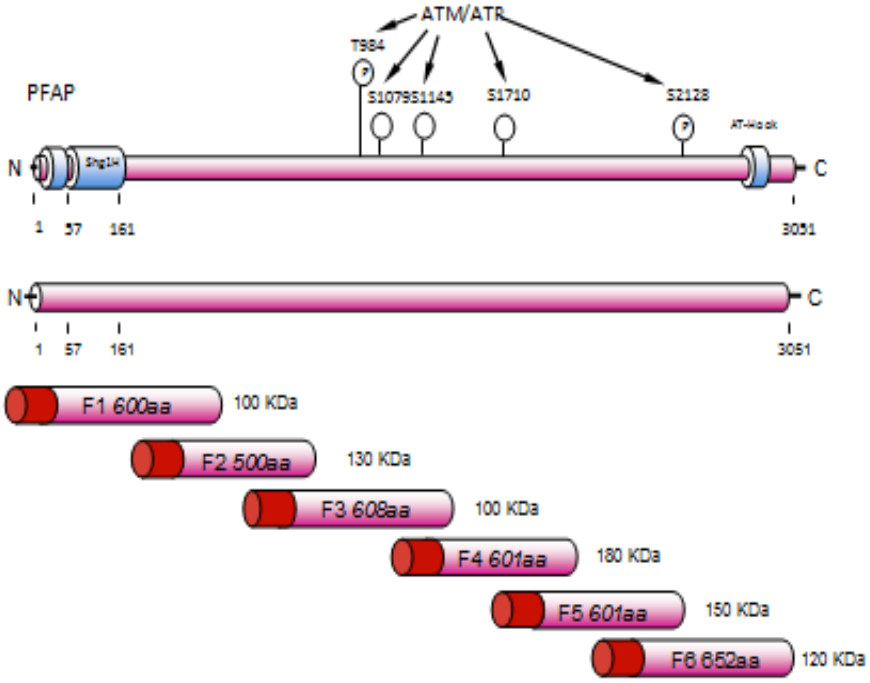

Figure 3.3: Schematic representation of PFAP and its GST-tagged fragments. *(Upper panel)* The positions of the known ATM/ATR phosphorylation sites within PFAP, as well as the identifiable domains, are indicated. (Lower panel) Approximate positions within PFAP, sizes in aa, and molecular masses of the six PFAP fragments. GST tags are denoted in red.

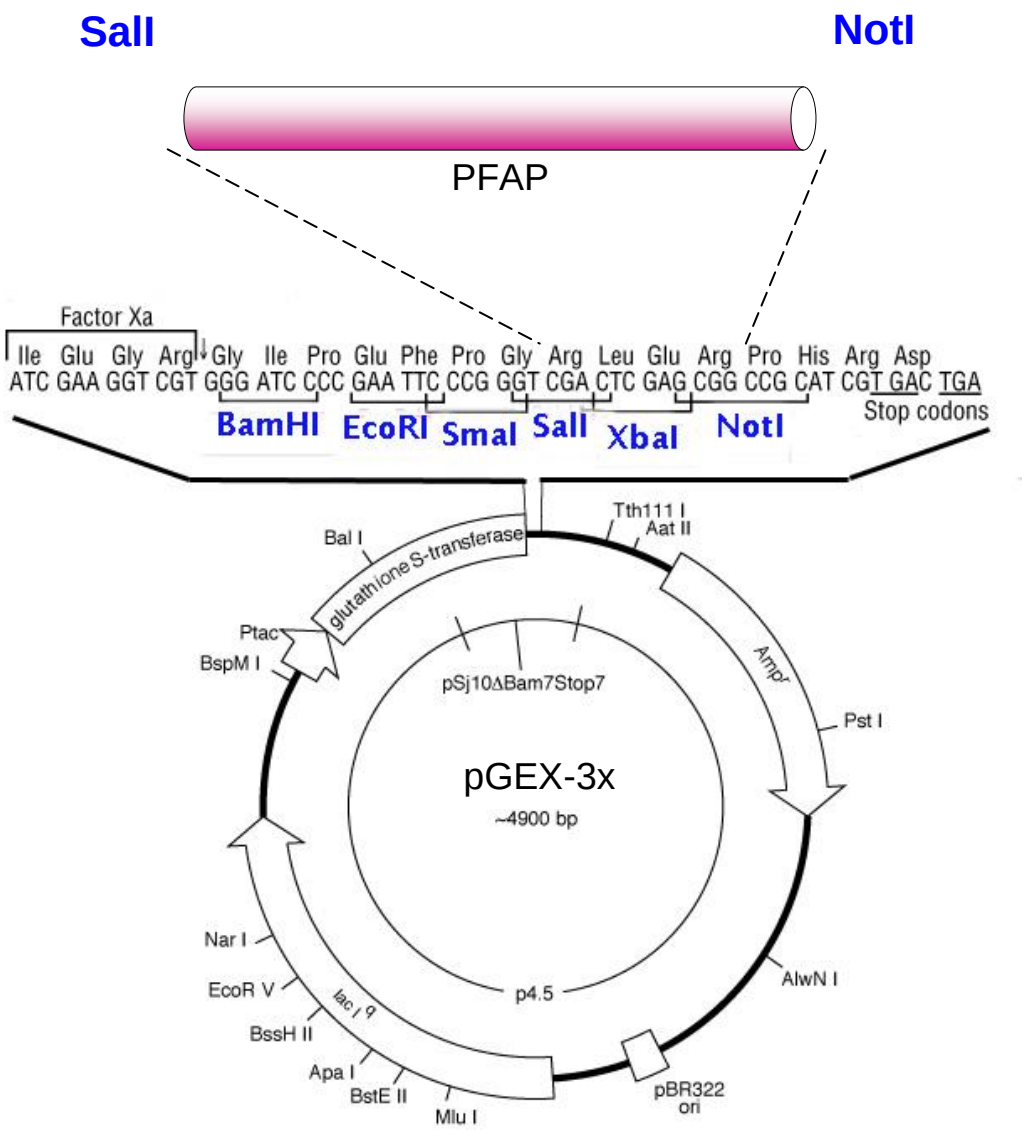

Figure 3.4: Cloning strategy to produce GST-tagged PFAP fragments. PFAP fragments were amplified from human cDNA using primers containing SalI or NotI restriction sites (not shown). Both the pGEX-3x vector and the amplified PFAP fragment were digested using SalI and NotI restriction enzymes, and PFAP fragments were ligated into pGEX-3x upstream of glutathione sepharose transferase.

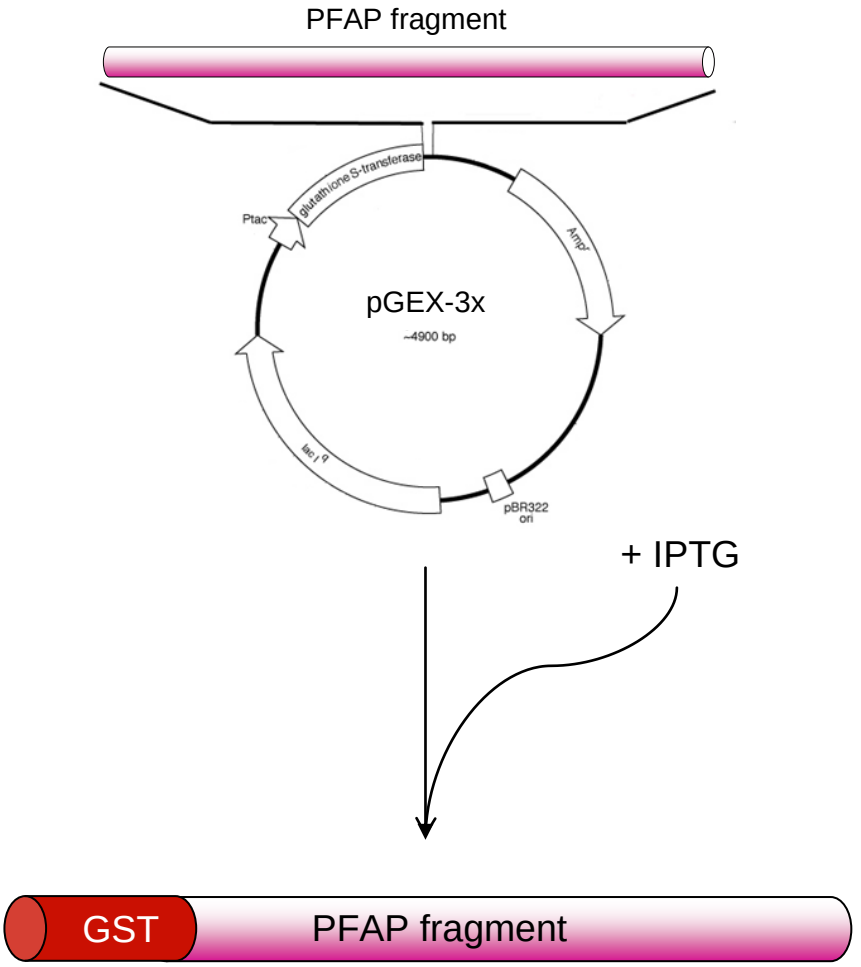

Figure 3.5: Strategy for expressing GST-tagged fragments of PFAP. Plasmids encoding PFAP fragments, or empty vector controls, were transformed into BL21 *E. coli (*not shown*)*. Addition of IPTG to the bacterial growth media stimulates transcription from the pGEX-3x promoter, allowing the expression of the tagged fragment.

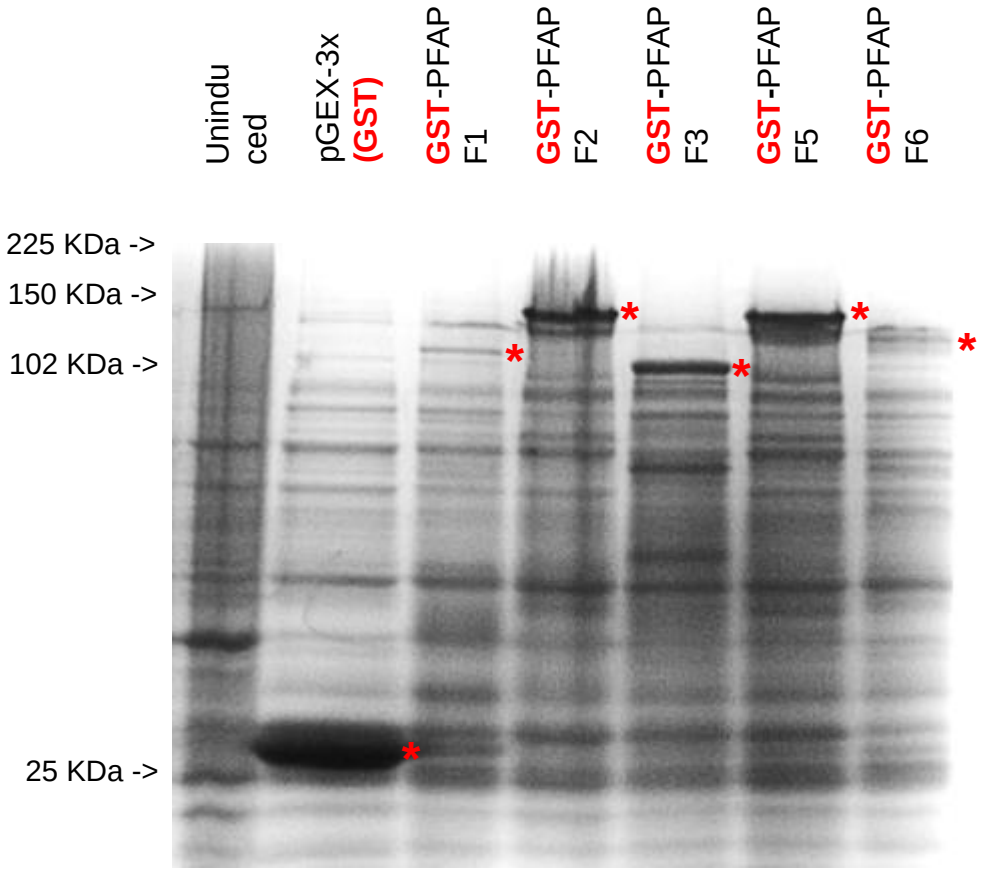

Figure 3.6: Small scale expression of GST-tagged fragments of PFAP. Plasmids encoding PFAP fragments, or empty vector controls, were transformed into BL21 *E. coli*. Single colonies were picked and grown in small 'starter cultures' in the absence (uninduced) or presence of IPTG. Bacterial pellets were collected 3 hours after induction, lysates prepared, and separated by SDS-PAGE. Gels were coomassie stained to reveal polypeptides. Induced proteins are denoted with an * and the position of molecular weight markers is indicated.

SDS-PAGE results shows that F1, F2, F3, F5, F6 and empty vector (negative control) were induced by IPTG as these fragments were expressed at their respective molecular weight, indicated by asterisks (Figure 3.6, lanes 3-8) and also shown in figure 3.3. On the other hand, uninduced culture was expressed at roughly 48KDa (molecular weight of PGEX 3X vector) (Figure 3.6, lane 2).

3.2.3 *in vitro* protein-protein interaction of the F1 fragment of PFAP

Amino terminus (N-terminal) of PFAP fragment contains sequences (55-164aa) that are homologous to BOD1 and yeast Shg1 (Shg1p) proteins. BOD1 (Bi-orientation defected 1), is a protein important for "chromosome segregation during mitosis" and belongs to the same Fam44 protein family as PFAP (Porter *et al*., 2007). On the other hand, Shg1 protein is a part of SET1C COMPASS histone methyltransferase complex in yeast, and is said to be functionally analogous to SET1 complex in humans (Nuland *et al*., 2013). Therefore, these sequences at the N-terminal of PFAP are termed as Shg1H domain. Shg1H domain is located in F1 fragment (Figure 3.7, upper panel). Thus, in vitro interaction of GST-tagged F1 fragment and SET1A was investigated, using GST-tagged F2 fragment as a negative control in some experiments.

3.2.4 Large scale purification of fragments F1 and F2

Protein purification was carried out to quantify the large amounts of F1 and F2 fragments to analyse protein interactions by GST pull down assays. Bacterial pellets of expressed F1 and F2 fragments were collected and lysates were prepared and passed over glutathione beads. Any unbound proteins were washed with excess glutathione and the purified proteins were eluted using elution buffer and dialysed. Purified proteins were separated by SDS-PAGE and subjected to brilliant coomassie blue stain. GST (empty pGEX 3X vector) was purified as a negative control. Purified protein fragments F1, F2 and GST indicated by asterisk in figure 3. 7.

3.2.5 PFAP F1 but not F2 interacts with SET1A

To analyse whether fragments F1 or F2 were sufficient to bind to SET1A in vitro, we incubated GST alone, GST-F1, GST-F2 with HeLa nuclear cell extracts. Protein complexes containing the GST-fusion proteins were then isolated using glutathione sepharose beads and the resulting complexes were washed and analysed by SDS-PAGE and western blotting with anti-SET1A and anti-SET1B antibodies.

Immunoblotting of GST pull down demonstrated that only SET1A and not SET1B was pulled down with F1 fragment of PFAP (Figure 3.8, lower panel). On the other hand, both SET1A and SET1B were not pulled down F2 fragment (Figure 3.8, lower panel).

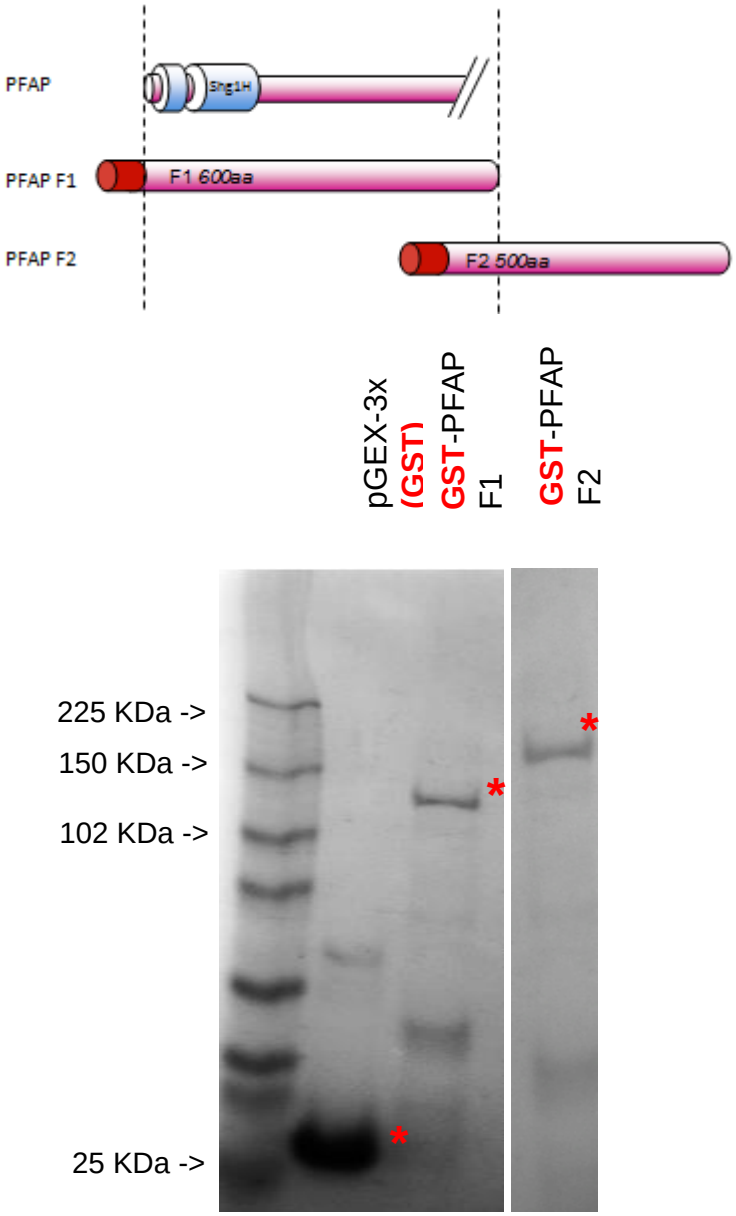

Figure 3.7: Purification of GST-F1 and F2 fragments of PFAP. (*Upper panel*) Schematic showing the position of the Shg1H domain of PFAP in relation to fragments F1 and F2. *(Lower panel*) Bacteria containing pGEX-3x, pGST-F1 or pGST-F2 were grown in large scale cultures, induced with IPTG, and lysates prepared. GST-tagged proteins were isolated from these lysates using Glutathione beads, and the bound proteins eluted with excess glutathione. Twenty-five microlitres of the purified proteins was analysed by SDS-PAGE and coomassie staining to ensure purification of the appropriate GST-tagged polypetide.

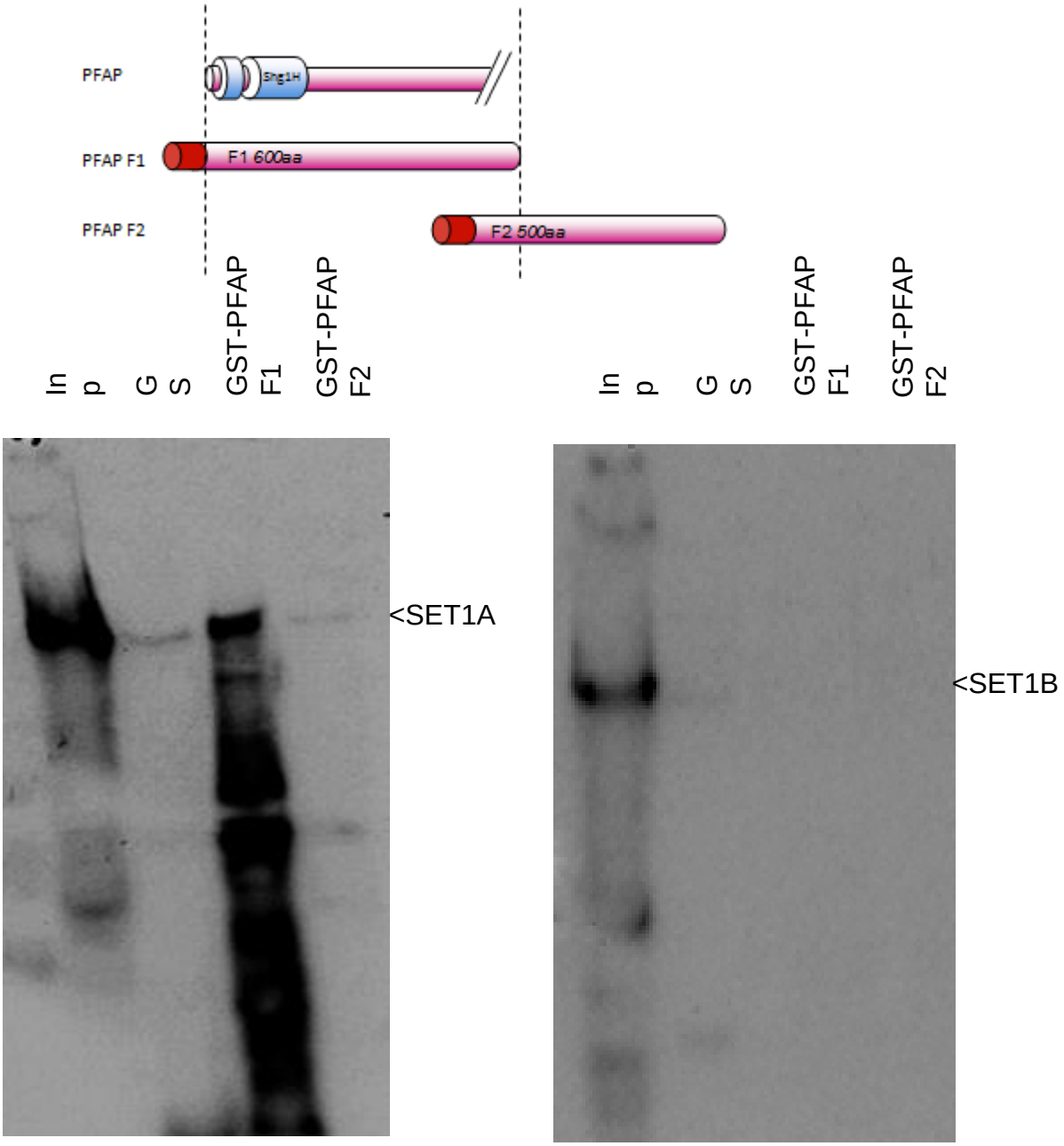

Figure 3.8: GST pulldown analyses of interactions between PFAP fragments and SET1A/SET1B. (*Upper panel*) Schematic showing the position of the Shg1H domain of PFAP in relation to fragments F1 and F2. (*Lower panel*) GST pulldown assays were performed as in the legend to Fig. 3.8. Protein complexes were analysed by Western blotting in conjunction with (A)anti-SET1A or (B) anti-SET1B. The presence of recombinant proteins in the isolated complexes was confirmed by coomassie staining (not shown).

GST pull down analyses suggests that endogenous SET1A protein interacts with F1 fragment and not F2 fragment of PFAP. On the other hand, endogenous SET1B protein is not interacting with either of the PFAP fragments. This confirms our hypothesis that only SET1A interacts with PFAP, possibly at the predicted Shg1H domain region of F1 fragment.

3.4 Co-localisation of FLAG-SET1A and Rad51

In response to mitomycin C, many DNA damage proteins such as Rad51, protein important for the homologous recombination relocalises within the nucleus to form distinct foci, representing sites where DNA repair takes place. These distinct nuclear foci are visualized by fluorescence microscopy. Since, it is known that SET1A is involved in the DNA repair. Therefore, we decided next to investigate co-localisation of SET1A with RAD51, to examine whether SET1A localises to the sites of ICL repair.

HeLa cell line expressing FLAG-tagged SET1A domain can express FLAG-SET1A when induced with doxycycline. FLAG-tagged version of SET1A was used instead of normal SET1A because the SET1A specific antibodies that we used for immunoblotting do not detect SET1A by immunofluorescence.

3.4.1 FLAG-tagged SET1A does not co-localise with Rad51 after MMC in HeLa cells

To investigate whether FLAG-tagged SET1A and Rad51 co-localised, HeLa cells expressing FLAG tagged SET1A were induced with doxycycline or left uninduced (doxycycline negative). Both induced and uninduced cells were either treated with 50 ng/ ml of mitomycin C (MMC) after 24 hours or left untreated. Cells were then fixed and stained with anti- rabbit Rad51 and anti-mouse FLAG tagged SET1A antibodies. The cells were examined by fluorescence microscopy (Nikon E600) (Figure 3.9).

In the absence of doxycycline, we would expect no FLAG-SET1A expression. However, FLAG-SET1A was expressed in the absence of doxycycline, in a pan-nuclear fashion. On the other hand, Rad51 was expressed, but did not form foci in the absence of MMC exposure. In doxycycline positive and MMC negative cells, we observed similar results, i.e. expression of both Rad51 and FLAG-SET1A and a very few distinct foci were visible. In cells treated with MMC after 24 hours, nuclear speckles were visible for FLAG-SET1A, and many distinct visible foci were formed by Rad51 (Figure 3.9). Non-specific nucleolar signals with Rad51 antibody which are not foci were observed in Rad51 panels in doxycycline induced and uninduced cells in the absence of MMC (Figure 3.9).
However, no co-localisation observed as both FLAG-SET1A and Rad51 exhibited non-overlapping sub-nuclear localization, suggesting both proteins were targeted to different genomic sites.

To confirm that FLAG-SET1A was expressed in the absence of doxycycline, we performed immunoblotting of HeLa FLAG SET1A using SET1A and FLAG antibodies in

cultures. In agreement with the immunofluorescence data, endogenous SET1A was weakly expressed in both doxycycline induced and uninduced cells, as expected (Figure 3.10). However, we also observed expression of the FLAG- tagged version of SET1A in both the absence and presence of doxycycline, suggesting the HeLa cells were 'leaky' i.e. that control of the exogenous FLAG-SET1A gene by the doxycycline-inducible promoter had somehow been lost. Therefore, we decided to use an alternative U2OS cell line expressing FLAG-SET1A.

3.4.2 FLAG-tagged SET1A does not co-localize with Rad51 after MMC in U2OS cells

3.4.2a FLAG-tagged expression in U2OS cells with PFA only

Prior to studying the potential co-localisation of SET1A and Rad51, we wanted to check that the cells were expressing FLAG-SET1A in a regulated, doxycycline- inducible fashion, and that it could be easily detected by IF. Therefore, U2OS cell lines expressing FLAG-SET1A under a doxycycline-inducible promoter were fixed with PFA (Paraformalaldehyde) only, before permeabilising the cells. This allowed the stabilisation of both soluble and insoluble FLAG-SET1A in its native location. Such experiments illustrated that FLAG-tagged SET1A was expressed in the nucleus of these cells only when induced with doxycycline, and that the FLAG-tagged SET1A does not express in the absence of doxycycline (Figure 3.11).

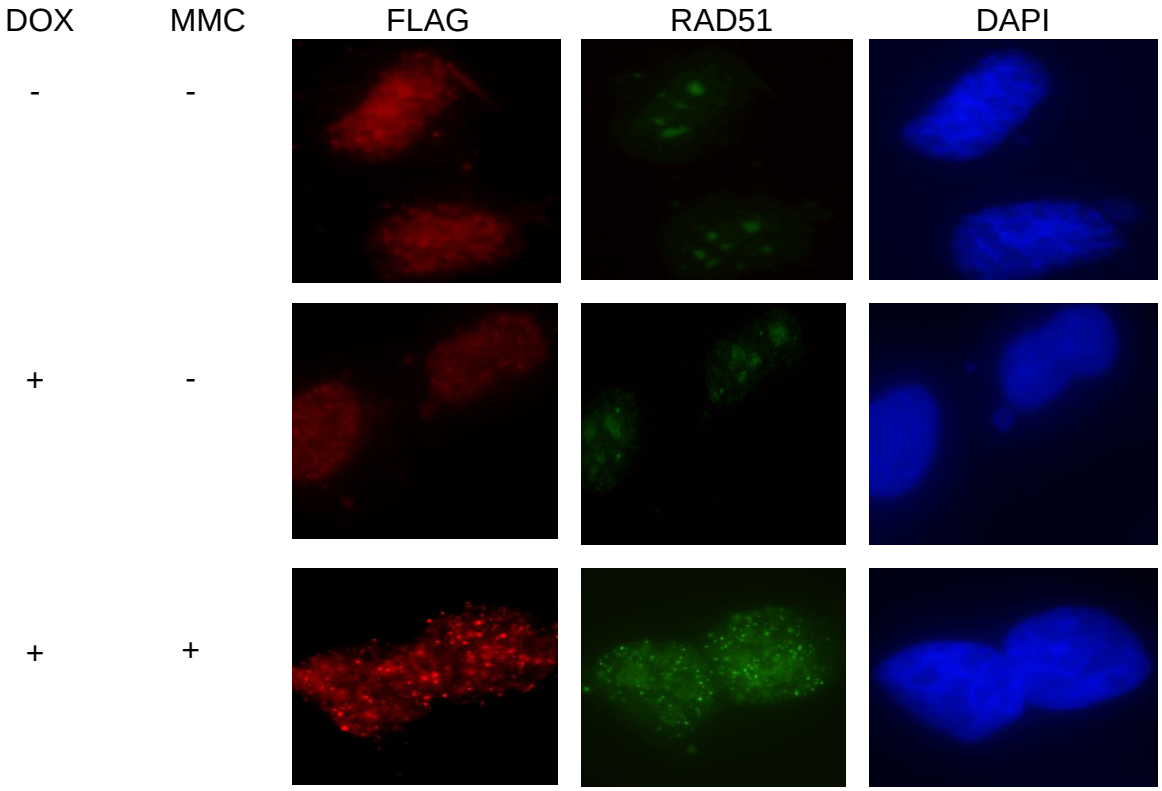

Figure 3.9: Immunofluorescence analyses of FLAG-SET1A localisation in HeLa cells. HeLa-FLAG-SET1A cells were cultured on glass coverslips, and treated with doxycycline or left untreated. After 24 h, cells were exposed where indicated to 50 ng/ml MMC for 24 h. Cells were pre-extracted, fixed and FLAG-SET1A and Rad51 detected by immunofluorescence using the appropriate antibodies. Anti-rabbit FITC and anti-mouse Alexa 595 were used as secondary antibodies. Nuclei were counter stained with 4,6-diamidino-2-phenylindole (DAPI) and cells observed using a fluorescence microscope (Nikon).

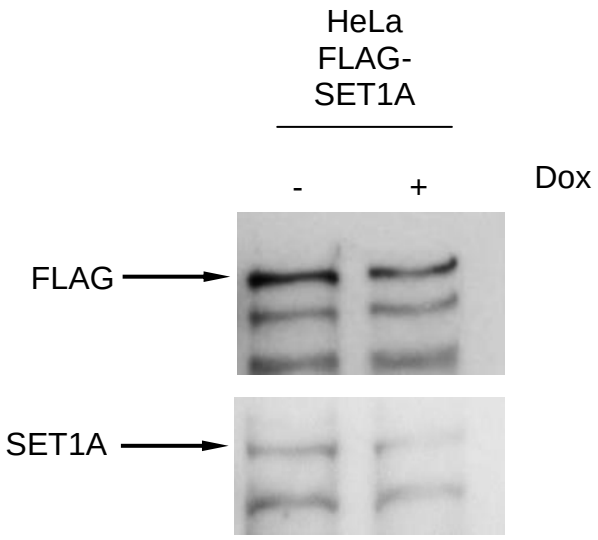

Figure 3.10: Immunoblot analyses of HeLa-FLAG-SET1A cells. HeLa-FLAG-SET1A cells were treated with doxycycline or left untreated for 48 h. Whole cell extracts were prepared and analysed by immunoblotting using antibodies to SET1A and FLAG. Data courtesy of Dr M.Higgs

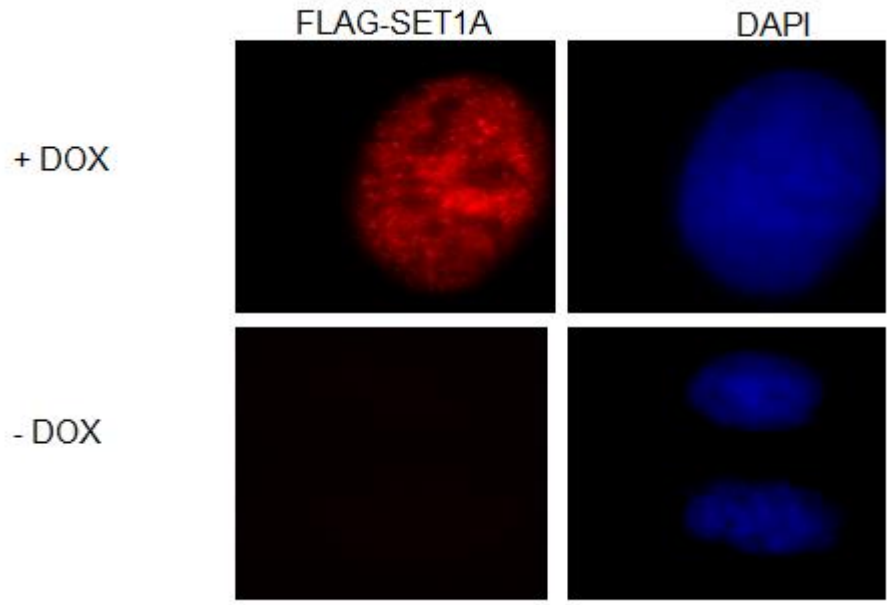

Figure 3.11: Immunofluorescence analyses of FLAG-SET1A localisation in U2OS cells. U2OS-FLAG-SET1A cells were cultured on glass coverslips, and treated with doxycycline or left untreated. Cells were fixed, permeabilised, and FLAG-SET1A detected by immunofluorescence using the appropriate antibody. Anti-mouse Alexa 595 was used as a secondary antibody. Images were captured as in Fig. 3.10. Data courtesy of Dr. M.Higgs.

3.4.1b U2OS cells treated with MMC

To examine the co-localisation of FLAG-SET1A and Rad51 in these cells, this time around cells were permeabilised using extraction buffer prior to fixing with PFA. This allowed the removal of soluble proteins from the cells, thus leaving only chromatin-bound structures that can be examined by fluorescence microscope.

Immunostaining of cells were carried out in the absence and presence of doxycycline, before and after exposure to MMC for 24 hours or 48 hours (Figure 3.12). Figure 3.12 indicates that, in the absence of both doxycycline and MMC, FLAG-SET1A was not expressed, and Rad51 forms few or no distinct foci. However, in MMC treated U20S cells after 24 hours and 48 hours in the absence of doxycycline, Rad51 foci were formed. On the other hand, in the presence of doxycycline cells, without MMC treatment, FLAG-SET1A was expressed, but formed no distinct foci. In contrast, Rad51 formed few visible foci in the presence of doxycycline cells without MMC treatment. Exposure of doxycycline-treated cells to MMC after 24 hours or 48 hours, FLAG-SET1A was expressed but formed no distinct foci, whilst Rad51 formed more distinct foci (Figure 3.12). These data confirms that although both FLAG-SET1A and Rad51 reveals distinct sub-nuclear localisation (Figure 3.12, merged panel), but the formation of Rad51 foci in FLAG-tagged SET1A cells, in the presence of MMC suggests that SET1A is involved in interstrand crosslink repair.

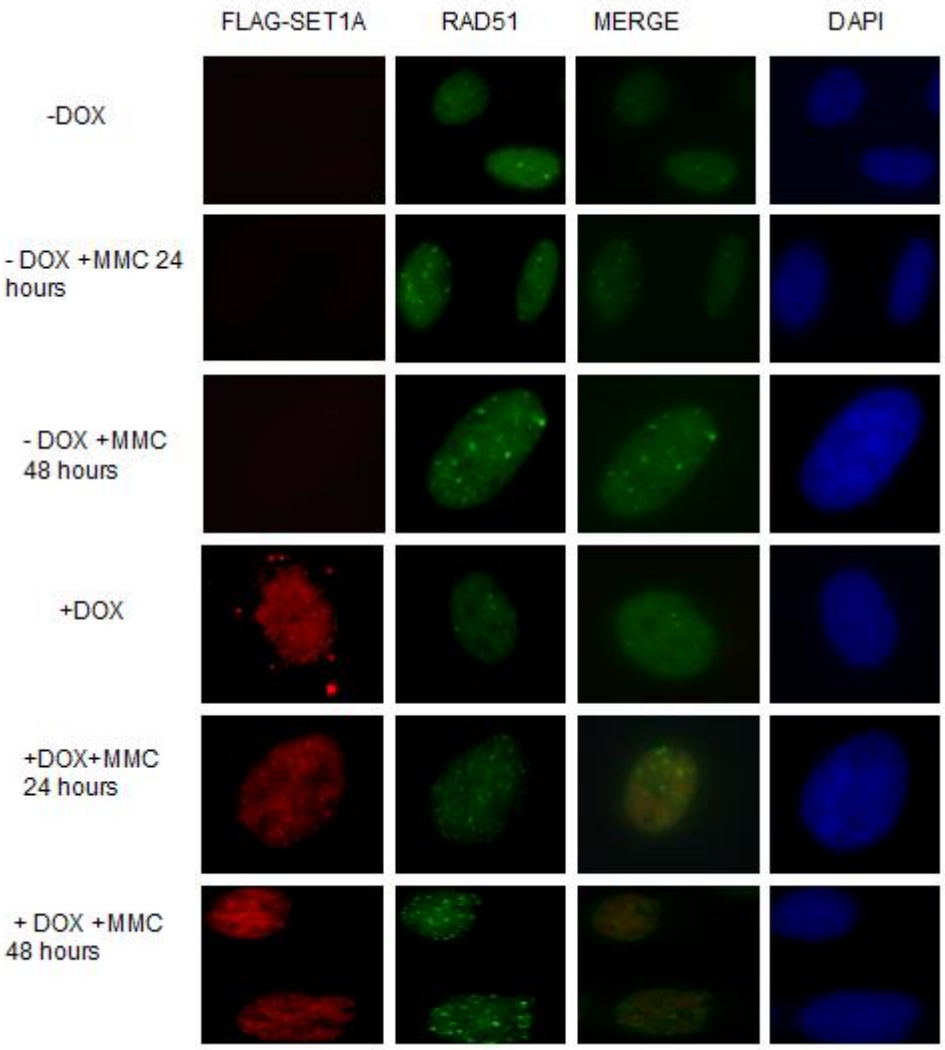

Figure 3.12: Immunofluorescence analyses of FLAG-SET1A localisation in U2OS cells. U2OS-FLAG-SET1A cells were cultured and treated as in Fig. 3.11. Twenty-four hours post addition of doxycycline, cells were exposed where indicated to 50 ng/ml MMC for a further 24 or 48 h. Cells were pre-extracted, fixed, and FLAG-SET1A and Rad51 detected by immunofluorescence as in Fig. 3.10.

3.5 Interaction of PFAP with other fanconi anaemia proteins

Our preliminary data suggests that PFAP is a component of the Fanconi anaemia pathway in particular acting downstream of the mono-ubiquitinated FANCD2-FANCI complex (data not shown). Therefore, we decided to examine PFAP interactions with Fanconi anaemia proteins such as FANCD2, FANCJ and FANCM.

PFAP was immunoprecipitated from HeLa nuclear extracts and the immune complexes were blotted for the presence of PFAP, FANCD2, FANCJ and FANCM. As a positive control, we also examined the immunoprecipitates for the presence of SET1A, which we have previously demonstrated to interact with PFAP by immunoprecipitation (see section 3.1). PFAP immunoprecipitated SET1A as expected (Figure 3.13). Surprisingly, FANCD2 and FANCJ was co-immunoprecipitated with PFAP, but was absent from immunoprecipitates from IgG (Figure 3.13, top panel and bottom left panel). Two bands were observed in co-immunoprecipitation of PFAP and FANCJ. The band detected at 140 KDa is the FANCJ band (Figure 3.1- bottom left panel). In contrast, FANCM was not detected by anti-FANCM antibodies in the input, or in the immunoprecipitates (Figure 3.13, lower right panel).

Although the above data suggests that PFAP interacts with FANCD2 and FANCJ in vitro, further work is required to confirm my observations. Unfortunately, time constraints precluded me from carrying out this work.

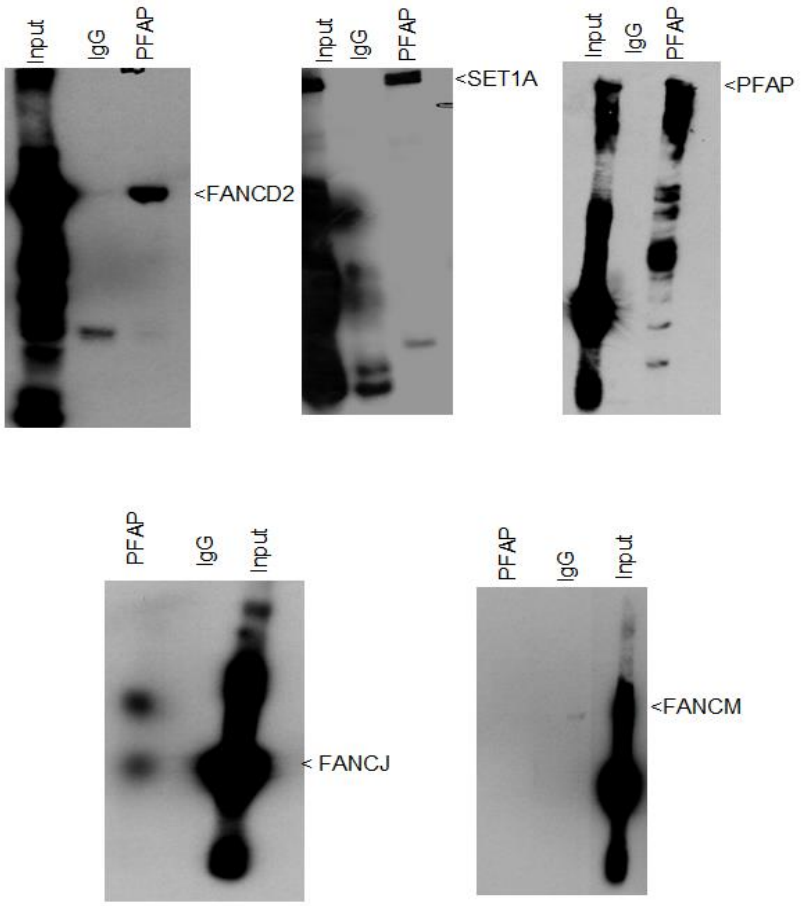

Figure 3.13: Analyses of *in vitro* interactions between PFAP and Fanconi anaemia proteins by co-immunoprecipitation. HeLa nuclear extracts were immunoprecipitated with either anti-PFAP antibodies. IgG was used as a negative control. Antibody-bound proteins were collected with Protein A beads, immune complexes washed and separated by SDS-PAGE, and then analyzed by Western blotting. Blots were probed with anti-FANCD2 (panel A), anti-

3.6 Cancer associated mutations might affect interaction of PFAP and SET1A

Since both PFAP and SET1A are involved in maintaining genetic stability, we hypothesized that the interaction of SET1A with PFAP may be disrupted by mutations of PFAP in tumorigenesis, and that this maybe the cause of genetic instability in these tumours. Therefore, we conducted screening of cancer-associated mutations in PFAP using COSMIC (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/) (Bamford *et al.*, 2004) and cbioportal (http://www.cbioportal.org/public-portal/) (Gao *et al.*, 2013; Cerami *et al.*, 2012) databases. Over two hundred cancer-associated mutations were spread throughout the length of PFAP were found in both databases (Figure 3.14)

In particular mutations within the F1 fragment were studied further, as the N-terminal region of F1 fragment of PFAP contains shg1H domain that we hypothesize to interact with SET1A. There were twenty one mutations found in F1 fragment of PFAP (Figure 3.15). Of these twenty one mutations, approximately one third of the mutations were located in the Shg1 domain of F1 fragment (Figure 3.16). Most of these were substitution mutations, with frameshift mutations at amino acid positions 137 (in the Shg1 domain) 335, 494, and 519 (rest of the F1 fragment of PFAP) (Figure 3.15-3.16).

In order to examine whether the identified mutations were likely to be disease-causing or deleterious to the function of PFAP, we used two prediction programs, Polyphen and Mutation Taster (Schwarz *et al.*, 2014), to predict the effect of the various amino acid substitutions identified in the F1 fragment of PFAP (Table 3.2).

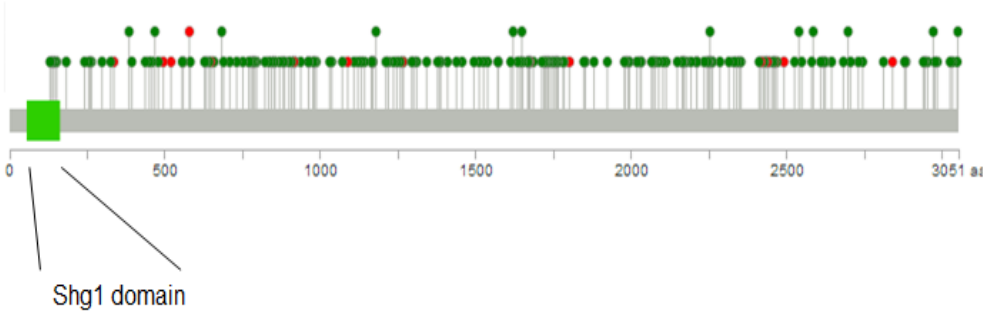

Figure 3.14: Schematic representation of cancer-associated mutations in PFAP. Mutations in PFAP were identified using the COSMIC **(Bamford *et al.,* 2004)** and CBioportal databases **(Gao *et al*. *Sci Signal*. 2013; Cerami *et al*. *Cancer Discov.* 2012)**. The positions and nature of the mutations are indicated: substitution missense mutations are in green, and non-sense or frameshift deletions are in red. The position of the Shg1H domain is indicated.

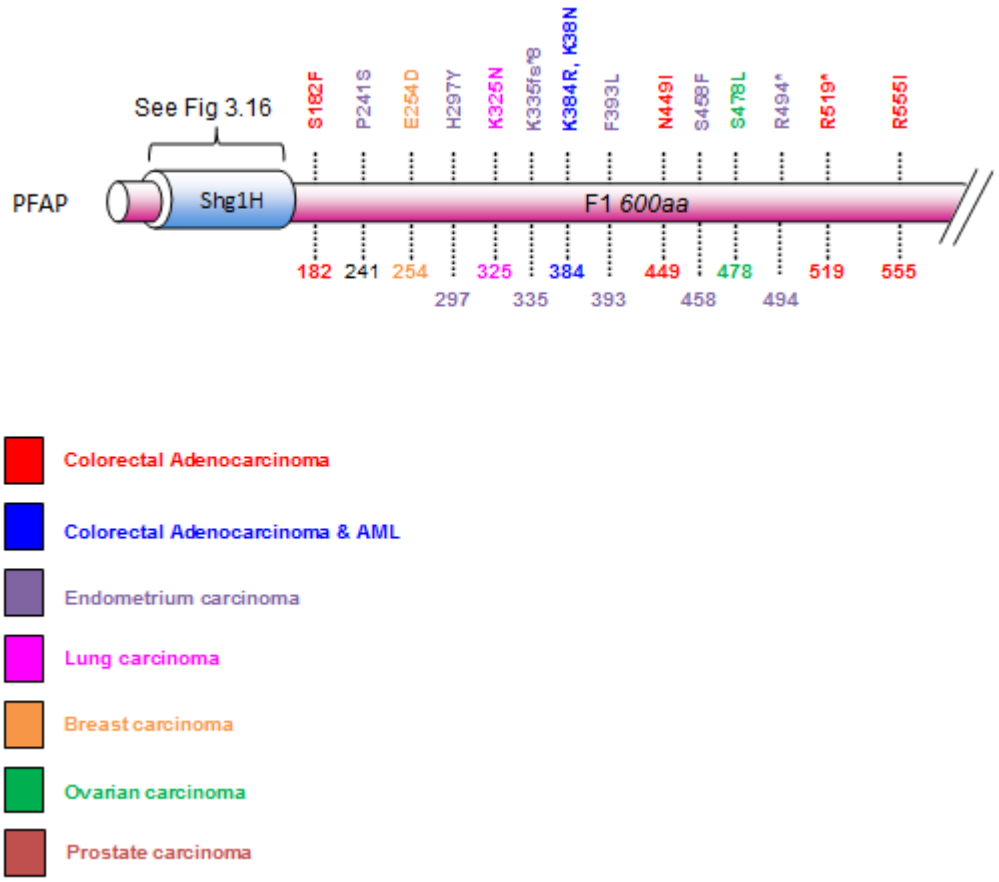

Figure 3.15: Schematic representation of cancer-associated mutations in the F1 fragment of PFAP. Mutations in the F1 fragment of PFAP (aa 1-600) were identified using the COSMIC database. The positions (in aa) and nature of the mutations is indicated, and the different colours denote the origin of the tumours in which they were found (*lower panel*).

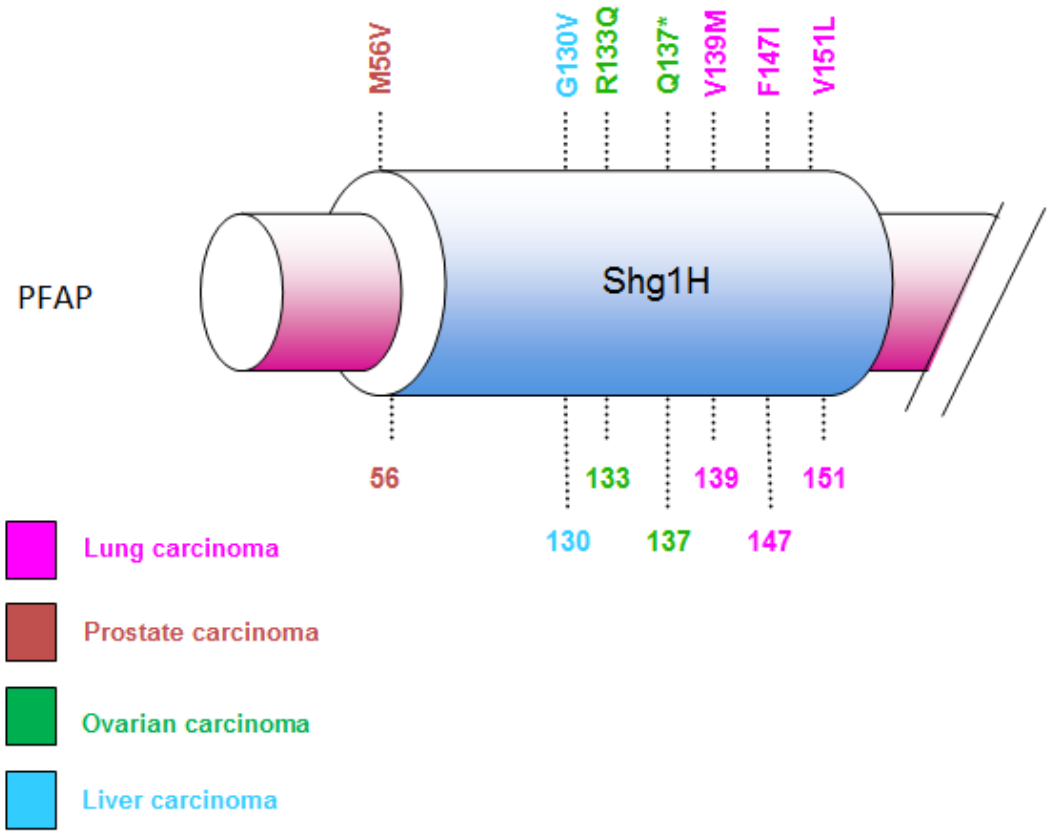

Figure 3.16: Schematic representation of cancer-associated mutations in the Shg1 homology domain of PFAP. Mutations in the Shg1H domain of PFAP (aa 55-155) were identified using the COSMIC database. The positions (in aa) and nature of the mutations is indicated, and the different colours denote the origin of the tumours in which they were found (*lower panel*).

Table 3.1: List of deleterious mutations within the Shg1 domain of PFAP. **(Adapted from Adzhubei *et al*., 2010; Schwarz *et al*., 2014).**

Mutations	**Cancer**	**Benign or deleterious mutations in Polyphen**	**Benign or deleterious mutations in Mutation Taster**
M56V	Prostate carcinoma	Benign	Benign
G130V	Liver hepatocellular carcinoma	Deleterious or disease causing	Deleterious or disease causing
R133Q	Ovarian carcinoma	Deleterious or disease causing	Deleterious or disease causing
Q137*	Ovarian carcinoma	Deleterious or disease causing	Deleterious or disease causing
V139M	Lung carcinoma	Deleterious or disease causing	Deleterious or disease causing
F147I	Lung carcinoma	Deleterious or disease causing	Deleterious or disease causing
V151L	Lung carcinoma	Deleterious or disease causing	Deleterious or disease causing

Table 3.1 demonstrates that apart from mutation at position 56, the rest of the mutations were predicted to be deleterious or harmful mutations that disrupt PFAP function.

Protein alignment of the PFAP F1 fragment with other mammals, amphibians, Shg1 protein in yeast conducted using ClustalW software (Larkin *et al*., 2007) , revealed that Shg1H domain was conserved in other species (Accession number provided in table 2.3) (Figure 3.17, 3.18). This alignment also shows that mutations found in PFAP Shg1H domain were also conserved in homologous proteins of other species. This suggests that these mutations are more likely to disrupt the function of PFAP and may also have an effect on PFAP and SET1A in vitro interaction.

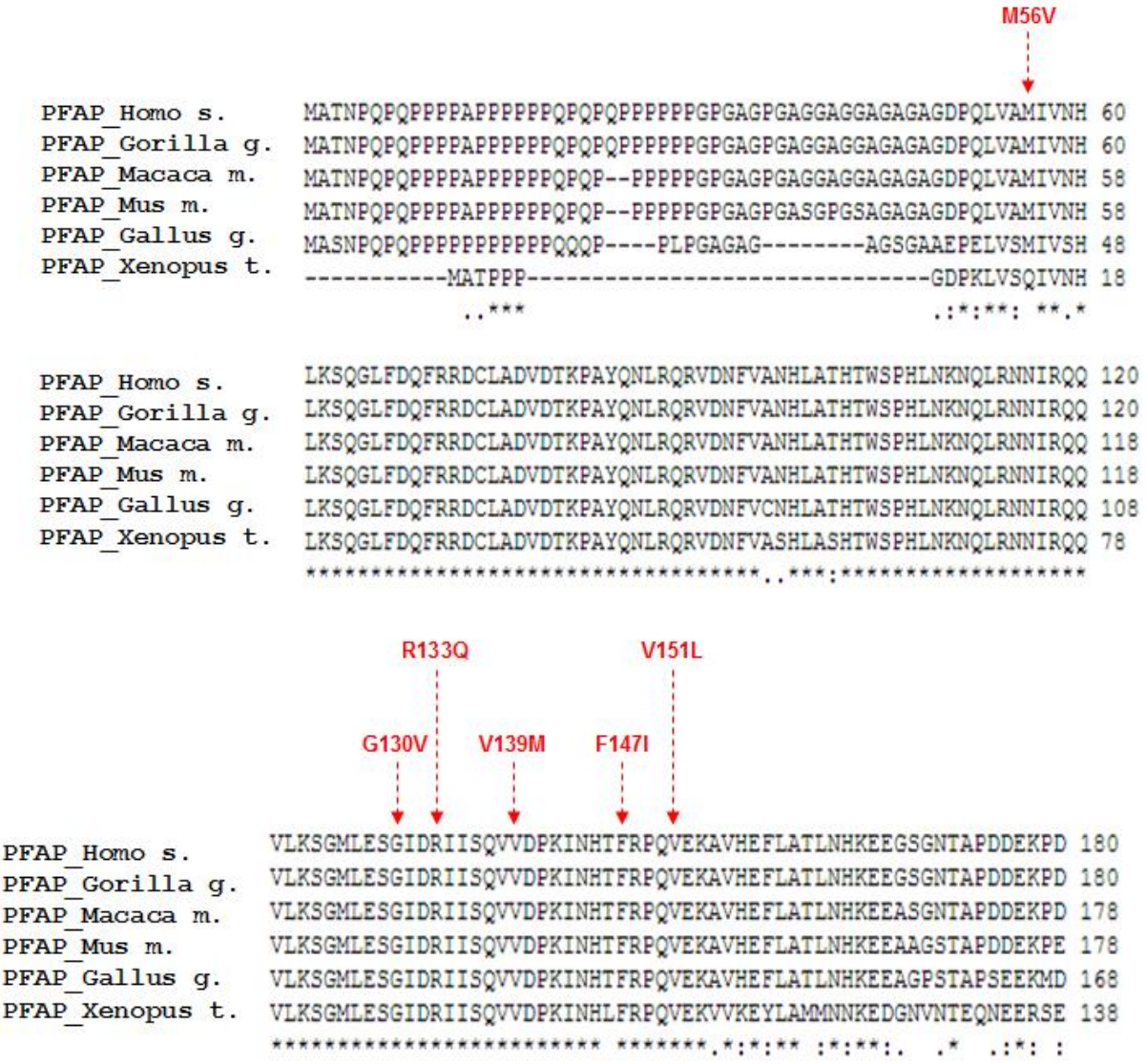

Figure 3.17: The position of cancer-associated mutations within the Shg1H domain of PFAP relative to amino acids conserved amongst mammalian/amphibian PFAP sequences. Amino acid sequences of PFAP sequences were obtained from the Refseq database and aligned using ClustalW **(Larkin *et al.*, 2007)**. Absolutely conserved amino acids are marked with a *. Positions of cancer-associated mutations are indicated with red arrows.

Figure 3.18: The position of cancer-associated mutations within the Shg1H domain of PFAP relative to amino acids conserved amongst mammalian/amphibian PFAP sequences. Amino acid sequences of PFAP sequences were obtained from the Refseq database and aligned using ClustalW. Absolutely conserved amino acids are marked with a *. Positions of cancer-associated mutations are indicated with red arrows.

4.0 Discussion

4.1 PFAP interacts with SET1A in vitro

PFAP (Putative Fanconi Anaemia Protein) is a large nuclear protein of unknown function, which belongs to the Fam44 family of proteins. Despite its size (3051aa), PFAP consists of few identifiable domains, which includes an N-terminal domain with homology to the COMPASS/HMT yeast component Shg1 (Shg1H domain), A-T hook and several identified ATM/ATR phosphorylation sites (Figure 1.5). Recent study carried out by van Nuland and co-workers (2013) have suggested that PFAP interacts with SET1B (Nuland *et al.*, 2013). However, unpublished data from our laboratory suggests that PFAP interacts with SET1A (Figure 3.2), the catalytic subunit of another histone methyltransferase. Thus, we hypothesised that PFAP interacts with SET1A rather than with SET1B as suggested by van Nuland *et al* (2013).

Immunoprecipitation of endogenous SET1A and SET1B with PFAP antibodies demonstrated that PFAP was pulled down with endogenous SET1A but not with SET1B (Figure 3.1). This suggests that PFAP strongly interacts with SET1A, but not with SET1B.

Our initial studies have shown that PFAP strongly interacts with SET1A and not with SET1B. However, cross-reactivities between immunoblotting SET1A and SET1B antibodies and immunoprecipitating anti-SET1B and anti-SET1A antibodies respectively (Figure 3.1- left and centre panels) suggesting that antibodies needs to be specific to detect specific protein of interests, otherwise it can lead to false negative results. For example, cross-reactivity can be interpreted as SET1A and SET1B interacting with each other, therefore, SET1B would be

accepted to be interacting with PFAP, which the results from our study does not show. Therefore, to prove this, specific antibodies are required to prevent the cross-reactivity.

PFAP was shown to be associated with BOD1 protein in the van Nuland study. The authors therefore, conclude that since BOD1 and SET1B interacts, PFAP is a part of SET1B histone methyltransferase complex. However, our data contradicts this claim. The other possible reason for this discrepancy could be that PFAP was also being pulled down with other non-catalytic subunits including WDR5, RBBP5, ASH2L and DPY30 which are shared by all SET/COMPASS complexes and Cfp1 and Wdr82 are shared by only SET1A/SET1B complexes (Figure 1.6) (van Nuland *et al*., 2013).

Since initial experiments strongly suggested that PFAP interacts with SET1A and not with SET1B, we therefore wished to identify the regions of PFAP responsible for binding to SET1A. To be able to answer this question, and since PFAP is a very large protein, six overlapping fragments of approximately 500-600 amino acids in length were generated. These fragments were introduced into pGEX 3X vector upstream of GST, to form GST-tagged pGEX encoded PFAP fragments, and bacterially expressed upon addition of IPTG. Once induced, only the GST-tagged F1 and F2 fragments were purified for further analyses of interactions of SET1A with PFAP.

Only the F1 fragment and F2 fragment were used to analyse interactions of PFAP with SET1A/SET1B. We chose these fragments as N-terminal region of PFAP contains sequences known as Shg1H domain similar to BOD1 and Shg1 protein. Therefore, we hypothesise that this so-called Shg1H domain of PFAP may contain sequences that interact with SET1A. The

Shg1H domain is located in the F1 fragment and is positioned between amino acids 55- 164. However, it does not mean that SET1A only interacts with F1 fragments; the other GST-tagged fragments of PFAP not analysed here may also bind SET1A.

The interaction between the GST-tagged PFAP fragments and SET1A was examined using GST pull down assays, using GST as negative control. GST pull down analysis demonstrated that endogenous SET1A was pulled down with F1 but not F2, whereas, SET1B was not pulled down by either of the fragments (Figure 3.8). These analyses further confirm that SET1A interacts with PFAP, particularly with F1 fragment of PFAP possibly at the Shg1 domain. Although, we hypothesise that the interaction occurs within the Shg1H domain, this needs to be proved experimentally by specifically examining the interaction of Shg1H domain of PFAP with SET1A to confirm our hypothesis.

One question which arises from our initial studies is whether the interaction between PFAP and SET1A is direct or is it mediated via another component of SET1A complex? It has been shown that SET1A and SET1B HMT complex share many conserved subunits such as Wdr5, Rbbp5, ASH2L, Wdr82, DPY30, Cfp1 (Figure 1.6) (Nuland *et al*., 2013). If the interaction of SET1A and PFAP was indirect (i.e. via other members of both HMT complexes) or mediated via BOD1 protein then we would have likely observed a strong interaction of PFAP with both SET1A and SET1B. Therefore, it is so likely that PFAP and SET1A interacts directly or there are other unidentified components.

The above data leads us to speculate as to the role of PFAP in histone methylation. From our data, we hypothesise that PFAP is involved in the ICL repair pathway. However, the mechanism(s) behind its role in this repair pathway are still unknown. Therefore, we can only speculate that PFAP is involved in the regulation of histone methylation based on the

results from the above experiments and from unpublished previous studies from our laboratory.

So far the function of PFAP is unknown. However, recent independent studies have raised a possibility of the role of PFAP in histone methylation. Eberl *et al* (2013) has shown that PFAP act as a "significant interactor" for K4me3 chromatin marks for the activation of lysine methylation on histone three (Eberl *et al*., 2013). Secondly, Nuland and co (2013) have demonstrated that PFAP interacts with SET/COMPASS like histone methyltransferase complexes - SET1A, SET1B, which are involved in the regulation of global levels of lysine methylation. On the other hand, Dehe and co (2006), showed that Shg1 protein (region of the yeast SET1C complex similar to Shg1H domain of PFAP) inhibits the level of di- and tri-methylation on H3K4, indirectly suggesting that PFAP may inhibit di-methylation and tri-methylation of lysine, (Dehe *et al*., 2006). Therefore, it becomes essential to investigate the role of PFAP in histone methylation and identify other interactors from the SET1A complex apart from endogenous SET1A protein.

Alternatively, the role of PFAP in methylation could also be investigated by several methods. Firstly, enzymatic activity of purified GST-tagged fragments of PFAP mainly F1 and F2 could be analysed by immunoblotting against anti-histone H3, methylated lysines. Secondly, we could possibly identify which GST-tagged PFAP fragment(s) interacts with histones using histone pull down assays, wherein GST-fusion proteins of PFAP fragments are incubated with recombinant or native histone octamer, precipitated using GST beads and analyse by SDS-PAGE. This will hopefully show that native histones are pulled down with GST-tagged PFAP fragments, and further analyse for histone lysine methylation marks by

immunoblotting. Or simply we can just measure the level of histone methylation in PFAP cell lines. These techniques could assist in elucidating the role of PFAP in histone methylation.

4.2 PFAP interaction with other Fanconi anaemia proteins

PFAP is believed to be involved in the fanconi anaemia pathway, in particular downstream of FANCI-FANCD2. Therefore, we wished to examine the interaction of PFAP with other Fanconi anaemia proteins. We chose to analyse interactions FANCD2, FANCJ and FANCM, as we had antibodies readily available for these proteins. Co-immunoprecipitation analyses demonstrated that FANCD2 and FANCJ interacted with PFAP, as both FANCD2 and FANCJ were pulled down by PFAP (Figure 3.13). However, immunoprecipitation of PFAP with FANCJ resulted in the detection of two bands. Out of these two bands, lower band expresses FANCJ protein. However, this need to be confirmed by running mutant FA-J cell line in parallel, as a negative control (Figure 3.13- bottom left panel).

These interactions were not surprising as FANCD2 (mono-ubiquitinated form of FANCD2) relocalises to the chromatin where it binds to BRCA2 and thus, facilitates the trigger of homologous recombination repair, which FANCJ is also believed to be part of. Further investigation of the interaction between FANCD2, FANCJ and PFAP must be carried out to confirm these findings and to ascertain the functions of this interaction. Future investigations should include reciprocal immunoprecipitation of FANCD2/ FANCJ to pull down PFAP, followed by co-localisation of FANCD2, FANCJ and PFAP.

4.3 FLAG-SET1A and Rad51 exhibits distinct sub-nuclear localization

In addition to examining the *in vitro* interaction of PFAP and SET1A, we decided to investigate co-localisation of FLAG-SET1A and RAD51, to see whether SET1A localises to ICL repair sites along with Rad51, which relocalises to nucleus to form distinct foci in response to MMC. Co-localisation was studied in HeLa cells and U2OS cells immunofluorescence staining. FLAG-tagged version of SET1A was used as antibodies for endogenous SET1A did not work properly. Immunofluorescence staining of both HeLa cells and U20S cells expressing FLAG-tagged version of SET1A demonstrated that Rad51 formed distinct foci in the presence of MMC in both cell lines (Figure 3.12), but no distinct foci were formed for FLAG-tagged SET1A in U20S cells (Figure 3.12), instead formed speckles in HeLa cells in doxycycline positive cells in the presence of MMC (Figure 3.10). Nevertheless, no co-localization was found as both FLAG-SET1A and Rad51 exhibit non-overlapping sub-nuclear localisation. The formation of many Rad51 foci in FLAG-tagged SET1A in doxcycline induced U2OS and HeLa cells after 24 and 48 hours of MMC treatment (Figure 3.9 and 3.12) indicates that SET1A is involved in ICL repair although, both SET1A and Rad51 proteins have distinct functions in ICL repair pathway.

The limitations of studying co-localisation using fluorescence microscope are bleaching and dealing with fixed or dead cells. Bleaching could result in loss of signals or activity. Therefore, bleaching can be prevented by reducing the length of the exposure or using more robust fluorophores that are less prone to bleaching. The other limitation is that immunofluorescence staining is always carried out in fixed or dead cells. Using fixed cells result in cross-linking of SET1A and Rad51, thus, resulting in false positive or false negative signals as a result of specific binding. These problems can be overcome by using alternative

approaches such as green fluorescent protein (GFP), which allow studying localization of proteins in living cells.

Alternative to fluorescence microscopy is to use confocal microscopy. Confocal microscopy provides better resolution, 3D visualisation and eliminates out-of-focus objects.

4.4 Cancer-associated mutations disrupts interaction of PFAP and SET1A in vitro

As both PFAP and SET1A are involved in the maintenance of genomic stability (our unpublished data), we, therefore, hypothesised that the interaction of SET1A with PFAP may be disrupted by mutation of PFAP in tumorigenesis, and that this maybe the cause of genetic instability in these tumours. Screening of COSMIC and cbioportal database identified over two hundred mutations; one third of these mutations were found in Shg1H domain of PFAP (Figure 3.16) and amino acids responsible for mutations in Shg1H domain were conserved in other mammalian and amphibian species (Figure 3.17). This suggests that these mutations may disrupt the in vitro interaction of PFAP and SET1A. As the crystal structure of PFAP has not been resolved, we cannot confirm that these mutations will have an impact on interaction of PFAP and SET1A using *in silico* analyses. However, we could confirm this by introducing these Shg1 mutations into GST-tagged F1 fragment of PFAP using site-directed mutagenesis and study the interaction between these mutant PFAP fragments and SET1A by the same techniques used in this study. Hopefully, this would show whether mutations has any effect on the in vitro interactions of SET1A and PFAP. These mutations might have an slight or drastic effect on the levels of histone methylation, but we don't know. These mutations may also affect the interactions of FANCD2 and PFAP, provided they both interact at Shg1H domain.

4.5 Future work

In future, if I am provided with an opportunity to continue with this project, firstly I would confirm that predicted Shg1H domain of the GST-tagged F1 PFAP fragment is the region where interaction between PFAP and SET1A. This would be done by generating GST-tagged Shg1H domain fragment by expressing in the same pGEX 3X vector, inducing protein expression, purifying and thus studying its in vitro interaction with PFAP using GST pull down assays, wherein GST-Shg1H domain would be used as a bait protein and SET1A as a prey protein.

The other way of confirming in vitro interactions between PFAP and SET1A is through co-localisation studies by immunofluorescence staining. Since, it is known that PFAP and SET1A interacts in vitro, therefore, it would be interesting know whether both the proteins which interacts in vitro, behaves in the same way in a cancer cell line treated with MMC or left untreated. When the cells are exposed to MMC, hopefully distinct foci becomes visible for both SET1A and PFAP, which suggests that both set of proteins are involved in ICL repair. On the other hand, if the foci of both proteins are observed at the same place, this will also demonstrate that both SET1A and PFAP are co-localised at the ICL repair sites.

Secondly, mutations found in Shg1H domain could disrupt the interaction between SET1A and PFAP. This can be confirmed by inserting these mutants into Shg1H domain by site-directed mutagenesis. In site directed mutagenesis, Primers designed will contain desired mutations and will be complementary to the template DNA around the mutation site, so it can hybridise with the DNA in the gene of interests. The single strand primer is extended using a DNA polymerase, which copies the rest of the gene. This mutant Shg1H gene is introduced

into sub-cloning vector and cloned. And then we would be able to express, purify and detected this mutants as GST-fusion products as before and study its interactions with SET1A by doing GST-pull down assays.

4.6 Summary

Interstrand crosslinks are highly toxic lesions caused by DNA crosslinking agents such as Mitomycin C. These interstrand crosslinks are repaired by the Fanconi anaemia pathway. PFAP (Putative fanconi anaemia protein) is a novel protein uncharacterised protein, which we propose it to be involved in Fanconi anaemia pathway, acting downstream of mono-ubiquitinated FANCI- FANCD2 (I-D2 complex). However, its role in this pathway still remains elusive. Van Nuland and co-worker showed that PFAP interacts with SET/COMPASS like histone methyltransferase called SET1B. However, unpublished data from our laboratory demonstrates that PFAP interacts with SET1A (another histone methyltransferase belonging to the same SET/COMPASS group) rather than SET1B. Therefore, we hypothesized that SET1A interacts with PFAP in vitro.

Firstly, in this study, we have demonstrated that the PFAP interacts with catalytic subunit of SET1A HMT called SET1A. This interaction is predicted to takes place at Shg1H domain in F1 fragment of PFAP. We have also speculated that any substitution mutations in Shg1 domain of F1 fragment of PFAP could disrupt the interaction of PFAP and SET1A.

Secondly, as SET1A is involved in DNA repair, therefore we wished to study whether SET1A localises to sites of ICL repair, as a result co-localisation of FLAG-tagged SET1A (SET1A) with Rad51 was studied. We did not find any co-localization, as both revealed non-

overlapping sub-nuclear localisation, suggesting that both proteins targets specific genomic sites and therefore, SET1A does not localise to the sites of ICL repair. However, with the formation of distinct foci in Rad51 panel when treated with MMC, illustrates that SET1A is involved in ICL repair.

Ultimately, in vitro interaction of PFAP and other fanconi anaemia proteins including FANCD2, FANCJ and FANCM were study as PFAP is believed to be involved in Fanconi anaemia pathway, in particular downstream of mono-ubiquitination of FANCI- FANCD2. Immunoprecipitation analyses demonstrated that FANCD2 and FANCJ interacts with PFAP, however, further investigations are required to confirm this findings.

To conclude, PFAP is a large nuclear protein with unknown function. However, two independent studies had raised the possibility that PFAP may play role in the regulating histone methylation, via interactions with the methylated forms of Lys4 of histone H3 (Eberl, 2013), and through a possible interaction with SET/COMPASS like histone methyltransferases SET1A/SET1B (Nuland *et al*., 2013). In this study, we have been successful in demonstrating in two key points:

Firstly, that PFAP interacts with SET1A and not with SET1B as shown by van Nuland (2013), possibly in the Shg1H domain in F1 fragment of PFAP.

Secondly, in vitro protein-protein interaction of PFAP with FANCD2 and FANCJ confirms that PFAP is involved in FA pathway, downstream of FANCD2-FANCI complex.

5.0 References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, SR.

Akkari, Y. M., Bateman R.L, Reifsteck, C.A., Olson, S.B. and Grompe, M. (2000). DNA replication is required to elicit cellular response to psoralen- induced DNA interstrand cross-links. *Molecular Cell Biology* 20(21), 8283- 8290.

Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, PA., Stratton, M.R. and Wooster, R. (2004). The COSMIC (Catalogue of Somatic Mutations in cancer) database and website. *British Journal of Cancer* 91(2), 355-358.

Bannister, A.J. and Kouzarides, T. (2005). Reversing histone methylation. *Nature* 436(7054), 1103- 1106.

Bernstein, C., Prasad, A.R., Nfonsam, V. and Bernstein, H. (2013). DNA Damage, DNA Repair and Cancer, New Research Directions in DNA Repair, Prof. Clark Chen (Ed.), ISBN: 978-953-51-1114-6, InTech, DOI: 10.5772/53919. Available from: http://www.intechopen.com/books/new-research-directions-in-dna-repair/dna-damage-dna-repair-and-cancer.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C. and Schultz, N. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* 2, 401- 404.

Chapman, J.R., Taylor, M.R. and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. *Molecular Cell* 47(4), 497-510.

Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A.R., Laghmani, el.H., Joenje, H., McDonald, N., de Winter, J.P., Wang, W. and West, S.C. (2007). Identification of FAAP24, a Fanconi anaemia core complex protein that interacts with FANCM. *Molecular Cell* 25(3), 331-343.

Ciccia, A., McDonald, N. and West, S.C. (2008). Structural and functional relationships of the XPF/MUS81 family of proteins. *Annual Reviews of Biochemistry* 77, 259- 287.

Ciccia, A. and Elledge, S.J. (2010). The DNA Damage Response: Making it safe to play with knives. *Molecular Cell* 40(2), 179-204.

Cole, R.S. (1973). Repair of interstrand cross-links in DNA induced by psoralen plus light. *The Yale Journal of biology and medicine* 46(5), 492.

Dehe, P.M., Dichtl, B., Schaft, D., Roguev, A., Pamblanco, M., Lebrun, R., Rodriguez-Gil, A., Mkandawire, M., Landsberg, K., Shevchenko, A., Shevchenko, A., Rosaleny, L.E.,

Tordera, V., Chavez, S., Stewart, A.F. and Geli, V. (2006). Protein Interactions within the Set1 Complex and Their Roles in the Regulation of Histone 3 Lysine 4 Methylation. *The Journal Of Biological Chemistry* 281(46), 35404 - 35412.

de Winter, J.P. and Joenje, H. (2008).The genetic and molecular basis of Fanconi anaemia. *Mutation Research* 668(1-2), 11-9.

Dillon, S.C., Zhang, X., Trievel, R.C. and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biology* 6(8), 227-239.

Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M. and Mann, M. (2013). A map of General and specialized chromatin readers in Mouse tissues generated by label-free interaction proteomics. *Molecular Cell* 49(2), 368- 378.

Fanconi, G. 1927 Familiaere infantile perniziosaartige Anaemie (pernizioeses Blutbild und Konstitution). *Jahrbuch Kinderheild* 117, 257–280

Gao, J., Aksoy, B.A., Dorgrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C. and Schultz, N. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioportal. *Science* 6(269), 1-19.

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

Hoeijmakers, J.H.J. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* 411(6835), 366-374.

Howlett, N.G., Harney, J.A., Rego, M.A., Kolling, F.W., 4th and Glover, T.W. (2009). Functional interaction between the Fanconi Anaemia D2 protein and proliferating cell nuclear antigen (PCNA) via a conserved putative PCNA interaction motif. *The Journal of Biological Chemistry* 284(42), 28935- 28942.

Huang, Y. and Li, L. (2013). DNA crosslinking damage and cancer- a tale of friend and foe. *Translational Cancer Research* 2(3), 144-154.

Ishiai, M., Kitao, H., Smogorzewska, A., Tomida, J., Kinomura, A., Uchida, E., Saberi, A., Kinoshita, E., Kinoshita-Kikuta, E., Koike, T., Tashiro, S., Elledge, S.J. and Takata, M. (2008). FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anaemia pathway. *Nature Structural & Molecular Biology* 15(11), 1138- 1146.

Jackson, S.P. and Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature* 461(7267), 1071- 1078.

Jenuwein, T. and Allis, C.D. (2001). Translating the histone code. *Science* 293(5532), 1074-1080.

Jorgensen, S., Schotta, G. and Sorensen, C.S. (2013). Histone H4 Lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Oxford Journals*, 1-10.

Knipscheer, P., Räschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Schärer, O.D., Elledge, S.J. and Walter, J.C. (2009). The Fanconi anaemia pathway promotes replication-dependent DNA interstrand crosslink repair. *Science* 326(5960), 1698- 1701.

Kottemann, M.C. and Smogorzewska, A. (2013). Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature Reviews* 493(7432), 356- 363.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21), 2947- 2948.

Lee, J.H. and Skalnik, D.G. CpG-binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys^4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. (2005). *The Journal Of Biological Chemistry* 280(50), 41725-41731.

Lee, J.H., Tate, C.M., You, J.S. and Skalnik, D.G. Identification and Characterization of the Human Set1B Histone H3-Lys[4] Methyltransferase Complex. (2007). *The Journal Of Biological Chemistry* 282(18), 13419-13428.

Lee, J.E., Wang, C., Xu, S., Cho, Y.W., Wang, L., Feng, X., Baldridge, A. Sartorelli, V., Zhuang, L., Peng, W. and Ge, K. (2013). H3K4 mono- and di- methyltransferase MLL4 is required for enhancer activation during cell proliferation. *eLife* 2, 1-39.

Lindahl, T. and Wood, R.D. (1999) Quality control by DNA repair. *Science* 286(5446), 1897-18905.

Martin, C. and Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nature Reviews Molecular Cell Biology* 6(11), 838-849.

Martin, S.A., Lord, C.J. and Ashworth, A. (2008). DNA repair deficiency as a therapeutic target in cancer. *Current Opinion in Genetics & Development* 18(1), 80-86.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R. 3rd, Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, Z., Lerenthal, Y., Shiloh, Y., Gygi, S.P. and Elledge, S.J. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316(5828), 1160-1166

McKinnon, P.J. (2009). DNA repair deficiency and neurological disease. *Nature Review Neuroscience* 10(2), 100-112.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America* 98(23), 12902-12907.

Moldovan, G.L. and D'Andrea, A.D. (2009). How the Fanconi Anaemia pathway guards the genome. *Annual Review of Genetics* 43, 223- 249.

Moldovan, G.L. and D'Andrea, A.D. (2012). To the rescue: The Fanconi Anaemia genome stability pathway salvages replication forks. *Cancer Cell* 22(1), 5-6.

Morrison, C., Shinohara, A., Sonoda, E., Yamaquchi-Iwai, Y., Takata, M., Weichselbaum, R.R. and Takeda, S. (1999). The essential functions of human Rad51 are independent of ATP hydrolysis. *Molecular Cell Biology* 19(10), 6891- 6897.

Niedernhofer, L.J., Lalai, A.S. and Hoeijmakers, J.H.J. (2005). Fanconi anaemia (cross)linked to DNA repair. *Cell* 123(7), 1191-1198.

Niedzwiedz, W., Mosedale, G., Johnson, M., Ong, C.Y., Pace, P. and Patel, K.J. (2004). The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair. *Molecular Cell* 15(4), 607-620.

Nuland, R.V., Smits, A.H., Pallaki, P., Jansen, P.W., Vermeulen, M. and Timmers, H.T. (2013). Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Molecular and Cellular Biology* 33(10), 2067-2077.

Patel, K.J. and Joenje, H. (2007). Fanconi anaemia and DNA replication repair. *DNA Repair* 6(7), 885-890.

Pecorino, L. (2012). Molecular Biology of Cancer. 3rd ed. Oxford University Press, Oxford, UK.

Porter, I.M., McClelland, S.E., Khoudoli, G.A., Hunter, C.J., Andersen, J.S., McAinsh, A.D., Blow, J.J. and Swedlow, J.R. (2007). Bod1, a novel kinetochore protein required for chromosome biorientation. *The Journal Of Cell Biology* 179(2), 187-197.

Schlacher, K., Wu, H. and Jasin, M. (2012). A Distinct Replication Fork Protection Pathway Connects Fanconi Anaemia Tumour Suppressors to RAD51-BRCA1/2. *Cancer Cell* 22(1), 106- 116.

Schwarz, J.M., Cooper, D.N., Schuelke, M. and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 11(4), 361- 362.

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. and Elledge, S.J. (2007). Identification

of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. (2007). *Cell* 129(2), 289- 301.

Spycher, C., Miller, E.S., Townsend, K., Pavic, L., Morrice, N.A., Janscak, P., Stewart, G.S. and Stucki, M. (2008). Constitutive phosphorylation of MDC1 physically links the MRE-11-RAD50-NBS1 complex to damaged chromatin. *The Journal Of Cell Biology* 181(2), 227-240.

Sulli, G., Di Micco, R. and d'Adda di Fagagna, F. (2012). Crosstalk between chromatin state and DNA damage response in cellular senescence and cancer. *Nature Reviews Cancer* 12(10), 709- 720.

Taniguchi, T., Garcia-Higuera, I., Andreassen, P.R., Gregory, R.C., Grompe, M. And D'Andrea, A. (2002). S-phase- specific interaction of the Fanconi anaemia protein, FANCD2, with BRCA1 and RAD51. *Blood* 100(7), 2414 - 2420.

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaquchi-Iwai, Y., Shinohara, A. and Takeda, S. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *The EMBO Journal* 17(18), 5497- 5508.

Tate, C.M., Lee, J.H. and Skalnik, D.G. (2010). CXXC Finger Protein 1 Restricts the Setd1A Histone H3-Lys4 Methyltransferase Complex to Euchromatin. *The FEBS Journal* 277(1), 210- 223.

Tischkowitz, M.D. and Hodgson, S.V. (2003). Fanconi anaemia. *Journal of Medical Genetics* 40(1), 1-10.

Walden, H.and Deans, A.J. (2014). The Fanconi anaemia DNA repair pathway: structural and functional insights into a complex disorder. *Annual Review of Biophysics* 43, 257- 278.

Ward, T.A., Dudasova, Z., Sarkar, S., Bhide, M.R., Vlasakova, D., Chovanec, M. and McHugh, P.J. (2012). Components of a Fanconi-Like Pathway Control Pso2-Independent DNA interstrand Crosslink Repair in Yeast. *PLoS Genetics* 8(8), 1-14.

Wang, X., Andreassen PR and D'Andrea A.D. (2004). Functional interaction of mono-ubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Molecular Cell Biology* 24(13), 5850- 5862.

Williams, H.L., Gottesman, M.E. and Gautier, J. (2012). Replication-independent repair of DNA interstrand crosslinks. *Molecular Cell* 47(1), 140- 147.

Zhang, Y. and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & Development* 15(18), 2343- 2360.