Investigation of transcriptional targeting and DNA lesion resolution of Activation Induced Deaminase by cofactor analysis.

Katharina L. Willmann

University College London

and

Cancer Research UK London Research Institute

PhD Supervisors: Svend K. Petersen-Mahrt & Jesper Q. Svejstrup

A thesis submitted for the degree of

Doctor of Philosophy

University College London

November 2011

Declaration

I, Katharina Willmann, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Katharina Willmann

November 2011

Abstract

Activation Induced Deaminase (AID) is a crucial protein in B cells, required for a functional humoral immune system. The regulation and the molecular mechanism of AID is the focus of my thesis.

As an enzyme, AID induces mutations and recombination in immunoglobulin genes, leading to increased antibody affinity for antigen and alterations in antibody-effector function. Catalytically, AID deaminates cytosine to uracil in single-stranded DNA, a DNA lesion that leads to a mismatch in double-stranded DNA. AID can be controlled via multiple pathways, including: a) targeting AID to the correct locus and b) regulating the processing of the AID-induced lesion. Proteins proposed to be involved in targeting AID include members of the transcription and chromatin regulation machinery, whereas a variety of DNA repair and DNA metabolism proteins have been shown to be involved in AID-induced lesion-processing. The physical and functional interactions of AID with these co-factors are not well characterised. In this thesis I determined the molecular details and biology of candidates for AID interaction.

First, I used a novel technique to uncover and characterise interacting partners of endogenous chromatin-bound AID from B cells. The presented work shows that the RNA polymerase II elongation associated protein PAF1 binds directly to AID, while knocking down members of the PAF complex in B cells inhibits AID-induced immunoglobulin diversification.

Secondly, I was able to identify a novel interaction of the DNA repair regulator PCNA with AID, both *in vitro* and *in vivo*. Importantly, PCNA function was proven to be susceptible to interference by AID and thus influencing lesion resolution.

Taken together, this thesis highlights the importance of cofactors for defining the AID molecular pathway, relevant to all known functions of AID within and possibly beyond the B cell reaction.

Acknowledgement

I would like to show my gratitude to those who have made this thesis possible. First and foremost, I am thankful to my supervisor Svend Petersen-Mahrt for his enthusiasm, his support, and all that he taught me along the way. I thank present and past lab members, Don-Marc Franchini, Maria Simon, Siim Pauklin, Kerstin Schmitz and Gopinath Rangam for advice and help in and outside the lab and for fruitful collaborations. I thank Jesper Svejstrup, Helle Ulrich, Simon Boulton and Stephen West for supervision and for making their support available when completing the last stages of my work. I am grateful to Bernardo Reina-San-Martin, Sara Milosevic and the whole Reina lab at IGBMC Strasburg for a successful collaboration on CH12. I thank CRUK facilities for their great work to support our research, especially Trevor at cell services for large scale cell growth. I would like to thank all my friends at the CRUK Clare Hall community as well as at LIF and IFOM Milan for always being there for scientific advice, reagent sharing, and especially, personal support and a laugh. You made the time in the lab enjoyable. I would like to thank Madita for a very special friendship. To Andres: Thank you for your love, encouragement and patience. I am very lucky. Lastly, I am eternally grateful to my family: my parents and my sister, for the tireless support and encouragement during many years.

Table of Contents

Table of figures

List of tables

Abbreviations

aa	Amino acid
AID	Activation Induced Deaminase
APOBEC	Apolipoprotein B-editing catalytic polypeptide-like subunit family
BER	Base Excision Repair
bp	Base pairs
C_H	Heavy chain constant region
ChIP	Chromatin immunoprecipitation
C region	Constant region
CSR	Class switch recombination
CTD	C-terminal domain
dA	Deoxyadenosine
dC	Deoxycytosine
DDT	DNA Damage Tolerance
dG	Deoxyguanosine
dT	Deoxythymidine
dsDNA	Double-stranded DNA
D	Diversity segment
DAPI	4',6-diamidino-2-phenylindole
DSIF	5,6-dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity-inducing factor
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
E. coli	*Escherichia coli*
FACS	Fluorescence activated cell sorting
FACT	Facilitates chromatin transcription
FCS	Fetal calf serum
FITC	Fluorescein Isothiocyanate
fmol	femtomol
HPRT	Hypoxanthine-guanine phosphoribosyltransferase gene
HR	Homologous recombination
HRP	Horseradish peroxidase
Ig	Immunoglobulin

iGC	Immunoglobulin gene conversion
IgH	Immunoglobulin heavy chain locus
IgL	Immunoglobulin light chain locus
Igκ	Immunoglobulin kappa light chain locus
IL-4	Interleukin 4
IP	Immunoprecipitation
IPTG	Isopropyl-b-D-thiogalactoside
IR	Infrared
J	Joining segment
kDa	Kilodalton
MMR	Mismatch Repair
MSH	DNA mismatch repair protein homologue of bacterial MutS
(m) RNA	(Messenger) ribonucleic acid
NER	Nucleotide Excision Repair
NES	Nuclear export signal
NLS	Nuclear localisation signal
NTP	Nucleotide triphosphate
NTS	Non-transcribed strand
ODA	Oligonucleotide deamination assay
PAF	Polymerase II associated factor
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase chain reaction
PE	Phycoerythrin
PIP	PCNA interacting peptide
pmol	picomol
PVDF	Polyvinyldifluoride
RAG	Recombination-activating gene
R	A or G
RFC	Replication factor C
RPA	Replication protein A
RL	*Renilla reniformis* luciferase

RNAPII	RNA polymerase II holoenzyme
RT	Room temperature
S	Switch region
SEC	Size exclusion chromatography
SDS	Sodium dodecyl sulfate
SHM	Somatic hypermutation
shRNA	Short hairpin ribonucleic acid
ssDNA	Single-stranded DNA
SUPTH	Suppressor of Ty homologue
TBE	Tris-Borate-EDTA
TGFβ	Transforming growth factor beta
TLS	Translesion synthesis
TS	Transcribed strand
TSS	Transcriptional start site
QRT-PCR	Quantitative real time PCR
UDG	Uracil-DNA Glycosylase (*E. coli*)
UNG	Uracil N-glycosylase (Eukaryote)
UV	Ultraviolet (light)
V region	Variable region
V	Variable gene segment
W	A or T
WRC	A/T-A/G-C
wt	Wild type

Chapter 1. Introduction

1.1 Immunity through Antibody diversification

1.1.1 B cells and the B cell reaction

The mammal immune system is based on cellular and humoral components. Cellular components develop from myeloid or lymphoid precursors in the bone marrow (Janeway, 2001). Cells of myeloid origin, such as macrophages, make up the innate immune system, whereas specific cells of lymphoid origin mediate the adaptive immune response, which is unique to vertebrates. In the adaptive immune response, specialised lymphoid cells produce humoral components to defend against foreign agents such as pathogenic microorganisms. Immune cells circulate the body through the blood stream, tissues and the lymphatic system until they activate their respective defence mechanisms. Lymphocytes that are part of the adaptive immune system are B cells, which mature in the bone marrow and are named after the Bursa of Fabricius in birds where they were discovered, and T cells, which mature in the thymus. Cells from innate and adaptive system communicate with each other to function effectively. For example, T helper cells have the ability to activate B cells and macrophages (Viret and Janeway, 1999). To this end, all lymphocytes express a large collection of receptors on their surface, which enable cell-cell communication and signal reception. The final stage of B cell maturation is aided by T cells and takes place in the germinal centres of lymphoid organs (Allen et al., 2007), where antigens are encountered. The antigen receptor of the B cells is one of the most important molecules of the adaptive immune system as it can recognise and bind an exceptionally broad variety of antigens, enabling the defence against an almost unlimited number of pathogens. The receptor, also called antibody or immunoglobulin, Ig, can be surface bound or secreted. Each B cell expresses one variety of antibody, and via clonal selection only those B cells that are useful to the immune system survive and proliferate (Burnet, 1957, Janeway, 2001).

1.1.2 Antibody maturation in B cells

Immunoglobulins are composed of four polypeptide chains, two disulphide linked heavy chains and two light chains that are attached to the heavy chains via further

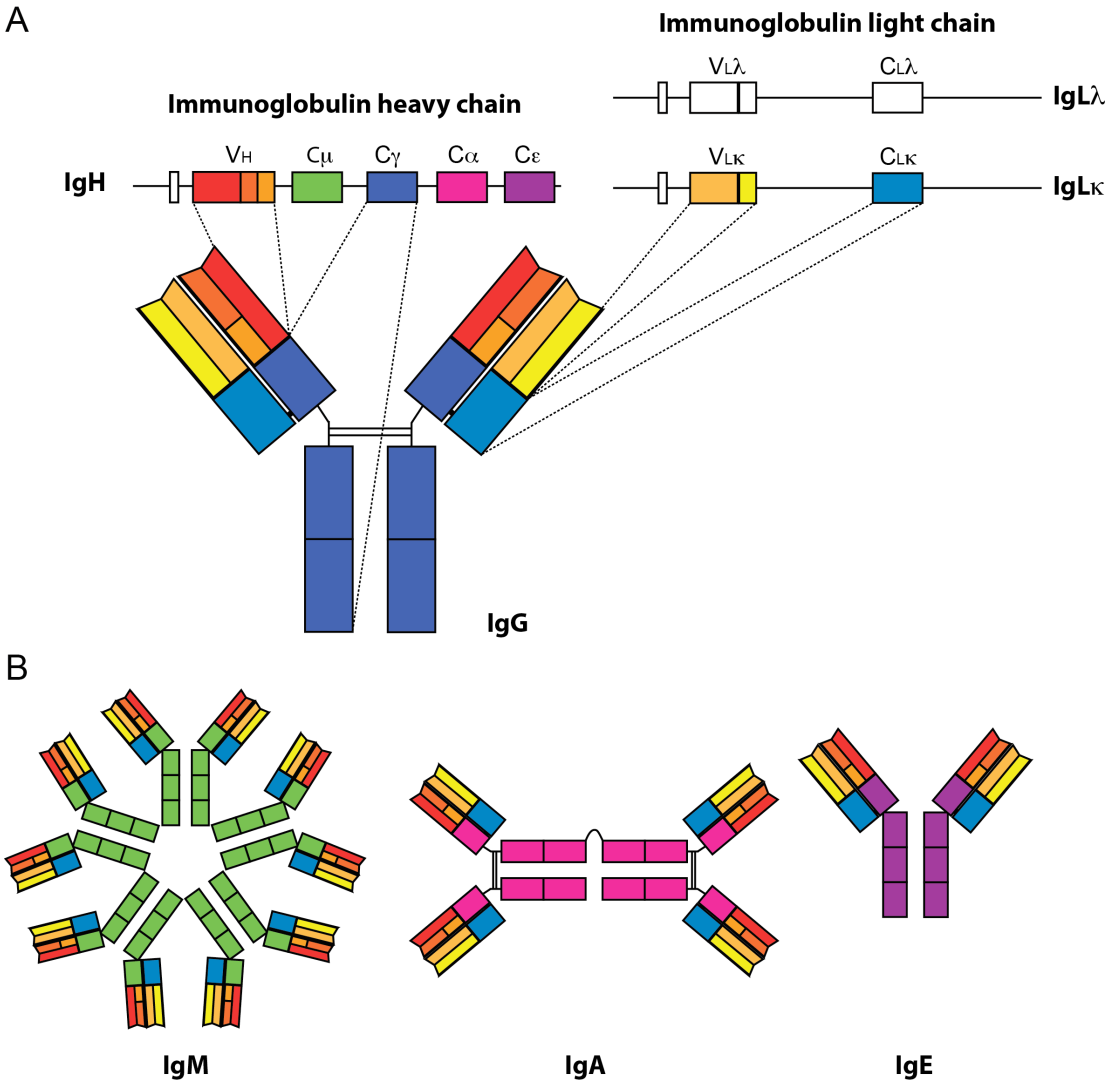


Figure 1. Organisation of the antibody molecules.

A, The antibody is composed of two heavy chains (dark blue) and two light chains (light blue). Both have variable (V) and constant (C) regions. Black bars mark disulphide bridges. **B**, The *Ig* heavy chain locus can produce transcripts that comprise constant regions for the isotypes IgM (green), IgG (dark blue), IgA (pink) and IgE (purple). The Ig light chain can stem either from the *IgLκ* or the *IgLλ* locus.

disulphide bridges (Figure 1 A). Although all vertebrates have immunoglobulins as part of their humoral defence, the mechanism of genetic maturation can vary between species. This description is restricted to the antibody maturation of mammals and, on occasion, of chickens (as indicated). While there are two genomic loci that can produce light chains (κ and λ), there is a single heavy chain locus (Figure 1 A). The heavy chain

can undergo genetic recombination (see details below) that allows for the production of five different sub-classes of heavy chains (μ, δ, γ, α, ε, Figure 1 A). The B cell can express any of the sub-classes of heavy chains, and the resulting classes of Igs [IgM, IgD, IgG, IgA and IgE (Figure 1 B)] define the Ig effector function (Janeway, 2001). A B cell always expresses membrane-bound IgM initially and can be induced to switch to any alternative secreted isotype. The predominant function of secreted IgM is to activate the complement system, a soluble component of the immune system to lyse bacteria. IgA is the main isoform secreted in body fluids, while IgG is the predominant form in the blood. IgG is primarily involved in opsonisation and inactivation of antigens and antigen carriers, as well as activating the complement system. IgE localises near epithelia and responds by sensitising mast cells, which can release proinflammatory signals.

B cell development progresses through several stages comprising a number of rearrangements of the immunoglobulin loci (reviewed in Cunningham-Rundles and Ponda, 2005) in order to generate the maximal useful possible variability of Ig. The cells originate in the bone marrow where they undergo antigen-independent development. Here, as pro-B cells, the first genomic rearrangement takes place, VDJ recombination of the antibody heavy chain (Figure 2 A and reviewed in Schatz and Ji, 2011). This process is dependent on the recombinases RAG1 and RAG2 that catalyse breakage and rejoining of DNA in a sequence specific way, whereby a variable (V) segment, a diversity (D) segment and a junction (J) segment are joint together to form a productive, recombined pre-B cell receptor (Figure 2 A). After an analogous VJ rearrangement of the κ or λ light chain in the pre-B cell, a complete IgM can be expressed on the surface. After clonal deletion of those B cells that recognise self antigens, the surviving immature B cells leave the bone marrow for the periphery, where upon antigen encounter, they undergo further steps of development and diversification. In the germinal centres of the lymphoid organs, (reviewed in Klein and Dalla-Favera, 2008) two reactions take place in order to optimise the properties of the antibody pool. Antigen contact induces somatic hypermutation (SHM) of *Ig* loci in the naïve B cells, introducing point mutations into the variable region of the *Ig* sequence to change the affinity for the antigen (Neuberger et al., 2003, Weigert et al., 1970). Through the process of antibody affinity maturation, only those B cell clones that have

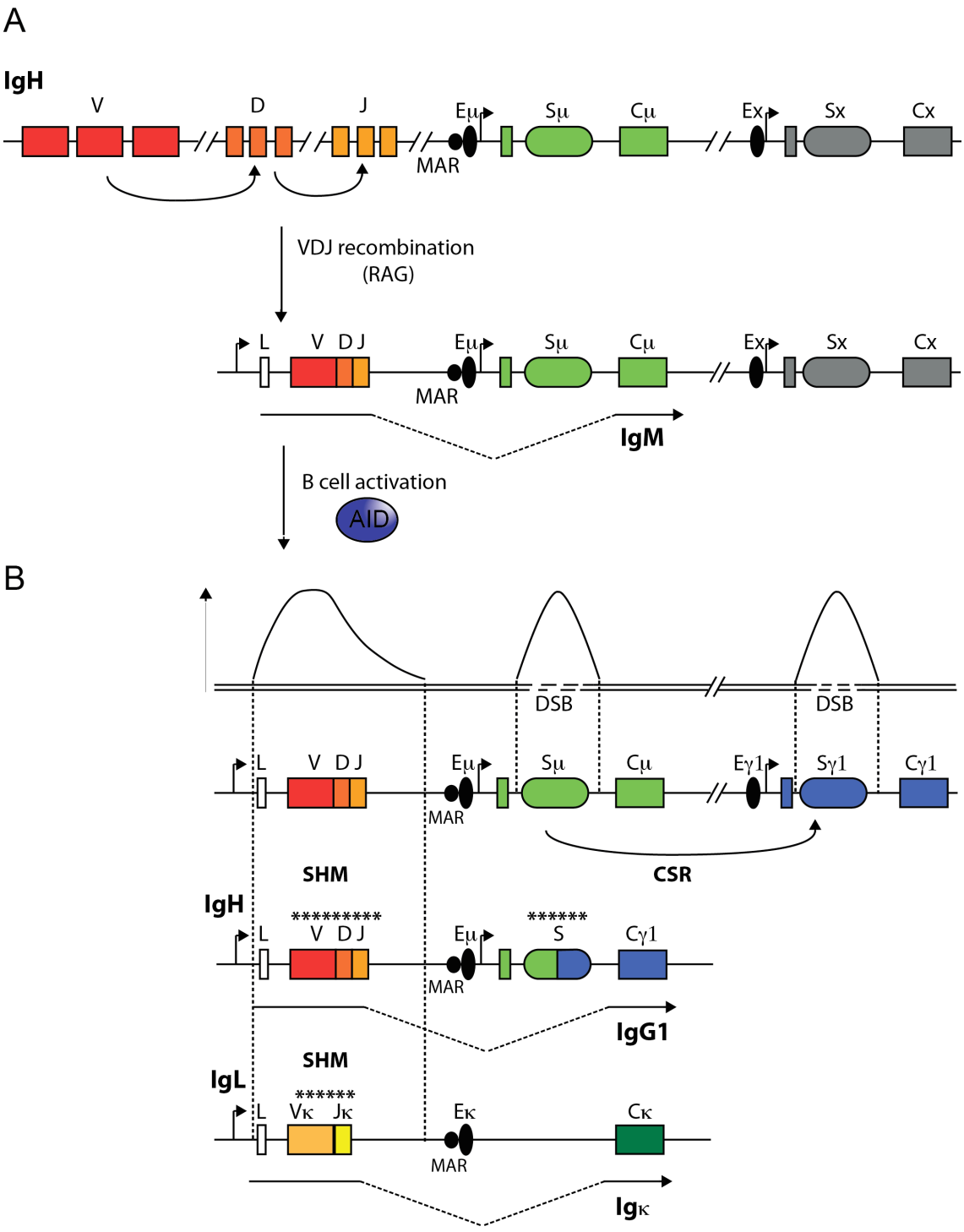

A
IgH
V
D
J
Eμ
Sμ
Cμ
Ex
Sx
Cx
MAR
VDJ recombination
(RAG)
L
V D J
IgM
B cell activation
AID
B
DSB
Eγ1
Sγ1
Cγ1
CSR
SHM
****
****
S
IgG1
IgL
Vκ Jκ
Eκ
Cκ
Igκ

Figure 2. Diversification events at the *Ig* locus.

A, the germline heavy chain locus (IgH) comprises variable (V) region segments (each with its own promoter), diversity (D) segments and junction (J) segments, followed by a matrix attachment region (MAR), the intronic enhancer (Eμ), and the switch region with several constant (C) genes. Each C region has a 5' switch region (Sx) and an intronic promoter and enhancer for the production of a sterile transcript. VDJ recombination in early antigen-independent B cell development results in a rearranged variable region (preceded by a leader peptide, L). Transcription from the V promoter results in IgM heavy chain expression (pre-B cell receptor). **B**, B cell activation upon antigen contact results in sterile switch transcripts and activation of AID. Mutations are introduced in the VDJ and S regions in a distribution shown as frequency curves. At S regions, AID lesions can lead to double strand breaks (DSB) that are required for CSR. AID lesions (shown in asterisks) lead to SHM at VDJ and to CSR between switch donor (Sμ) and acceptor (here, Sγ1). The intervening sequence is excised as circular DNA (not shown). Transcription from the V gene promoter leads to the expression of IgG1. The light chain locus undergoes SHM in the same way and its product (here, Igκ) completes the IgH chain.

(Modified from Pavri and Nussenzweig, 2011)

acquired an increased affinity for the antigen are selected, and prompted to undergo Class Switch Recombination (CSR). These steps are antigen and T helper cell dependent, and require effector molecules such as cytokines. The process of CSR (reviewed in Chaudhuri and Alt, 2004) includes induction of double strand breaks into switch regions of the *Ig* gene (Figure 2 B) and joining of the variable region to one of the constant regions, according to the signals received by the cytokines. CSR observed *in vivo* can also be recapitulated *in vitro*, through addition of cytokines to isolated B cells or B cell lines (reviewed in Chaudhuri and Alt, 2004).

1.2 The AID protein

1.2.1 Role of AID in immunity

Both SHM and CSR in the germinal centre are initiated by a small enzyme expressed upon activation of mature B cells, called Activation Induced Deaminase (AID) (reviewed in Petersen-Mahrt, 2005). This protein and its function will be the focus of this thesis. AID is also expressed in immature B cells of birds and some mammals when undergoing gene conversion of *Ig* genes (iGC), an alternative, templated sequence

diversification process (reviewed in Di Noia and Neuberger, 2007). AID is also expressed in non-immune tissues such as the germline and pluripotent tissues (Morgan et al., 2004). Individuals that lack AID or carry a mutated version of the gene develop a hyper IgM syndrome together with mild immunodeficiency (Revy et al., 2000), a phenotype recapitulated in knock-out mice (Muramatsu et al., 2000) and a consequence of the inability of activated B cells to undergo CSR and affinity maturation of Igs. The catalytic activity of AID to deaminate cytosines (dC) in DNA into uracils (dU) is necessary for these processes (Figure 3 A), and results in the controlled introduction of point mutations and other DNA rearrangements in the *Ig* locus.

1.2.2 Biochemical Properties

AID is a 24 kDa protein and founding member of the APOBEC family of nucleic acid deaminases. Although at the time of discovery AID was suggested to be an RNA editing enzyme based on sequence similarity to the related RNA deaminase APOBEC1, it was subsequently shown to deaminate DNA *in vitro* (Bransteitter et al., 2003, Chaudhuri et al., 2003, Dickerson et al., 2003) as well as in *E. coli* (Petersen-Mahrt et al., 2002). Enzymatic properties intrinsically restrict AID at the level of the number of lesions it can introduce and are thus an important point of regulation (reviewed in Peled et al., 2008). Catalytically, AID requires single stranded DNA (ssDNA) as a substrate, which *in vivo* can be provided by other DNA-associated enzymatic or metabolic processes (Odegard and Schatz, 2006). After deamination and reformation of the double stranded form, AID leaves a lesion in form of the DNA foreign base dU and a resulting mismatch, which can be further processed towards mutation or recombination. It was noticed that in hypermutated *Ig* genes, dC in the context of WRC (A/T-A/G-C) was more frequently targeted (Larijani et al., 2005, Rogozin and Kolchanov, 1992, Shapiro and Wysocki, 2002), and this was also found to be the preferred hot-spot motif for AID *in vitro* (Beale et al., 2004, Bransteitter et al., 2003, Yu et al., 2004).

In contrast to the cytoplasmic DNA deaminase APOBEC3G, which stays associated with its substrate after a first deamination and moves processively along the DNA towards the next target (Chelico et al., 2006, Coker and Petersen-Mahrt, 2007), AID dissociates from its substrate between deaminations *in vitro* (Coker and Petersen-Mahrt, 2007). Although the exact *in vivo* molecular mechanism at the target site still needs to

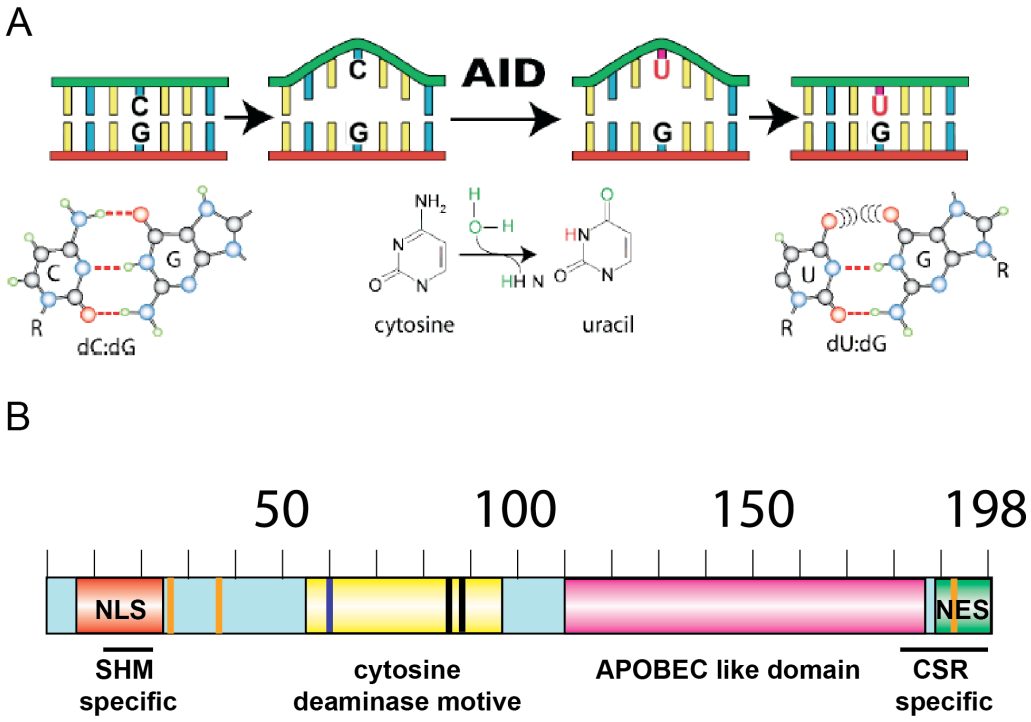


Figure 3. AID protein structure and properties.

A, Reaction catalysed by AID. Deamination of cytosine in ssDNA leads to a uracil-guanin basepair in dsDNA. **B**, The AID protein is depicted in light blue, with functional domains labelled and indicated in colours. Black bars indicate the SHM and CSR specific regions. Phosphorylated residues are drawn in orange, the active site in blue and the zinc co-ordinating residues in black.

be determined, protection of the genome from an overactive AID is likely to account for a distributive mode of action.

Structural models of AID are not available, although the crystal structure of APOBEC2, a related protein, has been solved and can at times be used comparatively. AID contains several functional domains (Figure 3 B). At the C-terminus, AID possesses a Nuclear Export Signal (NES). Also, the C-terminal portion of AID has been shown to be essential for CSR but not for SHM (Barreto et al., 2003, Shinkura et al., 2004, Ta et al., 2003). A domain bearing resemblance to AID's closely related proteins, APOBEC deaminases, stretches from amino acid (aa) 119 onwards. The cytosine deaminase domain lies between aa 55 and 94 and includes a zinc co-ordinating motif as well as the active site [histidine 56 (H56) and glutamic acid 58 (E58)]. Post-translational modifications of AID have been identified in primary B cells, including

phosphorylation at serine 38 (S38), S41, S43, (Pham et al., 2008) tyrosine 184 (Y184) (Basu et al., 2005), and threonine 140 (T140) (McBride et al., 2004).

AID may also be regulated by interactions with potential cofactors. A number of AID associated proteins have a role in subcellular distribution (Hasler et al., 2011, Maeda et al., 2010, Patenaude et al., 2009, Uchimura et al., 2011, Wu et al., 2005), or substrate accessibility, such as the ssDNA binding protein RPA (Chaudhuri et al., 2004). Recently, a number of interacting proteins were reported that are directly linked to transcription and chromatin regulation (Jeevan-Raj et al., 2011, Nambu et al., 2003, Okazaki et al., 2011, Pavri et al., 2010) or related to RNA processing (Basu et al., 2011, Conticello et al., 2008). One example, MDM2, interacts with AID through the C-terminal domain (MacDuff et al., 2006). However, interaction interfaces between AID and other proteins have only been mapped in a few cases (Conticello et al., 2008, Okazaki et al., 2011, Patenaude et al., 2009).

1.2.3 AID outside of Immunity

1.2.3.1 Anti-viral deaminases

The AID/APOBEC family consists of several deaminase members within the vertebrate lineage. APOBEC1 has a role in editing apolipoprotein B mRNA for lipid metabolism (Navaratnam et al., 1993, Teng et al., 1993), while APOBEC2, which is muscle-specific, and APOBEC4 do not seem to have catalytic activity (Liao et al., 1999, Rogozin et al., 2005, Sato et al., 2010). The APOBEC3 family of proteins have a role in innate immunity, as they are usually cytoplasmic anti-viral DNA targeting enzymes. Evolutionary, the AID/APOBEC protein family first emerged in vertebrates, with AID homologues found throughout bony fish, amphibians, birds and mammals (Conticello et al., 2005). Consequently, AID (together with APOBEC2) is thought to be the founding member of this family (Conticello et al., 2005), which has subsequently amplified and diversified further in mammals. In primates, the APOBEC3 family is the one with most numerous members. Mice have only one APOBEC3 protein, whereas humans possess at least eight APOBEC3 members (APOBEC3A-H). APOBEC3 proteins are able to restrict the propagation of retrotransposons and retroviruses (reviewed in Chiu and Greene, 2008). APOBEC3G is especially active against retroviruses such as the human immunodeficiency virus (HIV) through deamination, but also via a noncatalytic

function (Harris et al., 2003, Holmes et al., 2007, Newman et al., 2005). Briefly, APOBEC3G can be packaged into the viral particles of HIV, and mutate the viral single-stranded genome upon replication in the newly infected cell. That HIV has evolved a specific inhibitor peptide for APOBEC3G, *Vif* (Sheehy et al., 2002), emphasises the effectiveness of APOBEC3 proteins against retroviral proliferation.

Mutation of foreign nucleic acids may be the ancestral role of the AID/APOBEC precursor, with the need of family re-expansion once AID had specialised in *Ig* diversification. AID's cytoplasmic location and expression in the germline is reminiscent of this original purpose.

1.2.3.2 AID in epigenetics

Methylation of DNA has long been known to regulate chromatin state and accessibility. Inactive or silenced genes are associated with multiple methylation of cytosines (5meC) in the context of CpG, while in the surroundings of active promoters, CpG islands are kept methylation-free (Allis et al., 2006). Therefore, methylation of genes is a way for organisms to control gene expression, especially for embryonic development (Reik, 2007) or oncogene repression. Wide-spread demethylation occurs in early development of mammals, when methylation patterns are erased (Reik, 2007), but also occurs in cancer cells, which frequently have deregulated methylation patterns (Ehrlich et al., 1982, Feinberg and Vogelstein, 1983, Schar and Fritsch, 2011). However, an active demethylating enzyme has remained elusive (Fritz and Papavasiliou, 2010). On the other hand, the process of replacing methylated with non-methylated DNA is known to involve DNA repair related mechanisms (Gehring et al., 2009, Schar and Fritsch, 2011). Importantly, AID can deaminate 5meC *in vitro* (Bransteitter et al., 2003, Morgan et al., 2004) resulting in a thymidine (dT) residue. AID is expressed in cells undergoing active demethylation, such as oocytes (Morgan et al., 2004), and recently, genetic links between AID and local and genome wide CpG demethylation have been discovered (Bhutani et al., 2010, Popp et al., 2010, Rai et al., 2008). AID is required for genome wide demethylation in primordial germ cells (Popp et al., 2010), as well as for DNA demethylation during the reprogramming of heterokaryons (Bhutani et al., 2010). In the zebrafish embryo, epigenetic changes during development were also shown to be dependent on AID cooperating with enzymes from the base excision repair (BER) pathway (Rai et al., 2008).

1.2.4 Regulation of AID

1.2.4.1 Regulation of mRNA transcript levels

When AID was originally discovered, it was expressed in germinal centre B cells induced by cytokine and ligand stimulation (Muramatsu et al., 1999), signals normally provided by T-cells. Signalling is mediated by both stimulatory and inhibitory factors. Various classical B cell specific transcription factors play a major role in transcriptional control of AID, amongst them PAX5, a master regulator and pluripotency factor for the B cell lineage (Cobaleda and Busslinger, 2008, Medvedovic et al., 2011), the nuclear factor-κB (NF-κB) cascade, a stress and inflammation associated pathway (reviewed in Marusawa et al., 2011) as well as E-box proteins (de Pooter and Kee, 2010). These regulators are complemented by additional positive and negative factors (Dedeoglu et al., 2004, Gonda et al., 2003, Lee et al., 2006, Sayegh et al., 2003, Shapiro-Shelef et al., 2005, Xu et al., 2007). Furthermore, microRNAs affect AID expression. AID is negatively regulated by microRNA-155 as well as microRNA-181b (de Yebenes et al., 2008, Dorsett et al., 2008, Teng et al., 2008). When microRNA repression of AID is de-regulated, AID protein expression persists and increased *Ig* translocations are observed.

Hormones can directly control AID expression (Pauklin and Petersen-Mahrt, 2009, Pauklin et al., 2009). Specifically, oestrogen activates the AID promoter via the oestrogen receptor without a need for co-stimulatory signals (Pauklin et al., 2009). This finding may be of importance for AID expression outside the immune system, e.g. in ovarian, breast or prostate tissue (Morgan et al., 2004).

1.2.4.2 Regulation of protein

1.2.4.2.1 Post-translational modifications

Residues T140 and S38 on AID can be phosphorylated *in vitro* by serine/threonine kinases, PKA and various PKC isoforms (Basu et al., 2005, Chatterji et al., 2007, McBride et al., 2008, Pasqualucci et al., 2006). Phosphorylation of AID potentially contributes to regulation of localisation (Pasqualucci et al., 2006), activity (Pham et al., 2008), function in SHM and CSR (McBride et al., 2008, Pasqualucci et al., 2006) and interactions (Chaudhuri et al., 2004).

Sites of ubiquitination in AID have been mapped poorly so far, but it has been found that the majority of the AID ubiquitination for proteasomal degradation takes place in the nucleus (Aoufouchi et al., 2008), regulating active AID levels by protein turnover.

1.2.4.2.2 Subcellular localisation

The access of the AID protein to DNA is restricted primarily by localising AID to the cytoplasm, regardless if AID is expressed ectopically (Rada et al., 2002) or endogenously (Cattoretti et al., 2006, Greiner et al., 2005, Pasqualucci et al., 2004). Only limited amounts can be identified within the nucleus. Recent work has revealed a complex system of localisation and retention signals that lead to AID nucleocytoplasmic shuttling, which includes N-terminal and C-terminal peptide signal elements, as well as turnover of AID. However, this field remains controversial with contradicting reports, possibly reflecting different experimental systems that were examined. Indications for a classic bipartite N-terminal nuclear localisation signal (NLS) were found (Ito et al., 2004), while others suggested passive nuclear localisation (Brar et al., 2004, McBride et al., 2004). Recently, (Patenaude et al., 2009) showed evidence that an active but non-linear NLS exists for AID, complemented by an independent cytoplasmic retention mechanism (Hasler et al., 2011, Patenaude et al., 2009). The nuclear export of AID depends on the CRM1 export pathway (Brar et al., 2004, Ito et al., 2004, McBride et al., 2004). The picture was further enhanced by reports of a differential lifespan of nuclear and cytoplasmic AID, as it was shown that AID stability is differentially regulated in these compartments (Aoufouchi et al., 2008, Uchimura et al., 2011). However, the C-terminal domain is consistently reported to be important for cytoplasmic localisation, as well as for CSR (Barreto et al., 2003, Ta et al., 2003), and deletion mutants of this part of the protein show a relocalisation to the nucleus when overexpressed in fibroblasts (Ito et al., 2004, McBride et al., 2004). Thus, it is emerging that a balance of signals and pathways determines the subcellular concentrations of AID.

1.3 Molecular mechanisms of AID

1.3.1 Class Switch Recombination, Somatic Hypermutation and Gene Conversion

AID introduces dU into ssDNA, a lesion that represents a foreign base in DNA as well as a mismatch when dsDNA is reformed. This lesion can be processed by error-free (repaired) or error-prone mechanisms. *Ig* diversification relies on error-prone processing of the dU, leading to mutation or recombination (reviewed in Neuberger et al., 2003). Deamination by AID is the initiating event and central requirement that allows *Ig* diversification processes to occur (Muramatsu et al., 2000).

During SHM (reviewed in Di Noia and Neuberger, 2007), mostly un-templated point mutations are introduced into rearranged V regions, although insertions and deletions are occasionally observed (Goossens et al., 1998). Although deamination takes place at dC, mutations at C/G and A/T basepairs are observed with equal frequency. Transitions (interchange within purine or within pyrimidine bases, e.g. dC to dT) are slightly more frequent than transversions (interchange of a purine with a pyrimidine base, e.g. dC to dA) (reviewed in Di Noia and Neuberger, 2007). Various replication, DNA damage response and DNA repair pathways are known to be involved in generating different classes of mutations (Figure 4) (Neuberger et al., 2003, Odegard and Schatz, 2006). The most simple handling of the lesion that can occur is replication across the dU. Here, a dA nucleotide pairing with dU is inserted on the opposite strand by a replicative DNA polymerase, which has been termed phase Ia of SHM (reviewed in Neuberger et al., 2003). If the dU nucleotide is removed by the enzyme uracil-DNA glycosylase (UNG), an abasic site is created, so that replication across the site (via error-prone translesion polymerases) can give rise to both transition and transversion mutations, termed phase Ib of SHM. In addition to activating base-excision repair (BER) via UNG, a U/G mismatch can also recruit the mismatch repair (MMR) machinery (Cascalho et al., 1998, Rada et al., 2004, Wilson et al., 2005). This allows for mutations at A/T near the initiating U/G lesion, termed phase II, and is carried out through an error-prone patch repair process by DNA polymerase eta (Polη) (Delbos et al., 2007, Delbos et al., 2005) (Figure 4).

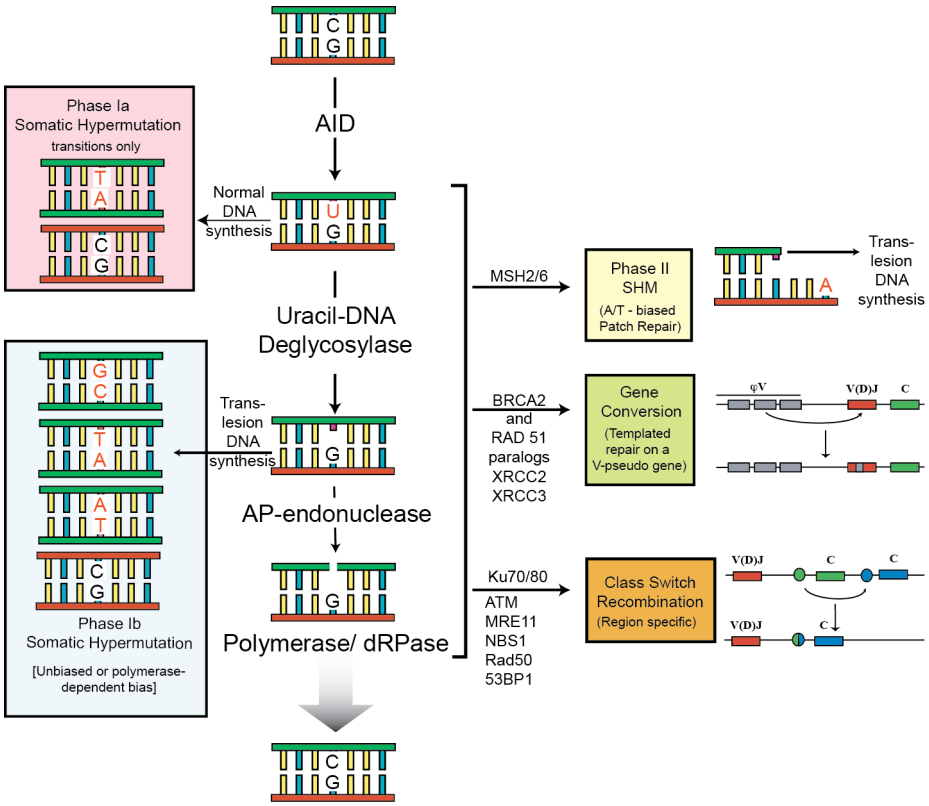


Figure 4. Model for AID-induced uracil lesion processing to SHM, iGC and CSR.

After deamination of dC by AID, the U/G mismatches can be processed by UNG, MMR proteins MSH2/6, or/and replicative or translesion polymerases. These DNA repair pathways involve downstream DNA repair factors that further process the lesions into point mutations for SHM, iGC or DSB for CSR, using non-homologous end joining. SHM can be divided into phases Ia, Ib and II, according to the mechanism and mutational outcome (described in the text).

(Modified from Neuberger et al., 2003)

Cross-talk between BER and MMR pathways can exist if an additional nick or abasic site (provided by UNG) is needed for the functioning of long patch repair in MMR (Jiricny, 2006, Schanz et al., 2009). It still needs to be determined if this second lesion is required to induce the full AID dependent MMR processes, as MMR at the *Ig* locus differs from global MMR in its mutagenic outcome.

Class Switch Recombination (reviewed in Chaudhuri and Alt, 2004) splits an expressed heavy chain variable region from a constant region and joins it to a different downstream constant region (Figure 2 B), deleting the interjacent DNA. As with SHM

(Yoshikawa et al., 2002), CSR can be induced on an artificial template by overexpressing AID in fibroblasts (Okazaki et al., 2002), indicating that it is executed by general cellular factors (Figure 4). Although CSR is not sequence specific, it is restricted to switch (S) regions that contain repetitive and guanine-rich regions and are of 2-10 kb in length.

Mutations at S regions share key features of SHM, notably, point mutations, inversions and deletions (Dunnick et al., 1993, Dunnick et al., 1989, Nagaoka et al., 2002, Petersen et al., 2001). Disruption of *Ung*, *Msh2*, *Msh6*, and *Exo1* in mice decrease CSR, indicating that these proteins process the dU lesion for double strand breaks in this process, too (Rada et al., 2004, Rada et al., 2002). CSR also depends on factors for nonhomologous end-joining, including Ku proteins and DNA-PKcs (Casellas et al., 1998). Furthermore, CSR requires a number of factors involved in the DNA damage response: the cell cycle regulator ATM (Lumsden et al., 2004, Reina-San-Martin et al., 2004), the phosphorylated histone variant γ-H2AX (Reina-San-Martin et al., 2003), NBS1, the regulatory component of the MRE11/RAD50/NBS1 (MRN) DNA damage response complex (Manis et al., 2004), and the p53-binding protein 1 (53BP1) (Reina-San-Martin et al., 2004, Reina-San-Martin et al., 2003, Ward et al., 2004). The DNA damage response is an early sensor of AID-mediated DNA lesions and co-ordinates downstream DNA repair pathways that lead to CSR (reviewed in Harrison and Haber, 2006).

In avian species as well as rabbits, cattle and pigs (Butler, 1998), antibody diversity is mainly mediated through the process of *Ig* gene conversion (iGC) (Figure 4) (Arakawa and Buerstedde, 2004, Reynaud et al., 1985, Reynaud et al., 1987). Due to reduced germline diversity in the V, D and J gene segment recombination in these species, rearranged variable regions are identical or similar in all B cells prior to the iGC step (Reynaud et al., 1985, Reynaud et al., 1987). The diversification process is based on templated introduction of an alternative DNA sequence into variable regions using adjacent V pseudogenes (ψV).

A chicken B cell lymphoma line that continuously undergoes iGC (Kim et al., 1990) has been used extensively to study this form of *Ig* diversification as well as the associated DNA repair processes. DT40 is a stable, bursal derived, fast proliferating, p53 null

lymphoma cell line carrying ALV deregulated *c-myc* (reviewed in Sale, 2004). In DT40, iGC is initiated by deamination of cytosines by AID (Arakawa et al., 2002, Harris et al., 2002), followed by base removal by UNG, which creates an abasic site that can be a substrate for cleavage followed by homologous repair (Arakawa and Buerstedde, 2004). This recombination process is mediated by RAD51 paralogues XRCC2, XRCC3 or RAD51B and BRCA2 (Hatanaka et al., 2005, Sale et al., 2001). iGC promotes the copying of new sequence information from an upstream donor pseudogene (ψV) into V regions. Non-templated mutations are also found in chicken *Ig* genes, (McCormack and Thompson, 1990, Reynaud et al., 1987) representing a very limited form of SHM that lacks phase II, i.e. mutation at A/T (Sale, 2004). A similar form of limited SHM can also be found in some human lymphoma cell lines (Denepoux et al., 1997, Harris et al., 2001, Poltoratsky et al., 2001). Although the basis of this finding is not well understood, it has been suggested that in DT40, translesion synthesis by Rev1 translesion polymerase across an AID-induced dU or abasic site would explain this reduced form of SHM (Simpson and Sale, 2003).

1.3.2 Targeting of AID activity

As described above, AID access to ssDNA is tightly restricted by a variety of mechanisms affecting expression levels, protein levels and AID activity. However, some loci are targeted by AID highly above background. This includes non-*Ig* as well as *Ig* loci. Although a variety of loci are targeted, *Ig* loci accumulate mutations at rates that are orders of magnitude (10-100 times) greater than non-Ig targets (Liu et al., 2008, Pasqualucci et al., 1998, Shen et al., 1998). Non *Ig* targets of AID in normal B cells include the genes *Bcl-6*, *c-myc*, *Pim1* (Liu et al., 2008, Pasqualucci et al., 1998, Shen et al., 1998), *Cd79a* (Gordon et al., 2003), *Cd79b* (Gordon et al., 2003), *Cd95* (Muschen et al., 2000), as well as a number of additional genes (Liu et al., 2008). Furthermore, genome-wide chromatin immunoprecipitation (ChIP) shows that AID can be detected occupying genes dispersed throughout the genome, in line with its promiscuous targeting (Yamane et al., 2011). Importantly, non *Ig* targets recapitulate the nature of mutations at *Ig* by carrying AID hotspot signatures (Storb et al., 2001) and localising to promoter proximal regions (Liu et al., 2008). The distribution of frequency of SHM in *Ig* genes resembles a bell shaped curve (depicted in Figure 2 B), with a peak closely

downstream of the promoter (Crews et al., 1981, Kim et al., 1981, Rada and Milstein, 2001, Rada et al., 1997). The first ~150 bp downstream of the transcription start site (TSS) are devoid of mutation and mark the 5' boundary (Rada et al., 1997). Mutation continues throughout a 1-1.5 kb region (Rada and Milstein, 2001 and references within), with a less sharp 3' boundary but excluding C regions. These AID targeting boundaries are retained even in the absence of *Ung* (Longerich et al., 2005) and thus any error-free repair, indicating that it is AID access that is restricted to this domain. Similarly well defined mutation regions and a comparable distribution can be observed in S regions when *Ung* and M*sh2* are absent (Xue et al., 2006).

1.3.2.1 Transcription and AID targeting

Transcription has long been associated with AID-induced immune diversification. The rate of transcription correlates with the rate of SHM (Peters and Storb, 1996), and germline transcription through the switch and the constant region precedes CSR (Stavnezer-Nordgren and Sirlin, 1986). In CSR, switch isotypes are chosen by transcriptionally activating the appropriate S regions through cytokines and other signals. Each S region responds to a specific cytokine combination (reviewed in Chaudhuri and Alt, 2004, Pavri and Nussenzweig, 2011). Thus, transcription has been recognised as a main signal for CSR (Shinkura et al., 2003). Antisense transcription has also been found in switch regions (Apel et al., 1992, Perlot et al., 2008), but it still needs to be determined if this contributes to DSB generation and/or CSR. It is not known if transcriptional targeting could be involved in AID dependent demethylation.

Early transgenic work demonstrated that the *Ig* promoter or enhancer elements are necessary for SHM (Betz et al., 1994). Furthermore, when a 750 bp transgene is inserted directly downstream of the promoter in the *IgH* locus, mutations are shifted upstream accordingly, staying associated with the promoter (Tumas-Brundage and Manser, 1997). These and other studies establish that transcription of a gene is necessary, but not sufficient for AID-induced immune diversification (Betz et al., 1994, Goyenechea et al., 1997, Klix et al., 1998, Kothapalli et al., 2011, Yamane et al., 2011). Additional factors such as location and chromatin configuration may also play a significant role. *In vivo*, AID equally deaminates both the transcribed (TS) and non-transcribed (NTS) strands of transcribed target DNA sequences (Rada et al., 2004). However, transcription supports deamination of the NTS when AID is expressed in

yeast, in bacteria and when a simplified *in vitro* system employing T7 polymerase is used (Chaudhuri et al., 2003, Dickerson et al., 2003, Gomez-Gonzalez and Aguilera, 2007, Petersen-Mahrt et al., 2002, Ramiro et al., 2003, Sohail et al., 2003). Furthermore, a slight strand bias of the final mutation patterns has been observed, arguing for a differential processing of mutations on either strand (Di Noia and Neuberger, 2007, Franklin and Blanden, 2008). The mechanism by which AID deaminates the transcribed strand (TS) is unknown, although different theories have been proposed. ssDNA on the TS could be generated during antisense transcription. Antisense transcripts have been detected at low levels in S regions, but it is not known if they participate in hypermutation. Antisense transcription can potentially deliver AID to the antisense RNA polymerase II machinery, or the required chromatin modifications could be deposited. More recently it has been proposed that targeting of the TS may be helped by the exosome complex (Basu et al., 2011).

The sequence of the target gene per se does not influence SHM, as V region replacements and artificial constructs can be substrates for SHM (Betz et al., 1994, Klotz et al., 1998, Peters and Storb, 1996, Yelamos et al., 1995). This is contrary to VDJ recombination, where specific sequence motifs play a major role. Although *cis* regulatory elements such as enhancers are needed to establish transcription, it is debated if they can regulate AID by direct recruitment. A definitive proof for sequence requirement has been lacking (Betz et al., 1994, Blagodatski et al., 2009, Inlay et al., 2006, Kim and Tian, 2009). E box transcriptional motifs have been suggested but may only act on the transcription levels (Conlon and Meyer, 2006, Michael et al., 2003, Staudt and Lenardo, 1991, Tanaka et al., 2010). Secondary structural motifs such as R-loops (Huang et al., 2007, Yu et al., 2003), G quadruplex structures (Duquette et al., 2004) or stem loops (Tashiro et al., 2001) have been suggested to play a role in CSR regulation, either to allow for AID targeting, substrate availability, or lesion resolution (Chaudhuri and Alt, 2004, Odegard and Schatz, 2006). In the V region for SHM, no such structures exist, although they have been postulated.

1.3.3 Resolution of the AID lesion

Initially, it had been assumed that AID's activity on the *Ig* loci is the key step in achieving the high specificity of SHM. Subsequent analysis identified that the way in

which the dU lesion is resolved contributes significantly to the overall outcome of SHM. Although AID seems to be targeted preferentially to the *Ig* locus, other loci also undergo AID dependent mutation. Importantly, once BER and MMR is removed from the system, those targets have a different mutation outcome and additional, normally repaired targets are revealed (Liu et al., 2008, Liu and Schatz, 2009). Moreover, excess levels of AID do not alter the much elevated level of targeting of the *Ig* locus compared to other loci (Robbiani et al., 2009). This indicates that not all loci are treated equally in terms of DNA repair fidelity. Therefore, the repair of deaminated dC generated by spontaneous DNA deamination versus those that are dependent on AID presence can differ considerably. AID itself may have a role in co-ordinating lesion resolution (Peled et al., 2008, Reynaud et al., 2003). This form of regulation is evident when comparing CSR versus SHM for specific AID cofactor association, which can trigger different response pathways. For instance, the AID associated protein KAP1, a chromatin regulator, is relevant for CSR but dispensable for SHM (Jeevan-Raj et al., 2011). Also, the C-terminal domain of AID itself is required for CSR but not for SHM, as described above.

1.3.3.1 DNA damage and DNA damage response pathways

The repair of DNA damage is essential for the survival of an organism. It has been estimated that daily, each cell experiences 20,000 to 30,000 DNA lesions of various types, with DNA repair efficiently processing all of them (Lindahl, 1993).

DNA damage encompasses a large variety of chemical and physical alterations of DNA and chromatin. Generated either by exogenous or endogenous (including AID) sources, these lesions include alkylations, oxidations, deaminations, depurinations, single stranded nicks, double strand breaks, as well as intra- and inter-strand crosslinks (reviewed in Lindahl, 1993). AID-induced lesions can initiate various DNA repair pathways, depending on the cellular context as well as the location of the lesion, however, the choice of which DNA repair pathway is used, is still poorly understood. Moreover, it is not understood how at AID targeted loci, DNA repair differs from conventional DNA repair in order to generate mutagenic outcomes. Below, pathways that play a role in the resolution of AID initiated dU lesions are described in their physiological function.

1.3.3.2 Pathways and proteins involved in dU lesion resolution

1.3.3.2.1 Base excision repair (BER)

BER recognises, excises, and accurately replaces specific forms of base modifications (reviewed in (Barnes and Lindahl, 2004). The first step of BER is based on the recognition of an inappropriate base, such as uracil, 3-methyladenine, 8-oxoguanine, or formamidopyrimidines. Each particular lesion is recognised by at least one specific DNA glycosylase (Lindahl and Wood, 1999) that has the ability to excise the altered base. The majority of these enzymes catalyse hydrolysis of the N-glycosylic bond, resulting in an abasic site with an intact DNA phosphodiester backbone. This is a substrate for apurinic endonuclease 1 (APE1), which both incises immediately 5′ to an abasic site and removes 3′-obstructive termini (Demple and Sung, 2005, Wilson and Barsky, 2001). In short-patch BER, DNA polymerase β (Polβ) replaces the excised nucleotide and removes the 5′-terminal abasic fragment left behind by APE1 incision (Bennett et al., 1997; Mol et al., 2000). In long-patch BER, polymerases δ and ε (Polδ and Polε) re-sythesise multiple nucleotides to replace the faulty stretch of DNA, and the displaced 5’ flap is cleaved by the FEN1 flap endonuclease. The remaining nick is sealed by either a XRCC1-DNA ligase III heterodimer, or DNA ligase 1 (LIG1) (McKinnon and Caldecott, 2007).

1.3.3.2.2 Mismatch repair (MMR)

MMR is designed to remove erroneous DNA structures such as mismatches and other bulky DNA lesions from newly synthesised DNA (reviewed in Jiricny, 2006). This process is based on the ability of mismatch-recognition components: the MSH2/MSH6 heterodimer, which will recognise mismatches, and the MSH2/MSH3 heterodimer, which will act on loop lesions (Jiricny, 2006). Repair is then activated by the recognition and binding of the mismatch complex by the MLH1-PMS2 heterodimer, as well as other DNA repair proteins, including the polymerase cofactor and sliding clamp PCNA and exonucleases. The machinery is thought to recognise and excise the newly synthesised strand utilising the presence of nicks (Iams et al., 2002). These boundary elements serve as a starting point for the degradation of several hundred basepairs of the error-containing strand by exonucleases (EXO1) and associated factors. After the mismatch has been removed, resynthesis of the degraded region by DNA polymerase,

followed by sealing of the remaining nick by DNA ligase, completes the repair process (Kunkel and Erie, 2005).

In AID dependent *Ig* diversification, a central role has been assigned to MSH2 and MSH6 as well as EXO1, but ablation of MSH3, MLH1 or PMS2 has only minor effects on the process. This suggests that MMR proteins may be co-opted into a related but mutagenic pathway at *Ig* loci (Bardwell et al., 2004, Cascalho et al., 1998, Ehrenstein et al., 2001, Frey et al., 1998, Martomo et al., 2004, Phung et al., 1999, Phung et al., 1998, Rada et al., 1998, Wiesendanger et al., 2000).

1.3.3.2.3 DNA polymerases and DNA sliding clamps

Mammals have four replicative DNA polymerases (Polα, Polγ, Polδ and Polε), whereas there is also a collection of other polymerases that perform non-replicative tasks (Weill and Reynaud, 2008). These functions involve the removal and repair of damaged bases (for instance, Polβ in BER) or the bypass of DNA lesions that block the progression of replication forks. The latter, also known as translesion synthesis (TLS) (Jansen et al., 2007), is performed amongst others by the Y family polymerases. These polymerases achieve replication directly across damaged DNA templates by means of a catalytic domain that can accommodate aberrant basepairs. Although this allows the polymerase to synthesise directly across DNA lesions, the inherent lack of proofreading activity renders TLS polymerases error-prone, even in the presence of an intact template. Of the several known TLS polymerases, only Y family polymerases Polη (which preferably misreads A/T bases) and Rev1 (a deoxycytidyl transferase) have been demonstrated to be involved in somatic hypermutation (Reynaud et al., 2009).

In vitro, most DNA polymerases are weakly processive without cofactors (Garg and Burgers, 2005, Hubscher et al., 2002). Auxiliary factors, such as the trimeric sliding clamp PCNA, are tethered to the DNA and used to reduce the dissociation of the polymerase from the template. Interaction between PCNA and the DNA polymerases increases the processivity of the polymerase by many orders of magnitude (Waga and Stillman, 1998b). Association of TLS polymerases with PCNA is dependent on monoubiquitination of lysine 164 (K164) of PCNA by the DNA damage-inducible Rad6 E3 ubiquitin ligase. The interaction between the PCNA and the TLS polymerase

is enhanced via the polymerases ubiquitin binging domain (Bienko et al., 2005, Guo et al., 2008, Guo et al., 2006, Kannouche et al., 2004).

A PCNA related trimeric DNA clamp is the 9-1-1 clamp, a heterotrimer consisting of RAD9, HUS1 and RAD1 (Eichinger and Jentsch, 2011). This complex serves predominantly as a DNA damage response and checkpoint protein. 9-1-1, too, can carry post-translational modifications and bind cofactors that are part of the DNA damage response. Notably, the 9-1-1 clamp has been examined for a role in AID dependent *Ig* diversification. In DT40, presence of 9-1-1 is required for iGC, but not for TLS (Saberi et al., 2008). This finding was confirmed by a mouse mutant of residue K185 in RAD1, a functionally homologous residue to PCNA K164, that showed no defect in TLS in *Ig* genes (Wit et al., 2011).

1.3.3.2.4 DNA damage tolerance pathway

DNA damage tolerance (DDT, also known as post-replication repair) is a mechanism for the bypass of DNA lesions without repair, and it is carried out by a set of either error-prone or error-free processes (Chang and Cimprich, 2009). DDT enables DNA replication to be continued even when DNA damage is present, by overcoming damage induced replication fork stalling. Thus, DDT is an essential component of the overall cellular response in surviving genotoxic stress. The error prone branch is based on TLS, and the error-free branch involves template switching. The pathways were initially identified as the Rad6 epistasis group, as they depend on the E2 ubiquitin conjugating enzyme Rad6 and the E3 ubiquitin ligase Rad18 (Hoege et al., 2002, Stelter and Ulrich, 2003). They are also dependent on PCNA, which fulfils a co-ordinative role in DDT (Moldovan et al., 2007). DNA damage induced monoubiquitination of K164 by RAD6 acts as a switch from the replicative to a TLS polymerase (Hoege et al., 2002, Bienko et al., 2005). Further K63-linked polyubiquitination of K164 PCNA stimulates template switching, enabling stalled replicative polymerases to bypass the lesion by switching transiently to the intact template strand of the sister chromatid (Chang and Cimprich, 2009).

1.3.4 AID and cancer

As AID is designed to introduce genomic instability into the cellular DNA, misregulation can easily lead to mutagenic and carcinogenic outcomes. This can occur

while AID is carrying out its natural function, leading to B cell malignancies, or when AID is activated out of context.

AID has been directly implicated in tumourigenesis of lymphocytes as well as cells outside of the immune system. Many B cell malignancies bear chromosomal translocations that join *Ig* loci and portions of proto-oncogenes as a result of aberrant CSR (Chiarle et al., 2011, Klein et al., 2011, Kuppers and Dalla-Favera, 2001, Ramiro et al., 2004, Robbiani et al., 2009). The *c-myc* oncogene frequently fuses to the *Ig* promoter at switch region breakage points in Burkitt's lymphoma (Pasqualucci et al., 2001, Rabbitts et al., 1984). This translocation may also be facilitated by the common nuclear localisation of *Ig* and *c-myc* in transcription factories (Osborne et al., 2007). Moreover, additional translocation partners of *Ig* and *c-myc* in lymphocytes tend to be actively transcribed genes (Chiarle et al., 2011, Klein et al., 2011). Mistargeted SHM has been reported in normal B cells as a by-product of AID activity as mentioned above, but has also been found to contribute to cancer development (Kuppers and Dalla-Favera, 2001, Muschen et al., 2000, Pasqualucci et al., 1998, Pasqualucci et al., 2001, Shen et al., 1998). Frequently, oncogenes such as *Bcl-6*, *Fas*, *c-myc*, *Pim1* are targeted. Apart from Burkitt's lymphoma, AID may also be required for tumour progression in other leukaemias such as chronic myeloid leukaemia (Klemm et al., 2009), where AID promotes clinical progression as well as drug resistance. Aside from AID mistargeting, AID overexpression may also contribute to oncogenesis. Systemic overexpression of AID clearly leads to tumour susceptibility in mice (Okazaki et al., 2003). Interestingly, B cells were less affected by tumour development, as they may have a natural protection against AID-induced lesions by expressing a specific DNA repair protein repertoire (Klemm et al., 2009). Indeed, in overexpression experiments, AID causes mutation in genes that are not usually mutated in normal B cells (Martin et al., 2002, Okazaki et al., 2003, Yoshikawa et al., 2002). As described above, AID can be activated via NF-κB. Apart from physiologically controlling AID expression, this pathway can also be tumourigenic (reviewed in Ben-Neriah and Karin, 2011). In *Helicobacter pylori* infections, AID is continuously activated via NF-κB and may contribute to gastric cancer (Matsumoto et al., 2007). Similarly, the PAX5 transcription factor has importance as an AID regulator (Gonda et al., 2003) but is also implicated in lymphomagenesis (Busslinger et al., 1996, Iida et al., 1996, Morrison et al., 1998,

Pasqualucci et al., 2001). Notably, *PAX5* is also a target for aberrant somatic hypermutation by AID (Pasqualucci et al., 2001). Hormones, specifically oestrogen, have long been known to be cancer causing agents (reviewed in Petersen-Mahrt et al., 2009). The finding that oestrogen directly activates AID (Pauklin et al., 2009) provides the most direct link between hormones and DNA mutagenesis to date.

1.4 Aims of the Study

There are important open questions in AID biology: How is AID targeted to induce DNA alterations at specific loci, and restricted from others? More specifically, which role does transcriptional targeting and differential lesion resolution play?

As co-factors of AID are likely to be important to regulate targeting, activity and lesion resolution, I used two approaches to gain insight into the AID interaction network:

Aim I: Characterising the AID interaction with transcription elongation factors

In order to understand complexes that AID associates with on chromatin, an unbiased biochemical approach was used to isolate AID and associated factors with a focus on the emerging connection between AID and the transcription machinery.

Aim II: Characterising the AID interaction with PCNA

To understand how AID can influence the DNA repair pathway downstream of lesion generation, the connection between AID and PCNA was examined.

1.4.1 Introduction for Aim I – transcription related mechanisms

1.4.1.1 The RNAPII *transcriptional cycle*

RNA polymerase II (RNAPII) transcription of mRNA is known to undergo five distinct phases, initiation, promoter escape, promoter-proximal pausing, elongation and termination (reviewed in Saunders et al., 2006). During each phase, phosphorylation of the heptamer repeats in the C-terminal domain (CTD) tail of the largest RNAPII subunit is of regulatory importance (reviewed in Buratowski, 2009). Phosphorylation can occur at serine 2 (S2) and serine 5 (S5) of the CTD, which allows for key proteins to associate and dissociate with the holocomplex (Buratowski, 2009). The transition from initiation to elongation takes place between the transcription start site (TSS) and 50 - 150 bp downstream. Processive elongation correlates with increasing phosphorylation of S2 of the CTD, while the earlier stages are associated with S5 phosphorylation. The CTD is fully S2 phosphorylated after approximately 1,000 bases downstream of the promoter of coding genes (Mayer et al., 2010).

1.4.1.2 Promoter-proximal pausing

Promoter proximal pausing of RNAPII can be defined as a state where the enzyme is recruited to a promoter, forms an initiation complex and clears the promoter to engage in transcription, but pauses proximal to the transcriptional start site. Upon gene-specific activation, RNAPII can be quickly activated to elongate and transcribe the gene. Promoter-proximal pausing represents a checkpoint before committing to productive elongation that is highly regulated (Saunders et al., 2006).

RNAPII pausing is well defined on heat shock genes in *Drosophila*, where it was shown for the first time at promoters of genes encoding heat shock proteins (Lis and Wu, 1993, Saunders et al., 2006). Other genes were also shown to be controlled downstream of RNAPII recruitment, such as *junb*, *c-myc* and *Igκ* (Aida et al., 2006, Krumm et al., 1992, Raschke et al., 1999, Strobl and Eick, 1992). The mechanism has now been recognised as a genome wide method of transcriptional control, for instance in development (Zeitlinger et al., 2007). In humans, stalled polymerase could also be detected genome-wide on genes (Kim et al., 2005), and was revealed to control oestrogen receptor regulated genes in human breast tissue (Aiyar et al., 2004).

Furthermore, in embryonic stem cells (Rahl et al., 2010) and activated B cells (Yamane et al., 2011), the vast majority of genes accumulate stalled RNAPII.

1.4.1.3 Transcription elongation control, elongation factors and transcriptional chromatin marks

Apart from transcription initiation complex assembly, processive elongation can be a rate limiting step in mRNA biogenesis, with numerous auxiliary factors being required (reviewed in Selth et al., 2010). Basal elongation factors are essential for elongation *in vitro*, and chromatin elongation factors help RNAPII transcribe through nucleosome templates.

Elongation barriers such as repetitive sequences, competition with other DNA metabolism processes or programmed pausing sites can induce RNAPII pausing, frequently in the promoter-proximal region as described (Gilmour and Fan, 2009, Lis, 2007). Regulated pausing requires specific factors, such as the 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor (DSIF) (Figure 5 A) (Swanson et al., 1991, Wada et al., 1998, Yamaguchi et al., 1999b) and the multisubunit negative elongation factor (NELF) (Yamaguchi et al., 1999a). DSIF is a heterodimer composed of Suppressor of Ty homologue 4 and 5 (SUPT4H and SUPT5H). Stalled RNAPII can be reactivated by the pTEFb kinase (Figure 5 A) (Fujinaga et al., 2004, Yamada et al., 2006), which phosphorylates SUPT5H and RNAPII, releasing the transcription complex. SUPT5H stays associated with RNAPII throughout the elongation stage, contacting additional elongation factors such as the PAF complex. The roles of the PAF complex are to communicate with transcriptional activators, the recruitment and activation of histone modification factors, and the recruitment of cleavage and polyadenylation factors to RNAPII (reviewed in Jaehning, 2010), but it is also a direct transcription elongation activator (Chen et al., 2009). This complex consists in humans of the subunits PAF1, CTR9, LEO1, CDC73, SKI8 and RTF1 (Figure 5 A). The PAF complex recognises phosphorylation of the CTD and of SUPT5H, and then initiates a cascade of histone modifications. The starting point of this cascade is Histone H2B ubiquitination, a mark set by the PAF complex by tethering the Histone H2 ubiquitination machinery (RAD6/BRE1) to the RNAPII machinery, promoting cotranscriptional chromatin remodelling (Kim et al., 2009). Another

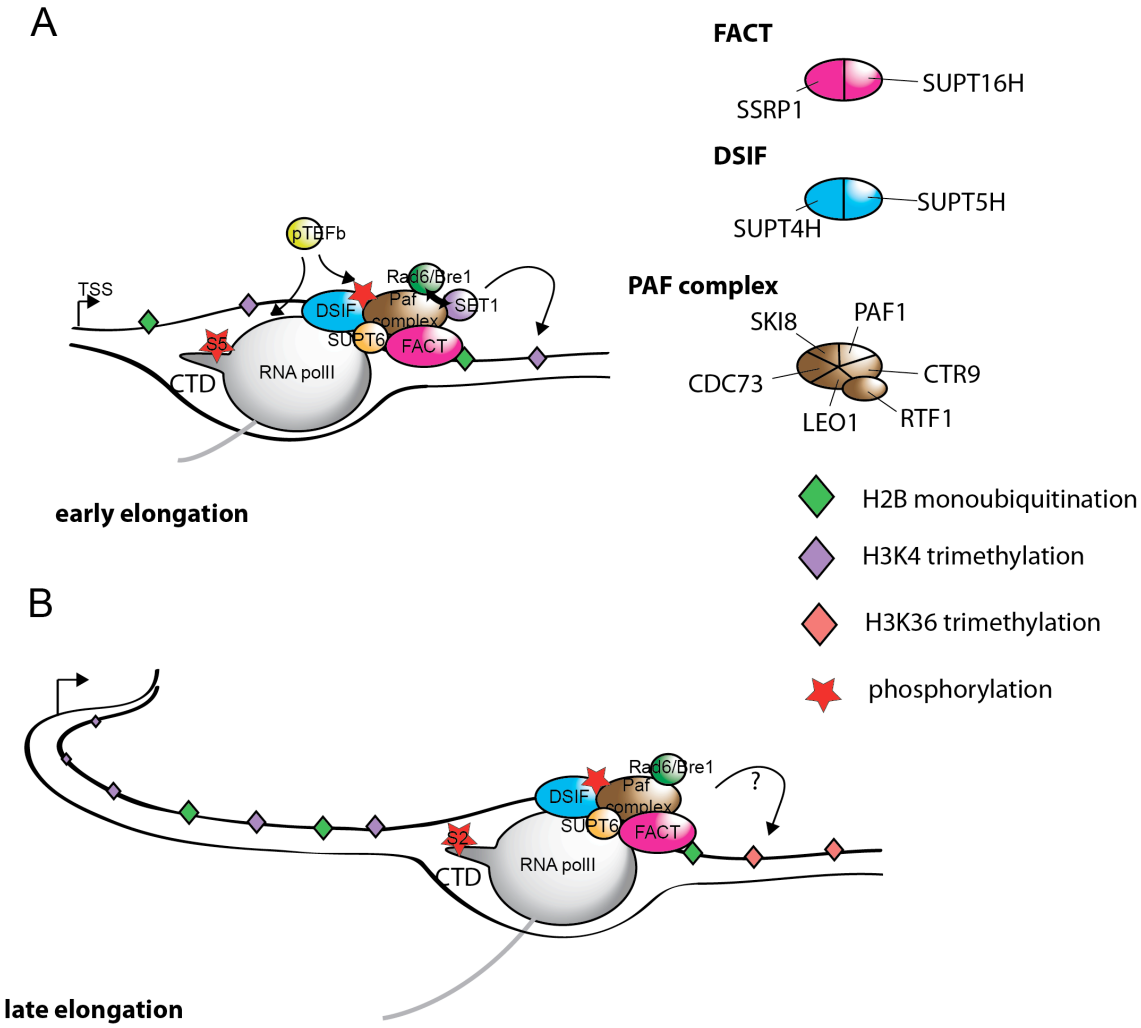


Figure 5. Transcription elongation and elongation promoting factors.

A, Elongating RNAPII proximal to the TSS is serine 5 phosphorylated at the CTD. The factors associated at this stage are indicated as ellipses. H2B monoubiquitination-dependent H3K4 trimethylation is the predominant chromatin mark. On the right hand side, the subunits of various elongation complexes are indicated. The RTF1 subunit is not as stably associated to the PAF complex as the other five subunits (Jaehning, 2010). **B**, RNAPII in later elongation stages is phosphorylated at serine 2 of the CTD, and the predominant chromatin mark is H3K36 trimethylation.

chromatin elongation factor, SUPT6H, is a histone chaperone and aids transcription through nucleosomes. The FACT complex (comprised of subunits SSRP1 and SUPT16H), also travels with the polymerase and has an ability to facilitate transcription by destabilising histone nucleosomes and by enhancing co-transcriptional chromatin marks (Pavri et al., 2006). The H3K4 trimethylation is a mark promoted by the FACT

complex and set by the PAF complex (Kim et al., 2009), which can also tether the enzyme complex SET1 to RNA polymerase. This mark is in general restricted to the start of a gene, and replaced by another co-transcriptional mark, H3K36 trimethylation, further towards the 3' end of the gene. The role of the PAF complex in H3K36 trimethylation is unknown in mammals (reviewed in Newey et al., 2009), but a dependency of this mark on the PAF complex and H2B monoubiquitination has been found in yeast (Krogan et al., 2003) and plants (Xu et al., 2001). There are numerous other cotranscriptional chromatin marks such as Histone H3 acetylation, which is strongly associated with active and accessible chromatin.

1.4.1.4 Links between AID dependant mutagenesis and transcription

1.4.1.4.1 Known connections to elongation factors

While the connection between AID targeting and transcription have been recognised for almost two decades, the molecular details of this connection are now emerging. Recent findings support the importance of transcription elongation and associated processes for AID targeting, as various RNAPII associated factors have been identified as cofactors for AID. These transcription elongation factors comprise basal elongation factors such as SUPT5H (Pavri et al., 2010) and exosome components [a nascent mRNA processing and degradation complex (Basu et al., 2011)], while elongation factors needed to transcribe through chromatin (chromatin elongation factors) such as SUPT6H (Okazaki et al., 2011) and the FACT complex (Stanlie et al., 2010) have also been implicated.

SUPT5H was identified by a shRNA screen for factors reducing CSR in B cells (Pavri et al., 2010), and was subsequently found to interact, although weakly, with AID. ChIP experiments showed the co-localisation of SUPT5H with AID on *Ig* genes and other potential AID targets, while confirming the connection between SUPT5H and stalled RNAPII. Finding stalled polymerase on *Ig* genes was in line with previous reports that document RNAPII pausing on *Igκ* (Raschke et al., 1999), on Sμ switch regions (Rajagopal et al., 2009) and other S core domains (Wang et al., 2009, Yamane et al., 2011). These data led to the conclusion that stalled RNAPII could be necessary to recruit AID, whereas it was not clear weather this would happen through a promoter-proximal pausing step or through extended pausing throughout the gene. Interestingly, SUPT5H was mapped throughout the *Ig* loci chromatin, even extending well into c

regions where AID-induced mutation/recombination is excluded (Pavri et al., 2010). Therefore, apart from simple RNAPII pausing, additional signals for AID recruitment, exclusion or activation are likely to occur.

1.4.1.4.2 Chromatin modifications on the *Ig* locus

Chromatin marks are another factor that is thought to influence AID targeting, activity and lesion resolution, either directly or indirectly (e.g. through transcription). In line with this, the histone chaperone SUPT6H has been shown to interact with AID, and its knock-down to have an adverse effect on CSR (Okazaki et al., 2011). Similarly, the FACT complex has been shown to be required for CSR, possibly by promoting the histone modification cascade that leads to H3K4 trimethylation of the *Ig* gene in B cells (Stanlie et al., 2010). This mark has been shown to be important for CSR in B cells and is enriched at the *Ig* locus (Wang et al., 2009). Knock-down of the enzyme complex that is able to transfer this mark to histones (SET1) has shown (Stanlie et al., 2010) a reduction both in H3K4 trimethylation at switch regions and in CSR. Both the H3K4 and the H3K36 co-transcriptional trimethylation marks are induced upon transcriptional activation of S-regions (Wang et al., 2009), whereby the H3K4 trimethylation domain overlaps better with the AID targeted region (Wang et al., 2009).

Other histone marks that have been correlated temporally and spatially to CSR and/or SHM include H2B phosphorylation on serine 14 (Odegard et al., 2005) and H4K20 methylation (Schotta et al., 2008), both DNA damage related marks; H3 acetylation (Kuang et al., 2009, Nambu et al., 2003, Wang et al., 2009), a hallmark of active chromatin; and H3K9 trimethylation, a mark generally associated with silenced genes (Chowdhury et al., 2008, Jeevan-Raj et al., 2011, Kuang et al., 2009, Wang et al., 2006, Wang et al., 2009).

1.4.2 Introduction for Aim II – biology of PCNA

1.4.2.1 PCNA is a major player in DNA replication, repair and damage response

As described above, processive DNA polymerases require cofactors like the sliding clamp PCNA, a trimeric ring protein that is loaded via the RFC clamp loader onto DNA at junction regions of ssDNA and dsDNA (Majka and Burgers, 2004) with the DNA threading through the ring (Figure 6 A) (Krishna et al., 1994).

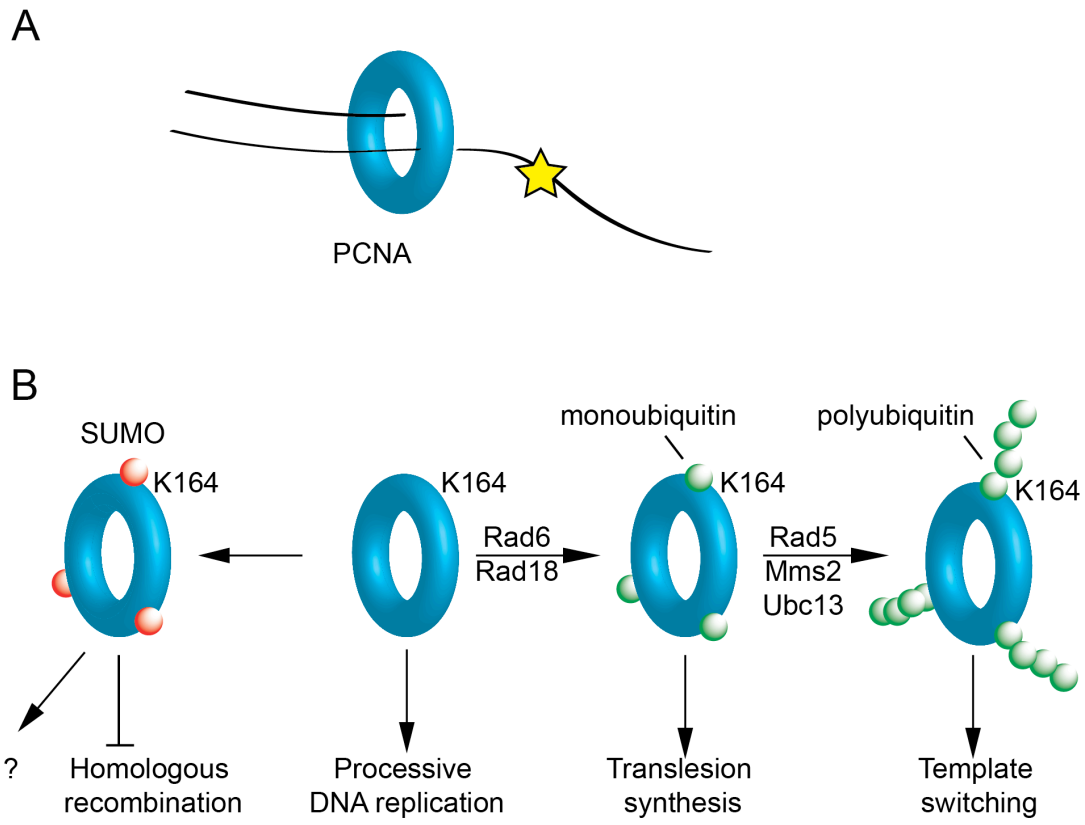


Figure 6. PCNA, its modifications and the pathways they trigger.

A, PCNA is a ring shaped trimer loaded on DNA to function as a sliding clamp, processivity factor and docking protein for DNA metabolism processes. DNA damage is depicted as a yellow star. **B**, PCNA can be sumoylated (red sphere), monoubiquitinated (green sphere) or polyubiquitinated (green chain). These modifications trigger a block of homologous recombination, translesion synthesis or template switching. Sumoylation may have additional unknown functions.

Outside DNA replication, PCNA plays an important role in DNA repair (Jonsson and Hubscher, 1997). All repair pathways that confer replacement of damaged DNA by DNA polymerases, including MMR, DDT, TLS, HR, long patch BER as well as nucleotide excision repair (NER) require PCNA as a cofactor. For instance, upon UV irradiation, PCNA immobilises on chromatin in the nucleus, enabling the repair of damaged DNA by contributing to chromatin bound repair complex formation (Bouayadi et al., 1997, Miura et al., 1992). A myriad of proteins involved in replication and the maintenance of DNA interact directly or indirectly with PCNA (Moldovan et al., 2007), making it a docking platform for a wide variety of pathways. Examples of interacting

proteins that are relevant for AID dependant processes are BER protein UNG, MMR protein MSH2, nucleases EXO1 and FEN1, and chromatin factors such as DNA methylase DNMT1.

1.4.2.2 PCNA modification and DDT

Proteins competing for PCNA binding can influence repair pathway choice, therefore this choice can be altered by modifications of PCNA. When PCNA is loaded onto the DNA near DNA damage, the modification of PCNA at K164, monoubiquitination, polyubiquitination or sumoylation can influence replication and DDT (Figure 6 B) (Moldovan et al., 2007, Ulrich, 2009). For instance, in yeast, PCNA is monoubiquitinated upon replication fork stalling or single strand gap exposure, caused by DNA lesions (reviewed in Ulrich, 2009), enabling TLS by recruiting error-prone translesion polymerases (Hoege et al., 2002, Bienko et al., 2005). Of note, in higher eukaryotes, TLS or polymerase switch can occur independent of PCNA K164 monoubiquitination (Edmunds et al., 2008, Hendel et al., 2011, Mirchandani et al., 2008). The DDT pathway has mostly been studied in yeast (Branzei et al., 2008), but the factors defining the modification state of PCNA and the responses that are triggered are more complex in vertebrates (Ulrich, 2009).

Monoubiquitination of PCNA seems to fulfil a similar role in higher eukaryotes as in yeast, activating TLS. However, in vertebrates not all PCNA monoubiquitination is Rad18-dependenent, and other E3 ligases can provide a residual pool of monoubiquitinated PCNA (Simpson et al., 2006, Terai et al., 2010, Arakawa et al., 2006).

In yeast and vertebrates, polyubiquitination of PCNA is a subpathway of the DDT pathway dependent on another set of ubiquitin conjugating enzymes (Mms2-Ubc13-Rad5 in yeast), which have more diversified human homologues (Motegi et al., 2008, Motegi et al., 2006, Oh et al., 1994, Xiao et al., 1998). This pathway is not well studied, but represents an error-free damage bypass pathway relying on template switching. The molecular mechanisms are still debated (Branzei et al., 2008, Broomfield et al., 2001), but the modification initiates a template switch using the intact sister chromatid and seems to proceed via a HR-related mechanism (Lehmann and Fuchs, 2006, Torres-Ramos et al., 2002, Vanoli et al., 2010).

In yeast, sumoylation of PCNA at K164 is damage-independent and has been shown to recruit the antirecombinogenic helicase Srs2 (Papouli et al., 2005, Pfander et al., 2005), possibly to protect cells in S-phase from toxic recombination intermediates. It has also been proposed to be an early step in the polyubiquitin-dependant HR DDT pathway (Branzei et al., 2008). An Srs2 homologue that would bind to sumoylated PCNA does not exist in vertebrates. Therefore, although sumoylation of PCNA could be proven in the DT40 system (Arakawa et al., 2006) and *Xenopus laevis* egg extract (Leach and Michael, 2005) the exact role of PCNA sumoylation in vertebrates is yet unknown.

1.4.2.3 PCNA K164 has a phenotype in AID dependent DNA rearrangements - The K164 mutants

Several studies have addressed the role of PCNA in AID dependent *Ig* rearrangement mechanisms (Arakawa et al., 2006, Langerak et al., 2007, Roa et al., 2008). As PCNA knock-out cells are not viable, these studies use a PCNA mutant (K164R) defective in modification, to observe the effects of disturbing various DNA repair and DDT pathways.

Studies in DT40 cells (Arakawa et al., 2006) and mice (Langerak et al., 2007, Roa et al., 2008) showed defects in AID-dependent *Ig* hypermutation in PCNA K164R mutants. In the mouse studies, where a classical phase II of SHM can be observed, K164R mutation leads to omission of mutations at A and T (corresponding to phase II mutations), while the overall mutation load was not (Roa et al., 2008) or little (Langerak et al., 2007) altered. This was interpreted as an effect of the disruption of PCNA's ability to interact with DNA Polη, the polymerase responsible for the bulk of phase II SHM. In DT40, the PCNA K164R mutation disrupts a G/C focused, rudimentary form of SHM. In contrast to the mouse models, in the DT40 mutant there was also a large overall drop in hypermutation activity (Arakawa et al., 2002).

On the effect of PCNA mutation in CSR, conflicting conclusions were made: While (Langerak et al., 2007) did not observe an effect on CSR, in (Roa et al., 2008) the K164R allele altered CSR.

Chapter 2. Materials and Methods

2.1 Materials

Note:

All chemicals were obtained from Sigma except where indicated.

All media or buffers marked with "*" were prepared by CRUK media kitchen.

2.1.1 Blocking solution for streptavidin plates

3% BSA, 1x Roche blocking reagent in 40 mM HEPES and 120 mM NaCl.

2.1.2 Buffer 2 (AID extraction buffer)

20 mM MES (pH 6.0), 100 mM NaCl, 50 mM KCl, 5 mM β-mercaptoethanol, 1.6 mM CHAPS, 300 mM L-arginine HCl, filtered and adjusted to pH 6.

2.1.3 Buffer A (PCNA loading buffer)

40 mM HEPES, 8mM $MgCl_2$, 0.1mg/ml BSA, 1mM DTT, 120 mM NaCl

2.1.4 Buffer C (HTNCre purification)

600 mM NaCl, 20 mM HEPES (pH 7.4)

2.1.5 Buffer D (HTNCre purification)

50% glycerol, 500 mM NaCl, 20 mM HEPES (pH 7.4)

2.1.6 Buffer R for AID activity assay

5 x Buffer R: 250 mM NaCl, 200 mM KCl, 15 mM $MgCl_2$, 200 mM Tris-HCl (pH 8), 5 mM DTT, 50% Glycerol (filtered and stored at 4 °C).

2.1.7 Coupling buffer

0.1 M $NaHCO_3$ (pH 8.3), 0.5 M NaCl.

2.1.8 DT40 lysis buffer

50 mM TrisHCl (pH 7.5), 100 mM NaCl, 50 mM KCl, 2 mM $MgCl_2$, 1mM EDTA, 10% Glycerol, 1 tablet of Complete Protease Inhibitors (Roche, UK) per 50 ml, 1 PhosStop phosphatase Inhibitor Cocktail tablet (Roche, UK) per 10 ml.

2.1.9 DT40 lysis buffer+T

DT40 lysis buffer + 0.3% Triton X-100.

2.1.10 DT40 lysis buffer+TB

DT40 lysis buffer+T + 150 U/ml Benzonase (25 U/µl, Novagen).

2.1.11 DT40 wash buffer

50 mM TrisHCl (pH 7.5), 200 mM NaCl, 100 mM KCl, 2 mM $MgCl_2$, 1mM EDTA, 10% Glycerol, 1 tablet of Complete Protease Inhibitors (Roche, UK) per 50 ml, 1 µg/ml 1 x FLAG peptide (CRUK peptide synthesis service)

2.1.12 *E. coli* lysis buffer

50 mM NaH_2PO_4, 500 mM NaCl, pH 8.0

2.1.13 Elution buffer

DT40 Lysis Buffer + 400 µg/ml 3xFLAG peptide (CRUK peptide synthesis service)

2.1.14 Freezing media

10% DMSO, 20% FCS in cell culture media

2.1.15 Fuscin formamide

0.1% fuscin dye (Sigma), 10 mM Tris HCl (pH 8), in formamide

2.1.16 Hypotonic lysis buffer

10 mM TrisHCl (pH 7.5), 2 mM $MgCl_2$, 3 mM $CaCl_2$, and 0.32 M sucrose supplemented with 1 Complete EDTA free Protease Inhibitor Cocktail tablet (Roche) per 50 ml and PhosStop phosphatase inhibitor cocktail (Roche Applied Sciences, UK).

2.1.17 IF blocking solution

PBSTS + 1% BSA +2% FCS, filtered.

2.1.18 20x KN

1M KCl, 2M NaCl

2.1.19 293 lysis buffer

20 mM Tris pH 8, 150 mM NaCl, 0.04 % SDS, 1 % NP40, PhosStop tablet (Roche), EDTA-free Protease inhibitor tablet (Roche)

2.1.20 PBS buffer*

3.3 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 136 mM NaCl (pH 7.2)

2.1.21 PBSTS

PBS + 0.1% Tween 20 + 0.02% sodium dodecyl sulphate

2.1.22 PCNA lysis buffer

50 mM NaH_2PO_4, (pH 8.0), 10 mM $NaHSO_3$ (pH 7.0), 500 mM NaCl, 5 mM β-Mercaptoethanol, 0.01% NP40, 10 mM imidazol, 5% Glycerol, Complete EDTA free Protease Inhibitor Cocktail tablet (Roche)

2.1.23 PCNA elution buffer

150 mM imidazol, 25 mM Tris-HCl (pH7.5), 1 mM β-Mercaptoethanol, 0.02% NP40, 5% Glycerol, 25 mM NaCl, Complete EDTA free Protease Inhibitor Cocktail tablet (Roche)

2.1.24 PTB buffer

50 mM Na_2HPO_4, 5mM Tris (pH 7.8)

2.1.25 PTB(A) buffer

50 mM Na_2HPO_4, 5mM Tris (pH 7.8), 500 mM NaCl, 15 mM imidazole

2.1.26 PTB(B) buffer

50 mM Na_2HPO_4, 5mM Tris (pH 7.8), 500 mM NaCl, 250 mM imidazole

2.1.27 PTB-Ta buffer

50 mM Na_2HPO_4, 5mM Tris (pH 7.8), 500 mM NaCl, 20 mM imidazole, 2M L-Tartaric acid

2.1.28 ph 4 wash solution

0.1 M acetic acid/sodium acetate, (pH 4), 0.5 M NaCl

2.1.29 ph 8 wash solution

0.1 M Tris-HCl, (pH 8), 0.5 M NaCl

2.1.30 5 x SDS sample buffer

10% SDS, 50% glycerol, 0.1% bromophenol blue, 50 mM DTT, 150 mM Tris-HCl (pH 6.8)

2.1.31 Southern blot denaturing buffer

0.5 M NaOH, 1.5 M NaCl

2.1.32 Southern blot hybridisation solution

200 mM $NaPO_4$ pH7.2, 1 μM EDTA, 1% BSA, 7% SDS, 15% Formamide

2.1.33 Southern blot SSC wash buffer

0.1x SSC + 0.1% SDS.

2.1.34 Southern blot transfer buffer

20x SSC stock*: 3 M NaCl, 0.3 M N_3citrate*$2H_2O$. pH adjusted to 7.0 (with 1 M HCl).

2.1.35 SYPRO fix solution

50% methanol, 7% acetic acid

2.1.36 SYPRO wash solution

10% methanol, 7% acetic acid

2.1.37 TBE electrophoresis buffer

10x TBE stock: 109 g of Tris base, 55 g Boric acid and 4.65 g EDTA was diluted in 1 litre dH_2O and pH was adjusted to 8.3 with HCl. The final concentration of TBE buffer was 89 mM Tris, 89 mM boric acid, 1.25 mM EDTA (pH 8.3).

2.1.38 TBS buffer

10x stock*: 1.54 M NaCl, 250 mM Tris-HCl, (pH 7.4).

2.1.39 TBS-T buffer

154 mM NaCl, 25 mM Tris-HCl (pH 7.4), 0.05% Tween-20 (pH 7.4).

2.1.40 TBS-Tri for IP

1x TBS*, 1% Triton X-100

2.1.41 TE buffer*

1 mM EDTA, 10 mM Tris-HCl (pH 7.5).

2.1.42 TEN buffer for AID activity assay*

1M NaCl, 10 mM EDTA, 50 mM Tris-HCl pH7.5

2.1.43 TNT Bead washing buffer

25 mM Tris (pH 8), 125 mM NaCl, 25 mM KCl, 2.5 mM MgCl2, 0.5 % NP40, and Complete Protease Inhibitors (Roche, UK).

2.1.44 2xTY*

16 g/l Bacto-tryptone (Difco, UK), 10 g/l Bacto-yeast extract (Difco, UK), 5 g/l NaCl. Autoclaved.

2.1.45 Tris-glycine Transfer Buffer for western blot

40 ml of 25x transfer buffer stock [12 mM Tris-base, 96 mM Glycine (pH 8.3)] + 200 ml methanol + 760 ml H_2O.

2.1.46 Txn 11 buffer

40 mM KCl, 20mM Tris-HCl (pH 7.9), 7 mM $MgCl_2$ 20μM $ZnCl_2$, 5mM DTT

2.1.47 Txn 12 buffer

40 mM KCl, 20mM Tris-HCl (pH 7.9), 0.2 mM $MgCl_2$ 20μM $ZnCl_2$, 5mM DTT

2.1.48 Western blot blocking solution

5% milk powder in TBS + 0.05% Tween 20

2.2 Methods

2.2.1 Biochemistry methods

2.2.1.1 Subcellular fractionation of DT40 cells

Large quantities of DT40 and DT40 AID-3FM were grown by CRUK cell service facilities using 5 litre spin paddle containers. Cells (1 - 4 x 10^{10}) were collected by centrifugation at 1,200 rpm in 1L bottles at 4 ˚C for 10 minutes and cell pellet was washed twice with 50 ml ice-cold 1 x PBS.

Cytoplasmic lysis: 5 times packed cell volume (PCV) (approx. 1 µl PCV = 10^6 cells) of Hypotonic Lysis Buffer was added to the cell pellet, resuspended gently and incubated for 12 min on ice. To the swollen cells, 10 % Triton X-100 was added to a final concentration of 0.3 %. The suspension was mixed and incubated for 3 min on ice, centrifuged for 5 min at 1,000 x g at 4 ˚C, and the supernatants were collected as cytoplasmic fraction. The nuclear pellets were washed once in hypotonic lysis buffer + 0.3 % Triton X-100, resuspended in 2 x PCV of DT40 Lysis Buffer + T and dounce homogenised with 30 strokes. The samples were incubated with gentle agitation for 30 min at 4 ˚C and ultracentrifuged at 33,000 x g for 30 min at 4 ˚C in a Beckmann ultracentrifuge (Ti45 or Ti70 rotor). The supernatants were collected as nucleoplasmic fraction. To extract chromatin, the pellets were resuspended in 2 x PCV of DT40 Lysis Buffer +TB containing Benzonase nuclease and dounce homogenized until resistance was lost. The samples were incubated at RT for 30 min and cleared by ultracentrifugation at 33,000 x g for 30 min at 4 ˚C in a Beckmann ultracentrifuge (Ti45 or Ti70 rotor). The supernatants were collected as chromatin fraction.

2.2.1.2 FLAG immunopurification of AID complex from chromatin

10-400 mg of a chromatin fraction obtained from DT40 cells as described above were pre-cleared with agarose beads for 1 hour at 4 °C. For analysis of association of AID with the PAF complex, salt concentrations in the chromatin fractions were adjusted to 200 mM NaCl and 100 mM KCl prior to IP by adding specific volumes of 20x KN. Concentration of Triton-X 100 was adjusted to 0.5 %. Anti-FLAG M2 affinity beads (Sigma, UK) were washed and equilibrated in DT40 Lysis Buffer + T. 1 µl of M2 beads (50 % slurry in DT40 Lysis Buffer + T) per 5 x 10^7 DT40 cells were incubated with chromatin fractions, for 3 - 4 hours at 4˚C on a rotator and collected for 3 min at 300 x g at 4 ºC. Beads were washed 5 times in 25 x bead volume of DT40 Wash Buffer (containing 1 µg/ml 1 x FLAG peptide N-DYDDDDK-C) and once with DT40 Lysis Buffer, each washing step was performed for 10 min at 4 ºC. M2 coupled protein complexes were eluted from M2 beads by two elution steps in 4 x bead volume of Elution Buffer (containing 400 µg/ml 3 x FLAG peptide, sequence N-MDYKDHDGDYKDHDIDYKDDDDK-C); the first elution for 1 hour at RT, the second over night at 4˚C.

2.2.1.3 Size exclusion chromatography of DT40 chromatin extract

One millilitre of DT40 chromatin extract was loaded onto a Superdex 200 - 10/300 GL column, which had been equilibrated in DT40 lysis buffer and calibrated with standard proteins using Äkta Explorer (GE Healthcare, UK). Fractions were collected at 1 ml volume steps using a 0.5 ml flow-rate, concentrated using StrataClean protein concentration resin (Stratagene), and analysed by western blot.

2.2.1.4 Mass spectrometry analysis

Samples were prepared by running proteins on 4-12% Bis-Tris NuPAGE Gels (Invitrogen). Gels were stained with SYPRO-Ruby (Invitrogen) using equipment that was pre-washed with methanol, under very clean and dust-free conditions. Gels were imaged, then bands were visualised using a UV transilluminator and bands of interest were cut out in small portions slices (1 - 2 mm). Samples were stored in ultrapure water until analysis. All mass spectrometry analysis was carried out at the CRUK central Protein Analysis and Proteomics Laboratory according to the below summarised protocol.

Polyacrylamide gel were prepared for mass spectrometric analysis using the Janus liquid handling system (PerkinElmer, UK) by alkylation and trypsin digestion, followed by extraction in formic acid and acetonitrile. Samples were analyzed by nano-scale capillary LC-MS/MS using a nanoAcquity UPLC (Waters, UK) coupled to a linear quadrupole ion trap mass spectrometer (LTQ XL/ETD, ThermoScientific, USA). LC-MS/MS data were searched against a protein database (UniProt Knowledge Base) using the Mascot search engine (Matrix Science, UK) (Perkins et al., 1999), and validated using the Scaffold program (Proteome Software Inc., USA) (Keller et al., 2002).

2.2.1.5 Preparation of human PCNA protein from E. coli

E. coli BL21-CODONPLUS(DE3)-RIL strain (Stratagene, USA) were transformed with pSPM588 (plasmid containing a C-terminal his-tagged human PCNA cDNA, expressed under the control of a T7 promoter, and a kanamycin resistance) . Bacteria were transformed according to the manufacturer's instructions and plated onto an appropriate selective LB plate. One fresh colony was picked and incubated over night at 37 °C in 10 ml LB with 50 μg/ml kanamycin. The next day, 1L of the same medium was inoculated with the over night culture and shaken at 37 °C until OD_{600} was 0.5.

Protein expression was induced by adding 1mM IPTG, and cultures incubated for 4 hours at 37 °C. Bacteria suspensions were centrifuged at 3500 g for 15 minutes at 4 °C. The cell pellet was resuspend in 50 ml of PCNA Lysis Buffer. Keeping the cells on ice, the suspension was sonicated for 12 minutes in total with a broad tip, amplitude 50%, and a duty cycle of 15 seconds on, 15 seconds off. Lysates were centrifuged at 35,000 rpm for 30 min at 4 °C in a 70 Ti rotor in a Beckman ultracentrifuge. The cleared lysate was incubated with 1 ml of prewashed Talon Metal affinity resin (Clontech) to bind the his-tagged PCNA for 90 min at 4 °C. The resin was washed with 50 ml of lysis buffer, then with 25 ml of elution buffer and the protein was eluted with 3 ml of elution buffer containing 150 mM imidazol. Eluted protein was dialysed against elution buffer and concentrated with a Millipore Amicon-15 (UFC 901024) concentrator by centrifugation at 3000rpm at 4 °C. The protein was loaded onto a Superdex 200 size exclusion column, which was equilibrated with the same buffer. The collected fractions containing PCNA trimer were pooled, snap-frozen, and stored at -80 °C. Aliquots were analysed on a 12% NuPAGE Bis-Tris gel (Invitrogen). To determine the protein concentration the gel was stained with SYPRO-Ruby (Invitrogen), scanned with a Typhoon Trio (GE Healthcare), and quantified using a range of BSA concentration as reference.

2.2.1.6 Preparation of human AID from E. coli

Human AID protein was prepared based on (Coker et al., 2006) with slight modifications. *E. coli* of the C41(DE3) strain (Miroux and Walker, 1996) were transformed with pSPM75 (a plasmid containing a C-terminal his-tagged human AID cDNA under the control of a T7 promoter with kanamycin resistance). Cells were plated onto an appropriate selective LB plate and incubated over night. One colony was picked and resuspended thoroughly in 1 ml selection-free 2x TY media. This was used to further inoculate 30 ml selection-free 2x TY, which is then grown, shaking at 37 °C until reaching an OD_{600} of 0.4. At this stage, 10 ml of the culture was used to inoculate each of three 2-liter flasks, each containing 1 litre of selection-free 2x TY, which was then grown, shaking at 37 °C, until reaching an OD_{600} of between 1.0 and 1.5. After immersion in iced H_2O for 15 minutes the culture was induced with a final concentration of 0.4 mM IPTG and in the presence of 10 μM $ZnCl_2$. Holes were made in the foil lids of the flasks to ensure that sufficient air exchange will take place during protein production. The cultures were incubated, shaking at 250 rpm at 16°C for 16 h.

The 3 L of bacteria were harvested by centrifugation at 3500 g, 10 min at 4 °C, the supernatant was removed, and the pellets were resuspended and combined in ice-cold H_2O. Bacteria were again centrifuged at 3500 g, for 10 min at 4 °C, before the supernatant was immediately removed and the pellet resuspended in 100 ml of buffer 2 along with six complete EDTA-free protease inhibitor cocktail tablets (Roche) and 150 mg RNaseA (Sigma), which had been heated previously to 80 °C for 10 min in buffer 2. Keeping the cells on ice, the suspension was sonicated with a broad tip, amplitude 50%, 12 minutes total, duty cycle 15 seconds on, 15 seconds off. The sonicated lysates were cleared by centrifugation at 100,000 g, for 30 min, at 4 °C in a Beckmann ultracentrifuge (Ti45 rotor). The supernatant was filtered using a pre-chilled 0.22 μm Millipore filter. The extract was then incubated with 2 ml of prewashed Ni-NTA beads (Qiagen) at 4 °C, rotating for 90 min. The beads were loaded in suspension onto a prewashed Poly-Prep chromatography column (Bio-Rad) and washed at 4 °C with 10 ml of 30 mM imidazole. The protein was eluted at 4 °C with 10 ml of 150 mM imidazole. The eluate was concentrated to approximately 150 μl using a centrifugal concentrator (Vivaspin) with a molecular weight cut-off of 10,000 d according to the manufacturer's instructions. The concentrated protein was snap-frozen in small aliquots and stored in liquid nitrogen for use within 10 months. Aliquots of the eluated proteins were analysed on a 12% NuPAGE Bis-Tris gel (Invitrogen). To determine the protein concentration the gel was stained with SYPRO-Ruby (Invitrogen), scanned with a Typhoon Trio (GE Healthcare), and quantified using a range of BSA concentrations on the gel as reference

2.2.1.7 Preparation of human APOBEC2 from E. coli

APOBEC2 protein was prepare in an analogous way to AID protein based on (Coker et al., 2006) with slight modifications.

2.2.1.8 Preparation of cell soluble Cre recombinase from E. coli

E. coli Tuner (DE3) pLysS strain (Novagen) were transformed according to the manufacturer's instructions with pTriEx-HTNCre [plasmid containing a His-TAT-NLS tagged Cre-protein cDNA, and an ampicillin resistance, (Peitz et al., 2002)], plated on ampicillin plates and incubated over night at 37 °C. A 50 ml starter culture (LB + 50 μg/ml carbenicillin + 34 μg/ml chloramphenicol + 1% glucose) was inoculated and

grown over night at 37 °C. On the next day, 2 litres LB + 100 μg/ml carbenicillin + 34 μg/ml chloramphenicol was inoculated with the starter culture and allowed to grow until OD_{600} was 1.5, then induced for 1 hour at 37 °C with 0.5 mM IPTG. Cells were harvested by centrifugation for 30 minutes at 3500 g at 4°C. The pellets were washed in ice-cold PTB and snap-frozen in a dry ice/ethanol bath for storage at -20 °C. On the day of purification, cells were thawn on ice and resuspended in 10 ml PTB containing 1 protease inhibitor tablet (Roche), 10 mg lysozyme and 20 μl Benzonase (25 U/μl, Novagen), followed by sonication in an ice cold glass beaker at 40% amplitude for 2 minutes total, duty cycle of 0.5 seconds on, 0.5 seconds off. After incubating the lysates on ice for 5 minutes, 10 ml PTB-Ta was added dropwise while stirring the lysate on ice, followed by a second 5 minute incubation on ice. After spinning for 30 minutes in a Beckmann ultracentrifuge at 30,000 g (Ti45 rotor) at 4 °C, the supernatant was gently mixed with 1.5 ml Nickel-NTA beads (Quiagen) for 1 hour. The suspension was transferred into a Poly Prep (BioRad) gravity flow column and washed with 2 x 5 bed volumes of PTB(A). HTN-Cre containing fractions were eluted with 3 bed volumes of PTB(B). The imidazol was removed from the eluates by dialysis against buffer C (2x for 30 minutes in 500 ml beaker at 4 °C) and further concentrated by dialysis against buffer D (2x for 30 minutes in 500 ml beaker at 4 °C). The resulting glycerol stock was aliquoted, snap frozen and stored at - 20 °C. The concentration was determined by PAGE and staining as described.

2.2.1.9 SYPRO Ruby staining

Protein gels were stained using the SYPRO Ruby system from Invitrogen according to the manufacturer's instructions. After electrophoresis, polyacrylamide gels were immersed in SYPRO fix solution for 30 minutes, the solution was changed and the incubation repeated. The gel was immersed in SYPRO Ruby stain and left agitating over night, followed by a 30-minutes wash in SYPRO wash solution. Before imaging, the gel was briefly washed in ultrapure H_2O.

2.2.1.10 Oligonucleotide deamination assay (AID activity assay)

2.2.1.10.1 Deamination reaction

1x Buffer R, 10 μg RNase A (Sigma) [heated previously to 80 °C for 10 min in 0.1 M Tris-HCl (pH 8)], 2.5 mM oligonucleotide SPM163 (5'biotin-

ATTATTATTATTAGCTATTTATTTATTTATTTATTTATTT-3'FITC) in 10 µl total volume were heated to 90 °C to denature the oligonucleotide, followed by quenching on ice and a pulse-spin. Purified deaminase was thawn on ice and the appropriate (as determined for each preparation by titration) amount was added to the ice-cold reaction mix for each sample in 0.5-ml tubes. The volume of deaminase added was generally between 0.05 and 0.5µl. The reaction was incubated at 37 °C for 15 min for optimal activity. Each reaction was terminated by adding 100 µl of H2O and then heating at 90 °C for 4 min before placing on ice.

The oligonucleotides were then purified using streptavidin-magnetic beads: 8 µl of resuspended streptavidin-magnetic bead slurry (Dynal Invitrogen) per reaction was washed twice in TEN buffer before resuspension in 800 µl TEN buffer. Bead suspensions were added to the reactions, mixed by rotating at RT for 15 min, followed by applying the tubes to the magnet supplied by the bead manufacturer. After carefully removing the supernatant, the procedure was repeated twice, each time washing with 750 µl TEN pre-heated to 70 °C. The final wash was with 100 µl TE, followed by resuspension of the beads in a UDG reaction mix (1µl 10x UDG buffer (NEB), 1 µl UDG (NEB), 8 µl H2O). The mix was incubated at 37 °C for 1 hour, after which the reaction was stopped by adding 20 µl fuscin formamide. After heating to 90 °C for 3 min, the sample was subjected to analysis by Urea-PAGE.

2.2.1.10.2 Urea-PAGE

Gel preparation: For a 17.5 % TBE-urea gel, a mixture of 4.7 g urea, 1 ml 10X TBE,1.27 ml H_20 and 4.38 ml 40% acrylamide was mixed on a shaker. 20 µl 25% APS and 7 µl TEMED were added and a minigel was poured. After polymerisation (about 1 hour) the gel was pre-run for 10 min at 200 V, then the samples were loaded and electrophoresis was carried out for 70 min at 200 V (e.g. for SPM 163, electrophoresis conditions depended on the length of the oligonucleotide and products). The fluorescent signal was detected using the FLA-5000 fluorescent scanner (Fuji).

2.2.1.11 Extended oligonucleotide deamination assay (AID activity assay on a PCNA coated substrate)

The assay was developed to combine the standard oligonucleotide deamination assay for AID with a PCNA loading assay [J. Parker and H. Ulrich, CRUK; (Parker et al.,

2008, Waga and Stillman, 1998a)]. The oligonucleotide KW5022 (5'BIO-TGAATTATTAGCTATTTAGATTTTTCATTTGCTGCTGTCAGCGTGGCACTGTT-DIG-C-3'FITC) was hybridised to its complement KW5023 (5'-AGTGCCACGCTGACAGCAGCAAATGA-3') by heating a mixture of both oligonucleotides at a 1:2 molar ratio (KW5022 to KW5023) to 95 °C and slowly cooling it in a PCR machine using the following program: 95 °C – 3 minutes; 80 °C – 15 minutes; 70 °C – 15 minutes; 65 °C – 15 minutes; 55 °C – 15 minutes; 40 °C – 10 minutes. Streptavidin plates (cat. No. 15120, Thermo Scientific) were allowed to reach room temperature and hydrated for 5 min with 200 µl 154 mM NaCl per well. Wells were blocked for 20 minutes at RT with 200 µl blocking solution followed by a wash with 250 µl buffer A. 2.5 pmol hybridised oligonucleotide was diluted in 90 µl buffer A and loaded into the wells. Plates were incubated for 1 hour at 4 °C to coat the plate with DNA. After 3 washes with 250 µl buffer A the first reaction mix (1 mM ATP, 0.5 µg anti-Digoxygenin Fab fragments (Roche), 50 pmol PCNA, 1 pmol RFC) in a total of 50 µl buffer A was applied to the wells. The reactions were incubated for 30 minutes at RT with occasional tapping, followed by one wash with 200 µl buffer A + 0.2 mM ATP at RT followed by one wash with 200 µl buffer A + 0.2mM ATP at 37 °C. The same loading mixture was re-applied, incubated and washed away 2 times as described followed by 2 additional washes with 200 µl buffer R + 0.2mM ATP.

Samples that were used for western analysis were stopped at this step by adding 100 µl 1.5x SDS buffer and 5 µl subject to PAGE followed by western blotting.

To the rest of the wells, a second reaction mix is applied consisting of 0.001% Triton, 0.3 µl AID (DEL99) and 20 µg RNAse A in a total of 20 µl buffer R. Wells were covered with parafilm and incubated in a humidified chamber for 1 hour at 37 °C. Next, 200 µl of TEN buffer heated to 70 °C was added to each well to stop the reaction. Wells were washed 2-times with TEN (pre-warmed to 70 °C) and once with TE buffer. After aspiration of all liquid the following UDG reaction mix was applied to each well: 0.001% Triton, 2.2 µl 10x UDG buffer (NEB), 2.2 µl UDG (NEB) in a total volume of 22 µl. Wells were covered and incubated in a humidified chamber for 1 hour at 37 °C, followed by addition of 30 µl of fuscin formamide and heating to 90 °C. The samples were analysed on a 12% TBE-Urea gel. Electrophoresis was for 60-70 minutes at 200 V (see section 2.2.1.10.2).

2.2.1.12 Reconstitution of a transcription elongation complex and NTS deamination by AID

RNA oligonucleotides were purchased from Dharmacon. DNA 90-mers (TS and NTS) were purchased from DNA technologies, Denmark. Radioisotope labelling of RNA or DNA oligonucleotides was performed using T4 polynucleotide kinase (Roche) according to the manufacturer's instruction (37 °C for 30 minutes) using γ–^{32}P-dATP, followed by phenol chloroform extraction of nucleic acid by using PhaseLock tubes (Eppendorf). After passing through a G25 spin column (GE Healthcare), nucleic acids are precipitated with 3x volumes of ethanol and 0.1 volumes of 3 M sodium acetate at -20 °C, washed twice with 95% ethanol (centrifuged at 15,000 rpm for 30 minutes between each wash), and resuspended in TE. After a second pass through a G25 column followed by ethanol/sodium acetate precipitation, the labelled nucleic acid is dissolved in TE pH 7.5.

Reconstitution reactions were carried out in the presence of 400-1000 counts per second labelled nucleic acid. 1 pmol of the 90 nucleotide long oligonucleotide (TS) was hybridised to 2 pmol complementary RNA by cooling the mixture from 65 ˚C to 25 ˚C. 3 pmol purified yeast RNAPII (gift from D. Hobson, CRUK) was added at 25 ˚C in Txn12 buffer followed by 6 pmol of the complementary 90-mer strand (NTS). The assembled complex was purified via a biotin label on the TS to remove unbound components using magnetic streptavidin beads according to the manufacturer's instructions (Promega, UK). The obtained complex on biotin beads was exposed to 0.12, 1.2 and 12 pmol of active AID before and after the addition of NTP's (0.58 mM or 0.29 mM) in Txn 11 buffer + 1mg/ml BSA. After phenol chloroform extraction and precipitation of the nucleic acids, a UDG reaction mix was applied analogous to ODA. Products were dissolved in a 90% formamide gel loading buffer and separated on a 8 % Polyacrylamide-TBE gel containing 8.3 M Urea using a BioRad 50 cm sequencing gel cell. The gel was dried, exposed to phosphoimager and scanned using a Typhoon Trio scanner (GE Healthcare).

2.2.1.13 In vitro PCNA ubiquitination

In vitro ubiquitination reactions were performed in 100 µl reaction volumes using the Biomol ubiquitination kit (UW9920-0001, Enzo life sciences, UK). A reaction had 10

µl 10 x Ubiquitination buffer, 2 µl 50 mM DTT, 5 µl Mg-ATP, 3 µl 20 x E1 conjugating enzyme, 3 µl E2 conjugating enzyme, 5 µl 20x Biotin-Ubiquitin, PCNA protein and water to 100 µl. 10 - 20 pmol AID, 10 - 20 pmol Apo2, 5 pmol PCNA and 5 pmol PCNA K164A of proteins were used in the assay. Proteins were purified as described above. Reactions were incubated at 37˚ C for 90 minutes before stopping with 4x SDS sample buffer and boiling for 3 minutes. Samples were analysed by Western blotting.

2.2.1.14 PCNA sumoylation assay

BL21-Gold (DE3) bacteria (Stratagene, UK) were transformed with vectors expressing his-tagged PCNA, the activating (E1) and conjugating (E2) enzymes, SUMO-I, and either an empty vector or a vector encoding for AID. Protein expression was induced by the addition of 1 mM IPTG. The cells (100 µl of culture) were centrifuged and resuspended in 250 µl of *E. coli* lysis buffer and diluted in SDS loading buffer. Proteins were separated on a 12 % NuPAGE Bis-Tris gel (Invitrogen), transferred on membrane and analysed by western blot.

2.2.1.15 Cross-linking antibodies to CNBr-activated Sepharose 4B

The cross-linking of antibodies was carried out using CNBr-activated Sepharose 4B (GE Healthcare, UK) according to the manufacturer instructions, using monoclonal anti-myc 9E10 control antibody (CRUK antibody services), monoclonal anti-AID hAnp52-1 (CRUK antibody services), monoclonal anti-PCNA PC10 (CRUK antibody services), and irrelevant mouse IgG2a control antibody (Serotec, USA). 300 µg sepharose powder was resuspended in 1 mM HCl yielding 1000 µl medium and washed for 15 minutes with abundant 1 mM HCl on a porous filter. The medium is briefly washed in coupling buffer, instantly activating the reactive groups therefore immediately transferring the medium to a tube containing the antibody in coupling buffer (5 mg Ab per ml medium). The mixture is rotated for one hour at RT or over night at 4 °C. Excess Ab is washed away with 5 medium volumes of coupling buffer. Any remaining active groups are blocked with 0.1 M Tris-HCl, pH 8, incubating for two hours. The medium is washed with at least three cycles of alternating pH (washing volume is 5x medium volume), using pH 4 wash solution and pH 8 wash solution. Finally, the medium is washed once in PBS and resuspended at a 50% slurry in PBS for

storage at 4 °C. 1ml of medium is treated in the same way but without antibody to use as preclearing material or control for IP.

2.2.1.16 Co-immunoprecipitation of AID expressed in E. coli

A pETDuet-1 derived vector for expressing human AID in *E.coli* was generated by using *NcoI*/*HindIII* restriction sites for cloning. Using this construct as a base, AID interacting proteins were cloned into the second multiple cloning site of the vector at the *NdeI/XhoI* restriction sites. The two proteins were co-expressed in BL21-CODONPLUS(DE3)-RIL cells (Stratagene, USA). Bacteria were transformed according to the manufacturer's instructions. One colony per construct was picked and dissolved in 2 ml 2x TY + 100 μg/ml Carbenicillin + 0.2% Glucose and incubated at 37°C for 1 hour. A 30 ml LB culture with 100 μg/ml Carbenicillin and 0.2% Glucose was inoculated with the started culture and protein expression was induced with IPTG (1 mM) in the presence of 0.1 mM ZnCl2 when OD_{600} was 0.6, for 3 h at 18°C. Cells were lysed by sonication in TBS-Tri for a total of 2 x 25 seconds on ice, 1 second on, 1 second off, 20% amplitude in a 2 ml tube. Debris was pelleted at 19,000 g and supernatant collected. For Paf1 interaction analysis, NaCl concentration was adjusted to 300 mM at this point. The lysate was pre-cleared with sepharose beads for 1 hour followed by over night incubation with sepharose-coupled monoclonal antibodies anti-AID hAnp52-1, anti-PCNA PC10 or control anti-myc 9E10. After 5 washes with TBS-Tri the immunoprecipitates were analysed by western blot using polyclonal anti-PCNA and monoclonal anti-AID, followed by Odyssey LICOR secondary antibodies. LICOR analysis of the blots was performed as recommended by the manufacturer.

2.2.1.17 Co-immunoprecipitation of AID expressed in TnT rabbit reticulocyte lysate

2.2.1.17.1 Paf1 interaction analysis

^{35}S labelled Paf1 was expressed using the TnT T7 Coupled Reticulocyte Lysate System (IVT) according to the manufacturer's instructions (Promega, USA), as follows. Paf1 was expressed from a PCR fragment harbouring the Paf1 cDNA and a T7 promoter. Per sample 500 ng of DNA template was combined with 1 μl of TnT retic buffer, 1 μl of ^{35}S labelled methionin, 0.5 μl of 1mM amino acid mixture (minus methionin), 0.5 μl RNaseOut (Invitrogen), 0.5 μl T7 polymerase, 12.5 μl TnT retic lysate, and topped up to

25 µl with water in the presence of 1mM $ZnCl_2$. The reaction was incubated at 30°C for 90 minutes followed by incubation with RNase A at 10 µg/ml for 15 minutes at 37°C. The labelled protein mixture was incubated with 100 ng of AID or 300 ng of APOBEC2 protein for 1 h at RT and 30 min at 4 ˚C. Proteins were isolated by prewashed and TBS-Tri + 2 % BSA blocked anti-AID hAnp52-1 or anti-myc (9E10) beads coupled to sepharose. After bead incubation for one hour at 4 ˚C, beads were washed 5 times in TBS-Tri + 2 % BSA, resuspended in SDS PAGE loading buffer, and separated on 12 % Bis-Tris polyacrylamide gels (Invitrogen, USA). Gels were dried, exposed to Fuji Imaging Plates, and visualized with a Fuji FLA-3000 scanner.

2.2.1.17.2 PCNA interaction analysis

^{35}S labelled AID (pSPM270) and control protein (Luciferase) were expressed from plasmids harbouring a T7 promoter using the TnT T7 Coupled Reticulocyte Lysate System (IVT) according to the manufactures protocol (Promega, USA) as described above. RecD purified protein was a gift from D. Wigley, CRUK. After *in vitro* translation, protein mixtures were added to 100 ng of PCNA or RecD protein for 1hour at RT, followed by 30 min at 4˚ C. Anti-PCNA (PC10 - Santa Cruz, USA) or control immunoglobulin was added for 30 min at 4˚ C. Mixtures were added to 20 µl Protein A/G Sepharose beads (Santa Cruz, USA) and further incubated for 30 min at 4˚ C. Bead complexes were washed 5 times in TnT bead washing buffer. Beads were resuspended in SDS gel loading buffer and analysed as above.

2.2.1.18 Western blotting

5 x SDS lysis buffer was added to 1 x to protein samples and denatured at 90 ºC for 5 minutes. Samples were loaded to 12% or 4-12% NuPAGE Bis-Tris gels (Invitrogen). 7 µl of Spectra Multicolour Broad Range Ladder (Fermentas) was electrophoresed simultaneously with the protein samples using MOPS SDS Running Buffer (Invitrogen, UK) at 200 V for 50-70 minutes at RT. The separated proteins were transferred onto Immobilon-P polyvinylidene fluoride (PVDF) membrane (Millipore, UK) in transfer buffer in a Novex MiniGel transfer apparatus (Invitrogen, UK) at 30 V for 150 minutes or 25 V for 240 minutes on ice. Membranes were blocked with 5% milk in TBS-T (blocking solution) for 30 minutes at RT, probed with primary antibody (in blocking solution) for 1hour at RT or 4 °C over night followed by washing the membrane with

Antibody	Manufacturer	Concentration/ dilution
Monoclonal anti-FLAG M2	F3165, Sigma, UK	1/3000
Monoclonal anti-FLAG M2 coupled to HRP	A8592, Sigma, UK	1/2000
Monoclonal anti-AID clone 4.26.1	CRUK antibody production service	1/1000
Monoclonal anti-AID clone hAnp52-1 (Conticello et al., 2008)	CRUK antibody production service	1/1000]
Rabbit polyclonal anti-beta Tubulin	ab6046, Abcam, UK	1/1000
Monoclonal anti-GFP clones 7.1 and 13.1	11814460001, Roche	0.4 µg/ml
Monoclonal anti-His	sc-8036, Santa Cruz, USA	1:1000
Monoclonal anti-beta-Actin	ab8226, Abcam, UK	1:2000
Monoclonal anti-Histone H3	ab10799, Abcam, UK	1:1000
Rabbit polyclonal anti-SUPT5H	ab26259, Abcam, UK	1/1000
rabbit polyclonal anti-PAF1	ab20662, Abcam, UK	1/1000
rabbit polyclonal anti-PAF1	A300-172A, Bethyl laboratories, USA	1/3000
rabbit polyclonal anti-SUPT6H	A300-801A, Bethyl laboratories, USA	1/2000
rabbit polyclonal anti-CDC73	sc-48770, Santa Cruz, USA	1/1000
mouse polyclonal anti-CDC73	ab43256, Abcam, UK	1/1000
rabbit polyclonal anti-CTR9	A301-395A, Bethyl laboratories, USA	1/3000
rabbit polyclonal anti-LEO1	ab70630, Abcam, UK	1/1000
Monoclonal anti-PCNA clone PC10	CRUK antibody production service	1 µg/ml
Rabbit polyclonal anti-PCNA	ab2426, Abcam, UK	1/1000
Goat anti-mouse Ig-HRP conjugate	P0449, DakoCytomation, UK	1/10,000
Goat anti-rabbit Ig-HRP conjugate	P0448, DakoCytomation, UK	1/10,000
Goat anti-mouse IR Dye 680	926-32220, LI-COR Biosciences	1/10,000
Donkey anti-rabbit IR Dye 800	926-32213, LI-COR Biosciences	1/10,000

Table 1. Antibodies used for western blotting.

TBS-T two times for 5 minutes at RT. The membrane was then incubated with HRP-conjugated secondary antibody in blocking solution for 1 hour at RT, washed three times with TBS-T for 7 minutes at RT, and bands detected with ECL Plus Western Blotting Detection System (GE Healthcare, UK) according to manufacturer's instructions. Primary and secondary antibodies used for western blotting are listed in Table 1.

Quantitative western blot was carried out using the LI-COR system, according to manufacturer's instructions. Secondary antibodies were coupled to IR Dye 680 or IR dye 800 and analysed with LI-COR Odyssey Infrared Reader.

2.2.2 Mammalian or chicken cell methods

2.2.2.1 Thawing cells

A cryotube with cells was thawed quickly at 37 °C by hand warming and adding warm media. Cells were immediately transferred to a 15 ml falcon tube with 10 ml warm culture medium in it and centrifuged at 1,200 rpm for 5 min followed by aspiration of the supernatant, leaving 100 µl of media on the pellet. The cells were resuspended by tapping the tube gently and transferred to a small sized flask with 6 ml culture medium.

2.2.2.2 Freezing cells

Cells were grown to approximately 10^6 cells / ml. 5-50 x 10^6 cells were centrifuged at 1000 rpm for 5-7 minutes and the supernatant aspirated. Cells were gently resuspended in 0.5-1 ml ice cold freezing medium and transferred into an ice cold cryotube. The cells were placed into a NALGENE Cryo Freezing Container containing isopropanol, and stored at -80 ºC for 72 hours before transferring the vials into liquid nitrogen.

2.2.2.3 293 cells

293 cells were maintained in DMEM + 10% heat inactivated fetal calf serum (30 min at 55 ºC). For splitting adherent cells, the culture media was aspirated, cells were rinsed with PBS and detached from the surface by trypsin treatment. Cells were then counted by haematocytometer and transferred to a new flask with fresh culture medium, splitting approximately 1:5 every 3 days.

2.2.2.4 CH12 cells

Murine B cell line CH12F3 (Nakamura et al., 1996) was a gift from B. Reina-san-martin. Cells were maintained at 37ºC at 5% CO_2 in RPMI (Invitrogen), 20 mM Hepes, 10% heat inactivated fetal calf serum (30 min at 55 ºC), 1mM Sodium Pyruvate, 50 µM beta-Mercaptoethanol freshly added. CH12 are split between 1:5 and 1:10 every 3-4 days.

2.2.2.5 Chicken DT40 cells

Chicken bursal derived DT40 cell line was cultured at 37 ºC with 5% CO_2. Cells were monitored on a daily basis. The requirement for feeding was estimated according to cell density, cell morphology and culture medium colour. Cells were counted by using a haematocytometer or a cell counter. Cells were maintained at a density of 5 x 10^5 and 2 x 10^6 cells / ml in RPMI 1640 +Glutamax media with 10% heat inactivated fetal calf serum (30 min at 55 ºC), 1% heat inactivated chicken serum (30 min at 55 ºC), 50 µM β-mercapthoethanol (Sigma, UK) and 10 U penicillin / streptomycin. Culture media was warmed up at 37 ºC prior to adding it to the cells.

Large volumes of DT40 cells were cultivated by CRUK cell services in spin paddle flasks of 2-5 litres volume in a 37 ºC room, using high bicarbonate E4 media (DMEM, 3g/l bicarbonate)+ 10% heat inactivated fetal calf serum (30 min at 55 ºC), 3% heat inactivated chicken serum (30 min at 55 ºC), 50 µM β-mercapthoethanol (Sigma, UK), and 10 U penicillin / streptomycin.

2.2.2.6 Retrovirus production and retroviral transduction of shRNA constructs into CH12

29mer shRNAs were cloned in the Origene HuSH pGFP-V-RS retroviral vectors (OriGene, USA, Table 2) and used together with pCL-Amphotropic helper vector (suitable for CH12 retroviral transduction) to produce infective virus from 293 host cells. Host cells in log phase were prepared for transfection by plating 150,000 cells in 2 ml DMEM media per well in a 6-well tissue culture plate. The next day, 293 cells were transfected at 60-70% confluency. The DNA (very clean) was diluted to 1 µg/µl solution. Per construct and well a transfection mix is prepared using FuGene 6 reagent (Roche) at a 6 to 1 ratio: 94 µl OptiMEM medium + 6 µl FuGene 6 + 1 µl retroviral shRNA construct + 1 µl pCL-Ampho. The mixture was vortexed and pulse-spun,

Target gene	sequence	Retroviral vector
aid	5'-ACCAGTCGCCATTATAATGCAA-3'	pLMP (B. Reina-San-Martin Laboratory, France)
leo1	5'-GTGGCAGTGACAATCACTCTGAACGGTCA-3'	HuSH pGFP-V-RS (GI529047, OriGene, USA)
supt5h	5'-GCTTGGCTACTGGAACCAGCAGATGGTGC-3'	HuSH pGFP-V-RS (GI336016, OriGene, USA)
cdc73	5'- GACGTGCTCAGCGTCCTGCGACAGTACAA-3'	HuSH pGFP-V-RS (GI561719, OriGene, USA)
paf1	5'-GGTGACGGAGTTTACTACAATGAGCTGGA-3'	HuSH pGFP-V-RS (GI518753, OriGene, USA)
ctr9	5'-GATGAGGATTCCGACAGTGACCAGCCGTC -3'	HuSH pGFP-V-RS (GI528588, OriGene, USA)
Untargeted control shRNA	5'- GCACTACCAGAGCTAACTCAGATAGTACT -3	HuSH pGFP-V-RS (TR30007, OriGene, USA)

Table 2. shRNA sequences used for knock-down in CH12.

followed by a 15-minute incubation at RT. The 100 µl mix was added carefully and dropwise to the 293 cells and the plate shaken gently. Cells were returned to an incubator designated for virus work and left over night. After 24 hours, virus production starts. GFP expression was confirmed under the microscope and transfection efficiency estimated. CH12 cells were split for next day. At 48 hours post transfection, $1x10^6$ CH12 cells per construct/well were infected using the viral supernatant from the 293 host cells. To this end, $1x10^6$ CH12 cells were placed in a 15 ml tube and spun down. The supernatant was aspirated shortly before virus is added. Viral supernatants were harvested after 48 hours using a sterile syringe fitted with a 0.45 µm filter, to remove debris and host cells, into 15 ml tubes. Hepes (Invitrogen) was added to a 20 mM final concentration and polybrene to 10 µg/ml. The mixture was topped up to 2 ml and CH12 cells were resuspended in the virus mixture, followed by transferring the cell suspension to 6-well tissue culture plates. Plates were centrifuged for 1.5 hours at 1150 g at RT without brake, then transferred to an incubator designated for viral work. At this stage

host cells could be trypsinised or scraped to accurately determine the transfection rate using FACS.

After 24 hours post infection viral supernatants were disposed of appropriately and CH12 cells washed 1x with media, followed by resuspension in puromycin-containing culture media (0.5 μg/ml). Half of the cells were saved and plated into non-selective media to compare death rates. In the case of the pLMP control retroviral vector, transduction efficiency could be analysed by monitoring GFP expression pre-selection and post-selection by FACS.

2.2.2.7 Stimulation of class switch recombination in CH12

Assays are carried out in 6-well tissue culture plates. 200,000 CH12 cells growing in log phase are used per well. CH12 are diluted to 200,000 cells/ml in media and 1 ml aliquoted into each well. Cytokines are diluted into media at a 2x concentration. Final concentrations should be: 5 ng/ml Interleukin 4 (recombinant murine IL-4 Peprotech 214-14), 200 ng/ml anti-CD40 (functional grade purified anti mouse CD40 clone HM40-3, eBIOSCIENCES 16-0402-85), 0.3 ng/ml TGF-β (R&D systems Europe 240-B-002). 1 ml of cytokine solution is added to each well containing cells and shaken gently before incubating for 24 hours at which point a first flow cytometric analysis for IgA positive cells should be carried out, followed by a second one at 72 hours post stimulation.

2.2.2.8 Flow cytometry (FACS)

Sample preparation: Where cells were antibody stained prior to flow cytometry (as in switching analysis), a 96-well round bottom plate was used for the procedure. Depending on cell density, 200-1000 μl of CH12 suspension was taken for each sample. The cells were washed in ice-cold PBS and spun at 1200 rpm at 4 ºC followed by another wash in PBS + 1% FCS. After vortexing the plate briefly to dissolve the pellet, 50 μl of antibody mix was added and cells resuspended. The antibody used for anti-IgA staining was: goat anti-mouse IgA-PE (Southern Biotech 104009), 1:1000 in PBS + 1% FCS. The antibody used for anti-IgM staining was: Rat Anti-Mouse IgM-FITC (BD Biosciences 553437), 1:1000 in PBS + 1% FCS. Cells were stained for 10 min at RT in the dark. After washing 2x with PBS + 1% FCS, cells were resuspended in 500-300 μl in PBS + 1% FCS and transferred to FACS tubes. For GFP analysis, cells were washed

and then resuspended in 500-300 µl PBS + 1% FCS. TOPRO DNA dye (Invitrogen) was added to all samples at 30 nM final concentration.

For IgA switching analysis, a non-stimulated and a non-stained control was included, whereas for GFP expression analysis, a non-GFP expressing control line was included.

Flow cytometry was carried out on a BD FACSCalibur with 4-colour option. Live cell population was selected by using FSC-H and TOPRO (FL4-H), excluding TOPRO-positive cells. Live cells were analysed for PE signal (FL2-H) or GFP/FITC signal (FL1-H). Data were acquired using BD CellQuest Pro software, and analysed using FlowJo 9.2 software.

2.2.2.9 Isolation of RNA

1 x 10^6 cells were spun down in a 1.5 ml tube, the supernatant discarded and cell pellet immediately resuspended in 500 µl Trizol Reagent (Invitrogen), pipetting up and down to break up cells. The samples were incubated for five minutes at RT to facilitate lysis. 100 µl chloroform was added, followed by vortexing for 15 seconds and incubating at RT for five minutes. Samples were spun at 12,000 g for 15 minutes at 4 °C to separate phases. The upper aqueous phase was transferred to a new tube, and 250 µl isopropanol was added and the tube mixed, followed by a 10-minute incubation at RT. Next, the samples were spun at 12,000g for 30 minutes at 4 °C. The resulting pellet was washed with 1ml of 70% ethanol in nuclease-free water, followed by another spin at 12,000g for five minutes at 4 °C. The supernatant was removed and pellet was dried in the air. RNA was resuspended in 20 µl nuclease-free water and stored at -80 °C.

2.2.2.10 cDNA synthesis

First-strand cDNA synthesis was carried out using SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's instructions. 1 µl (300ng) random hexamer primers were combine with1 µg total RNA and 1 µl 10 mM (each) dNTP mix, topped up to 12 µl with sterile nuclease-free water and heated to 65 °C for five minutes followed by chilling the mixture on ice. After brief centrifugation 4 µl 5x first strand buffer, 2 µl 0.1 M DTT and 1 µl RNasin RNase inhibitor (Promega) were mixed and incubated at 25 °C for two minutes. Next, 1 µl of SuperScript II enzyme was added and mixed, followed by an incubation at 25 °C for 10 minutes and an incubation at 42 °C for 50 minutes. A control sample omitting the addition of the reverse transcriptase was

Target	Primer sequence	PCR program
cd79b	5'-CCACACTGGTGCTGTCTTCC-3' 5'-GGGCTTCCTTGGAAATTCAG-3'	50 cycles: 94 °C 10 seconds 60 °C 20 seconds 72 °C 30 seconds
hprt	5'-GTTGGATACAGGCCAGACTTTGTTG-3' 5'-GATTCAACTTGCGCTCATCTTAGGC-3'	50 cycles: 95 °C 15 seconds 60 °C 30 seconds 72 °C 30 seconds
Germ line transcript GLTα	5'-CAAGAAGGAGAAGGTGATTCAG-3' 5'-GAGCTGGTGGGAGTGTCAGTG-3'	50 cycles: 94 °C 10 seconds 61 °C 30 seconds 72 °C 20 seconds
Germ line transcript GLTμ	5'-ACCTGGGAATGTATGGTTGTGGCTT-3' 5'-TCTGAACCTTCAAGGATGCTCTTG-3'	50 cycles: 94 °C 10 seconds 60 °C 20 seconds 72 °C 30 seconds
aid	5'-GAAAGTCACGCTGGAGCCG-3' 5'-TCTCATGCCGTCCCTTGG-3'	50 cycles: 94 °C 7 seconds 61 °C 30 seconds 72 °C 10 seconds
leo1	5'- GAGGAGCAAGACCAGAAGTCAG -3' 5'-TGTCGCTGTCTGCTTCGGAATC-3'	50 cycles: 94 °C 10 seconds 60 °C 20 seconds 72 °C 20 seconds
supt5h	5'- TGCACTGCAAGAAGCTGGTGGA -3' 5'-GCTCATAGGAGTGAAGCCACCA-3'	50 cycles: 94 °C 10 seconds 61 °C 20 seconds 72 °C 30 seconds
cdc73	5'- GAGAGAGTGTGGAGGACAAGAAC -3' 5'-GCACGACCTTCTTCTCTGGCTT-3'	50 cycles: 94 °C 10 seconds 61 °C 20 seconds 72 °C 30 seconds
paf1	5'- GGAGGAAGAGATGGAGGCTGAA -3' 5'- CACTTGCCTCATCTCTGTCACC -3'	50 cycles: 94 °C 10 seconds 60 °C 20 seconds 72 °C 30 seconds
ctr9	5'- GTGACACCTACTCTATGCTGGC -3' 5'- TGGCAGCATACAGGTTCTTGGC -3'	50 cycles: 94 °C 10 seconds 60 °C 20 seconds 72 °C 30 seconds

Table 3. Primers and programs used for qRT-PCR.

included and processed in an identical way to control for amplification from contaminating DNA in PCR.

The reaction was inactivated at 70 °C for 15 minutes, topped up with TE to 100 µl and the cDNA stored at -20 °C.

2.2.2.11 Quantitative real time PCR (qRT-PCR)

The real-time PCR reaction were set up in triplicates as follows: 2 µl cDNA template, 5 µl SYBR Green JumpStart Taq ReadyMix (Sigma), 1 µl 5µM primers and 2 µl H_2O were mixed in wells of a 96-well qRT-PCR plate and subject to PCR according to Table 3. cDNA template was used at a 1/125 dilution. Standard curves were prepared by making serial dilutions of control cDNA from cells expressing scrambled shRNA. Transcript quantities were calculated relative to standard curves and normalised to CD79b or HPRT mRNA. Data were acquired and analysed in a LightCycler 480 qRT-PCR machine (Roche) and corresponding software.

2.2.2.12 Isolation of genomic DNA

Genomic DNA was isolated from cells using PUREGENE products Genomic DNA Isolation (Gentra Systems, USA) according to manufacturer's recommendations. 50 µl of TE (pH 8.0) were added to DNA pellets and left to rehydrate over night at RT. Samples were stored at -20 ºC.

2.2.2.13 Southern blotting

The southern blotting probe was prepared by PCR amplifying a part of the AID gene using Pfu Turbo DNA polymerase and PCR parameters: primer annealing 57ºC, 30 seconds, polymerization 72ºC. 60 seconds, 35 cycles. Primers used were: 5'-CTTATGACTGTGCCCGACATG-3' and 5'-CTGCTCCAGAGAGGACAC-3'. Radioactive probe was generated using α-^{32}P dCTP together with the PrimeIt labelling kit (Stratagene, USA,) according to the manufacturer's instruction. The probe was purified using a 50 G nick column (GE Healthcare). Genomic DNA was digested with *KpnI* and subject to gel electrophoresis on a 0.7% agarose gel at 4ºC. The gel was immersed for 45 minutes in 0.5 M NaOH and 1.5 M NaCl to denature the DNA, followed by immersion into 20 x SSC. Capillary transfer blotting was performed using hybond N+ membrane (Amersham, UK). After transfer, the membrane was fixed by placement on 3MM paper soaked in 0.4M NaOH for 20 minutes. After washing the

membrane in 2X SSC it was transferred to a rotating hybridisation bottle together with prewarmed hybridisation solution for pre-hybridisation (1 hour). The radioactive probe was incubated with the membrane in fresh hybridisation solution at 65ºC over night. After washing at least four times with southern blot washing buffer, the membrane was exposed to a Fuji IP plate and readout by a Fuji FLA-5000 scanner.

2.2.2.14 Midi-preparation of plasmid DNA for transfection into cells.

Plasmids were transformed into the bacterial subcloning strain XL1-Blue. A single bacterial colony was inoculated into 50 ml LB medium containing antibiotics and incubated for 14-16 hours at 37 ºC with shaking (225 rpm).

Isolation of plasmid DNA was then carried out using PureYield Plasmid Midiprep System, containing a washing step to eliminate endotoxins, (Promega, UK) according to the manufacturer's instructions. The DNA was eluted with 600 µl of nuclease-free water. Plasmid DNA was stored at -20 ºC.

2.2.2.15 Generation of recombinant DT40 lines

pKW5001-5005 were generated as derivatives of the plasmid described in (Pauklin et al., 2009). Briefly, a 3' fragment of the chicken AID gene lacking the last 33 nucleotides, containing *PshAI* and a *SpeI* restriction sites was generated by PCR using *Pfu* turbo polymerase, primers used were 5'- AGAGACCAGTGTCCTCTCTGGA – 3' and 5'- NNNACTAGTATCATCTACTTCATACAGTGG – 3', and cloned via the *PshAI* and a *SpeI* restriction sites into pSPM509-5014, plasmids based on (Pauklin et al., 2009), Appendix Figure 29, encoding an AID targeting vector carrying *Renilla* luciferase (RL), GFP, 3 x FLAG (3F), 3 x FLAG 2 x TEV 3 x MYC (3FM) or 3 x MYC (3M) tagged AID.

1 x 10^7 DT40 cells were spun down at 90 g for 5-10 minutes, resuspended in RPMI growth medium to give a 500 µL final volume. 25-50 µg of *Not I* linearised targeting construct was added to the culture solution and transferred to a 4 mm electrode gap Gene Pulser Cuvette. The cuvette was placed on ice for 10 minutes and electroporated using the following parameters: voltage = 550 V, capacitance (FD) = 25 µF, resistance = ∞, followed by placing the culture on ice for five minutes. The cells were transferred to a flask, containing 10 ml filter-sterilized culture medium followed by two hour incubation, whereafter 10 more ml media was added and the culture incubated over

night. The next day, the culture was pelleted and resuspended in 10 ml puromycin containing media (1 μg/ml) and dispersed into 96 well flat bottom plates for single cell cloning (200 μl cell suspension per well). After 5-10 days, surviving clones appeared as colonies. Clones were fed or split every day until reaching high density. Construct integration was monitored by southern blotting and appropriate clones were expanded, for performing selection marker excision. To this end, 3 x 10^5 cells for each recombinant line were washed in RPMI growth medium and resuspended in 300 μl RPMI growth medium containing 6 μM purified HTN-Cre protein, followed by incubation for 3 hours. 300 μl RPMI growth media and 30 μl FCS containing 10% chicken serum was added to the cells and they were incubated further over night. The next day, cells were diluted 1:8300 in RPMI growth media and plated at 100 μl cell suspension per well in a 96 well flat bottom plates. Colonies were grown and expanded as describes and subject to southern blot analysis to determine successful recombinants.

2.2.2.16 Lipofectamine transfection of 293 cells

Transfection of 293 cells with Lipofectamine 2000 (Invitrogen, UK) was performed as suggested by the manufacturer. The day before transfection, cells were plated in a density so that the next day, cells would have a confluency of 90-95%. The media used for Lipofection was Opti-MEM Reduced Serum Medium (Invitrogen). Cells were grown in Opti-MEM supplemented with 10% FBS. Approximately 5 hours before transfection, fresh media with FCS but without antibiotics was added to the cells. Lipofectamine was diluted in Opti-MEM medium without serum and incubated for 5 min at RT. Plasmid DNA was diluted in Opti-MEM medium without serum and combined with Lipofectamine, followed by incubation at RT for 20 min. The medium in plates was substituted with serum free medium and the DNA-Lipofectamine complexes were added to the wells. The cells were incubated for 6 hour at 37 ºC and the medium was once again substituted with medium containing 10% FCS. After incubation for 24 hours, fresh media with serum and antibiotics were added to the cells.

2.2.2.17 Luciferase Interaction Assay

Renilla luciferase - AID fusion proteins (full length and N-terminal as well as C-terminal truncations) were generated by cloning the human AID ORF/peptides downstream of the Renilla luciferase gene. 10 cm plates of 293T cells were transfected

with the indicated DNA using Lipofectamine 2000 (Invitrogen) according to the manufacturers instructions. 48 hr later, cells were lysed in 293 Lysis buffer, centrifuged and supernatants harvested. Lysate volumes were adjusted to equivalent luciferase counts based on the total lysate counts as determined by a Glomax luminometer (Promega, USA). Lysates were incubated with rotation over night at 4˚ C in the presence of antibody: PC10 antibody for endogenous PCNA (CRUK antibody service), M2-FLAG antibody (Sigma) for transfected FLAG-tagged PCNA, or an IgG2a antibody as a control (Serotec, USA). Protein A/G sepharose beads (Santa Cruz, USA) were added for 1 hr at 4˚ C, beads washed and resuspended in 30 µl 293 Lysis buffer. 20 µl were added to 20 µl of 1x Renilla Luciferase Assay Buffer (Promega, USA), followed by addition of 75 µl luciferase substrate and luciferase reading in the Glomax luminometer. Per cent binding was determined from counts from each sample as a proportion of the total lysate counts.

2.2.2.18 AID/APOBEC chimera GFP-IP

Chimera GFP-fusion plasmids were a gift from M. S. Neuberger and S. Conticello (Conticello et al., 2008). 293T cells were transfected with the indicated DNA using Lipofectamine 2000 (Invitrogen) according to the manufacturers instructions. Cells were lysed in 293 lysis buffer. Chimera protein expression levels in total lysates were checked by Western blotting and adjusted to equivalent levels with mock lysate. Lysates were incubated with control or FLAG-PCNA lysates and a polyclonal anti-GFP antibody (Abcam, ab290), rotating over night at 4˚ C. Subsequently, protein A agarose beads were added and rotated a further 1 hr at 4˚ C. Following 3 x washes, proteins bound to the beads were analysed by Western blotting. Chimera proteins were detected using a monoclonal anti-GFP antibody, while PCNA was detected using M2-anti FLAG antibody.

2.2.2.19 Microscopy

Immunofluorescence studies of DT40 cells were carried out based on (Szuts and Sale, 2006). Irradiation of DT40 cells was performed by spinning down 2 ml of DT40 culture at 400 g onto poly-l-lysine covered coverslips in 24-well tissue culture dishes, dipping them in PBS, blotting the edges and immediately placing them under a UV lamp with a

measured flow rate to receive a dose of 10 J/m^2. Afterwards they were placed back in warm DT40 media for a 40-minute recovery before spinning once more.

After treatment, coverslips were washed in PBS and incubated in 0.2% Triton in PBS for 5 minutes to extract soluble proteins. After another wash in PBS, 4% paraformaldehyde is added for 5 minutes to fix cells, followed by a wash in PBS. IF blocking solution was added for 10 minutes and primary antibody diluted in IF blocking solution was applied to the coverslip.

Fixed cells were stained using mouse monoclonal anti-PCNA PC10 (CRUK antibody services) 1:1000 and rabbit polyclonal anti-GFP (abcam ab290) 1:2000, and coverslips incubated for 1.5 hours at 37 °C in a humidified box in the dark. After a 10-minute immersion in PBST to wash, a 1:1000 staining mix of secondary antibodies (goat anti-mouse coupled to Alexa Fluor 594; goat anti-rabbit coupled to Alexa Fluor 488, Invitrogen) was applied and coverslips incubated for 1 hour at 37 °C in a humidified box in the dark. After another 10-minute wash in PBSTS coverslips were mounted upside down on mounting medium containing DAPI (ProLong Gold, Invitrogen) and sealed with nail varnish.

Images were acquired on a Zeiss Axio Imager M1 microscope using a Plan Apochromat 100×/1.4 oil objective lens (Zeiss) equipped with an ORCA-ER camera (Hamamatsu) and controlled by Volocity 4.3.2 software (Improvision).

Chapter 3. Analysis of AID cofactors on chromatin

3.1 Results

3.1.1 Biochemical identification of proteins associating with AID on chromatin of B cells

The purpose of this part of the thesis was to identify and characterise novel interacting partners of AID, as well as to pursue the connection between AID and the transcription machinery in more detail. To isolate proteins that interact with AID on chromatin, where AID activity takes place, two conditions were required to be fulfilled: First, AID had to be purified from a source that naturally expresses AID and undergoes AID dependent *Ig* diversification. Second, the purification had to be enriched for chromatin bound AID. An initial mass spectrometric analysis of the protein complexes that associate with AID on chromatin was performed by S. Pauklin and S. Petersen-Mahrt, CRUK, and is the starting point of this aspect of my thesis work.

Cell lines suitable for this analysis had been engineered in the lab (Pauklin et al., 2009), using the DT40 B cell line as a parent. Briefly, in these lines, one allele of endogenous AID was tagged at the C-terminus with various peptide tags, either 3xFLAG peptides (3F) or the combination of 3xFLAG peptide - 2xTEV cleavage sites - 3xMyc peptides (3FM). This yielded cells expressing tagged AID protein at endogenous amounts (DT40-3F and DT40-3FM). Even though it is known that the C-terminus of AID plays an important role in sub-cellular localisation, the introduction of the mono-allelic C-terminal tags did not change AID's subcellular localisation or immune diversification activity (data not shown).

To characterise proteins that associate with chromatin bound AID, a subcellular fractionation protocol for DT40 adapted from a recently developed method for isolation of chromatin bound protein-complexes (Aygun et al., 2008) was used. As mentioned above, in diversifying B cells AID is usually sequestered in the cytoplasm, thus to overcome this problem, it was required to grow large amount of cells ($1 - 2 \times 10^{10}$). Cell lysates were sub-fractionated into cytoplasm, nucleoplasm and chromatin fraction and the fractions were subject to FLAG peptide immunoprecipitation (IP) of AID followed

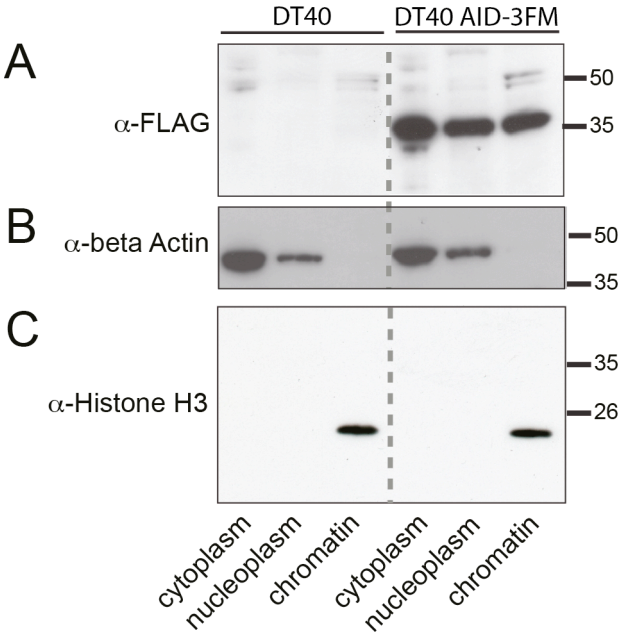


Figure 7. Subcellular fractionation of DT40.

Analysis of DT40 subcellular fractions. 15 μg of each fraction was loaded (not representative of the total protein amount in each fraction, total chromatin bound protein is approx. 2% of total protein in all fractions) **A**, Lysates of the wild type (DT40) and the recombinant strain (DT40 AID-3FM) are probed for FLAG expression in all 3 fractions, as well as probed for markers of **B**, cytoplasm (beta-Actin) and **C**, chromatin (Histone H3).

by one-dimension SDS-PAGE and mass spectrometry identification. A representative fractionation experiment can be seen in Figure 7. AID was present in each fraction, and chromatin AID was estimated to be less than 2 % of total AID-3FM.

Data obtained from large scale mass spectrometric analysis of AID-3F chromatin fraction comprised 151 peptide hits corresponding to 52 proteins (Figure 8 A). Non-tagged control cells were used to generate a control IP dataset, which was used to filter peptide hits as unspecific associations. By this method, hits were reduced to 75 specific peptide hits corresponding to 25 proteins. An extended list of mass spectrometry identified proteins can be viewed in Appendix Table 4. The majority of the AID associated proteins on chromatin were part of mRNA processing (Figure 8 B). 15% of peptides could be assigned to RNA polymerase alone. More than 80 % of all isolated peptides were interconnected in a functional and physical interacting network (as classified using Ingenuity Systems Pathway Analysis software, data not shown),

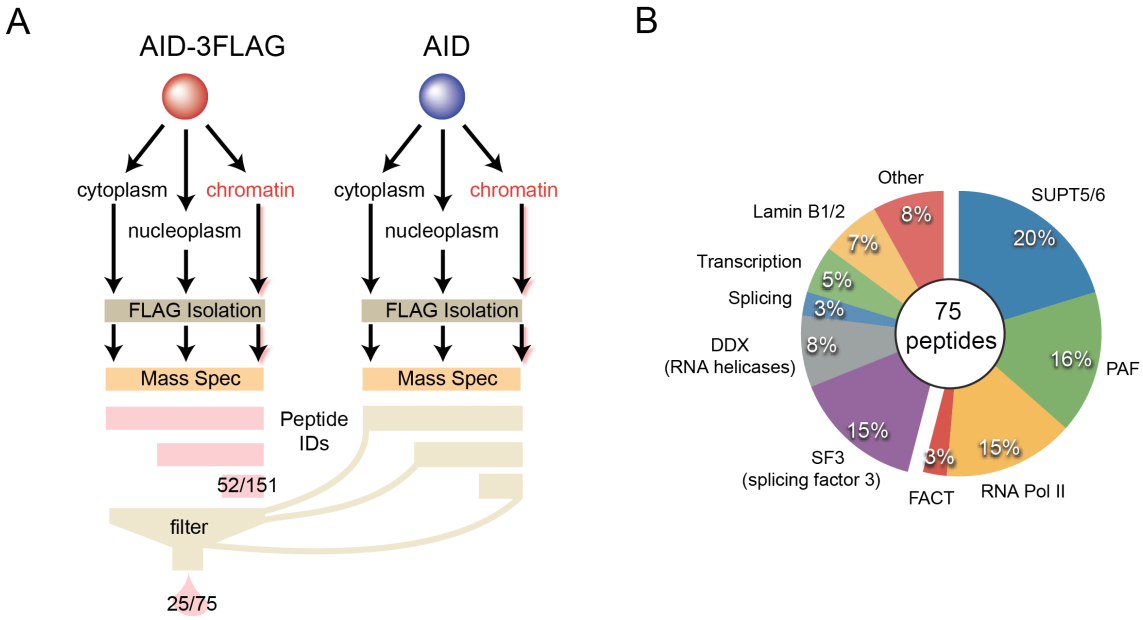


Figure 8. Summary of a mass spectrometric analysis of AID isolated from B cell chromatin.

A, Schematic of complex isolation and analysis. DT40 cells expressing tagged AID (red) and untagged AID (blue - control) were fractionated into cytoplasm, nucleoplasm and chromatin. Each fraction was then subjected to anti-FLAG bead IP and mass spectrometry analysis. 151 peptides corresponding to 52 proteins were identified from the chromatin fraction only. The equivalent Mass Spec data set from control cells was used as a filter to remove false positives (yellow lines/boxes), leaving 75 peptides (25 proteins). **B**, Analysis of filtered peptide hits of the chromatin fraction. The most abundant peptides identified were functionally grouped and groups are displayed in a pie chart as percent of total peptides (75) that could be assigned to each complex. The right 54% of the pie chart comprises proteins directly involved in RNAPII elongation.

(in collaboration with Siim Pauklin)

clustering around RNAPII. They include RNAPII elongation factors (SUPT5H, PAF complex components - PAF1, CTR9, LEO1), chromatin modifiers (SUPT6H, DNA topoisomerase I, FACT complex components - SSRP1 and SUPT16H), exosome components (SUPT6H), and RNAPII pausing and re-initiation factors (PAF complex, SUPT5H, and SUPT6H) (Figure 8 B); with some proteins serving multiple functions. Furthermore, another 20 peptides comprised proteins involved in RNA metabolism (splicing associated factors and RNA helicases). The high percentage (54 %) of peptides that co-isolate with AID and are part of the same biological process (RNAPII

elongation) indicate that this isolation and analysis procedure had identified key AID interacting proteins from chromatin of immune diversifying cells.

In line with these findings, recent work from other laboratories has identified SUPT6H, SUPT5H and the exosome to interact directly or indirectly with AID and be involved in CSR (Pavri et al., 2010, Basu et al., 2011, Okazaki et al., 2011), while the identification of the FACT complex in the mass spectrometry data supports the discovery that SSRP1 and SUPT16H are essential for CSR (Stanlie et al., 2010). In our analysis, the PAF complex especially caught our attention as several subunits of this complex (CTR9, PAF1 and LEO1) were identified. This could indicate that a functional form of the complex, in conjunction with AID, could carry out an important role in B cell diversification, and provide the missing link between SUPT5H function and the FACT complex activity.

Taken together, these data prompted me to reproduce the initial mass spectrometry results, confirm AID interaction with the found RNAPII associated proteins, and elucidate the role of RNAPII pausing and elongation factors for AID function.

To validate the interactions, the AID-3FM DT40 line was subjected to the described fractionation protocol, and FLAG immunoprecipitates analysed by replica mass spectrometry (data not shown) and by western blot. The association of AID with PAF1 could be validated by using two different antibodies (Figure 9 A), as well as the association with another subunit of the PAF complex, LEO1 (Figure 9 B). Although not identified in the original mass spectrometry analysis, CDC73 showed significant association with AID on chromatin (Figure 9 C). Importantly, interaction of AID with the aforementioned SUPT5H (Figure 9 D) and SUPT6H (Figure 9 E) could also be confirmed. It is important to note that the chicken genome is not fully characterised, and its annotation is incomplete. Therefore, the number of proteins identified may be underestimated. Furthermore, there are only a limited number of characterised antibodies that recognise the chicken homologues of candidate proteins – as seen by the seemingly weak interaction between SUPT5H and AID. This is most likely due to the poor cross-reactivity of the SUPT5H-specific antibody between species. Nevertheless, this interaction was confirmed by multiple mass spectrometry analysis (3 out of 5). Taken together, this work has for the first time identified and verified AID-associated

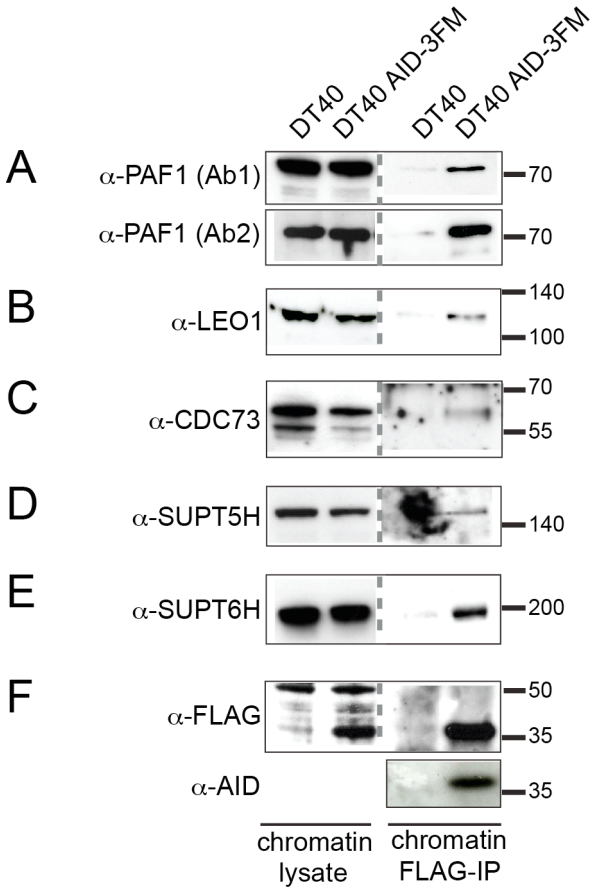


Figure 9. Western blot analysis of interaction candidates of chromatin-bound AID.

Co-IP of RNAPII associated proteins with AID-3FM purified from DT40 chromatin. AID-3FM was immunoprecipitated from chromatin fractions prepared as in Figure 7, and eluted proteins were probed by Western-blotting. Input chromatin lysate from DT40-3FM and DT40 control cell lines prior to FLAG-IP are shown on the left. **A**, anti-PAF1 (using two alternative polyclonal antibodies); **B**, anti-LEO1; **C**, anti-CDC73; **D**, anti-SUPT5H; **E**, anti-SUPT6H; **F**, AID expression and IP. AID is detected by an anti-FLAG antibody and an anti-AID antibody. Anti-AID antibody was unable to detect chicken AID at physiological levels in DT40 lysates.

complexes on chromatin in diversifying B cells. Furthermore, it highlights the connection between AID and RNAPII transcription elongation.

3.1.2 AID is found in a high molecular complex in chromatin of B cells

To further characterise chromatin-AID in its physiological conformation, we determined the possible size of the AID associated complex. The DT40 AID-3FM chromatin fraction was separated using a Superdex 200 column for size exclusion

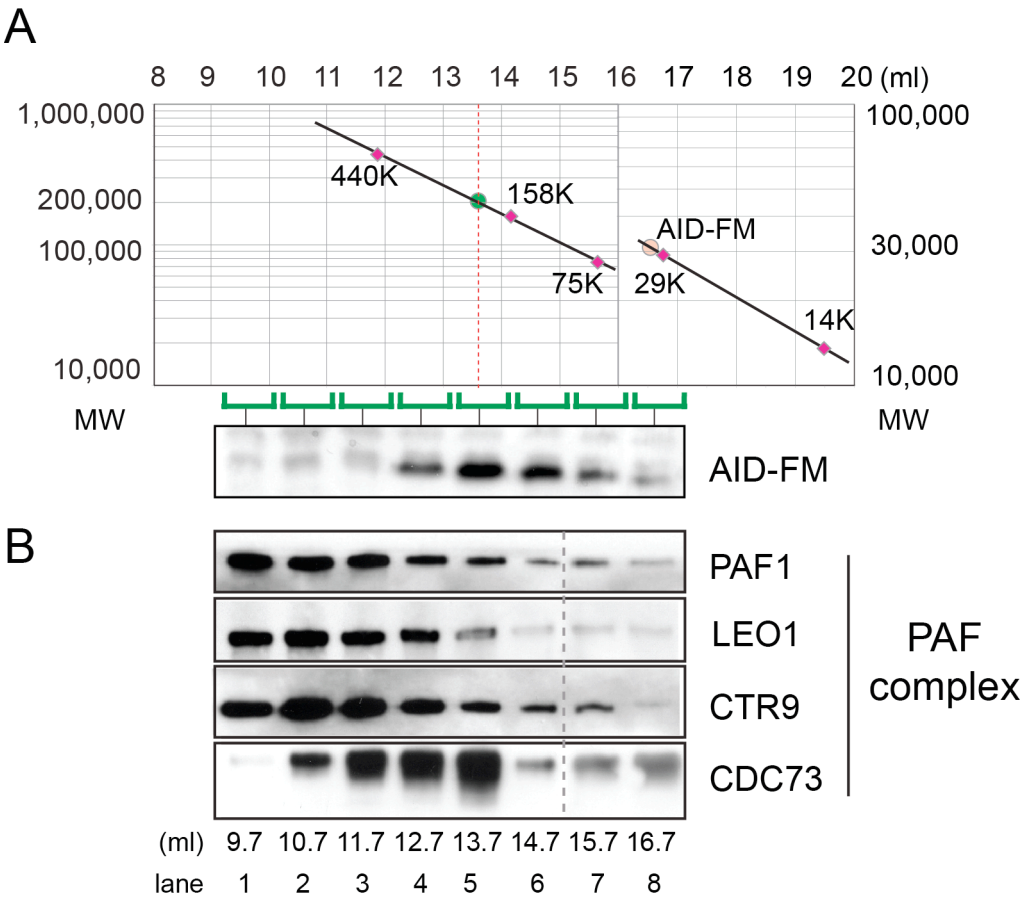


Figure 10. Size exclusion chromatography profile of DT40 chromatin bound fraction.

The chromatin fraction from DT40 AID-3FM cells was loaded onto a Superdex 200 column. **A**, The elution profile of standard proteins is plotted in the graph. Red circle: Theoretical elution volume of AID-3FM (29 kD). Green circle: Peak fraction of AID-3FM (200 kDa). Eluted fractions (elution volume indicated) were analysed by western blot probing for the presence of AID-3FM **B**, Western blot analysis of the fractions as in A probing for PAF1, LEO1, CTR9, and CDC73.

(in collaboration with Gopinath Rangam)

chromatography (SEC). The eluted fractions were collected and analysed by western blot for the presence of AID. The obtained profiles could be compared to the elution profile of standard proteins to assign molecular sizes to fractions. AID was identified as part of a 200 kD protein complex (120 kD to 300 kD - based on standard proteins),

while only a minor fraction of AID eluted at its theoretical monomeric size of 27 kD (Figure 10 A). This is the first demonstration that chromatin bound AID resides in a large heteromeric complex. Earlier observations of AID as a component of a significantly larger complex by Jeevan-Raj et al., 2011 and others were not chromatin specific and were based on cell extracts in which AID was overexpressed as a transgene rather than expressed at endogenous AID levels.

Although the peak of AID elution is fairly sharp, it remains to be determined whether AID resides in one or more species of complex, and how stable the associations with its direct interacting partners are. Certainly, AID does not form one highly stable complex with all partners identified by co-IP, or with one fully identified partner complex. For instance, the human PAF complex is larger than 400 kDa (Kim et al., 2010). To see what type of an overlap with the AID complex we could identify, we analysed the SEC profile of the PAF complex in DT40 chromatin extract. Western blot analysis of the chromatin SEC fractions (Figure 10 B) demonstrated that PAF1, LEO1, and CTR9 co-migrate in a large (>400 kD) complex (elution ml 9.7-11.7), only with the peak trailing fractions overlapping with the AID peak (elution ml 13.7, lane 5). The DT40 PAF complex's migration through SEC is almost identical to that of the PAF complex isolated from HeLa cells (Kim et al., 2010), indicating that the overall complex configuration is conserved in *Ig* diversifying cells. Although CDC73 is known to be part of the PAF complex, it had a slightly different elution profile, with its peak fraction more closely matching that of the AID peak fraction (lane 5).

Taken together, an overlap of a small subset of the AID and PAF protein complexes would be consistent with the data, alternatively, subunits of the PAF complex could overlap with AID and form a complex (e.g. CDC73).

3.1.3 A direct interaction of PAF1 with AID?

Considering the data from our SEC experiment, it was possible that AID interacted with transcription elongation complexes via binding of one subunit. Therefore, interaction of individual candidate proteins with AID was tested using an *E. coli in vitro* IP system. Here, both proteins were co-expressed from the same plasmid in bacteria, one of them immunoprecipitated, and co-immunoprecipitates detected by western blot. The cloned human candidate cDNAs were C-terminally FLAG tagged while human AID was

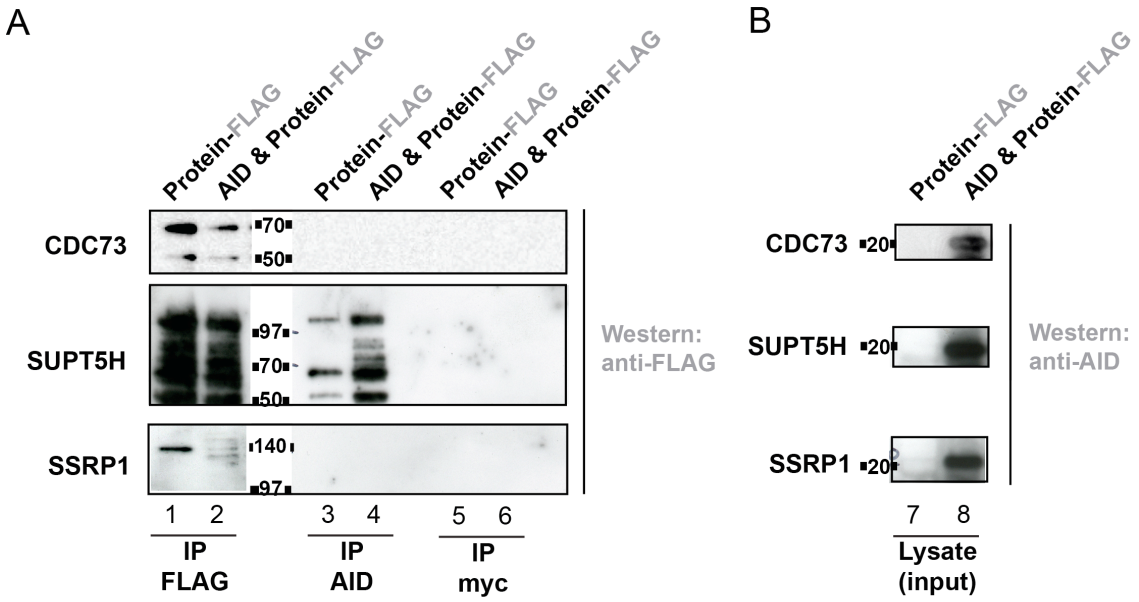


Figure 11. Transcription elongation factors as candidates to interact with AID *in vitro*.

A Candidate RNAPII associated proteins for direct AID interaction were expressed with a FLAG tag in *E. coli* together with AID (CDC73, SUPT5H and SSRP1 as indicated on the left). Lysates were used for IP with anti-FLAG antibody (lanes 1 - 2), anti-AID antibody (lanes 3 - 4) or anti-myc control antibody (lanes 5 - 6) and lysates containing PAF1-FLAG only served as controls. Precipitates were analysed by western blotting using anti-FLAG. **B**, AID expression was confirmed for each experiment in the lysates, shown in a separate blot.

untagged. After protein immunopurification, the co-IP's were analysed by anti-FLAG western blot (Figure 11). The first proteins that were tested were SUPT5H [a protein for which indication of interaction existed (Pavri et al., 2010)], a subunit of the FACT complex (SSRP1) and a subunit of the PAF complex that had co-eluted with AID in the SEC experiment (CDC73). The presence of FLAG-tagged protein could be confirmed by FLAG IP (Figure 11 A, lanes 1 and 2), and AID interaction would be indicated by detection of co-IP of FLAG-tagged protein (Figure 11 A, lanes 3 and 4). In the case of SUPT5H, interaction could not be shown above a background signal (lane 3 versus lane 4 and Figure 11 B, lane 7 versus lane 8). SSRP1 and CDC73 showed no co-IP signal (lane 4). Of all the proteins tested, only FLAG-PAF1 was co-isolated with AID (Figure 12 A). This association was dependent on AID (lanes 4 – 6, lanes 11-13) and did not

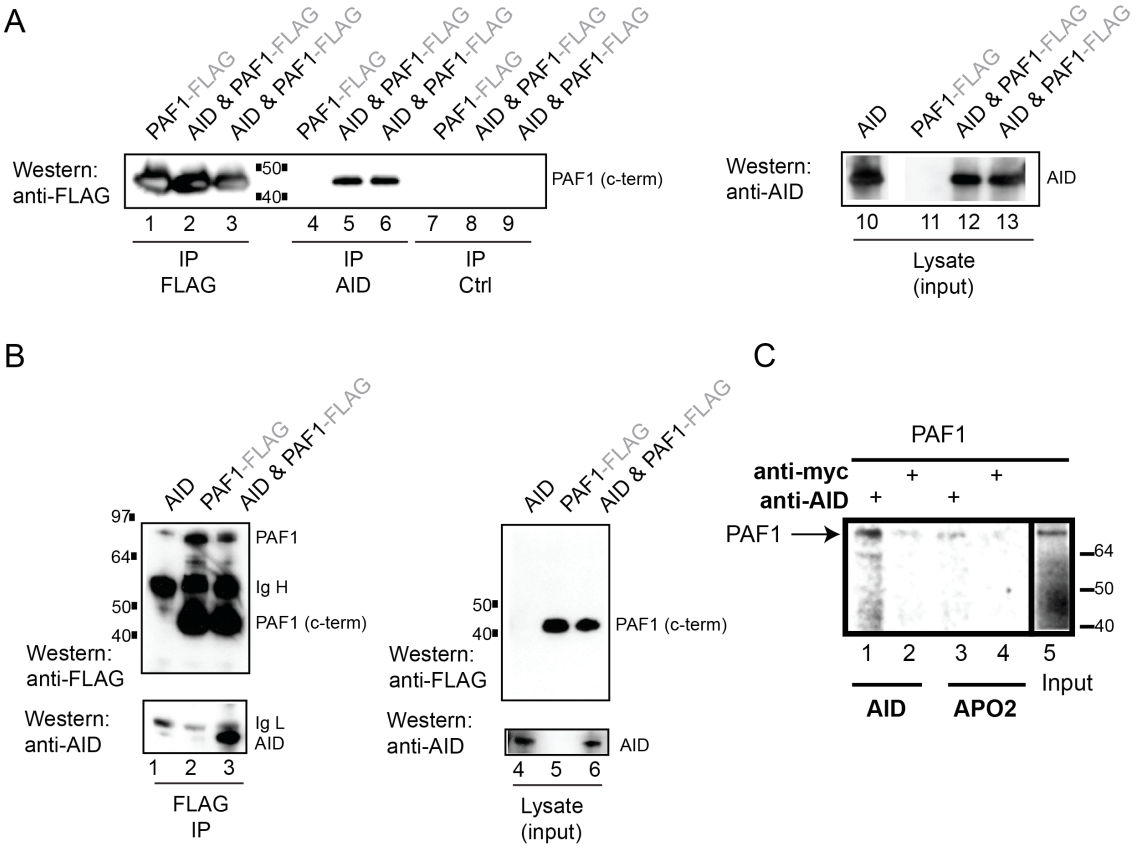


Figure 12. AID and PAF1 interact directly *in vitro*.

A, B; AID and PAF1 were expressed from the same plasmid in *E.coli*. PAF1 fragment is ~50 kDA.
A, PAF1 co-immunoprecipitates with anti-AID. IP was performed as in Figure 11. Lysates for expression of both proteins were made in duplicates, and lysates containing PAF1-FLAG only served as controls. AID expression in the lysates was confirmed, as before. **B**, AID co-immunoprecipitates with PAF1. Reciprocal IP was performed using anti-FLAG beads and *E.coli* lysates as in A. (Top panels) Western blot of IP (lane 1 - 3) and lysates (4 - 6) using anti-FLAG. (Bottom panels) The same IP and lysate as on top probed with anti-AID. Lysate with AID only (lane 1 and 4), with PAF1-FLAG (lane 2 and 5), and with AID and PAF1-FLAG (lane 3 and 6). **C**, AID and PAF1 are interaction partners in a cell free system. Pull down assays were performed with ^{35}S-labelled *in vitro* translated PAF1 and recombinant AID or APOBEC2 (APO2) purified from *E.coli*. Pull down samples were analysed on SDS page followed by autoradiography. Assays performed with control antibody (lane 2 & 4) or APO2 protein (lane 3 & 4) served as controls.

occur in the absence of AID specific antiserum (lanes 7 - 9). Analysis of the FLAG-IP of PAF1 indicated that equal amounts of PAF1 could be immunoprecipitated in all *E. coli* samples (lane 1 - 3). Notably, PAF1 is processed as a shortened fragment in bacteria, even when rare codon strains are used (data not shown). The observed fragment runs near 50 kDa in polyacrylamide gel electrophoresis, while human PAF1 has a calculated molecular weight of 60 kDa but can migrate near 80 kDa (Rozenblatt-Rosen et al., 2005). Small amounts of full length PAF1 protein can be seen after AID IP in Figure 12 B. Although we could identify full length PAF1 in the FLAG IP, the 50 kDa C-terminal FLAG tagged fragment was the prominent band in the lysate. All indicated PAF1 bands in the blot (full length and C-terminal) were verified to contain PAF1 peptides by mass spectrometry. A reciprocal IP experiment was done confirming AID interaction with the C-terminal part of PAF1 (Figure 12 B, lane 3).

The interaction of AID and PAF1 in a system that excludes any other eukaryotic proteins supports a direct interaction. However, to confirm the interaction between AID and PAF1 in a different system, we performed pull down analysis with *in vitro* produced proteins from rabbit reticulocyte lysate. As shown in Figure 12 C, *in vitro* transcribed and translated radiolabelled PAF1 was pulled down by purified recombinant AID protein (lane 1), while APOBEC2 protein (lane 3 and 4) or a control antibody (lane 2) did not have the same effect. Together, these data indicate that AID directly interacts with the PAF complex via the PAF1 protein.

3.1.4 shRNA knock-down of transcription elongation factors have an effect on Class Switch Recombination in B cells

Because the PAF complex associates with AID *in vitro* and *in vivo*, it was determined whether depletion of members of the PAF complex could have a biological effect on AID dependent recombination processes. As a system, the murine B cell lymphoma line CH12 was used for readout of AID function, monitoring AID dependent CSR. The CH12 subclone CH12F3 (called CH12 here) expresses the IgM isotype on the cell surface and can be induced to switch to the IgA isotype *in vitro* using the cytokine combination IL-4, TGF-β and anti-CD40 (Nakamura et al., 1996). Compared to primary B cells, CSR in CH12 occurs at a much higher frequency. In Figure 13 A, FACS

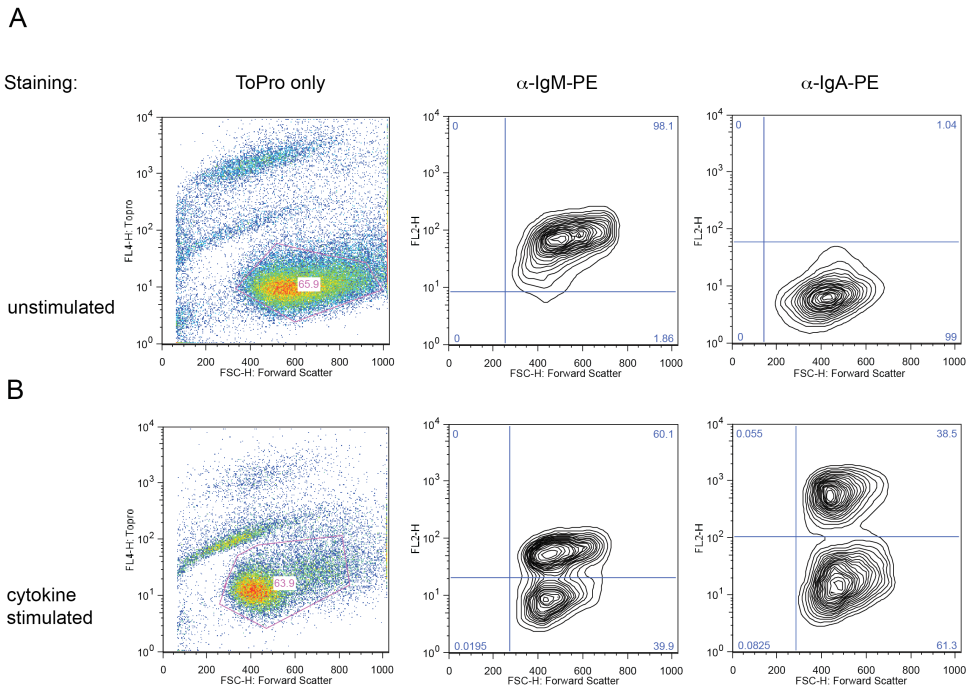


Figure 13. Class Switch Recombination assay in the CH12 B cell line.

A, A control wt CH12 population analysed by flow cytometry. Cells were stained with ToPro dead cell stain in all cases and either for IgM surface expression or IgA surface expression. Live cells were selected in the coloured plot (x-axis: FSC-forward scatter; y-axis: ToPro dye signal). The live subpopulation was used for observation of surface Ig expression by monitoring levels of antibody (PE) staining (contour plots). The percentage of negative or positive staining is indicated in the respective quadrants. **B**, A cytokine stimulated CH12 population analysed by flow cytometry. Cells were stained and analysed as in A.

profiles of a representative CH12 population is shown. The live cell population is determined by ToPro staining. When IgM as well as IgA staining is applied, the FACS plots show that this B cell line is largely IgM^+ and IgA^-. When culturing the cells with cytokines, within 72 hours generally 40-60% of the population undergoes CSR from IgM to IgA resulting in surface expression changing from IgM^+ to IgA^+ (Figure 13 B). Depletion of AID and SUPT5H (Pavri et al., 2010) was used as controls. In order to deplete PAF complex subunits in CH12 B cells, a retroviral system was used to express short hairpin RNA specific for the target mRNA. First, a packaging retroviral vector

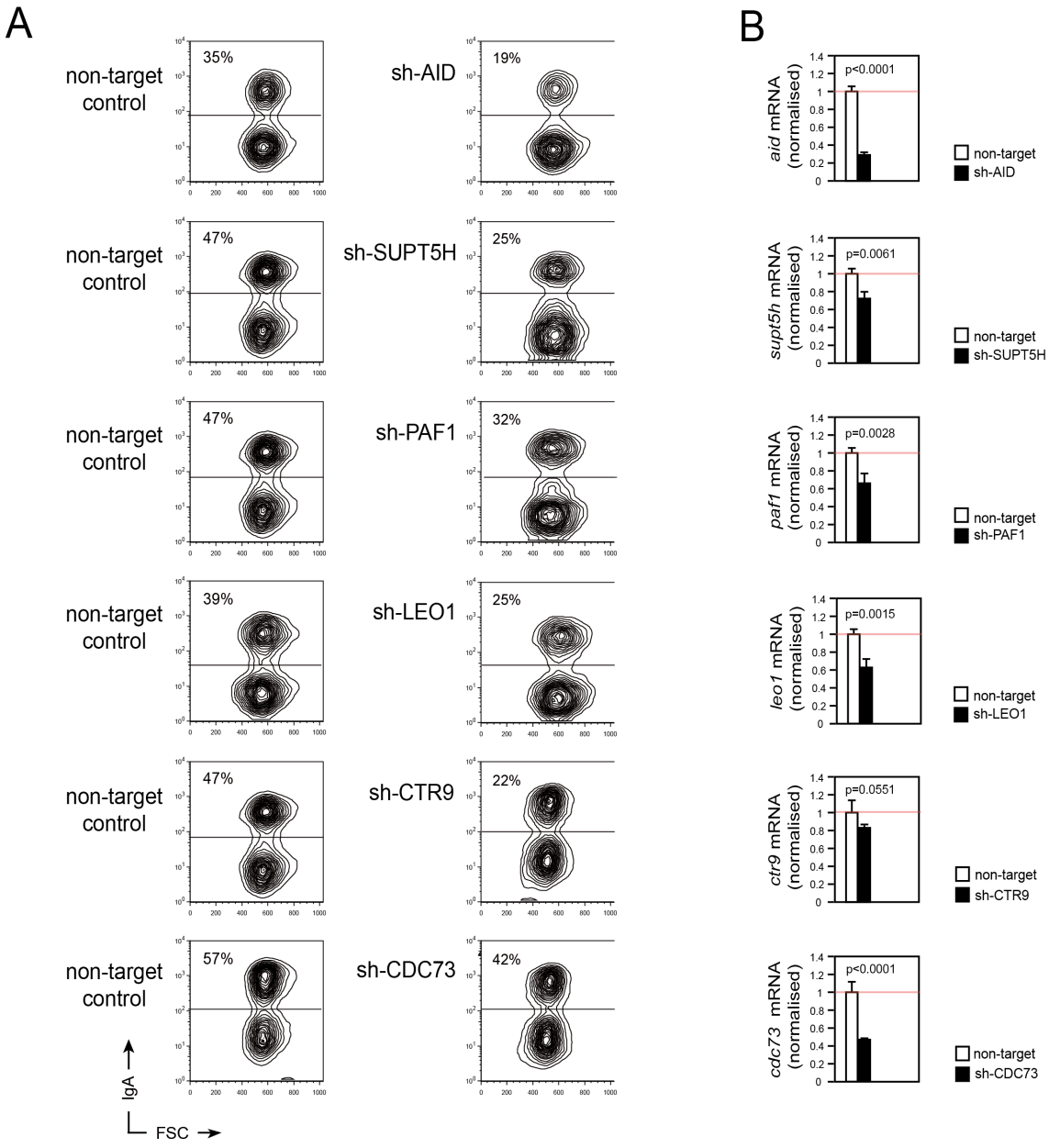

A
non-target control
35%
sh-AID
19%
non-target control
47%
sh-SUPT5H
25%
non-target control
47%
sh-PAF1
32%
non-target control
39%
sh-LEO1
25%
non-target control
47%
sh-CTR9
22%
non-target control
57%
sh-CDC73
42%
IgA
FSC
B
aid mRNA (normalised)
p<0.0001
non-target
sh-AID
supt5h mRNA (normalised)
p=0.0061
non-target
sh-SUPT5H
paf1 mRNA (normalised)
p=0.0028
non-target
sh-PAF1
leo1 mRNA (normalised)
p=0.0015
non-target
sh-LEO1
ctr9 mRNA (normalised)
p=0.0551
non-target
sh-CTR9
cdc73 mRNA (normalised)
p<0.0001
non-target
sh-CDC73

Figure 14. shRNA knock-down of PAF complex members in CH12 results in impaired CSR.

A IgA surface expression was monitored by flow cytometry (FACS) using stimulated CH12 cells expressing shRNAs specific for *supt5h*, members of the PAF complex (*paf1*, *leo1*, *ctr9* and *cdc73*) or a non-target shRNA negative control (representative plots shown). An shRNA specific for *aid* was used as a positive control. Numbers within the FACS plots indicate the percentage of IgA positive cells in the live population upon cytokine stimulation (left panels). **B**, Quantitative RT-PCR of transcripts in knock-down cells (black bars) relative to the corresponding non-target shRNA controls (white bars). *aid*, *supt5h*, *paf1*, *ctr9* and *cdc73* mRNAs are analysed in each respective knock-down population used for CSR assay. cDNA was generated by reverse transcribing mRNA. Transcript C_t values were normalised to *cd79b* or *hprt* mRNA abundance and are shown relative to the levels in the non-target shRNA negative control, set to 1. Error bars: standard deviation. *p* values, Student's t-test.

(In collaboration with Sara Milosevic)

[pCL-Ampho (Naviaux et al., 1996)] together with a vector carrying a shRNA hairpin expression cassette (HuSH pGFP-V-RS, provided by OriGene, USA; or pLMP shRNA vector for AID) was transfected into host cells (293T cells). pCL-Ampho expresses retroviral proteins, including an amphotropic envelope protein, which are sufficient to assemble infective retroviral particles that self-inactivate after the first viral cycle. pGFP-V-RS carries a GFP marker for monitoring host cell transfection, a shRNA hairpin, a puromycin expression cassette, and flanking long terminal repeats that serve as a signal for packaging the interjacent DNA into the viral particles. For AID shRNA introduction, a pLMP backbone was used analogous to pGFP-V-RS. Next, viral supernatant produced by the host cells was used to infect the target cells (CH12) and successful transformants were selected using puromycin resistance. Finally, the capacity of transduced B cells to undergo CSR to IgA was determined by flow cytometry in an assay analogous to the experiment in Figure 13. Transduced cell populations were stimulated to undergo CSR *in vitro* (Figure 14 A). The efficiency of shRNA dependent gene knock-down was determined by quantitative real-time PCR (qRT-PCR) (Figure 14 B). Consistent with previous reports, it was found that knock-down of *aid* and *supt5h* resulted in a significant reduction in the efficiency of CSR (Figure 14 A). Remarkably, knock-down of *paf1*, *leo1*, and *ctr9* resulted in a similar reduction in the efficiency of

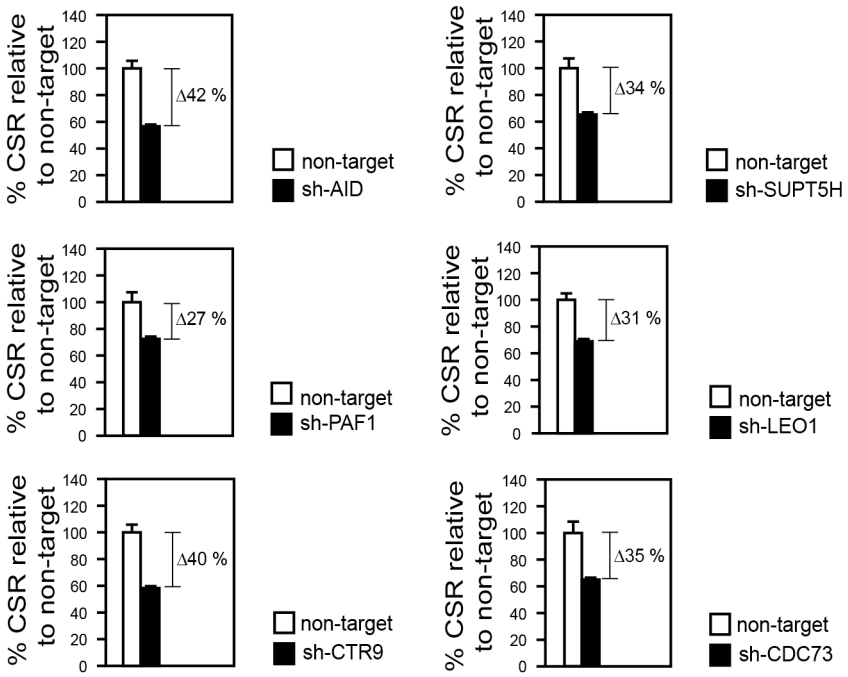


Figure 15. CSR reductions in multiple knock-down experiments.

Average percentage (+ standard deviation) of CSR in knock-down cells (black bars) relative to the non-target shRNA control (white bars) from three independent experiments. CSR in cells expressing the non-target shRNA control was set to 100 %. The % difference in CSR efficiency between non-target and target shRNA knock-down is indicated. Error bars: standard deviation.

(In collaboration with Sara Milosevic)

CSR (Figure 14 A). These results could be observed over 3 independent experiments with reductions ranging from 27 % to 42 % (Figure 15). As the PAF complex is part of the RNAPII transcription machinery, the knock-down of its individual subunits could have broader influences on the cell than just altering AID's function at the *IgH* locus during CSR. In order to minimise the selection pressure of general transcription profile alterations that could occur, cells were subject to CSR assay after no more than 5 days post infection. Furthermore, the effect of gene knock- down on switch region transcription and *aid* mRNA expression was monitored in samples used for CSR assays (Figure 16). As observed in previous studies, in *aid* and *supt5h* knock-down cells, germline transcription at the μ (donor) (Figure 16A) or α (acceptor) (Figure 16 B)

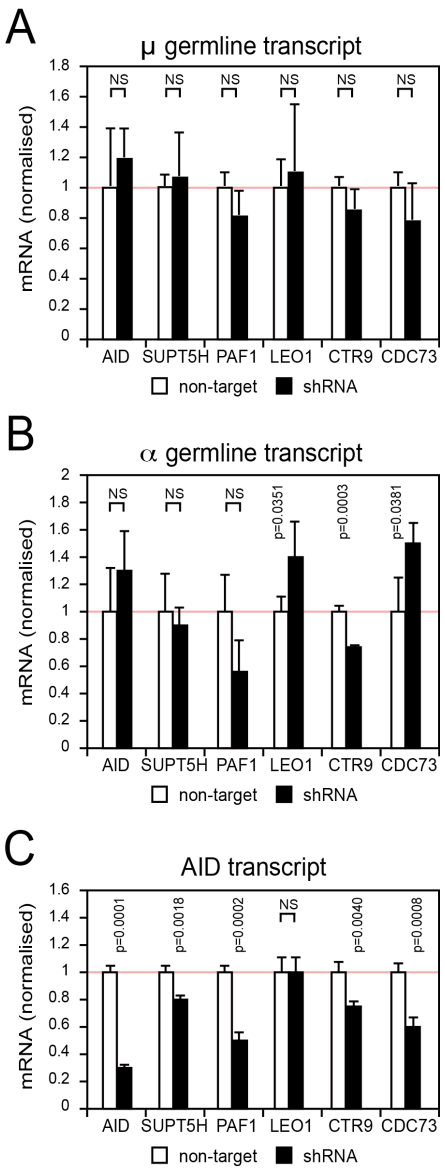


Figure 16. Switch region and AID transcript levels in shRNA knock-down cells

Quantitative RT-PCR of transcripts in knock-down cells (black bars) relative to the corresponding non-target shRNA controls (white bars). *aid*, *supt5h*, *paf1*, *ctr9* and *cdc73* knock-down cells used for CSR assay in Figure 14 are analysed. All transcript C_t values were normalised to *cd79b* or *hprt* mRNA abundance and are presented relative to the levels in the non-target shRNA negative control, set as 1. Error bars: standard deviation. *p* values, Student's t-test. NS: not significant. **A,** Analysis of μ germline transcript in knock-down cells as indicated. **B**, Analysis of α germline transcript in knock-down cells as indicated. **C**, Analysis of *aid* transcript in knock-down cells as indicated.

(In collaboration with Sara Milosevic)

switch regions were unaffected. While transcription at the donor switch region was also unaffected by the knock-down of any of the PAF complex subunits (Figure 16 A), it was found that knock-down of *leo1*, *ctr9* and *cdc73* resulted in altered levels of germline transcription at the acceptor switch region, with a significant reduction of transcription in the *ctr9* knock-downs (Figure 16 B). Furthermore, knock-down of *paf1*, *ctr9* and *cdc73* resulted in a significant reduction in the level of *aid* mRNA (Figure 16 C). Importantly however, knock-down of *leo1* did not reduce AID mRNA expression (Figure 16 C), nor reduce the levels of germline transcripts (Figure 16 A and B), yet CSR was significantly reduced (Figure 14 A and Figure 15). This indicates that LEO1, as a PAF complex component, can directly influence AID's activity in CSR, even if the knock-down was incomplete. The direct role in CSR of the other members of the PAF complex was not possible to delineate, as their effect could also be indirect by altering the expression of CSR relevant transcripts.

In this project, I was able to show that on chromatin AID associates with the PAF complex, likely mediated through the direct biding of AID to PAF1, and that this complex/interaction plays a significant role during CSR.

Chapter 4. Analysis of AID cofactor PCNA

4.1 Results

Taking a different approach than in the first part of the study, here, a candidate protein was selected for AID association based on previous knowledge on AID function and *Ig* diversification cofactors. Partners of AID will help control AID-induced DNA alterations by regulating targeting of AID (addressed in the previous part of the study) or by regulating the outcome of the AID-induced lesion. It is well established that AID-induced lesions can be processed via BER or MMR related processes, including employment of specialised polymerases (TLS). BER, MMR and TLS are dependent on polymerases and, thus, on a processivity cofactor, the predominant one being the DNA sliding clamp PCNA. PCNA has also been shown to have a role in *Ig* diversification *in vivo*. Moreover, it has the ability to act as a switch between a set of DNA damage response pathways through variable modification at its K164 residue, comprising monoubiquitination, polyubiquitination and sumoylation. Therefore, I explored PCNA as a candidate AID interacting protein.

4.1.1 The interaction between AID and PCNA

4.1.1.1 In vitro association of AID and PCNA

In an initial experiment, AID's ability to interact with PCNA was tested in a cell-free system similar to the system used in 3.1.3 (Figure 17A). For this experiment, AID or a control protein (luciferase) was expressed and labelled with ^{35}S radionuclide using a TNT rabbit reticulocyte lysate system. The protein mixture was then incubated with recombinant PCNA protein or control protein (recD protein, another DNA binding protein, used to control for unspecific indirect binding via DNA). After IP with anti-PCNA or isotype control antibody, a co-precipitation can be observed only in those samples where PCNA and anti-PCNA were used. Experiments in our as well as other laboratories have shown that AID can unspecifically adsorp to gels and resins such as agarose (Metzner et al., 2010), therefore another system (employing a different bead product) was used to corroborate the interaction between AID and PCNA. To this end, the *E. coli* co-expression system described before was used. AID and PCNA were co-expressed from the same plasmid in *E. coli* followed by co-IP. Western blot analysis of

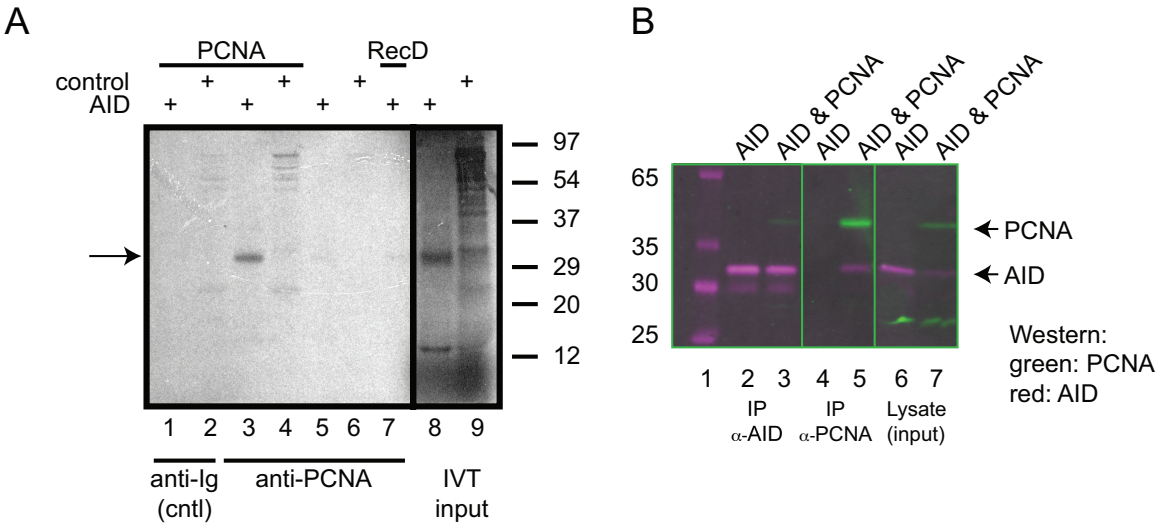


Figure 17. AID interacts with PCNA *in vitro*.

A, *In vitro* co-IP. Radiolabelled AID or luciferase control proteins are expressed using rabbit reticulocyte lysate and subjected to incubation with either purified recombinant his-PCNA, RecD control protein or without protein. IPs are performed using anti-PCNA antibody or IgG isotype control antibody, and analysed on SDS page followed by autoradiography. **B**, Co-IP in E. coli. AID only or AID and PCNA are simultaneously expressed in *E. coli*, followed by IP of AID or PCNA. Using LICOR infrared co-identification in Western-blot, AID protein gives a magenta signal and PCNA a green signal in inputs and reciprocal IP's.

IP and lysates was performed using the two-colour infrared LICOR Odyssey detection system. Here too, a specific PCNA and AID interaction could be identified (see Figure 17 B). The lysates showed AID (magenta signal) and PCNA (green signal) expression (lanes 6 and 7), and were subject to IP using anti-AID as well as anti-PCNA antibodies. A PCNA specific band was identified as a co-precipitate in anti-AID samples (lane 3). This interaction was also identified with a reciprocal IP, when using an anti-PCNA antibody for IP, an AID association was identified (lane 5).

Together, these data indicate that AID and PCNA interact directly *in vitro*.

4.1.1.2 In human cells

To show the interaction in higher eukaryotes, AID and PCNA expression vectors were co-transfected into 293 human embryonic fibroblasts. FLAG-tagged PCNA was used together with a *Renilla reniformis* luciferase-AID fusion (RL-AID). Following anti-FLAG-IP or anti-PCNA-IP, AID binding was measured by the relative amount of

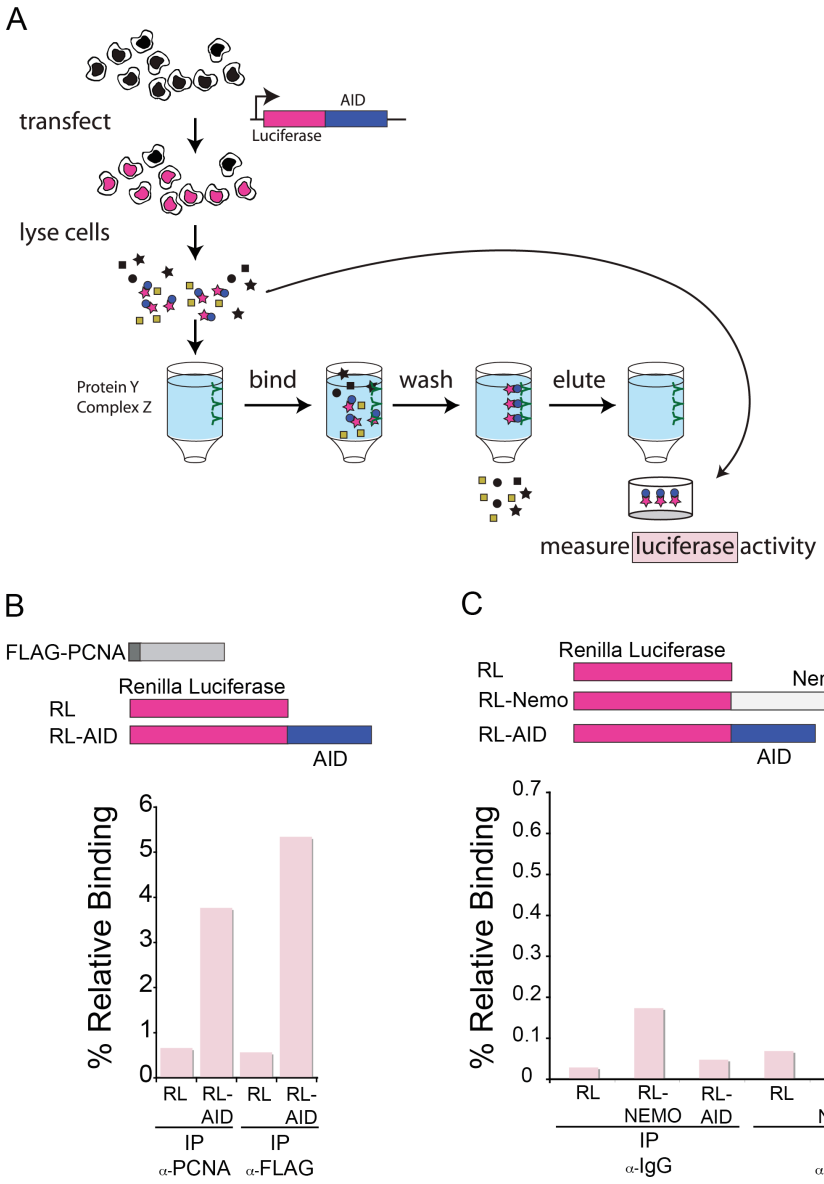

A
transfect
AID
Luciferase
lyse cells
Protein Y
Complex Z
bind
wash
elute
measure luciferase activity
B
FLAG-PCNA
Renilla Luciferase
RL
RL-AID
AID
% Relative Binding
RL
RL-AID
RL
RL-AID
IP
α-PCNA
IP
α-FLAG
C
Renilla Luciferase
RL
RL-Nemo
Nemo
RL-AID
AID
% Relative Binding
RL
RL-NEMO
RL-AID
RL
RL-NEMO
RL-AID
IP
α-IgG
IP
α-PCNA

Figure 18. AID and PCNA interaction in 293 fibroblasts using luciferase fusion.

A, Schematic representation of the RL-AID interaction assay. The RL-AID protein is expressed in cells that are lysed and incubated with immobilised antibody targeted to bind PCNA. After washing and eluting, the luciferase activity of the eluate is detected by an enzymatic luciferase assay as a measurement of bound RL-AID. **B**, 293 fibroblasts transfected with FLAG-PCNA and *Renilla* luciferase or RL-AID fusion construct are used for IP. After anti-PCNA or anti-FLAG IP, bound luciferase is measured by luciferase activity assay. **C**, 293 fibroblasts transfected with RL-AID fusion construct or control constructs only are used for co-precipitation. Control constructs comprise *Renilla* luciferase only, as well as RL fused to the small protein NEMO (RL-NEMO). After anti-PCNA and control IP, bound luciferase is measured as in B.

(In collaboration with Maria Simon)

bound luminescence (Figure 18 A), whereby the ratio of bound luminescence versus luminescence in the input was used to judge enrichment.

Anti-FLAG-IP or anti-PCNA-IP showed luminescence binding, while controls (RL only or control IPs) did not (Figure 18 B). Because the anti-PCNA IP showed AID interaction, and because PCNA is an abundant nuclear protein, it was tested if AID interacted with endogenous PCNA. To this end, only RL-AID was transfected. As with the co-transfection experiments, AID could be identified to interact with PCNA (Figure 18 C). These data demonstrate that AID can interact with PCNA *in vivo*.

4.1.1.3 In B cells

To confirm the interaction of AID and PCNA in cells where physiological AID function can be observed, co-IP experiments were performed using the previously described DT40 derivatives, DT40 AID-3FM and DT40 AID-3F. Tagged AID was recovered from whole cell lysates using immobilised M2 anti-FLAG antibody, and co-precipitated PCNA was detected using western blot (see Figure 19 A). Notably, the quantitative difference in PCNA signal parallels the amount of AID-FLAG recovery (Figure 19 A), rather than the PCNA expression (Figure 19 B). This data indicates that AID and PCNA interact under physiological conditions.

4.1.1.4 Interacting domains in PCNA

To identify a domain or stretch of amino acid that would confer AID binding in the PCNA protein, a truncation series of FLAG-PCNA (Figure 20 A) was generated. Using

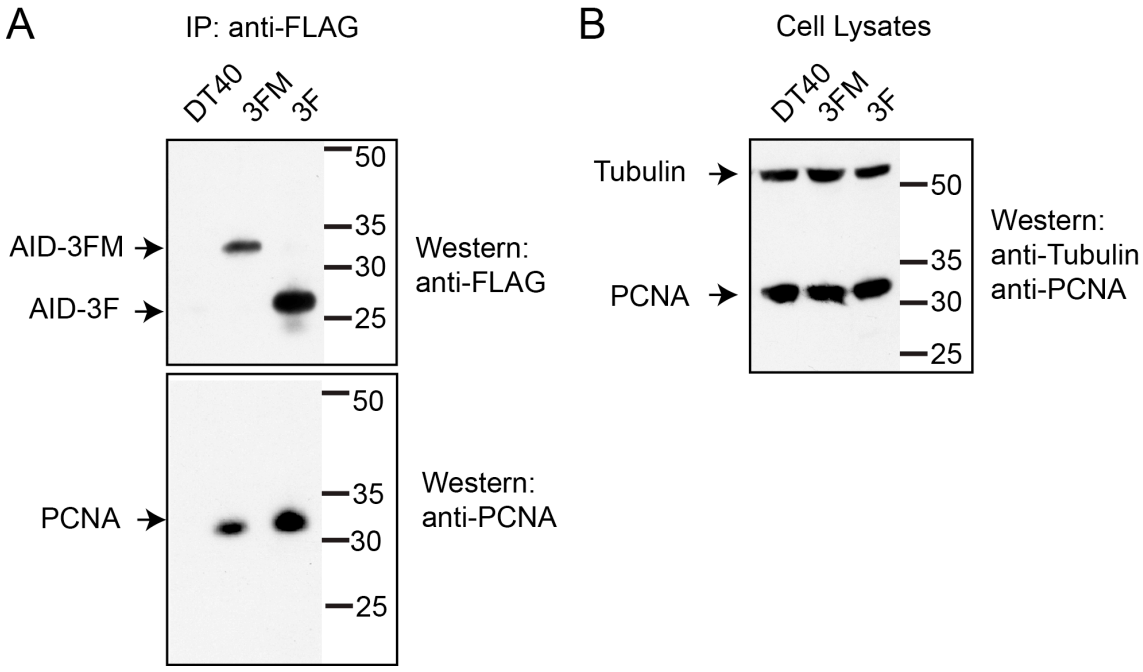


Figure 19. Endogenous interaction of PCNA and AID in B cells.

A, FLAG-IP samples from control cells (DT40 parental wt strain), DT40 AID-3FM and DT40 AID-3F are run on a western blot and probed with an anti-FLAG or an anti-PCNA antibody. **B**, Expression levels of PCNA in the lysates. Western blots probed with an anti-PCNA antibody and with an anti-tubulin antibody (loading control).

the above described 293 expression and luciferase assay system, a decrease in AID binding to PCNA with C terminal truncations near the critical residue K164 in PCNA could be identified (Figure 20 A). This suggests a C- terminal AID interacting domain on PCNA. Because of the strong structural requirements for PCNA trimerisation and function, further truncations may create an artificial situation and are likely to corrupt protein structure. This led me to test a small number of PCNA point mutations, located within the potential region of interaction (Figure 20 B). Please refer to Appendix Figure 31 for the localisation of all mutants and residues of interest in a 3 dimensional model of human PCNA.

The point mutant plasmids were constructed using the *E. coli* co-expression vectors as in Figure 17. Mutants were expressed together with AID, then anti-AID IP was performed and co-IP monitored. To ensure western blot signals were linear, the LICOR Odyssey system based on infrared signal emitting secondary antibodies was used. A reduced affinity to AID of the tested PCNA mutants could not be observed when

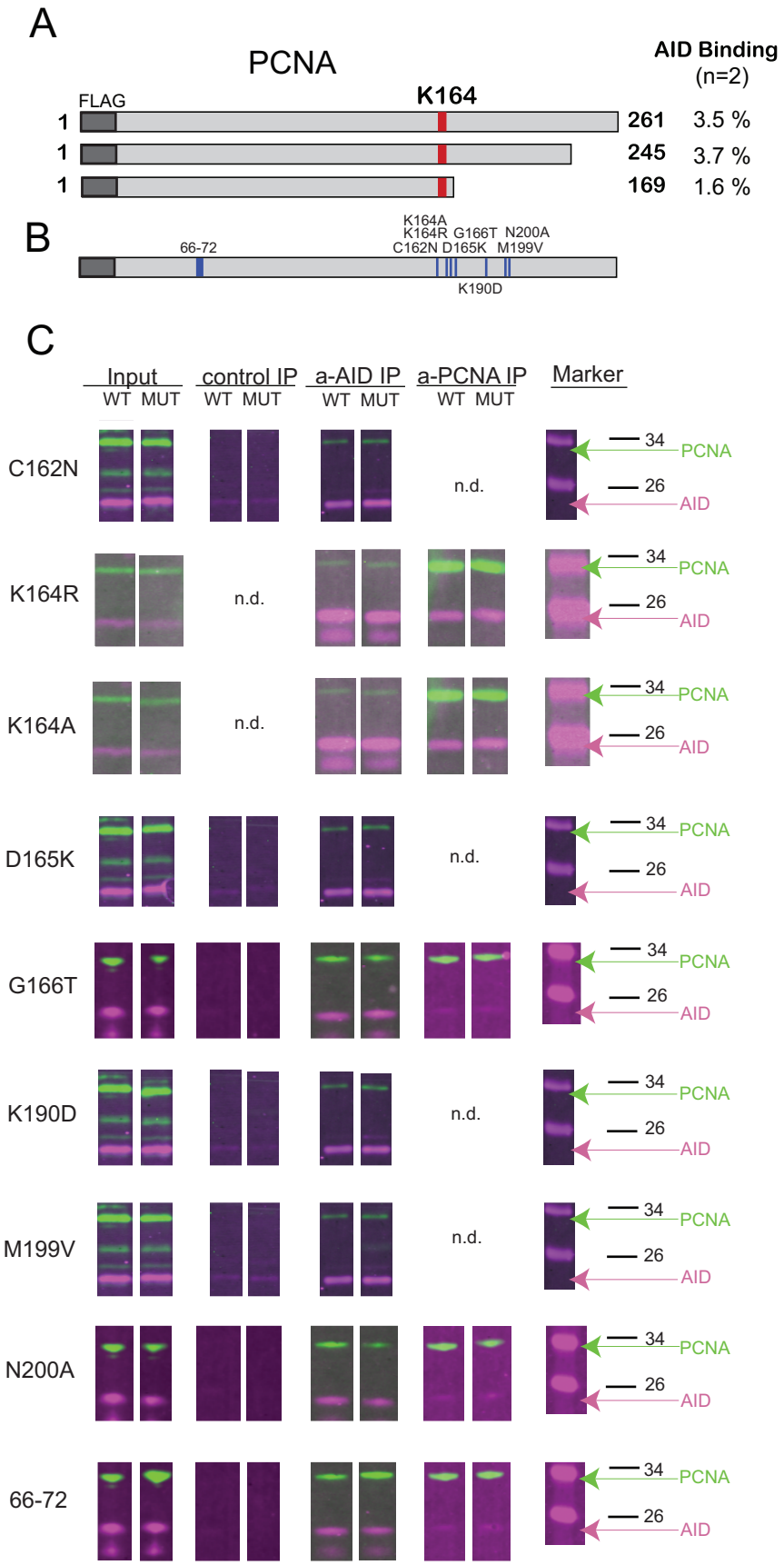

A
PCNA
AID Binding
(n=2)
FLAG
K164
1
261
3.5 %
1
245
3.7 %
1
169
1.6 %
B
66-72
K164A
K164R G166T N200A
C162N D165K M199V
K190D
C
Input
control IP
a-AID IP
a-PCNA IP
Marker
WT MUT
C162N
K164R
K164A
D165K
G166T
K190D
M199V
N200A
66-72
n.d.
34
26
PCNA
AID

Figure 20. AID interacts with the C-terminal part of PCNA.

A, Summary of luciferase interaction assay data showing the region of PCNA required for interaction with AID. FLAG-PCNA full length and truncations and RL-AID were transfected into 293 cells. Binding was analysed as in Figure 18. At least 2 experiments were used. **B**, Localisation of the point mutants in PCNA tested for interaction using the *E. coli* assay. Mutations are indicated by amino acid abbreviations. Mutant 66-72 is a 5 amino acid mutant converting the indicated stretch of the protein into the *Xenopus laevis* homologue sequence. **C**, Test for loss of interaction between AID and PCNA point mutants. An assay as described in Figure 17 was used. The respective wild type control is shown for each mutant IP.

compared to the wild type control (Figure 20 C), even when higher stringency (high salt) washes were employed (data not shown). These data identify no single amino acid to confer PCNA binding to AID, though most likely the interaction domain lies close to the site of modification on PCNA (K164) towards the C-terminus.

4.1.1.5 Interacting domains in AID

A similar approach was taken to identify PCNA binding domains in AID. The AID crystal structure has not been solved, possibly a consequence of the protein's tendency for insolubility and domain instability (SPM lab members, personal communication). However, several AID domains have been functionally characterised, and we noticed a motif reminiscent of a PIP (PCNA interacting peptide) box near aa 135 (Figure 21 A). Therefore, a truncation series of AID was generated and tested for PCNA interaction using the RL-AID 293T overexpression system described above.

The summary of these results can be viewed in Figure 21 B; each truncation was tested at least 3 times. Using truncations from the N terminus and C terminus of RL-AID, no single stretch of amino acids was identified that conferred PCNA binding, as more than one region of AID could mediate residual association. Deletions at the N terminus of AID indicated that the PCNA AID interaction could be placed between aa 120-140, whereas deletions from the C terminus identified a second binding region throughout the N terminus up to aa 80, with a minor loss of binding when deleting aa 80 - 130, supporting the C-terminal deletions.

To investigate binding domains further, a different approach was used in order to address the mentioned structural caveats of truncation studies. Previously described chimeric GFP tagged AID derived proteins were utilised. These chimeras were

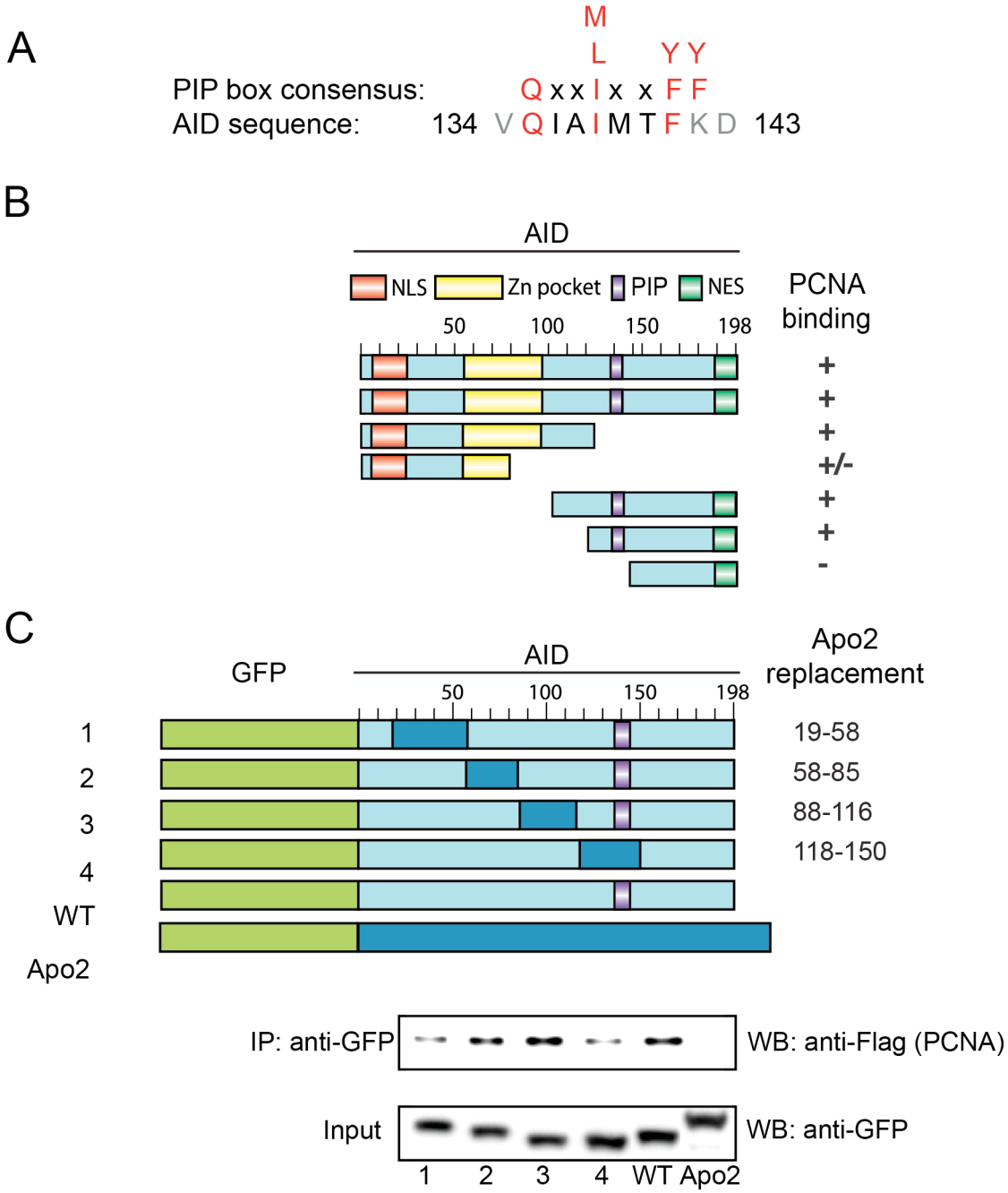

A
M
L Y Y
PIP box consensus: Q x x I x x F F
AID sequence: 134 V Q I A I M T F K D 143
B
AID
NLS
Zn pocket
PIP
NES
PCNA binding
50
100
150
198
+
+
+
+/-
+
+
-
C
GFP
AID
Apo2 replacement
50
100
150
198
1
2
3
4
WT
Apo2
19-58
58-85
88-116
118-150
IP: anti-GFP
WB: anti-Flag (PCNA)
Input
WB: anti-GFP
1
2
3
4
WT
Apo2

Figure 21. AID interacts with PCNA via multiple domains.

A, The putative PIP box in the AID protein. The consensus sequence of a PIP box after (Xu et al., 2001) is shown highlighted in red. The reminiscing motif in human AID is depicted for comparison. **B**, Luciferase Interaction Assay data showing AID truncations binding to PCNA. The data represents the summary from at least 3 independent transfections. RL-AID fusion truncations were transfected as in Figure 18, and endogenous PCNA was immunoprecipitated with anti-PCNA antibody (PC10). Luciferase activity was determined from the input and the bound fraction, and the ratio was used to judge binding efficiency. **C**, PCNA binding of AID-APOBEC2 chimeras. 293T cells were transfected with GFP-chimeras 1-4, GFP-AID (wt) and GFP-APOBEC2, together with FLAG-PCNA. Chimera protein expression levels were adjusted to equal levels prior to IP with a rabbit polyclonal anti-GFP antibody. IP's were analysed by western blot, detecting PCNA with an anti-FLAG antibody and GFP fusion proteins with a mouse monoclonal anti-GFP antibody.

(In collaboration with Maria Simon and Heather Coker)

comprised of APOBEC2 (a DNA deaminase family protein without DNA deaminase activity) portions and AID portions (Conticello et al., 2008). The chimeras have been used to characterise other AID interacting proteins (Conticello et al., 2008, Patenaude et al., 2009). This system allowed for individual parts of AID to be replaced by the corresponding APOBEC2 peptides, potentially without structural loss, and tested for interaction (Figure 21 C). The control (full length APOBEC2-GFP) is not able to bind PCNA (Figure 21 C). The data show that replacing the N terminus proximal amino acids 19-58 (Chimera 1) as well as amino acids 118-150 (Chimera 4) of AID with APOBEC2 peptides reduced AID-PCNA interaction, highlighting similar regions for interaction as the truncation experiments (Figure 21 C)

In conclusion, these data establish that AID interacts with PCNA through multiple domains, possibly through two linear epitopes or through a specific three-dimensional conformation. The cryptic PIP box in AID may contribute to PCNA binding, but is certainly not the only region involved in the interaction. A range of PCNA binding proteins have been identified that also rely on contacting PCNA additionally via peptides outside the PIP box (Bruning and Shamoo, 2004).

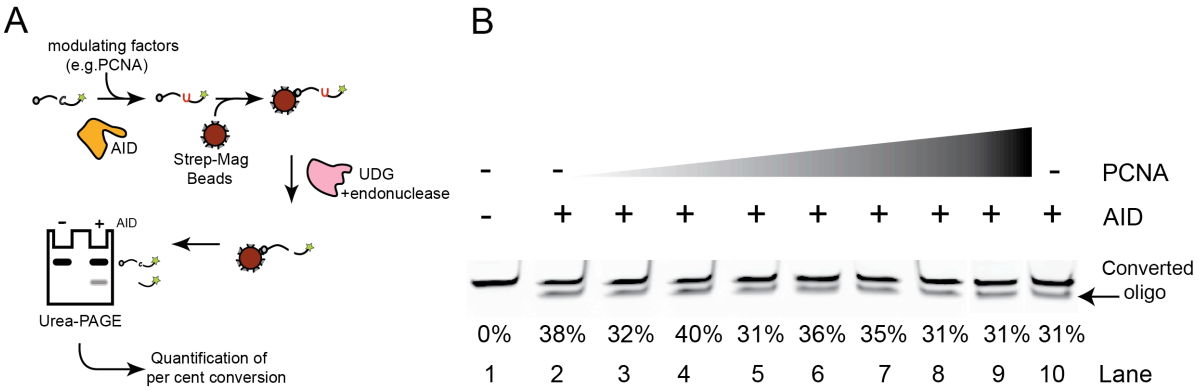


Figure 22. Influence of the presence of soluble PCNA on AID activity

A, Schematic of the *in vitro* AID oligonucleotide deamination assay (ODA). A biotinylated and fluorescein-conjugated single-stranded oligonucleotide containing an AID hotspot motif is incubated with AID and purified using the biotin label. This is followed by processing with UDG/endonucleases, with the resulting products visualised and quantified on Urea-PAGE. **B**, Presence of increasing amounts of PCNA does not alter AID oligonucleotide conversion. 2.5 pmol DNA were incubated for 15 min with ~1pmol of active AID yielding between 38% and 31% conversion (lanes 2 and 10). An increasing amount of PCNA was added (3, 9, 18, 36, 72, 144 pmols in lanes 3, 4, 5, 6, 7, 8, 9, respectively). Representative experiment.

4.1.1 Does PCNA influence AID's catalytic activity?

Several PCNA interacting enzymes have been shown to be stimulated by PCNA, e.g. DNA polymerases (Bambara and Jessee, 1991, Prelich et al., 1987, Tan et al., 1986), APE2 (Unk et al., 2002), FEN1 (Li et al., 1995, Wu et al., 1996) and DNMT1 (Iida et al., 2002). To test weather the presence of PCNA could influence the enzymatic activity of AID, we utilised our *in vitro* AID activity assay (Coker et al., 2006, Petersen-Mahrt and Neuberger, 2003), also known as oligonucleotide deamination assay (ODA). As outlined in Figure 22 A, 1 pmol of active AID is incubated with 2.5 pmol of a biotinylated and fluorescein-conjugated single-stranded oligonucleotide, which contains a cytosine in the context of the preferred AID sequence (AGC) (Beale et al., 2004, Morgan et al., 2004). After deamination, the uracil containing oligonucleotide is purified using biotin-streptavidin binding to magnetic beads. This is followed by incubation of the immobilised oligonucleotide with commercial UDG enzyme, which also contains low levels of endonucleases that cleave the abasic sites generated by UDG. The resulting products are separated by denaturing Urea-PAGE and visualised using a fluorescence scanner. The ratio of cleaved to non-cleaved product is quantified

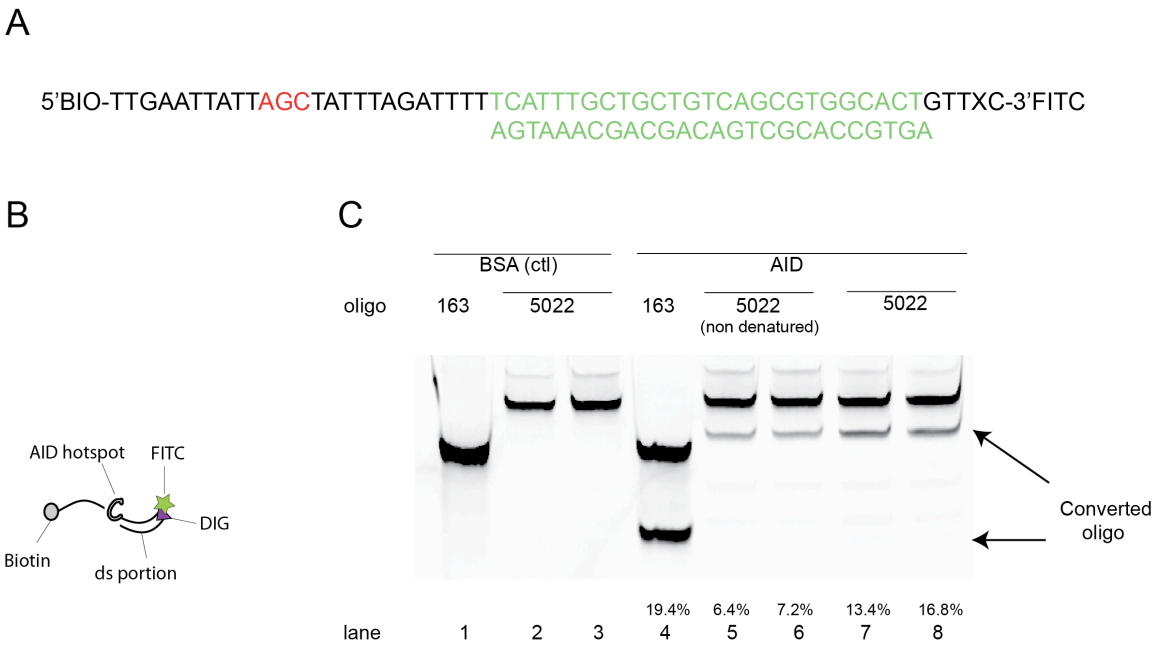


Figure 23. Description of oligonucleotide 5022.

A, Sequence of the oligonucleotide 5022 used for the loaded PCNA assay and its features. "BIO" represents a biotin label, "FITC" represents a fluorescein label, "X" represents a digoxygenin labelled nucleotide, highlighted in red is the AID hotspot motif and highlighted in green is the double stranded portion. **B**, Schematic of the oligonucleotide 5022. **C**, Testing deamination of the oligonucleotide 5022 in the standard ODA. The single stranded form of 5022 is used together with the standard oligonucleotide (163) as a control in an assay as described in Figure 22 A. 5022 is heat denatured before the AID reaction as in the standard assay in lanes 7 and 8, while this step has been omitted in lanes 5 and 6.

and the per cent conversion calculated. Figure 22 B shows the deamination reaction of AID without PCNA, and with increasing concentrations of PCNA present at the time of deamination. The presence of soluble PCNA does not have a positive, or negative, effect on the ability of AID to deaminate cytosines in this ODA (Figure 22 B). This result was confirmed with a different batch of purified AID (data not shown), as well as with lower AID levels to ensure that a stimulatory effect was not concealed by saturation kinetics. It is assumed that most of the PCNA present in the experiment shown in Figure 22 is in solution. Although it has been suggested that PCNA can load onto linear DNA substrate by diffusion, this process is inefficient and requires high concentrations of PCNA (Burgers and Yoder, 1993). The clamp loader RFC actively loads PCNA onto DNA in an ATP dependent manner so that the toroid will encircle DNA (Tsurimoto and Stillman, 1989, Tsurimoto and Stillman, 1991). Because some

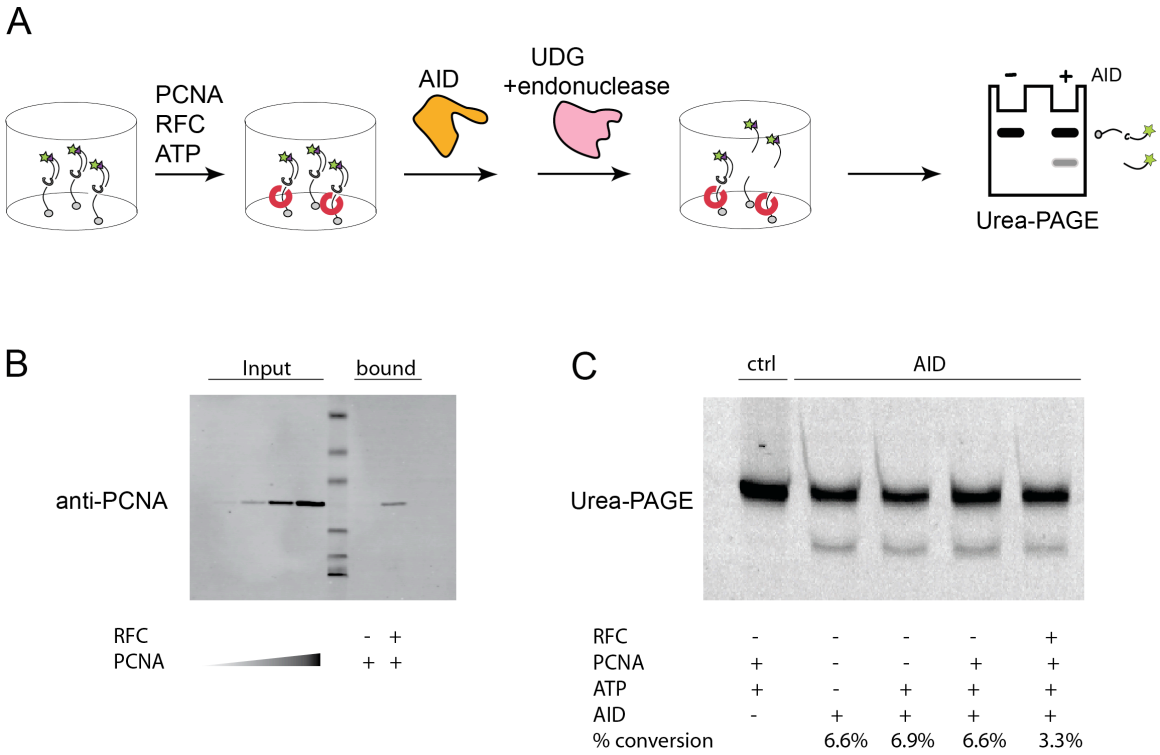


Figure 24. Influence of the presence of loaded PCNA on AID activity.

A, Outline of the modified *in vitro* AID oligonucleotide deamination assay. A biotinylated and fluorescein-conjugated oligonucleotide containing an AID hotspot motif, dsDNA portion and digoxygenin is bound to a biotinylated plate and subject to incubation with PCNA, RFC, ATP, AID and UNG. As before, products can be visualised on Urea-PAGE. **B**, RFC dependent immobilisation of PCNA on assay plates. With this method, up to 150 fmol PCNA can be bound to each assay well. Anti-PCNA western blot analysing input and assay wells with or without RFC. **C**, Presence of loaded PCNA does not alter AID oligonucleotide conversion. Each well was incubated with ~4 pmol of active AID.

enzymes are stimulated specifically when PCNA encircles DNA (Ducoux et al., 2001, Gomes and Burgers, 2000, Li et al., 1995, Tom et al., 2000, Tom et al., 2001), I wanted to tested if DNA bound PCNA would alter AID activity. To this end, the standard ODA was modified to include a digoxygenin label in the oligonucleotide in order to bind a bulky anti-digoxygenin antibody to its terminus. This was to ensure that during the reaction, loaded PCNA would not slide off the DNA. As RFC loads PCNA preferably to boundaries between single stranded and double stranded PCNA (Tsurimoto and Stillman, 1990), the oligonucleotide was long enough to accommodate a double

stranded portion. The sequence of the used oligonucleotide (5022) can be viewed in Figure 23 A, and a schematised depiction in Figure 23 B. The oligonucleotide can be deaminated by purified AID using the standard ODA, albeit at a lower efficiency (Figure 23 C, lanes 4 versus 7 and 8). This efficiency is further reduced when the normal denaturation step preceding the AID reaction is omitted (lanes 5&6 versus 7&8). This was tested because the hybridised form of the oligonucleotide will not allow this step in the modified assay. Streptavidin coated plates were used to immobilise PCNA on biotin containing oligonucleotides (Figure 24 A). First, streptavidin plates were coated with oligonucleotide 5022. Next, PCNA was immobilised to DNA using RFC (Figure 24 B) by applying three consecutive rounds of a loading mix [recombinant human PCNA, ATP and recombinant yeast RFC complex (J. Parker, CRUK)] to the DNA coated plates. This resulted in a typical yield of 10-150 fmol immobilised PCNA per well, showing that the basic steps of the assay are working as well as that yeast RFC is able to load human PCNA. The DNA-protein complexes were incubated with AID, followed by a UDG reaction. As the results in Figure 24 C show, loaded PCNA (dependent on RFC), does not have a stimulatory effect on AID activity. A weak inhibitory effect is possible, but could not be shown to a sufficient resolution. Such an inhibitory effect can be readily explained with PCNA occupying AID hotspot motifs.

Limitations of this assay are the level of efficiency of oligonucleotide conversion by the end of the procedure that does not resolve small changes well, as well as the small amount of loaded PCNA in the reaction (up to 150 fmol PCNA on up to 2.5 pmol DNA = 5%). Nevertheless, these preliminary data support that AID activity is not stimulated by PCNA in the configuration that was tested.

4.1.2 AID can interfere with PCNA modification

As there was no indication that PCNA could influence AID *in vitro*, we wanted to test if the opposite –AID altering PCNA function- was the case, as it is conceivable that AID is needed for regulation or recruitment of DNA repair at AID lesions.

PCNA's functions in DNA repair are regulated by post-translational modifications. AID's interaction with PCNA lies near K164, the site for ubiquitination and sumoylation. Because of this, it was established, initially *in vitro*, if AID had a functional role in influencing these post-translational modifications on PCNA. An *in*

vitro mono-ubiquitination assay was used, based on commercially available purified components of the ubiquitination machinery. This allowed the monitoring of PCNA ubiquitination via western blotting as a slower migrating band. First, a number of provided E2 ubiquitin-conjugating enzymes were tested for their ability to ubiquitinate purified human PCNA *in vitro*, with UBC5B and UBC5C acting on PCNA (Figure 25 A). Subsequently, AID protein was added at near stochiometric amounts to PCNA in the reaction, which contained UBC5B as well as E1 activating enzyme, ATP and free ubiquitin. Via western blotting, PCNA modification was monitored by observing a higher molecular weight PCNA band (Figure 25 B). Using APOBEC2 as a control protein, the data show that recombinant AID represses *in vitro* ubiquitination of PCNA, possibly by directly binding the substrate PCNA.

To monitor if AID could also inhibit modification of K164 PCNA in a cell-based assay, PCNA and AID were co-transfected into an *E. coli* sumoylation system (Lyst et al., 2006, Mencia and de Lorenzo, 2004). Here, E1 and E2 enzymes as well as his-tagged SUMO I were transfected into a host *E. coli* together with the PCNA and AID expression vectors. Again, western blot analysis showed that PCNA was readily sumoylated in the presence of E2 conjugating enzyme in this *E. coli* system (Figure 25 C). Monitoring the sumoylation of PCNA over time in the presence or absence of AID demonstrated that even in the context of a cell, AID is able to inhibit PCNA K164 modification (Figure 25 C).

In conclusion, inhibition of modification of PCNA by AID is likely to be relevant to the biological function of the PCNA-AID interaction.

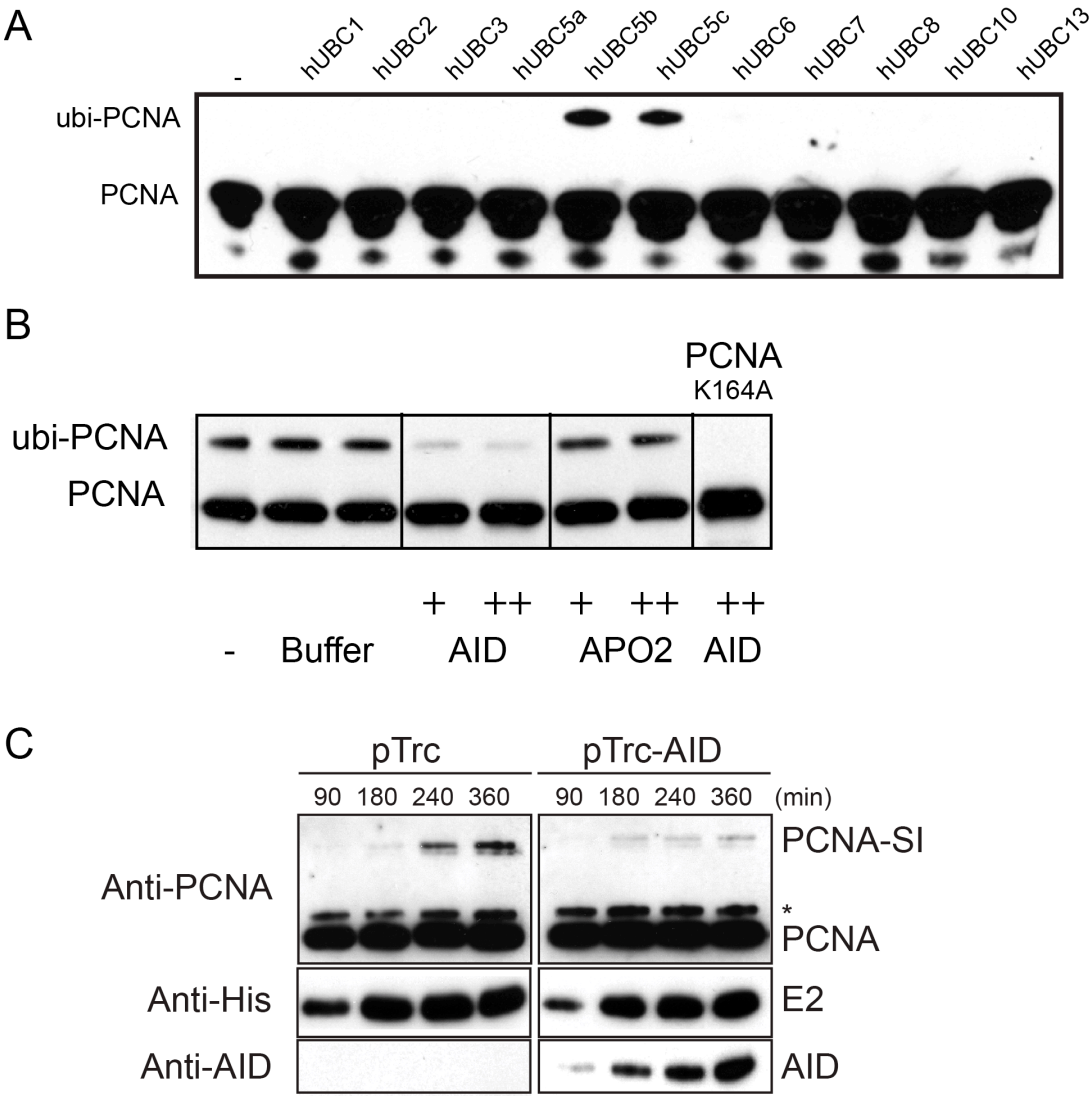


Figure 25. AID inhibits modifications of PCNA.

A, *In vitro* monoubiquitination of PCNA by various E2 conjugating enzymes. Purified human PCNA is *in vitro* ubiquitinated in the presence or absence of 11 different E2 conjugating enzymes. Reactions were incubated for 90 min at 37 ˚C and higher molecular (modified) PCNA was detected via western blotting. **B**, AID reduces the *in vitro* ubiquitination of PCNA by Ubc5b. Purified PCNA (or PCNA K164A) was subject to ubiquitination in the presence of purified APO2 or AID protein. **C**, AID inhibits PCNA sumoylation in an *E. Coli* expression system. Sumoylation of PCNA was induced by co-expressing PCNA, SUMO I and the E1/E2 enzymes in bacteria. The effect of AID expression on PCNA sumoylation was assessed by co-transfecting the expression vector for AID (pTrc-AID) or the empty vector (pTrc) as a control. Cells were induced for the time period indicated. Subsequently the culture was centrifuged, resuspended and gel separated. Blots were probed with an anti-AID antibody, an anti-His-tag-HRP antibody for E2 detection or with an anti-PCNA antibody.

(In collaboration with Maria Simon and Don-Marc Franchini)

Chapter 5. Discussion

AID and its biological functions has continuously surprised researchers from the moment of its discovery. Starting from the early assumption that this small enzyme was a simple RNA editing molecule, it was discovered that it actively mutated DNA, that it is highly regulated, and that its function is interlinked with many DNA associated pathways, and even plays an important role outside the immune system in epigenetic reprogramming. Yet one of the most intriguing question remains unanswered: How is AID targeted to induce DNA alterations at specific loci, and restricted from others? In particular, which role do transcriptional targeting and differential lesion resolution play in these varying outcomes? Although progress has been made (Odegard and Schatz, 2006), additional work is required for the years to come.

Therefore, to understand AID biology one needs to understand the co-factors of AID that will regulate and modulate its targeting, its activity, and its lesion resolution.

For these reasons, I set out to identify and characterise interacting partners of AID.

5.1 Aim I

In Aim I I pursued a currently emerging connection between AID and the transcription machinery, by identification and analysis of chromatin associated AID cofactors.

As highlighted in the introduction, transcription has long been associated with AID-induced immune diversification. Recent findings support the importance of transcription elongation and associated processes for AID targeting. Furthermore, chromatin marks seem to be factors that influence AID targeting, activity and lesion resolution, either directly or indirectly.

5.1.1 A novel biochemical study of AID on chromatin

This study was undertaken in order to analyse chromatin bound AID in an unbiased way. Interaction studies in this particular fraction can be seen as a survey of proteins that AID associates with on DNA, the place of most interest for studying biological AID function.

This allowed for the first time identification and characterisation of proteins that are associated with AID on chromatin in their physiological environment. The method had

two key properties: One, the endogenous AID protein was tagged with a FLAG epitope, and secondly, we adapted a recently developed method for specific isolation of chromatin bound protein-complexes (Aygun et al., 2008).

The chromatin AID-interactome consisted of RNAPII core and associated proteins, splicing factors, RNA helicases, chromatin modifiers and RNAPII elongation complexes. Furthermore, the factors/complexes FACT, SUPT5H, PAF and SUPT6H were identified, which have been implicated in AID function by us and others.

The exceptionally high occurrence of factors in our analysis that form part of the same interaction network (RNAPII elongation), even from the same complex (PAF complex), indicates that biologically meaningful associations have been retrieved.

This tight concordance of having validated work of others as well as identified a major complex, led me to confirm and validated the initial results via biochemical and functional assays. I focused on the PAF complex, a novel AID interacting factor that has several functional and physical connections to RNAPII and its associated factors (as described in the introduction). The PAF complex plays a central role in transcriptional control, and serves as a molecular link between early transcription elongation, as marked by pausing factor DSIF (SUPT5H and SUPT4H), and downstream extended chromatin elongation modifications dependent on FACT (SSRP1 and SUPT16H) and SUPT6H. Identification of SUPT6H as an AID associated protein in our study supports previous findings (Okazaki et al., 2011), as well as the recent discovery that the exosome processing machinery is important for CSR (Basu et al., 2011), because SUPT6H has been shown to co-purify with the exosome (Andrulis et al., 2002).

5.1.1.1 Function-specific AID associations

The AID complex interactions were isolated from DT40 cells, which undergo iGC but do not undergo CSR. Although previous work excluded SUPT6H from a function in SHM (Okazaki et al., 2011), the biochemical findings in my study place SUPT6H with V region diversification complexes. This difference may be due to AID overexpression in the system used in (Okazaki et al., 2011). However, the biochemical association in DT40 and the requirement during CSR would argue for a general importance of the complex regardless of cellular function. To further test this hypothesis and extend it to other elongation factors (especially the PAF complex), respective knock-down or

knock-out populations of DT40 and/or other hypermutating cell lines should be assayed for AID dependent diversification.

Another approach to examine CSR versus SHM specific cofactors of AID is to take advantage of the C-terminus of AID and its importance for CSR (Barreto et al., 2003, Ta et al., 2003). Thus, a C-terminal deletion mutant of AID could be used to identify CSR specific interaction partners of AID, as well as identify factors for V region diversification, or those that are shared. For these reasons, I generated DT40 cell lines that lack the last 10 amino acids of AID (AIDΔC), the region responsible for CSR specificity and nuclear export, yet contain the same peptide tags as the described wt tagged AID lines (see Appendix 6.1.1). Surprisingly, the AIDΔC mutants shows no accumulation of nuclear AID as predicted from similar mutants in fibroblasts (Ito et al., 2004, McBride et al., 2004), and additionally shows overall reduced AID levels (see Appendix Figure 28). Although there is some uncertainty about the unexpected phenotype of the truncation mutant in DT40, a novel tool was generated that is bound to help to dissect if different subpopulations of AID form different functional complexes.

5.1.1.2 Biochemical characterisation of AID chromatin associations

Protein-protein interactions can vary in strength ranging from loose associations to tight partners. Although we have been able to identify a specific, direct interaction of AID with PAF1 *in vitro*, the kinetics of the interaction still need to be determined *in vivo*. Our mass spectrometry analysis, as well as indications from other laboratories, have identified a large number of possible AID interacting proteins (see Introduction). So far, AID seems to prefer weak interactions, and many of the reported interactions could not be found in our mass spectrometry analysis, including PCNA. Each of these factors may require individual purification conditions. Others have found weak interactions for AID, such as in the case of SUPT5H (Pavri et al., 2010). In my hands, this interaction could barely be detected in the *in vitro E. coli* system, although I may have been using a more stringent isolation method. Considering the small size of AID and its multiple regulatory steps, it is likely that AID will have a number of associated proteins/complexes within the cell, possibly some of them serving as adapters for AID subpathways, rather than specific tight partner interactions. The 200 kDa complex that AID was found to reside in represents the main form of chromatin AID in B cells, but it

remains to be determined if this is the active form of chromatin AID, and if this complex is one homogenous species. Several subunits of the PAF complex (PAF1, CTR9 and CDC73), partially overlap with the 200 kDa AID complex, thus AID associating with one or more subunits of the complex would be consistent with the data. Most likely PAF1 participates in this complex. However, it is likely that the complex has further components that associate directly or indirectly with AID, e.g. from amongst our mass spectrometry identified associations. Because AID would engage in several processes in the nucleus (sequestering, transport, targeting and recruitment) it would have to be able to switch partners easily. By interacting with several partners of a specific cellular process (e.g. of the RNAPII machinery), tethering to this process may be reinforced and seems to predominate on chromatin. Associations could be mediated through AID's hydrophobic sequence and/or its interaction with RNA (Bransteitter et al., 2003, Nonaka et al., 2009).

The difficulties in purifying AID in a stringent manner as a homogenous population with few specific, tight interaction partners may lead to identification of artificial interactions if not validation properly. Therefore, my work ensured a specific isolation method combined with verification under physiological conditions.

5.1.2 The PAF complex and its interaction with AID

The direct interaction with PAF1 and specific effect on CSR by *leo1* knockdown, could be due to direct as well as indirect effects. The somewhat surprising finding that even moderate knock-down of the transcripts of the PAF complex members had a substantial effect on CSR could be partially explained by interference with transcription of *Ig* or *aid*. However, another explanation would be a destabilisation of the PAF complex when one subunit is depleted, as it has been observed in HeLa cells when knocking down *ctr9* or *ski8* by RNAi. Not only levels of the target protein decreased, but also protein levels of other subunits of the PAF complex (Zhu et al., 2005). Another conceivable scenario explaining the loss of transcription at *Ig*, for instance upon *paf1* knock-down (3.1.4, Figure 16), comprises a direct effect on transcription by AID. One could envision not only facilitation of AID deamination by PAF, but also recruitment of efficient elongation to *Ig* via the AID-PAF1 interaction, thus rendering AID a "transcription factor" (from discussions with K. Schmitz and S. Petersen-Mahrt, IFOM Milan).

For further experiments using the shRNA system, it would be desirable to improve the knock-down efficiency. One way to achieve this would be to select retrovirally infected CH12 populations not by puromycin resistance but flow cytometry. Hereby, shRNAs should be cloned into a retroviral vector such as pLMP, which carries a co-expressing GFP marker. Disadvantages of puromycine selection are the requirement of a selection over several days potentially resulting in stressed cells or incomplete selection. This can be overcome by selecting highly fluorescent clones early after transduction, likely to produce more shRNA molecules. Once these problems are solved, a number of questions can be addressed: To examine the specific role of the PAF complex in B cells, levels of potentially affected histone modifications in PAF complex knock-down populations could be monitored, specifically, it would be interesting to confirm the role of the PAF complex in H3K4 trimethylation, [although in mammals there are redundant pathways that can set this mark on chromatin (Shilatifard, 2008)]. The role of the PAF complex in H3K36 trimethylation could also be addressed, as in mammals, it is still unknown if there is a dependency of this mark on the PAF complex and H2B monoubiquitination (Newey et al., 2009). Apart from monitoring global changes in histone modification patterns in knock-down cells, changes of local histone modification and transcription machinery composition could be addressed in more detail on single genes, whereby the *Ig* loci evidently will be of special interest. Firstly, it would be important to map a distribution of the PAF complex in relation to AID, SUPT5H and RNAPII by ChIP along the *Ig* loci in normal cells and cells inducing class switching. On the other hand, PAF complex knock-down populations could be valuable to shed light on further details such as the occupancy of histone modifications implicated in AID dependent processes. Apart from this, occupancy of RNAPII, associated elongation factors such as FACT, SUPT6H, SUPT5H, and AID can be addressed in dependence of the presence of the PAF complex. As a caveat, successful AID ChIP experiments in the past have been scarce, possibly due to lack of ChIP grade antibody for AID. Difficulties are illustrated by a recent genome wide study (Yamane et al., 2011), where background had to be controlled for by including AID knock-out ChIP data in the analysis.

Biochemically, PAF knock-down may alter the size of the AID complex as well as the PAF complex, which could be analysed by size exclusion chromatography of chromatin

fraction in knock-down cells. However, this experiment may be restricted by the number of cells needed. Changes to the complexes analysed in wt DT40 cells (3.1.2 Figure 10) may be expected upon knock-down, if a substantial amount of the observed AID complex contains PAF1 and/or the PAF complex. As for the PAF complex itself, a destabilisation when knocking down one subunit is likely (Zhu et al., 2005).

In summary, this work demonstrates the involvement of the PAF complex in AID dependent processes, most likely through a direct interaction between AID and PAF1. The data described in this study extends the current model of AID gaining access to DNA by stalled RNA polymerase to a more complex model, where AID is intimately and specifically linked with RNAPII while pausing, restarting and elongating, requiring the correct chromatin environment defined by histone modification cascades.

The PAF complex has many roles in transcription, including elongation activation, chromatin regulation and mRNA biogenesis. In yeast, the PAF complex is nonessential and only a subset of genes may be affected by PAF complex absence, whereas in vertebrates, the PAF complex is essential (Wang et al., 2008). Mouse mutants of the oncogene an tumour suppressor gene *cdc73* are embryonic lethal as PAF controls many developmental processes in mammals (reviewed in Newey et al., 2009). Therefore, a study of a specific effect of PAF complex knock-out on B cells *in vivo* would require a conditional knock-out mouse model.

5.1.3 RNAPII pausing and elongation to enhance AID activity

A model for AID recruitment to active transcription requires a landing platform, e.g. SUPT5H associated with a paused or stalled RNAPII complex, on which a mutasome containing AID and the PAF complex could land. While on the chromatin, the FACT complex, SUPT6H and exosome complex associate with AID, modify the local chromatin structure, and allow for AID to gain access to ssDNA during RNAPII transcription and mRNA biogenesis (Figure 26).

Mutations at the *Ig* locus start 50-150 bp after the TSS, and end after 1.5 - 2.0 kb. As laid out below, my work provides further insight into the establishment of the AID dependent mutation domain (5' boundary), while the understanding of molecular mechanism for the 3' boundary remains less clear.

5.1.3.1 AID recruitment by paused RNAPII and the roles of SUPT5H and PAF

In general, SUPT5H associates with pausing polymerases, while the presence of the PAF complex is required for a processively elongating polymerase (usually after restart of promoter proximal pausing) (Chen et al., 2009). Although a direct binding of AID to PAF1 was shown in this study, it is not clear if this interaction is for targeting AID to the paused polymerase or for retention of AID during elongation. There are several possibilities of how AID associates with RNAPII: First, loading could occur during formation of the initiation complex. This is substantiated, as in higher vertebrates, PAF complex has been shown to bind to RNAPII along the entirety of the gene (Kim et al., 2010, Zhu et al., 2005). AID may only gain access to ssDNA by promoter-proximal pausing and elongation, requiring a form of pausing throughout the transcription elongation process. Secondly, AID could be loaded at pausing sites and immediately be active on its substrate, while staying associated with polymerase throughout the targeted region. Finally, AID could associate and dissociate with pausing polymerase throughout the *Ig* gene. Fixation of mutations could be regulated on the level of subsequent steps that involve repair proteins (see below). Recent genome wide ChIP data for AID in B cells (Yamane et al., 2011) indicates that AID co-localises with polymerase along the *Ig* gene, supporting AID being present throughout transcription from an early stage on. Interestingly, recent work identified the PAF complex to have a synergistic function with SUPT5H in negatively regulating zebrafish blood developmental genes (Bai et al., 2010). Here, the PAF complex acts as a negative factor for RNAPII elongation, i.e. PAF can be a stalling factor. Furthermore, the *Ig* genes are rich in associated polymerase uniformly, unlike a classical promoter proximal stalled gene (Rahl et al., 2010, Zeitlinger et al., 2007), which has a pronounced peak at the TSS accompanied by less detectable RNAPII within the transcription unit (Pavri et al., 2010, Yamane et al., 2011). Although from these profiles one cannot conclude at which rate RNAPII moves on *Ig*, efficient elongation does not seem to take place, as in this case RNAPII detection levels would be lower. Instead, impaired elongation and polymerase accumulation would fit the current data (Rajagopal et al., 2009, Raschke et al., 1999, Wang et al., 2009). Taken together, these data indicate, that AID enters the RNAPII complex early via the PAF1 complex and exerts its function with the help of constant transcriptional

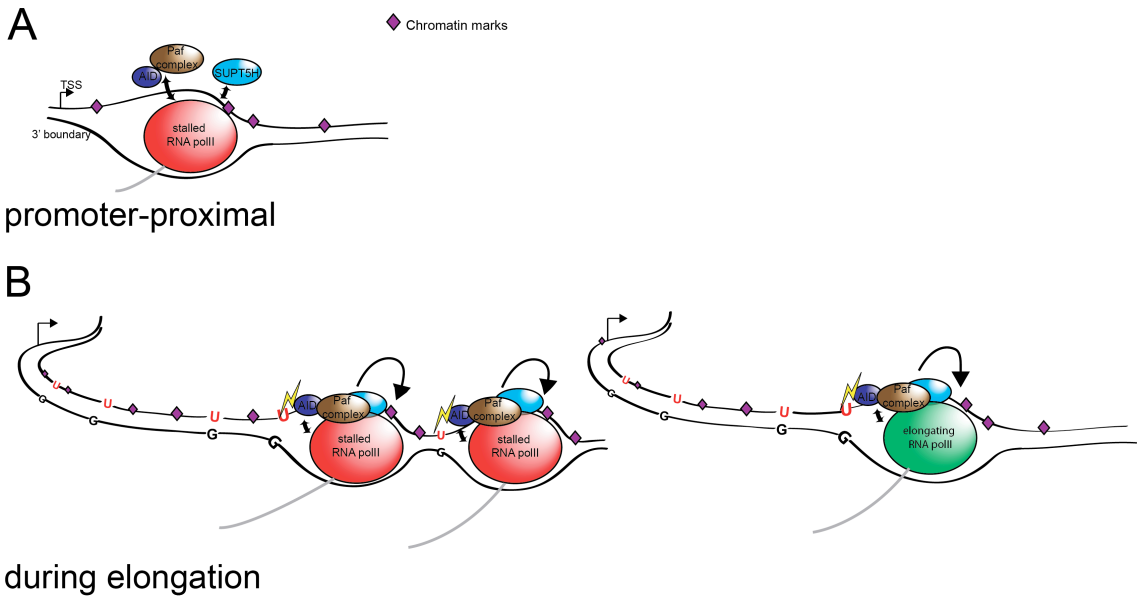


Figure 26. Model of elongation factors associated with AID during *Ig* gene transcription.

A, Promoter-proximal pausing mediated by SUPT5H is attributed to AID targeted genes. AID and PAF complex may associate with polymerase at this stage. **B**, Two alternatives of the RNAPII configuration that is potentially required for AID recruitment or activation. Scenario 1) stalled elongation; 2) processive elongation.

pausing mediated, amongst others, by SUPT5H (Figure 26 B). Further *Ig* gene ChIP and biochemical studies will elucidate the significance of these relationships. Importantly, it is unlikely that the exact molecular mechanisms for AID targeting and efficiency will be delineated from studies on genome wide associations. Dissection of a unique protein and chromatin composition at the *Ig* locus will be more informative, and may lead to understand how mistargeted genes recapitulate mutation.

Biochemically, *in vitro* transcription reconstitution experiments mediating AID deamination in the context of a eukaryotic RNAPII transcription bubble could enhance our understanding of the molecular mechanisms at AID targets. As can be seen in the Appendix (6.1.3) I have made progress towards the setup of such an *in vitro* system that merits further investigation (Appendix Figure 30). Here, RNA elongation competent transcription bubbles were reconstituted *in vitro* using purified DNA, RNA and eukaryotic RNAPII components. These complexes can be exposed to AID while transcription or transcriptional arrest is taking place to draw biochemical conclusions

about co-transcriptional accessibility of cytosines for AID. Subsequently, purified transcription elongation factors could be added to the reactions to delineate the optimal transcription elongation complex composition for AID.

5.1.3.2 Chromatin environment

During transcription coupled AID processes at AID targeted loci, the chromatin environment as shaped by the PAF complex, as well as downstream influence of AID on the resolution pathways will be of importance. An Ig locus specific RNAPII elongation complex (including SUPT5H, AID and the PAF complex) could establish a defined set of marks in order to promote mutability at this locus, e.g. by enhancing repeated pausing and re-start, thereby providing optimal substrate accessibility for AID. As described, during transcription the PAF complex helps set the right co-transcriptional chromatin marks, serving as docking platform for the H2B ubiquitination machinery, as well as for setting the mark H3K4 trimethylation (Jaehning, 2010). As described, H3K4 trimethylation serves as an important mark in CSR (Stanlie et al., 2010, Wang et al., 2009), but is generally restricted to the 5' end of a gene, and replaced by H3K36 trimethylation towards the 3' end of the gene. In a recent ChIP study (Yamane et al., 2011), AID was genome-wide associated with regions marked by H3K4 trimethylation (as well as other marks for open chromatin), but much less so with H3K36 trimethylation marks. Importantly, both these marks are induced upon transcriptional activation of S-regions at the *Ig* locus (Wang et al., 2009), but the H3K4 trimethylation domain is extended, while onset of the H3K36 trimethylation mark is pushed back towards the 3' end. This correlates roughly with the cease of mutational load/AID activity (Wang et al., 2009). If H3K36 is either a prerequisite or an indicator of efficient elongation in the body of the gene, it is not surprising that it would be excluded from the mutated part of *Ig* where pausing is predominant. The PAF complex could help establish this particular distribution, as its role in H3K4 trimethylation in higher eukaryotes is well established and at least in yeast, a connection to H3K36 trimethylation has been documented. Past work has implicated histone modification during SHM, but much less is known about epigenetic requirements for SHM than for CSR. However, the isolation of most of the required components for setting the H3K4 trimethylation mark during transcription from DT40 would imply a similar requirement during V region diversification such as SHM.

5.1.3.3 The influence of RNAPII pausing on downstream resolution pathways

Furthermore, RNAPII pausing may have a dual function, first for AID loading, and second, it may enhance AID's resident time at the target locus. The extent of RNAPII pausing may allow for AID to recruit DNA repair factors outside their normal physiological pathway to enhance the mutability at the target locus. Recent observations in our lab show that to enhance its effect on DNA alterations, AID needs to recruit DNA repair factors to its own lesion (Franchini et al., submitted see discussion for Aim II). This could be a means by which the AID-induced uracils become mutagenic, considering an excess load of thousands of spontaneous dUs that the genome suffers daily (Lindahl, 1993), but that are repaired error-free. Supporting this is the observation that some loci, after AID targeting, are more efficiently repaired than others (Liu et al., 2008). It is therefore possible that there is a link between enhanced RNAPII pausing and regulation of AID lesion resolution. Those loci that are deaminated by AID but repaired efficiently (Liu et al., 2008) may have AID loaded during elongation but may be missing 'extra' pausing for DNA repair pathway channelling steps. AID's universal presence near promoters (Pavri et al., 2010, Yamane et al., 2011) may only become mutagenic if AID can recruit or alter DNA repair factors. This hypothesis is substantiated by this thesis work, showing that AID binds the DNA damage response protein PCNA in B cells, potentially altering DNA repair pathway choice upon lesion recognition (discussed below, Figure 27). If AID needs to stay close to its lesion in order to influence downstream processing, the model of a tight association of AID on chromatin during transcription would be supported, as opposed to a continuous dissociation and association model.

5.1.4 Conclusion

This work has provided biochemical and genetic insight into understanding the targeting of AID to the *Ig* locus. The approach to isolate physiological AID containing protein complexes only from chromatin has identified a novel component, the PAF complex, as well as verified the significance of previously identified factors (SUPT5H, SUPT6H, and FACT) in AID biology. Furthermore, the data described here extends the current model of AID gaining access to DNA by stalled RNA polymerase to a more complex model, where AID is intimately and specifically linked with RNAPII in the

phase of pausing and elongation, requiring the correct chromatin environment defined by histone modification cascades. Future work will build on these discoveries to define AID targeting further. Interestingly, the finding that AID interacts with PAF1 and the RNAPII elongation complex is reminiscent of a model put forth 15 years ago by Storb and colleagues, where an unknown mutator (now known to be AID) would bind to initiating RNAPII and travel along the machinery during transcription elongation (Peters and Storb, 1996).

5.2 Aim II

The approach to study PCNA as a potential interacting partner for AID was based on a candidate approach and guided by previous knowledge on AID function and cofactors.

In this study it was shown that AID interacts with PCNA in several systems, most likely in a direct manner. As AID and PCNA were shown to interact in a system that does not contain other eukaryotic proteins, it is doubtful that the interaction is mediated through RPA, a reported AID interacting protein (Chaudhuri et al., 2004) and a cofactor of PCNA (Davies et al., 2008, Moldovan et al., 2007). Moreover, it has been possible to narrow down domains of interaction within AID as well as within PCNA.

Although only preliminary, PCNA does not seem to globally recruit AID or influence AID function. In a classical, as well as a newly developed ODA that enables monitoring AID activity on a substrate carrying loaded PCNA, AID catalysis was not stimulated by PCNA. Using a immunofluorescence approach, PCNA foci also don't seem to recruit or tether AID (supported by preliminary experiments in Appendix Figure 28). Therefore, one possibility is that this interaction is of importance after AID has carried out its catalytical function. The kinetics of events at an AID lesion may develop in the following way: AID is targeted to its site of action by transcription dependent mechanisms, followed by recruitment of PCNA for post lesion control of resolution. Importantly, an inhibitory effect of AID on the ability of PCNA to be modified at residue K164 could be observed in two different systems.

5.2.1 PCNA and its influence in *Ig* diversification

5.2.1.1 PCNA and repair proteins in SHM

SHM requires a DNA glycosylase for mutation generation outside phase Ia (Transitions at dC), as dU can be converted to an abasic site, which becomes a substrate for other DNA repair processes. UNG, the principal function of which is to remove dUs from the genome after DNA replication, is currently the only uracil dependent DNA glycosyslase required for SHM, as over-expression of another uracil glycosylase (SMUG1) cannot rescue the SHM phenotype (Di Noia and Neuberger, 2007). UNG is also the only glycosylase found to associate with PCNA (Moldovan et al., 2007). Hence it is possible

that the interaction of PCNA and AID produce a complex at the site of the lesion, and only PCNA interacting BER proteins are able to have access to the lesion. PCNA also associates directly with the MMR protein MSH2. Thus, components of both, MMR and BER, can be influenced by PCNA (possibly in competition). Cross-talk between both pathways exists in the form of nicks or abasic sites, generated by UNG as substrates for MMR reactions (Schanz et al., 2009). It is possible, that indirect recruitment of repair factors by AID through PCNA as well as repair balance shift can have a direct influence on Ig diversification.

5.2.1.2 The influence of PCNA K164 on SHM

PCNA has a role in AID dependent *Ig* diversification, but the details of this mechanism have not been fully elucidated. As mentioned in the introduction, mutational shift in *Ig* hypermutation in PCNA K164 mutants had been interpreted as inability of PCNA to recruit polymerase for phase II mutations (mainly at A/T). However, looking closely at the mutation patterns in the three systems examining the role of PCNA K164 mutations (Arakawa et al., 2006, Langerak et al., 2009, Langerak et al., 2007, Roa et al., 2008) revealed an effect on phase I of SHM. In the mutant mice, aside from the drop in mutations at A/T, overall mutation frequency is not (Roa et al., 2008) or weakly (Langerak et al., 2007) altered. Importantly, a compensatory increase of transition mutations at C and G bases in the transcribed strand (increased mutations at G) could be identified. Furthermore, despite an altered rudimentary form of SHM in DT40 (forced by pseudogene removal), PCNA K164 mutants retain this aspect of the phenotype (increased transitions at G/C) (Arakawa et al., 2006). In DT40, this is recapitulated in RAD18 mutants, the major ubiquitin ligase for PCNA. The knock-out cell line shows a similar trend in that transition mutations at C/G are enhanced. However, RAD18 is not the only ligase to induce PCNA ubiquitination (Simpson et al., 2006).

This complex phenotype, including the remarkable shift in mutagenesis favouring phase Ia suggests that PCNA has an early role in SHM, making it a potential central protein for all classes of SHM mutations.

On the other hand, TLS polymerases are not exclusively dependent on monoubiquitination of K164 to mediate associate with PCNA, as shown for both Rev1 (Edmunds et al., 2008) and as Polη (Haracska et al., 2006, Nikolaishvili-Feinberg et al.,

2008). Both in K164 mutation and in Rev1 knock-out in DT40, there is a deleterious effect on Rev1 dependent mutations (Rev 1 preferentially catalyses C to G and G to C mutations). This indicates that Rev1 may require ubiquitination of PCNA in that system (Arakawa et al., 2006), while in mice this dependency could not be seen, as Rev1 specific mutations did not (Roa et al., 2008) or only slightly (Langerak et al., 2007) change in a PCNA K164 background. These discrepancies might be due to the different configuration of SHM in the two systems.

5.2.1.3 PCNA modification and TLS in SHM

The data in this study indicate that AID could repress PCNA ubiquitination, which could prevent TLS DNA polymerase association. As TLS polymerases (e.g. Rev1) also play a role in phase Ib of SHM in DT40 (Ross and Sale, 2006), and ubiquitination of PCNA enhances TLS polymerase activity (Garg and Burgers, 2005), AID's ability to alter this interaction would allow for a shift in the balance between phase Ia and the TLS polymerase dependent phases Ib and II. If K164 modifications are repressed, the dU becomes a template for TLS polymerase independent phase Ia more easily. As replication across dU would be fixed as a mutation, whereas TLS polymerase dependent repair can be error free, the phase Ia pathway may operate at a higher mutagenic frequency than pathways relying on any repair proteins. The increase in mutations at G/C in the K164 phenotypes described above would support this.

It will have to be determined in detail if AID is able to exclude specific polymerases from accessing dU lesions and if there could be an influence on UNG.

The influence of AID on PCNA modifications could be investigated further *in vivo*. It is possible that an overexpression of AID could alter the overall pattern of the modified PCNA forms, and possibly alter the cell's sensitivity to DNA damage by blocking PCNA modification.

5.2.1.4 PCNA and SHM in the cell cycle

The cell cycle can also influence PCNA K164 modification and with it, protein association. For instance, Rev1 activity that is independent of monoubiquitination of PCNA is thought to have a role during S phase at stalled DNA replication forks via a non-catalytic function (Edmunds et al., 2008). During DNA repair in G1, PCNA is mono-ubiqutinated at K164 by the RAD6/RAD18 damage response pathway (Hoege et

al., 2002). It has also been hypothesised that the different types of mutations occurring in SHM arise in different phases of the cell cycle (Franklin and Blanden, 2008, Liu and Schatz, 2009, Weill and Reynaud, 2008). UNG activity peaks at S-phase and supports that mutations at C/G (BER dependent) are fixed here, while MMR-like processing and thus mutations at A/T might be enhanced outside of S-phase (Weill and Reynaud, 2008), or at least after passage of the replication fork (Sale et al., 2009). In (Faili et al., 2002) it has been suggested that AID-induced lesions are generated in G1, while the final resolution towards mutations of phase II of SHM would occur in S-phase, but this hypothesis has not been proven.

It will be interesting to determine when in the cell cycle the AID-PCNA interaction is most prominent.

5.2.1.5 AID can influence the lesion resolution

Work in our lab recently detailed the involvement of polymerase-based repair pathways in AID lesion resolution (Franchini et al, 2011 submitted). An *in vitro* system was developed that can reveal processing of an AID lesion using a DNA repair competent *Xenopus laevis* oocyte extract (reviewed in Garner and Costanzo, 2009). In this system AID is targeted to a supercoiled plasmid to induce lesions in stretches of ssDNA. When added to the oocyte extract, repair can be monitored by incorporation of biotinylated cytosine. Strikingly, if the system was supplied biotinylated adenosine instead of cytosine, a base that cannot be deaminated by AID, base incorporation was still detected. Therefore, incorporation of bases away from the original AID lesions could be observed (similar to phase II of SHM). This proves patch repair activity during the process of AID-induced lesion resolution.

Another important observation made using this *in vitro* system is, that AID needs to stay associated with its own lesion to efficiently induce repair. If AID is removed from the plasmid prior to extract incubation, repair response is impaired, suggesting a repair recruitment function or a physical interaction with an effector protein of AID. This hypothetical protein has not yet been identified in this system, but may well be part of a PCNA complex.

5.2.2 AID, PCNA and DNA demethylation

Through its interaction partners and function as a docking platform on DNA, PCNA is well known to also have a role in various epigenetic mechanisms. For instance, through interaction with the histone chaperone CAF-1, correct nucleosome placement is ensured during replication and NER repair (Mello and Almouzni, 2001). Another epigenetic role for PCNA is to maintain DNA methylation patterns. Methylation of DNA is known to be performed by DNA methyltransferases 1, 3a and 3b, with DNMT1 designated as a "maintenance" methyltransferase that can reinstate methylation status of hemimethylated DNA during DNA replication, as well as function in the DNA repair response (Chuang et al., 1997, Iida et al., 2002, Ha et al., 2011). On the other hand, DNA repair has been implicated to play an important role in CpG DNA demethylation (reviewed in Gehring et al., 2009). The molecular mechanisms of this reaction has remained elusive, and a variety of mechanisms for active demethylation have been proposed (Fritz and Papavasiliou, 2010). Although PCNA had been associated with a possible DNA demethylation activity (Vairapandi et al., 2000), no clear mechanistic link had been identified between active DNA demethylation and PCNA. As outlined in the introduction, recent work has indicated that some DNA demethylation is dependent on AID (Bhutani et al., 2010, Morgan et al., 2004, Popp et al., 2010, Rai et al., 2008). Interestingly, AID is active on methylated cytosines *in vitro* and in the oocyte extract repair system (Franchini et al., 2011 submitted). Here, methylated DNA substrate also required to retain AID in order to recruit efficient repair. Moreover, an *in vivo* system (Franchini et al., 2011 submitted) that recapitulated AID-induced DNA demethylation of the H19 gene was established in mice. The data suggested that AID tethered to a gene initiates demethylation through a processive polymerase mechanism, which is likely to involve PCNA.

Demonstrating a direct link between AID and PCNA, provides a mechanism for AID-induced local demethylation. After gaining access to the target, AID would deaminate cytosines or methylated cytosines in single stranded DNA. Reformation of the double stranded DNA and continuous association of AID with the mismatch lesion would allow for PCNA binding and recruitment of PCNA associated factors (e.g. UNG, MSH2, or EXO1). Association of other MMR factors could induce demethylation via processive polymerisation away from the lesion for 500 - 1500 bp (Jiricny, 2006). It still

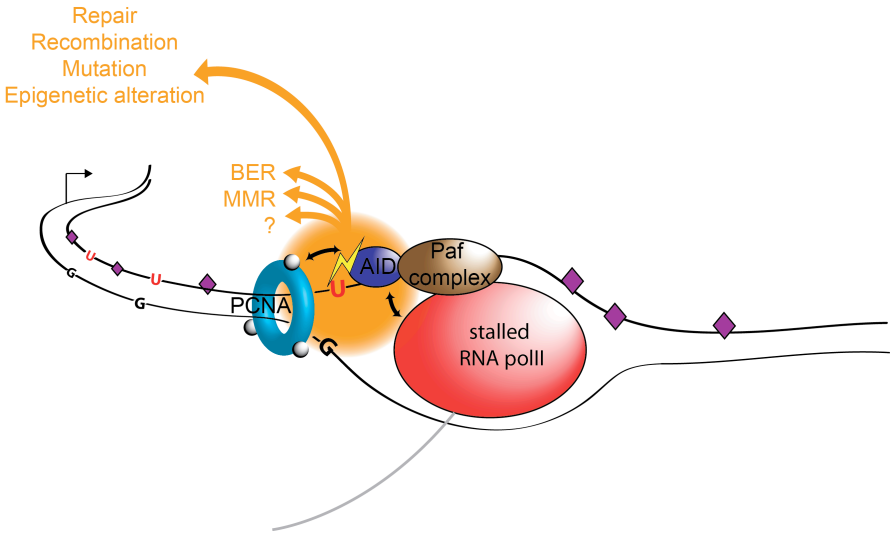


Figure 27. Model of interplay between AID targeting and AID lesion resolution.

AID may be targeted to RNAPII, followed by deamination and interaction with PCNA (blue ring), influencing downstream dU processing.

needs to be determined if during AID-dependent MMR, as in normal MMR, a second lesion (Jiricny, 2006, Schanz et al., 2009) is needed to induce the full repair (or demethylation).

5.2.3 Transcriptional pausing and AID resident time: An effect on PCNA?

Enhanced AID resident time at the site of the lesion could facilitate assembly of a mutasome complex containing various DNA repair factors. As described earlier, RNAPII transcriptional pausing is likely to achieve this. This work introduces a novel player in AID dependent DNA alterations described in Chapter 3, PAF1. Interestingly, Rad6, the ubiquitin conjugating enzyme that is stimulated by PAF1, also functions in PCNA monoubiquitination.

RNAPII pausing may have a dual function, first for AID loading, and second it may enhance AID's resident time at the target locus. It is possible that enhanced RNAPII pausing could both enhance the ssDNA accessibility for AID (Figure 26) as well as cause AID to reside near its own lesion for a prolonged time and induce assembly of a mutasome (Figure 27). Hereby, in a yet poorly understood way, AID could influence the composition and activity of the mutasome complex through its interaction with PCNA in a modification dependent manner.

5.2.4 Conclusion

The immunoglobulin loci seem to have accumulated a large number of different targeting mechanisms for AID as well as the correct lesion resolution machinery. Perhaps, these loci most efficiently combine properties that attract factors needed for programmed mutation, which may explain its outstanding level of targeting by the mutasome above all other loci. If components could accumulate dependent on AID and on each other, the result would be a self-amplifying system with very efficient and specific targeting and lesion resolution. Support for such a model comes from data that shows that UNG and MMR components accumulate on *Ig* dependent on each other and AID and data showing RPA accumulation on *Ig*. Importantly, data from this study highlights the importance of AID as the central mutasome protein reaching to interacting partners for targeting (transcription machinery) as well as lesion resolution (PCNA). These and the other interactions reported for AID would support a self accumulating system that can be mistargeted at a lower efficiency and have a slightly different compositions for each of the tasks AID encounters.

Chapter 6. Appendix

6.1 Work in progress

6.1.1 Can AID be visualised on chromatin together with immobilised nuclear PCNA?

Some of PCNA functions can be observed by imaging, as local concentrations of PCNA in the nucleus and immobilisation on chromatin can be detected during replication and upon genotoxic stress (Miura et al., 1992). Certain proteins (e.g. CAF1) that utilise PCNA can be visualised to colocalise with PCNA upon UV irradiation (Green and Almouzni, 2003). Observing physiological AID under such conditions has proven far more challenging, partially due to lower expression levels than PCNA and AID's mainly cytoplasmic localisation. On the other hand, there are reports towards a relocalisation effect of DNA damage on AID (Brar et al., 2004).

This rationale led me to test if our DT40 tagged AID tools could be utilised to probe the interaction of AID and PCNA and to possibly identify a function for relocalising AID. To this end, DT40 cells were used that carried a GFP-tagged allele of AID, analogous to DT40 lines described (S. Pauklin, CRUK). The cells were irradiated with UV light to enhance PCNA immobilisation on chromatin. To visualise chromatin bound immobilised PCNA (as commonly observed), triton extraction was used (Szuts and Sale, 2006), since omitting this step leads to a diffuse nuclear staining of PCNA (Appendix Figure 28 A). Without triton extraction, nuclear PCNA was detected in most cells in a non-synchronised DT40 population, with a slight variation in intensity. The high proportion of stained cells could be explained by the fast growth rate of DT40 that makes the population rich in S-phase cells that have PCNA upregulated. The AID staining without triton extraction confirms AID to be localised predominantly in the cytoplasm (Rada et al., 2002). When staining is performed using the detergent step, soluble proteins are extracted from the cell structure and only proteins bound to the cell matrix remain. In Appendix Figure 28 B, PCNA foci can be detected inside the nucleus (red signal). It could not be delineated if these were replicative or damage-related concentrations of

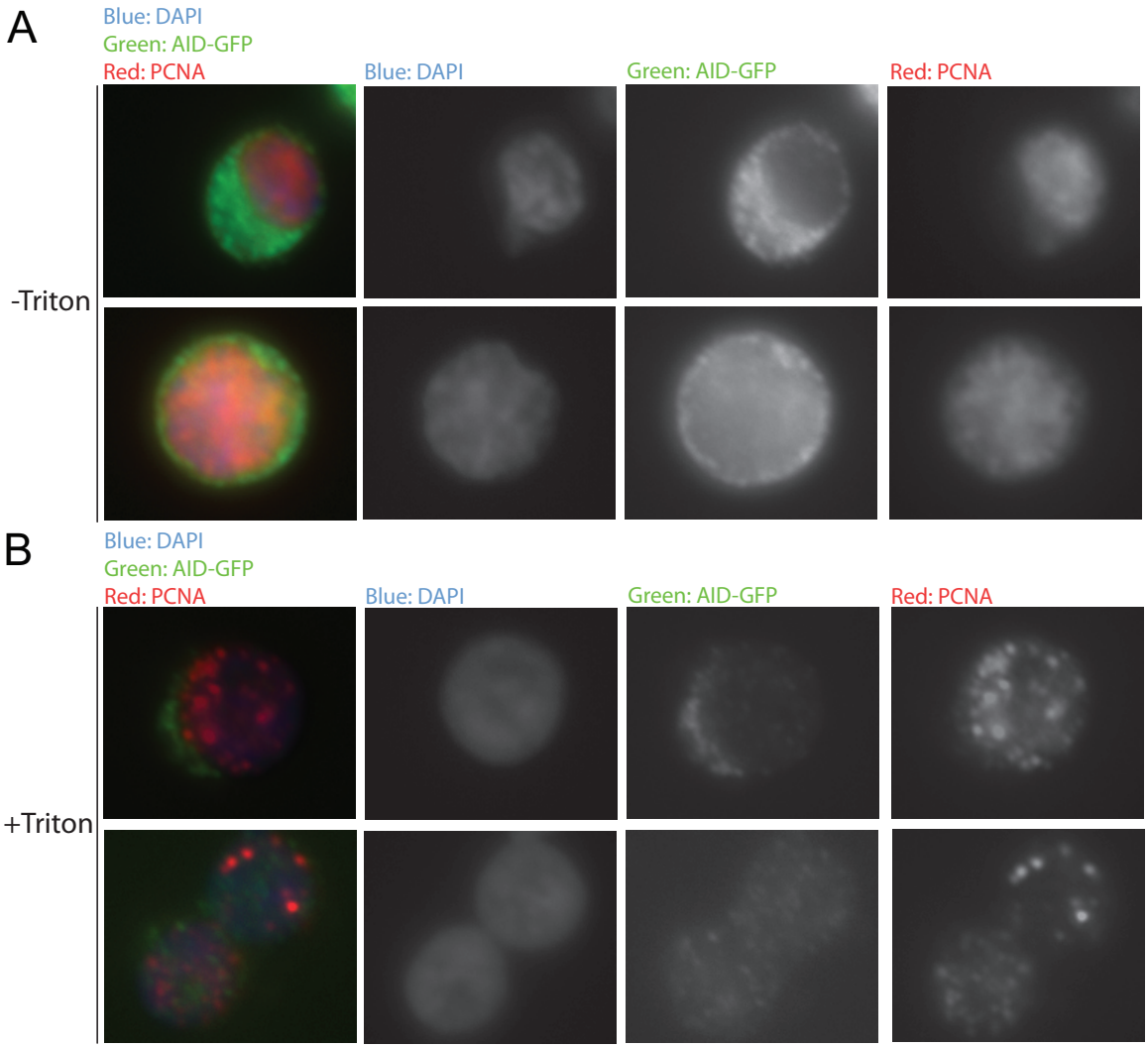


Figure 28. AID does not show colocalisation with Triton insoluble PCNA foci in DT40.

A, DT40 cells expressing endogenously GFP-tagged AID were UV irradiated and fixed, and without detergent extraction, immunofluorescent staining was performed using anti-GFP and anti-PCNA antibodies. Two representative examples of stainings are shown as 3-channel overlays (DAPI signal, blue; anti-GFP signal, green; anti-PCNA signal, red), together with the singled out channels in monochrome. **B**, DT40 cells were treated and stained as described in A, but with the inclusion of a Triton extraction step in the immunofluorescence protocol.

PCNA as unirradiated DT40 cells showed similar levels of PCNA foci in my hands. In the case of AID, nuclear staining seems to be absent after Triton extraction (representative cells shown from different angles), but some residual signal can be detected from the cytoplasmic pool of AID. This is not surprising as only an estimate of 2% of total AID resides in the nucleus (Chapter 3). Interestingly, a cytoplasmic retention mechanism for AID has been invoked repeatedly (Patenaude et al., 2009, Wu et al., 2005), and association of AID with cytoskeletal proteins such as tubulin have been suggested by our own and other laboratories' mass spectrometry studies (S. Pauklin, CRUK; Wu et al., 2005). I could not identify significant overlay of PCNA and AID accumulations (Appendix Figure 28 B), as there seemed to exist no insoluble AID accumulations inside the nucleus under conditions that visualise PCNA foci in this compartment.

In conclusion, these preliminary data show no visible recruitment of AID to PCNA foci under the tested conditions, but it cannot be excluded that this could take place in a localised manner at an undetected scale or under different stimulating conditions.

6.1.2 An approach to analyse function-specific AID using AIDΔC mutants

Another approach to the AID interaction project was in the context of different functions of AID. More specifically, during CSR, SHM, iGC or epigenetic reprogramming, the targeting of AID as well as the lesion resolution differs, pointing towards a differential composition of the respective complex. It is well established that CSR and SHM, although related, use different sets of DNA repair proteins in order to exert their function, and that the C-terminus of AID is especially important for CSR (Barreto et al., 2003, Ta et al., 2003), whereas SHM and enzymatic activity are independent of the C terminus. Thus, there may be common subsets of proteins interacting with AID irrespectively of the C-terminal domain being present. For these reasons, I generated DT40 cell lines that would lack the last 10 amino acids of AID (AIDΔC). AIDΔC would be expected to accumulate in the nucleus, providing enrichment in the AID population that will form complexes on DNA. Furthermore, comparing the binding partners of full length AID and AIDΔC may allow a classification in SHM or CSR relevant factors. Generation of the cell lines was analogous to the lines used for the biochemical analysis described in Chapter 3 (see Appendix Figure 29). The cloned AID locus was modified resulting in a truncated and tagged AID. Subsequently, endogenous AID was targeted in DT40 parental cells and recombinant clones selected (Appendix Figure 29 A and B). AID was tagged with well-proven peptide tags such as FLAG-tag and MYC tag. For the purpose of *in vivo* observations, I additionally generated GFP-tagged as well as *Renilla* luciferase tagged AID cell lines. Clones of the obtained cell lines were analysed visually and biochemically (Appendix Figure 29 C and D). Surprisingly, Figure 29 D shows that AID is not enriched in the nuclear fraction of the AIDΔC mutant as expected. Looking at several recombinant clones using FACS, immunofluorescence and western blot (Appendix Figure 29 C and D), preliminary observations show reduced levels of overall AID protein in the AIDΔC mutant. Both of these phenomena were not expected, as similar mutants when overexpressed in other cellular systems (fibroblasts), did not show this phenotype (Ito et al., 2004, McBride et al., 2004). However, the regulatory layers controlling AID shuttling and stability are complex and possibly not fully understood (Patenaude et al., 2009). The effect of the AID truncation could also differ with exogenous or endogenous expression in B cells. In line with this, a differential reaction has been observed between

overexpression and endogenous systems when blocking nuclear export by leptomycin B. This treatment can enrich AID in the nucleus, although at a much higher efficiency in AID overexpressing cells than in endogenously AID expressing B cells (Patenaude et al., 2009). Both above observations, lower nuclear and lower overall AID protein levels compared to wild type cells, could be explained by a higher nuclear turnover of AID protein, which has also been observed before (Aoufouchi et al., 2008, Uchimura et al., 2011). This would reduce the levels of nuclear, but also cytoplasmic AID, which is a pool for nuclear import.

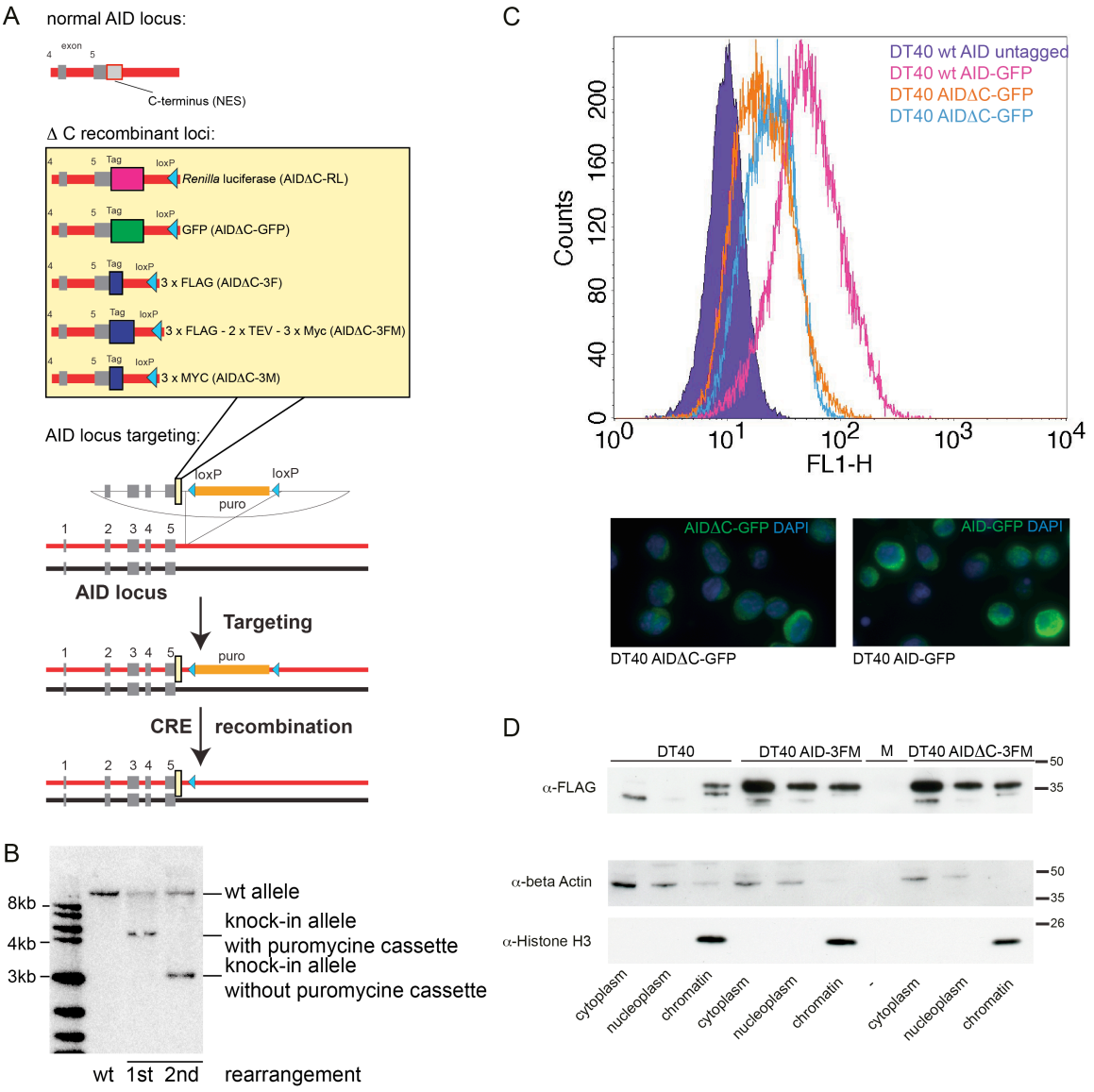

A
normal AID locus:
exon
C-terminus (NES)
Δ C recombinant loci:
Renilla luciferase (AIDΔC-RL)
GFP (AIDΔC-GFP)
3 x FLAG (AIDΔC-3F)
3 x FLAG - 2 x TEV - 3 x Myc (AIDΔC-3FM)
3 x MYC (AIDΔC-3M)
AID locus targeting:
loxP
puro
AID locus
Targeting
CRE
recombination
B
8kb
4kb
3kb
wt allele
knock-in allele
with puromycine cassette
knock-in allele
without puromycine cassette
wt 1st 2nd
rearrangement
C
DT40 wt AID untagged
DT40 wt AID-GFP
DT40 AIDΔC-GFP
DT40 AIDΔC-GFP
Counts
FL1-H
AIDΔC-GFP DAPI
AID-GFP DAPI
DT40 AIDΔC-GFP
DT40 AID-GFP
D
DT40
DT40 AID-3FM
M
DT40 AIDΔC-3FM
α-FLAG
α-beta Actin
α-Histone H3
cytoplasm
nucleoplasm
chromatin

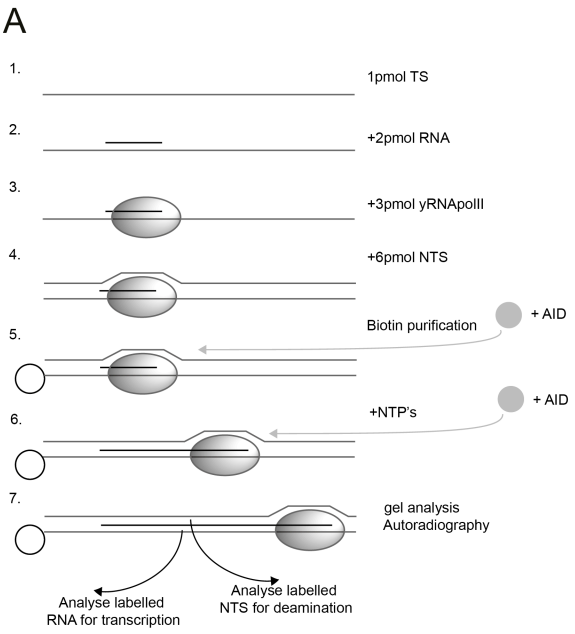
Figure 29. Generation and analysis of DT40 cell lines harbouring a endogenously tagged AID mutant lacking the C terminus.

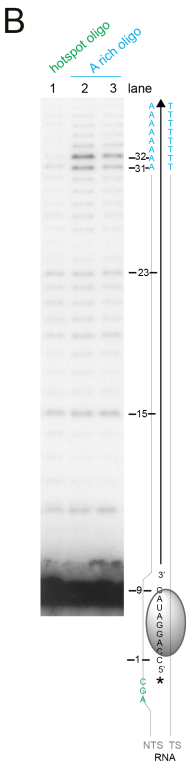
A, Schematic of the generation of the five DT40 lines. The genomic AID locus was modified to firstly, lack the final 10 amino acids of the protein, and secondly, to carry the C-terminal tags *Renilla* luciferase (RL), GFP, 3 x FLAG (3F), 3 x FLAG 2 x TEV 3 x MYC (3FM) or 3 x MYC (3M). The respective constructs were transfected into DT40 cells for homologous recombination at the AID locus, and puromycine resistant clones were selected. Positive clones were treated with cell soluble, purified HTN-CRE protein for removal of the puromycine resistance. **B**, representative southern blot analysis of the AID loci in wt, intermediate recombinants and final recombinant cell lines. Southern blot results were used for identifying recombinant clones. **C**, Analysis of the newly generated DT40 AIDΔC-GFP cell line in comparison to the DT40 wt, and AID-GFP cell lines (S. Pauklin, G. Bachmann, CRUK). Upper panel, flow cytometric analysis of GFP fluorescence in 2 recombinant clones (orange and blue profiles) showing a lower GFP expression then the DT40 wt AID-GFP cell line (pink profile), with the DT40 wt as a control (Lacking GFP, purple profile). Lower panel, immunofluorescent imaging of a DT40 AIDΔC-GFP population (left) and an AID-GFP population (right). AID-GFP - green, DAPI - blue. **D**, Fractionation analysis of DT40 AIDΔC-3FM versus DT40 wt and AID-3FM cell line (Pauklin et al., 2009). Western blot analysis using anti-FLAG, anti-beta actin, and anti-Histone H3; DT40 - left, AID-3FM -middle, and AIDΔC-3FM - right). Fractions were not loaded representative of the total protein amount in each fraction, but 15 μg protein was loaded in all lanes.

6.1.3 *In vitro* transcription assays

To understand the molecular link between AID and the RNAPII complex, *in vitro* reconstitution studies were used. AID deamination facilitated by transcription has been reconstituted *in vitro* before using phage-coded T7 RNA polymerase (Chaudhuri et al., 2003) (Basu et al., 2011) or *E. coli* RNA polymerase (Besmer et al., 2006). Using T7 polymerase transcription, (Canugovi et al., 2009) could reconstitute AID deamination facilitated by transcriptional pausing. However, the RNA polymerases used in these studies are much simpler enzymes than the eukaryotic RNAPII and function without the eukaryotic cofactors that have now been shown to be crucial to AID function. Therefore, these conditions reflect an artificial situation. This is confirmed by a need for an overwhelming presence of RPA in these reactions. Therefore, we developed a biochemical assay that addresses the mechanistics of transcription coupled DNA deamination using eukaryotic enzymes. To this end, I undertook pilot studies based on a system described in (Saeki and Svejstrup, 2009) and by D. Hobson, and J. Svejstrup, CRUK (personal communication) (Appendix Figure 30 A). Here, NTP dependent RNA elongation competent transcription bubbles were reconstituted *in vitro* (Appendix Figure 30 B). Two different substrates show RNA elongation by yeast RNAPII after complex formation and NTP addition. As shown in Appendix Figure 30 B (lanes 2 and 3), it is also possible to arrest transcription by introducing obstacle sequences or nucleotides. These complexes can be exposed to AID while transcription or transcriptional arrest is taking place (Appendix Figure 30 C). When AID was applied to a transcription complex substrate, the NTS is largely protected from degradation (lanes 3, 4, 5 versus 10, 11, 12) as opposed to a control single stranded NTS. However a few product bands appear (lanes 3, 4, 5) in comparison to lane 2 (without AID). Some products are dependent on AID as well as UDG (e.g. stars in lane 5, vs lane 6), possibly representing true deamination, deglycosylation and cleavage events. Because of multiple dC residues along the oligonucleotide, the length of the products cannot be interpreted as an indication towards a site of AID deamination. However, digestion of the single stranded substrate (right hand side) reveals unspecific nuclease activity in the protein preparations. Lane 13 shows a UDG independent degradation of DNA by contamination of AID. Although the UDG preparation (commercial) does not degrade the single stranded DNA substrate, it is

known to contain endonucleases able to work on abasic and possibly other substrates. This leads to the full degradation of the substrate when both preparations are combined. For exact delineation of activities responsible for individual products, different protein preparations need to be used. In the future, an AID mutant that lacks deamination catalysis, E58Q, (Chen et al., 2008, Holden et al., 2008, Ramiro et al., 2003) will be included to control for nuclease activities. After substrate oligonucleotides as well as protein preparations are optimised, it will be possible to address the co-transcriptional accessibility of cytosines for AID and to draw biochemical conclusions. It may be possible to induce paused RNAPII and even to mimic traffic by reconstituting transcription bubbles with several colliding RNA polymerases (D. Hobson, CRUK) to study the effect on AID accessibility to ssDNA. Subsequently, purified transcription elongation factors could be added to the reactions to delineate the optimal transcription elongation complex composition for AID accessibility to its substrate.


A
1.
1pmol TS
2.
+2pmol RNA
3.
+3pmol yRNApolII
4.
+6pmol NTS
Biotin purification
+ AID
5.
+NTP's
+ AID
6.
7.
gel analysis
Autoradiography
Analyse labelled
RNA for transcription
Analyse labelled
NTS for deamination


B
hotspot oligo
A rich oligo
1 2 3 lane
32
31
23
15
9
1
NTS TS
RNA

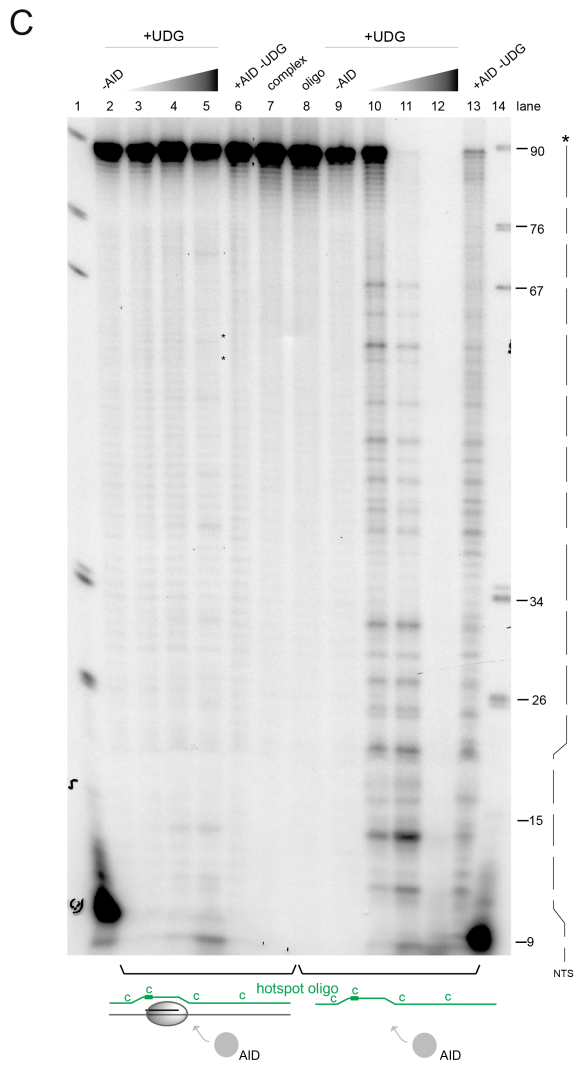

C
+UDG
+UDG
-AID
+AID -UDG
complex
oligo
-AID
+AID -UDG
1 2 3 4 5 6 7 8 9 10 11 12 13 14 lane
90
76
67
34
26
15
9
NTS
hotspot oligo
AID
AID

Figure 30. *In vitro* reconstitution of transcription coupled deamination.

A, schematic of the *in vitro* reconstitution assay based on (Saeki and Svejstrup, 2009). 1 pmol of a 90 nucleotide long oligonucleotide representing the transcribed strand (TS) is hybridised to 2 pmol complementary RNA. 3 pmol purified yeast RNAPII (gift from D. Hobson, CRUK) is added followed by 6 pmol of a complementary 90-mer strand representing the non-transcribed strand (NTS). The assembled complex was purified via a biotin label on the TS. The obtained complex was exposed to AID before and after the addition of NTP's, which results in start of transcription. When using a radiolabelled RNA, progression of transcription in the assay can be monitored (as in B), whereas when using a radiolabelled NTS, deamination can be monitored if an additional UDG digestion step is included (as in C). **B**, Two different TS-NTS pairs are tested for complex formation and RNA elongation. One pairs was designed to contain an AID hotspot motif (AGC) in the potentially single stranded region of the transcription bubble (green); the other pair was designed to contain an A rich stretch towards the end of the transcribed region known to arrest or destabilise the polymerase (barrier) (blue). Radiolabelled RNA was used to visualise elongated RNA products, forming after complex formation and NTP addition (lanes 1-3). Product in the length of 31 and 32 nucleotides accumulates in lanes 2 and 3 at the normal NTP concentration, 0.58mM, (lane 3) as well as at half the NTP concentration (lane 2). **C**, Left part, AID was applied to a transcription complex substrate as formed in A; right part, AID was applied under the same conditions to the respective single stranded NTS substrate. Radiolabelled NTS was used to visualise products. In both conditions, the 90-mer tested in B, carrying a hotspot at the site of the transcription bubble, was used (green). Lanes 3, 4, 5 versus 10, 11, 12 show products after increasing exposure to AID followed by UDG digestion. Lanes 2 and 9 show controls without AID. Lanes 6 and 13 show controls without UDG. Lane 7 shows a sample of the untreated complex, and lane 8 shows a sample of the untreated NTS. Stars in lane 5 indicate possible specific products.

6.2 Additional information

Identified Proteins (232)	Name	Accession Number	kDa	NC	NA	CC	CA
Transcription elongation factor SPT5 - Gallus gallus (Chicken)	SPT5	SPT5H_CHICK	120 kDa	0	0	0	11
Putative uncharacterized protein - Gallus gallus (Chicken)	SF3A1	Q5ZM84_CHICK	89 kDa	0	4	0	8
U5 small nuclear ribonucleoprotein 200 kDa helicase - Homo sapiens (Human)	U5-200 helicase	U520_HUMAN (+1)	245 kDa	2	6	1	7
Elongation factor 2 - Gallus gallus (Chicken)	EF2	EF2_CHICK	95 kDa	2	6	0	6
RNA polymerase-associated protein CTR9 homolog - Homo sapiens (Human)	CTR9	CTR9_HUMAN (+5)	134 kDa	0	0	0	6
DNA-directed RNA polymerase II subunit RPB2 - Homo sapiens (Human)	RNA pol II (sub2)	RPB2_HUMAN (+5)	134 kDa	0	0	0	6
DNA-directed RNA polymerase II subunit RPB1 - Homo sapiens (Human)	RNA pol II (sub1)	RPB1_HUMAN (+6)	217 kDa	1	0	0	5
RNA polymerase II-associated factor 1 homolog - Bos taurus (Bovine)	PAF1	PAF1_BOVIN (+5)	60 kDa	0	0	0	4
Transcription elongation factor SPT6 - Homo sapiens (Human)	SPT6	SPT6H_HUMAN (+2)	199 kDa	0	0	0	4
Lamin-B2 - Gallus gallus (Chicken)	Lamin B2	LMNB2_CHICK	68 kDa	0	0	0	4
Putative uncharacterized protein - Gallus gallus (Chicken)	PolyA BP like with 4 RRM	Q5ZL53_CHICK (+9)	71 kDa	5	6	0	3
Splicing factor 3B subunit 1 - Homo sapiens (Human)	SF3B	SF3B1_HUMAN (+2)	146 kDa	0	4	1	3

Putative uncharacterized protein - Gallus gallus (Chicken)	SF RS7	Q5ZMI0_CHICK	26 kDa	0	0	2	3
Putative uncharacterized protein - Gallus gallus (Chicken)	DHX15	Q5F3A9_CHICK (+1)	88 kDa	1	2	0	3
Putative uncharacterized protein - Gallus gallus (Chicken)	SF RS6	Q5ZJ59_CHICK (+10)	40 kDa	0	2	2	2
DNA-directed RNA polymerases I, II, and III subunit RPABC1 - Bos taurus (Bovine)	RNA pol I,II,III subunit (25)	RPAB1_BOVIN (+8)	24 kDa	0	0	3	2
Nucleophosmin - Gallus gallus (Chicken)	Nucleophosmin	NPM_CHICK	33 kDa	1	1	0	2
Small nuclear ribonucleoprotein-associated protein B' - Gallus gallus (Chicken)	snRNP BP	RSMB_CHICK (+1)	25 kDa	2	0	0	2
DNA topoisomerase I - Gallus gallus (Chicken)	DNA Topo I	P79994_CHICK	91 kDa	0	0	0	2
Putative uncharacterized protein - Gallus gallus (Chicken)	DDX3X	Q5F491_CHICK (+46)	72 kDa	0	1	0	2
cDNA FLJ77459, highly similar to Homo sapiens interleukin enhancer binding factor 3, 90kDa, mRNA - Homo sapiens (Human)	NF-AT 90 IL3F	A8K6F2_HUMAN	77 kDa	0	1	0	2
RNA polymerase-associated protein LEO1 - Homo sapiens (Human)	LEO	LEO1_HUMAN (+2)	75 kDa	0	0	0	2
Putative uncharacterized protein - Gallus gallus (Chicken)	MCM5	Q5ZKL0_CHICK	82 kDa	0	0	0	2
WD40 repeat-containing protein SMU1 - Gallus gallus (Chicken)	SMU1	SMU1_CHICK (+12)	57 kDa	0	0	0	2
Nuclease-sensitive element-binding protein 1 - Gallus gallus (Chicken)	YB1	YBOX1_CHICK (+28)	36 kDa	0	1	0	1
Putative uncharacterized protein - Gallus gallus (Chicken)	BUB3	Q5ZHW3_CHICK	37 kDa	0	0	1	1

Putative uncharacterized protein - Gallus gallus (Chicken)	DEK	Q5F422_CHICK	42 kDa	0	1	1	1
U2 small nuclear ribonucleoprotein B" - Homo sapiens (Human)	U2snRNPB2	RU2B_HUMAN (+6)	25 kDa	1	0	1	1
RNA-binding protein HuA - Gallus gallus (Chicken)	HuA	Q9PW24_CHICK (+13)	36 kDa	1	0	0	1
Putative uncharacterized protein - Gallus gallus (Chicken)	EIF6	Q5ZL92_CHICK	27 kDa	1	0	0	1
Pre-mRNA-processing factor 6 - Homo sapiens (Human)	PRP6	PRP6_HUMAN (+12)	107 kDa	0	1	0	1
Putative uncharacterized protein - Gallus gallus (Chicken)	THRAP3	Q5ZMK4_CHICK	72 kDa	0	0	0	1
U4/U6 small nuclear ribonucleoprotein Prp4 - Bos taurus (Bovine)	PRP4	PRP4_BOVIN (+12)	58 kDa	0	0	0	1
ACTL6A protein - Bos taurus (Bovine)	ACTL6A	A4IFJ8_BOVIN (+14)	47 kDa	0	0	0	1
FACT complex subunit SPT16 - Homo sapiens (Human)	SPT16	SP16H_HUMAN (+6)	120 kDa	0	0	0	1
ATP-dependent RNA helicase DDX1 - Gallus gallus (Chicken)	DDX1	DDX1_CHICK (+15)	82 kDa	0	0	0	1
FACT complex subunit SSRP1 - Gallus gallus (Chicken)	SSRP1	SSRP1_CHICK (+13)	80 kDa	0	0	0	1
Uncharacterized protein C17orf85 homolog - Gallus gallus (Chicken)	C17orf85	CQ085_CHICK (+4)	69 kDa	0	0	0	1
Putative uncharacterized protein - Gallus gallus (Chicken)	IK	Q5ZHM2_CHICK (+28)	65 kDa	0	0	0	1
Lamin-B1 - Gallus gallus (Chicken)	LaminB1	LMNB1_CHICK	67 kDa	0	0	0	1
Nuclear protein matrin 3 - Gallus gallus (Chicken)	Martin3	Q8UWC5_CHICK	101 kDa	0	6	0	0
Splicing factor 3a subunit 2 - Gallus gallus (Chicken)	SF3a2	Q66VY4_CHICK (+3)	35 kDa	0	1	2	0

116 kDa U5 small nuclear ribonucleoprotein component - Gallus gallus (Chicken)	U5snRNPS1	U5S1_CHICK	109 kDa	1	2	0	0
Cohesin complex subunit - Gallus gallus (Chicken)	cohesin	Q8AWB8_CHICK (+15)	142 kDa	0	3	0	0
Putative uncharacterized protein - Gallus gallus (Chicken)	ABCE1	Q5ZJX6_CHICK (+11)	67 kDa	0	3	0	0
Nuclear cap-binding protein subunit 1 - Gallus gallus (Chicken)	NCBP1	NCBP1_CHICK	93 kDa	0	3	0	0
Putative uncharacterized protein - Gallus gallus (Chicken)	PDHX	Q5F3G9_CHICK	53 kDa	1	1	0	0
DnaJ homolog subfamily C member 10 precursor - Xenopus laevis (African clawed frog)	DNAJC10	DJC10_XENLA	91 kDa	0	2	0	0
LOC398542 protein - Xenopus laevis (African clawed frog)	ARS2-b	Q6INH5_XENLA	98 kDa	0	2	0	0
Putative uncharacterized protein - Gallus gallus (Chicken)	hnRNP M	Q5ZL80_CHICK	76 kDa	0	2	0	0
Dynein light chain 2, cytoplasmic - Bos taurus (Bovine)	DyneinLC2	DYL2_BOVIN (+12)	10 kDa	0	1	0	0
Cytoskeleton-associated protein 5 - Homo sapiens (Human)	CKAP5	CKAP5_HUMAN (+9)	226 kDa	0	1	0	0
Serine hydroxymethyltransferase, cytosolic - Oryctolagus cuniculus (Rabbit)	GLYC	GLYC_RABIT (+45)	53 kDa	0	1	0	0
Putative uncharacterized protein - Gallus gallus (Chicken)	RiboS26L	Q5ZM66_CHICK (+66)	13 kDa	0	1	0	0
NIK- and IKBKB-binding protein - Bos taurus (Bovine)	NIBP	NIBP_BOVIN (+6)	127 kDa	0	1	0	0
Putative uncharacterized protein - Gallus gallus (Chicken)	DROSHA	Q5ZIR3_CHICK (+11)	155 kDa	0	1	0	0

Eukaryotic translation initiation factor 3 subunit C - Homo sapiens (Human)	EIF3C	EIF3C_HUMAN (+7)	105 kDa	0	1	0	0
Eukaryotic translation initiation factor 3 subunit G - Bos taurus (Bovine)	EIF3G	EIF3G_BOVIN (+33)	36 kDa	0	1	0	0
Nuclear calmodulin-binding protein - Gallus gallus (Chicken)	NCBP1	Q9YHD2_CHICK	84 kDa	0	1	0	0
Transport protein particle subunit TMEM1 - Homo sapiens (Human)	TMEM1	TMEM1_HUMAN (+13)	142 kDa	0	1	0	0
Staphylococcal nuclease domain-containing protein 1 - Homo sapiens (Human)	SND1	SND1_HUMAN (+20)	102 kDa	0	1	0	0
Putative uncharacterized protein - Gallus gallus (Chicken)	PLK1	Q5ZJ36_CHICK	67 kDa	0	1	0	0
Putative uncharacterized protein - Gallus gallus (Chicken)	GTPBP4	Q5ZM18_CHICK (+17)	74 kDa	0	1	0	0

Table 4. Full list of peptides recovered from mass spectrometric analysis of AID isolated from B cell chromatin.

Indicated is the description, name and accession number of the identified proteins as well as the number of peptide hits in the nucleoplasm control IP sample (NC), the nucleoplasm FLAG tagged AID IP sample (NA), the chromatin control sample (CC) and the chromatin FLAG tagged AID IP sample (CA).

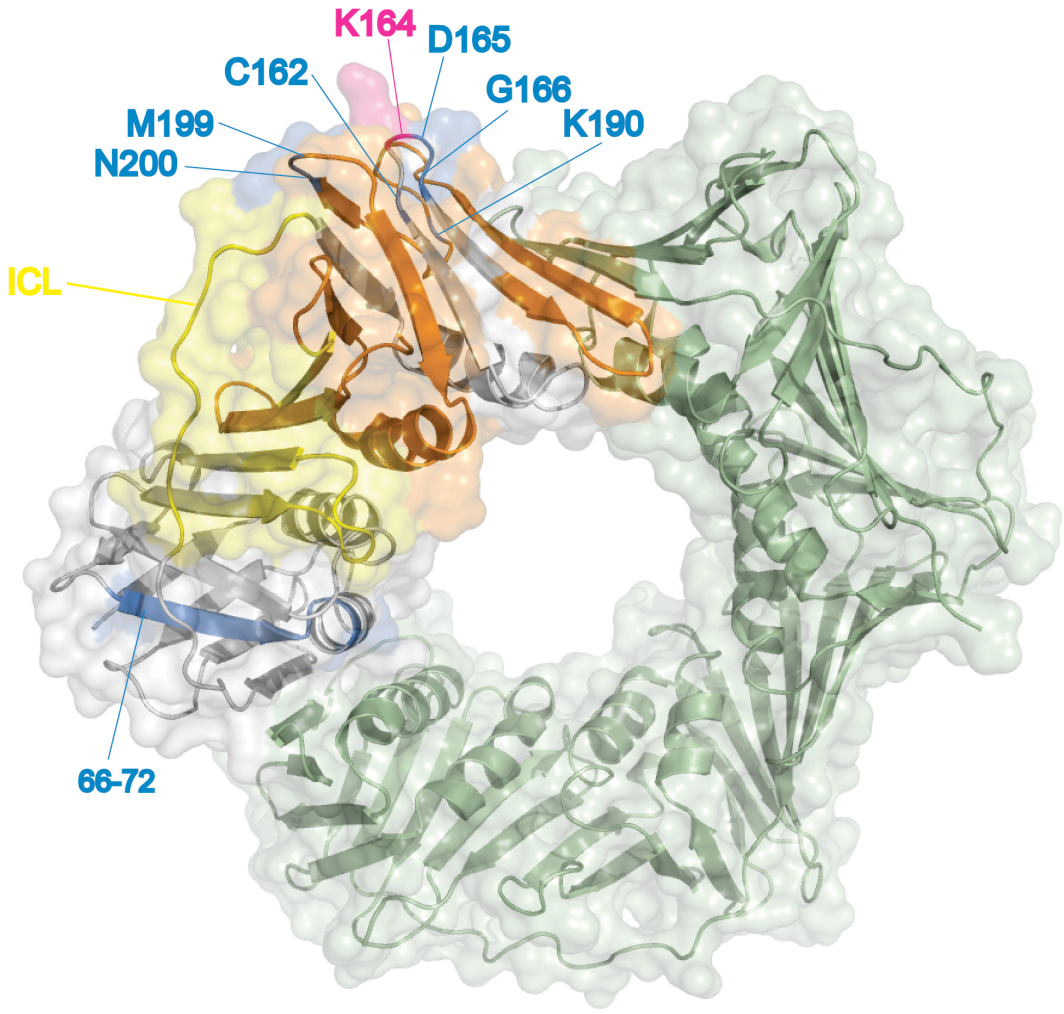


Figure 31. Structure model of the human PCNA trimer.

A transparent surface model as well as an underlying ribbon structure model of the PCNA trimeric ring is depicted (2 monomeres depicted in light green, one monomer depicted in light grey indicating residues of interest with colours). Residues mutated in the experiments described in Figure 20 B are coloured in blue, the crucial residue lysine 164 is coloured in pink. The C terminus of the PCNA monomer deleted in the experiments in Figure 20 A is coloured in orange. The area of PCNA interaction with PIP box motives lies at the interdomain connecting loop (ICL) and the groove below and is coloured in yellow (including some C terminal residued coloured in orange). The model was generated with the MacPyMOL software (DeLano Scientific LLC) using PDB file 1VYM (native human PCNA).

Reference List

AIDA, M., CHEN, Y., NAKAJIMA, K., YAMAGUCHI, Y., WADA, T. & HANDA, H. 2006. Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the junB gene. *Mol Cell Biol,* 26**,** 6094-104.

AIYAR, S. E., SUN, J. L., BLAIR, A. L., MOSKALUK, C. A., LU, Y. Z., YE, Q. N., YAMAGUCHI, Y., MUKHERJEE, A., REN, D. M., HANDA, H. & LI, R. 2004. Attenuation of estrogen receptor alpha-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor. *Genes Dev,* 18**,** 2134-46.

ALLEN, C. D., OKADA, T. & CYSTER, J. G. 2007. Germinal-center organization and cellular dynamics. *Immunity,* 27**,** 190-202.

ALLIS, C. D., JENUWEIN, T. & REINBERG, D. 2006. *Epigenetics,* Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

ANDRULIS, E. D., WERNER, J., NAZARIAN, A., ERDJUMENT-BROMAGE, H., TEMPST, P. & LIS, J. T. 2002. The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. *Nature,* 420**,** 837-41.

AOUFOUCHI, S., FAILI, A., ZOBER, C., D'ORLANDO, O., WELLER, S., WEILL, J. C. & REYNAUD, C. A. 2008. Proteasomal degradation restricts the nuclear lifespan of AID. *J Exp Med,* 205**,** 1357-68.

APEL, T. W., MAUTNER, J., POLACK, A., BORNKAMM, G. W. & EICK, D. 1992. Two antisense promoters in the immunoglobulin mu-switch region drive expression of c-myc in the Burkitt's lymphoma cell line BL67. *Oncogene,* 7**,** 1267-71.

ARAKAWA, H. & BUERSTEDDE, J. M. 2004. Immunoglobulin gene conversion: insights from bursal B cells and the DT40 cell line. *Dev Dyn,* 229**,** 458-64.

ARAKAWA, H., HAUSCHILD, J. & BUERSTEDDE, J. M. 2002. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science,* 295**,** 1301-6.

ARAKAWA, H., MOLDOVAN, G. L., SARIBASAK, H., SARIBASAK, N. N., JENTSCH, S. & BUERSTEDDE, J. M. 2006. A role for PCNA ubiquitination in immunoglobulin hypermutation. *PLoS Biol,* 4**,** e366.

AYGUN, O., SVEJSTRUP, J. & LIU, Y. 2008. A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. *Proc Natl Acad Sci U S A,* 105**,** 8580-4.

BAI, X., KIM, J., YANG, Z., JURYNEC, M. J., AKIE, T. E., LEE, J., LEBLANC, J., SESSA, A., JIANG, H., DIBIASE, A., ZHOU, Y., GRUNWALD, D. J., LIN, S., CANTOR, A. B., ORKIN, S. H. & ZON, L. I. 2010. TIF1gamma controls erythroid cell fate by regulating transcription elongation. *Cell,* 142**,** 133-43.

BAMBARA, R. A. & JESSEE, C. B. 1991. Properties of DNA polymerases delta and epsilon, and their roles in eukaryotic DNA replication. *Biochim Biophys Acta,* 1088**,** 11-24.

BARDWELL, P. D., WOO, C. J., WEI, K., LI, Z., MARTIN, A., SACK, S. Z., PARRIS, T., EDELMANN, W. & SCHARFF, M. D. 2004. Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat Immunol,* 5**,** 224-9.

BARNES, D. E. & LINDAHL, T. 2004. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu Rev Genet,* 38**,** 445-76.

BARRETO, V., REINA-SAN-MARTIN, B., RAMIRO, A. R., MCBRIDE, K. M. & NUSSENZWEIG, M. C. 2003. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell,* 12**,** 501-8.

BASU, U., CHAUDHURI, J., ALPERT, C., DUTT, S., RANGANATH, S., LI, G., SCHRUM, J. P., MANIS, J. P. & ALT, F. W. 2005. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature,* 438**,** 508-11.

BASU, U., MENG, F. L., KEIM, C., GRINSTEIN, V., PEFANIS, E., ECCLESTON, J., ZHANG, T., MYERS, D., WASSERMAN, C. R., WESEMANN, D. R., JANUSZYK, K., GREGORY, R. I., DENG, H., LIMA, C. D. & ALT, F. W. 2011. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell,* 144**,** 353-63.

BEALE, R. C., PETERSEN-MAHRT, S. K., WATT, I. N., HARRIS, R. S., RADA, C. & NEUBERGER, M. S. 2004. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol,* 337**,** 585-96.

BEN-NERIAH, Y. & KARIN, M. 2011. Inflammation meets cancer, with NF-kappaB as the matchmaker. *Nat Immunol,* 12**,** 715-23.

BESMER, E., MARKET, E. & PAPAVASILIOU, F. N. 2006. The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. *Mol Cell Biol,* 26**,** 4378-85.

BETZ, A. G., MILSTEIN, C., GONZALEZ-FERNANDEZ, A., PANNELL, R., LARSON, T. & NEUBERGER, M. S. 1994. Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. *Cell,* 77**,** 239-48.

BHUTANI, N., BRADY, J. J., DAMIAN, M., SACCO, A., CORBEL, S. Y. & BLAU, H. M. 2010. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature,* 463**,** 1042-7.

BIENKO, M., GREEN, C. M., CROSETTO, N., RUDOLF, F., ZAPART, G., COULL, B., KANNOUCHE, P., WIDER, G., PETER, M., LEHMANN, A. R., HOFMANN, K. & DIKIC, I. 2005. Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science,* 310**,** 1821-4.

BLAGODATSKI, A., BATRAK, V., SCHMIDL, S., SCHOETZ, U., CALDWELL, R. B., ARAKAWA, H. & BUERSTEDDE, J. M. 2009. A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation. *PLoS Genet,* 5**,** e1000332.

BOUAYADI, K., VAN DER LEER-VAN HOFFEN, A., BALAJEE, A. S., NATARAJAN, A. T., VAN ZEELAND, A. A. & MULLENDERS, L. H. 1997. Enzymatic activities involved in the DNA resynthesis step of nucleotide excision repair are firmly attached to chromatin. *Nucleic Acids Res,* 25**,** 1056-63.

BRANSTEITTER, R., PHAM, P., SCHARFF, M. D. & GOODMAN, M. F. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-

stranded DNA but requires the action of RNase. *Proc Natl Acad Sci U S A,* 100**,** 4102-7.
BRANZEI, D., VANOLI, F. & FOIANI, M. 2008. SUMOylation regulates Rad18-mediated template switch. *Nature,* 456**,** 915-20.
BRAR, S. S., WATSON, M. & DIAZ, M. 2004. Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *J Biol Chem,* 279**,** 26395-401.
BROOMFIELD, S., HRYCIW, T. & XIAO, W. 2001. DNA postreplication repair and mutagenesis in Saccharomyces cerevisiae. *Mutat Res,* 486**,** 167-84.
BRUNING, J. B. & SHAMOO, Y. 2004. Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1. *Structure,* 12**,** 2209-19.
BURATOWSKI, S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell,* 36**,** 541-6.
BURGERS, P. M. & YODER, B. L. 1993. ATP-independent loading of the proliferating cell nuclear antigen requires DNA ends. *J Biol Chem,* 268**,** 19923-6.
BURNET, F. M. 1957. Clonal Selection Theory: A modification of Jerne's theory of antibody production using the concept of clonal selection. *Aust. J. Sci,.* 20, 67–69
BUSSLINGER, M., KLIX, N., PFEFFER, P., GRANINGER, P. G. & KOZMIK, Z. 1996. Deregulation of PAX-5 by translocation of the Emu enhancer of the IgH locus adjacent to two alternative PAX-5 promoters in a diffuse large-cell lymphoma. *Proc Natl Acad Sci U S A,* 93**,** 6129-34.
BUTLER, J. E. 1998. Immunoglobulin diversity, B-cell and antibody repertoire development in large farm animals. *Rev Sci Tech,* 17**,** 43-70.
CANUGOVI, C., SAMARANAYAKE, M. & BHAGWAT, A. S. 2009. Transcriptional pausing and stalling causes multiple clustered mutations by human activation-induced deaminase. *FASEB J,* 23**,** 34-44.
CASCALHO, M., WONG, J., STEINBERG, C. & WABL, M. 1998. Mismatch repair co-opted by hypermutation. *Science,* 279**,** 1207-10.
CASELLAS, R., NUSSENZWEIG, A., WUERFFEL, R., PELANDA, R., REICHLIN, A., SUH, H., QIN, X. F., BESMER, E., KENTER, A., RAJEWSKY, K. & NUSSENZWEIG, M. C. 1998. Ku80 is required for immunoglobulin isotype switching. *EMBO J,* 17**,** 2404-11.
CATTORETTI, G., BUTTNER, M., SHAKNOVICH, R., KREMMER, E., ALOBEID, B. & NIEDOBITEK, G. 2006. Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood,* 107**,** 3967-75.
CHANG, D. J. & CIMPRICH, K. A. 2009. DNA damage tolerance: when it's OK to make mistakes. *Nat Chem Biol,* 5**,** 82-90.
CHATTERJI, M., UNNIRAMAN, S., MCBRIDE, K. M. & SCHATZ, D. G. 2007. Role of activation-induced deaminase protein kinase A phosphorylation sites in Ig gene conversion and somatic hypermutation. *J Immunol,* 179**,** 5274-80.
CHAUDHURI, J. & ALT, F. W. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol,* 4**,** 541-52.
CHAUDHURI, J., KHUONG, C. & ALT, F. W. 2004. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature,* 430**,** 992-8.

CHAUDHURI, J., TIAN, M., KHUONG, C., CHUA, K., PINAUD, E. & ALT, F. W. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature,* 422**,** 726-30.

CHELICO, L., PHAM, P., CALABRESE, P. & GOODMAN, M. F. 2006. APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. *Nat Struct Mol Biol,* 13**,** 392-9.

CHEN, K. M., HARJES, E., GROSS, P. J., FAHMY, A., LU, Y., SHINDO, K., HARRIS, R. S. & MATSUO, H. 2008. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature,* 452**,** 116-9.

CHEN, Y., YAMAGUCHI, Y., TSUGENO, Y., YAMAMOTO, J., YAMADA, T., NAKAMURA, M., HISATAKE, K. & HANDA, H. 2009. DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev,* 23**,** 2765-77.

CHIARLE, R., ZHANG, Y., FROCK, R. L., LEWIS, S. M., MOLINIE, B., HO, Y. J., MYERS, D. R., CHOI, V. W., COMPAGNO, M., MALKIN, D. J., NEUBERG, D., MONTI, S., GIALLOURAKIS, C. C., GOSTISSA, M. & ALT, F. W. 2011. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell,* 147**,** 107-19.

CHIU, Y. L. & GREENE, W. C. 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol,* 26**,** 317-53.

CHOWDHURY, M., FOROUHI, O., DAYAL, S., MCCLOSKEY, N., GOULD, H. J., FELSENFELD, G. & FEAR, D. J. 2008. Analysis of intergenic transcription and histone modification across the human immunoglobulin heavy-chain locus. *Proc Natl Acad Sci U S A,* 105**,** 15872-7.

CHUANG, L. S., IAN, H. I., KOH, T. W., NG, H. H., XU, G. & LI, B. F. 1997. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science,* 277**,** 1996-2000.

COBALEDA, C. & BUSSLINGER, M. 2008. Developmental plasticity of lymphocytes. *Curr Opin Immunol,* 20**,** 139-48.

COKER, H. A., MORGAN, H. D. & PETERSEN-MAHRT, S. K. 2006. Genetic and in vitro assays of DNA deamination. *Methods Enzymol,* 408**,** 156-70.

COKER, H. A. & PETERSEN-MAHRT, S. K. 2007. The nuclear DNA deaminase AID functions distributively whereas cytoplasmic APOBEC3G has a processive mode of action. *DNA Repair (Amst),* 6**,** 235-43.

CONLON, T. M. & MEYER, K. B. 2006. The chicken Ig light chain 3'-enhancer is essential for gene expression and regulates gene conversion via the transcription factor E2A. *Eur J Immunol,* 36**,** 139-48.

CONTICELLO, S. G., GANESH, K., XUE, K., LU, M., RADA, C. & NEUBERGER, M. S. 2008. Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. *Mol Cell,* 31**,** 474-84.

CONTICELLO, S. G., THOMAS, C. J., PETERSEN-MAHRT, S. K. & NEUBERGER, M. S. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol,* 22**,** 367-77.

CREWS, S., GRIFFIN, J., HUANG, H., CALAME, K. & HOOD, L. 1981. A single VH gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of the antibody. *Cell,* 25**,** 59-66.

CUNNINGHAM-RUNDLES, C. & PONDA, P. P. 2005. Molecular defects in T- and B-cell primary immunodeficiency diseases. *Nat Rev Immunol,* 5**,** 880-92.

DAVIES, A. A., HUTTNER, D., DAIGAKU, Y., CHEN, S. & ULRICH, H. D. 2008. Activation of ubiquitin-dependent DNA damage bypass is mediated by replication protein a. *Mol Cell,* 29**,** 625-36.

DE POOTER, R. F. & KEE, B. L. 2010. E proteins and the regulation of early lymphocyte development. *Immunol Rev,* 238**,** 93-109.

DE YEBENES, V. G., BELVER, L., PISANO, D. G., GONZALEZ, S., VILLASANTE, A., CROCE, C., HE, L. & RAMIRO, A. R. 2008. miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J Exp Med,* 205**,** 2199-206.

DEDEOGLU, F., HORWITZ, B., CHAUDHURI, J., ALT, F. W. & GEHA, R. S. 2004. Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. *Int Immunol,* 16**,** 395-404.

DELBOS, F., AOUFOUCHI, S., FAILI, A., WEILL, J. C. & REYNAUD, C. A. 2007. DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J Exp Med,* 204**,** 17-23.

DELBOS, F., DE SMET, A., FAILI, A., AOUFOUCHI, S., WEILL, J. C. & REYNAUD, C. A. 2005. Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. *J Exp Med,* 201**,** 1191-6.

DEMPLE, B. & SUNG, J. S. 2005. Molecular and biological roles of Ape1 protein in mammalian base excision repair. *DNA Repair (Amst),* 4**,** 1442-9.

DENEPOUX, S., RAZANAJAONA, D., BLANCHARD, D., MEFFRE, G., CAPRA, J. D., BANCHEREAU, J. & LEBECQUE, S. 1997. Induction of somatic mutation in a human B cell line in vitro. *Immunity,* 6**,** 35-46.

DI NOIA, J. M. & NEUBERGER, M. S. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem,* 76**,** 1-22.

DICKERSON, S. K., MARKET, E., BESMER, E. & PAPAVASILIOU, F. N. 2003. AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med,* 197**,** 1291-6.

DORSETT, Y., MCBRIDE, K. M., JANKOVIC, M., GAZUMYAN, A., THAI, T. H., ROBBIANI, D. F., DI VIRGILIO, M., SAN-MARTIN, B. R., HEIDKAMP, G., SCHWICKERT, T. A., EISENREICH, T., RAJEWSKY, K. & NUSSENZWEIG, M. C. 2008. MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity,* 28**,** 630-8.

DUCOUX, M., URBACH, S., BALDACCI, G., HUBSCHER, U., KOUNDRIOUKOFF, S., CHRISTENSEN, J. & HUGHES, P. 2001. Mediation of proliferating cell nuclear antigen (PCNA)-dependent DNA replication through a conserved p21(Cip1)-like PCNA-binding motif present in the third subunit of human DNA polymerase delta. *J Biol Chem,* 276**,** 49258-66.

DUNNICK, W., HERTZ, G. Z., SCAPPINO, L. & GRITZMACHER, C. 1993. DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res,* 21**,** 365-72.

DUNNICK, W., WILSON, M. & STAVNEZER, J. 1989. Mutations, duplication, and deletion of recombined switch regions suggest a role for DNA replication in the immunoglobulin heavy-chain switch. *Mol Cell Biol,* 9**,** 1850-6.

DUQUETTE, M. L., HANDA, P., VINCENT, J. A., TAYLOR, A. F. & MAIZELS, N. 2004. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev,* 18**,** 1618-29.

EDMUNDS, C. E., SIMPSON, L. J. & SALE, J. E. 2008. PCNA ubiquitination and REV1 define temporally distinct mechanisms for controlling translesion synthesis in the avian cell line DT40. *Molecular cell,* 30**,** 519-29.

EHRENSTEIN, M. R., RADA, C., JONES, A. M., MILSTEIN, C. & NEUBERGER, M. S. 2001. Switch junction sequences in PMS2-deficient mice reveal a microhomology-mediated mechanism of Ig class switch recombination. *Proc Natl Acad Sci U S A,* 98**,** 14553-8.

EHRLICH, M., GAMA-SOSA, M. A., HUANG, L. H., MIDGETT, R. M., KUO, K. C., MCCUNE, R. A. & GEHRKE, C. 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res,* 10**,** 2709-21.

EICHINGER, C. S. & JENTSCH, S. 2011. 9-1-1: PCNA's specialized cousin. *Trends Biochem Sci*.

FAILI, A., AOUFOUCHI, S., GUERANGER, Q., ZOBER, C., LEON, A., BERTOCCI, B., WEILL, J. C. & REYNAUD, C. A. 2002. AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat Immunol,* 3**,** 815-21.

FEINBERG, A. P. & VOGELSTEIN, B. 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature,* 301**,** 89-92.

FRANKLIN, A. & BLANDEN, R. V. 2008. The strand bias paradox of somatic hypermutation at immunoglobulin loci. *Trends Immunol,* 29**,** 167-72.

FREY, S., BERTOCCI, B., DELBOS, F., QUINT, L., WEILL, J. C. & REYNAUD, C. A. 1998. Mismatch repair deficiency interferes with the accumulation of mutations in chronically stimulated B cells and not with the hypermutation process. *Immunity,* 9**,** 127-34.

FRITZ, E. L. & PAPAVASILIOU, F. N. 2010. Cytidine deaminases: AIDing DNA demethylation? *Genes Dev,* 24**,** 2107-14.

FUJINAGA, K., IRWIN, D., HUANG, Y., TAUBE, R., KUROSU, T. & PETERLIN, B. M. 2004. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol,* 24**,** 787-95.

GARG, P. & BURGERS, P. M. 2005. DNA polymerases that propagate the eukaryotic DNA replication fork. *Critical reviews in biochemistry and molecular biology,* 40**,** 115-28.

GARNER, E. & COSTANZO, V. 2009. Studying the DNA damage response using in vitro model systems. *DNA Repair (Amst),* 8**,** 1025-37.

GEHRING, M., REIK, W. & HENIKOFF, S. 2009. DNA demethylation by DNA repair. *Trends in genetics : TIG,* 25**,** 82-90.

GILMOUR, D. S. & FAN, R. 2009. Detecting transcriptionally engaged RNA polymerase in eukaryotic cells with permanganate genomic footprinting. *Methods,* 48**,** 368-74.

GOMES, X. V. & BURGERS, P. M. 2000. Two modes of FEN1 binding to PCNA regulated by DNA. *EMBO J,* 19**,** 3811-21.

GOMEZ-GONZALEZ, B. & AGUILERA, A. 2007. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc Natl Acad Sci U S A,* 104**,** 8409-14.

GONDA, H., SUGAI, M., NAMBU, Y., KATAKAI, T., AGATA, Y., MORI, K. J., YOKOTA, Y. & SHIMIZU, A. 2003. The balance between Pax5 and Id2 activities is the key to AID gene expression. *J Exp Med,* 198**,** 1427-37.

GOOSSENS, T., KLEIN, U. & KUPPERS, R. 1998. Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. *Proc Natl Acad Sci U S A,* 95**,** 2463-8.

GORDON, M. S., KANEGAI, C. M., DOERR, J. R. & WALL, R. 2003. Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). *Proc Natl Acad Sci U S A,* 100**,** 4126-31.

GOYENECHEA, B., KLIX, N., YELAMOS, J., WILLIAMS, G. T., RIDDELL, A., NEUBERGER, M. S. & MILSTEIN, C. 1997. Cells strongly expressing Ig(kappa) transgenes show clonal recruitment of hypermutation: a role for both MAR and the enhancers. *EMBO J,* 16**,** 3987-94.

GREEN, C. M. & ALMOUZNI, G. 2003. Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo. *EMBO J,* 22**,** 5163-74.

GREINER, A., TOBOLLIK, S., BUETTNER, M., JUNGNICKEL, B., HERRMANN, K., KREMMER, E. & NIEDOBITEK, G. 2005. Differential expression of activation-induced cytidine deaminase (AID) in nodular lymphocyte-predominant and classical Hodgkin lymphoma. *J Pathol,* 205**,** 541-7.

GUO, C., TANG, T. S., BIENKO, M., DIKIC, I. & FRIEDBERG, E. C. 2008. Requirements for the interaction of mouse Polkappa with ubiquitin and its biological significance. *J Biol Chem,* 283**,** 4658-64.

GUO, C., TANG, T. S., BIENKO, M., PARKER, J. L., BIELEN, A. B., SONODA, E., TAKEDA, S., ULRICH, H. D., DIKIC, I. & FRIEDBERG, E. C. 2006. Ubiquitin-binding motifs in REV1 protein are required for its role in the tolerance of DNA damage. *Mol Cell Biol,* 26**,** 8892-900.

HA, K., LEE, G. E., PALII, S. S., BROWN, K. D., TAKEDA, Y., LIU, K., BHALLA, K. N. & ROBERTSON, K. D. 2011. Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery. *Human molecular genetics,* 20**,** 126-40.

HARACSKA, L., UNK, I., PRAKASH, L. & PRAKASH, S. 2006. Ubiquitylation of yeast proliferating cell nuclear antigen and its implications for translesion DNA synthesis. *Proc Natl Acad Sci U S A,* 103**,** 6477-82.

HARRIS, R. S., BISHOP, K. N., SHEEHY, A. M., CRAIG, H. M., PETERSEN-MAHRT, S. K., WATT, I. N., NEUBERGER, M. S. & MALIM, M. H. 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell,* 113**,** 803-9.

HARRIS, R. S., CROOM-CARTER, D. S., RICKINSON, A. B. & NEUBERGER, M. S. 2001. Epstein-Barr virus and the somatic hypermutation of immunoglobulin genes in Burkitt's lymphoma cells. *J Virol,* 75**,** 10488-92.

HARRIS, R. S., SALE, J. E., PETERSEN-MAHRT, S. K. & NEUBERGER, M. S. 2002. AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Curr Biol,* 12**,** 435-8.

HARRISON, J. C. & HABER, J. E. 2006. Surviving the breakup: the DNA damage checkpoint. *Annu Rev Genet,* 40**,** 209-35.

HASLER, J., RADA, C. & NEUBERGER, M. S. 2011. Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1alpha (eEF1A). *Proc Natl Acad Sci U S A*.

HATANAKA, A., YAMAZOE, M., SALE, J. E., TAKATA, M., YAMAMOTO, K., KITAO, H., SONODA, E., KIKUCHI, K., YONETANI, Y. & TAKEDA, S. 2005. Similar effects of Brca2 truncation and Rad51 paralog deficiency on immunoglobulin V gene diversification in DT40 cells support an early role for Rad51 paralogs in homologous recombination. *Mol Cell Biol,* 25**,** 1124-34.

HENDEL, A., KRIJGER, P. H., DIAMANT, N., GOREN, Z., LANGERAK, P., KIM, J., REISSNER, T., LEE, K. Y., GEACINTOV, N. E., CARELL, T., MYUNG, K., TATEISHI, S., D'ANDREA, A., JACOBS, H. & LIVNEH, Z. 2011. PCNA Ubiquitination Is Important, But Not Essential for Translesion DNA Synthesis in Mammalian Cells. *PLoS Genet,* 7**,** e1002262.

HOEGE, C., PFANDER, B., MOLDOVAN, G. L., PYROWOLAKIS, G. & JENTSCH, S. 2002. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature,* 419**,** 135-41.

HOLDEN, L. G., PROCHNOW, C., CHANG, Y. P., BRANSTEITTER, R., CHELICO, L., SEN, U., STEVENS, R. C., GOODMAN, M. F. & CHEN, X. S. 2008. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature,* 456**,** 121-4.

HOLMES, R. K., MALIM, M. H. & BISHOP, K. N. 2007. APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci,* 32**,** 118-28.

HUANG, F. T., YU, K., BALTER, B. B., SELSING, E., ORUC, Z., KHAMLICHI, A. A., HSIEH, C. L. & LIEBER, M. R. 2007. Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. *Mol Cell Biol,* 27**,** 5921-32.

HUBSCHER, U., MAGA, G. & SPADARI, S. 2002. Eukaryotic DNA polymerases. *Annual review of biochemistry,* 71**,** 133-63.

IAMS, K., LARSON, E. D. & DRUMMOND, J. T. 2002. DNA template requirements for human mismatch repair in vitro. *J Biol Chem,* 277**,** 30805-14.

IIDA, S., RAO, P. H., NALLASIVAM, P., HIBSHOOSH, H., BUTLER, M., LOUIE, D. C., DYOMIN, V., OHNO, H., CHAGANTI, R. S. & DALLA-FAVERA, R. 1996. The t(9;14)(p13;q32) chromosomal translocation associated with lymphoplasmacytoid lymphoma involves the PAX-5 gene. *Blood,* 88**,** 4110-7.

IIDA, T., SUETAKE, I., TAJIMA, S., MORIOKA, H., OHTA, S., OBUSE, C. & TSURIMOTO, T. 2002. PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. *Genes to cells : devoted to molecular & cellular mechanisms,* 7**,** 997-1007.

INLAY, M. A., GAO, H. H., ODEGARD, V. H., LIN, T., SCHATZ, D. G. & XU, Y. 2006. Roles of the Ig kappa light chain intronic and 3' enhancers in Igk somatic hypermutation. *J Immunol,* 177**,** 1146-51.

ITO, S., NAGAOKA, H., SHINKURA, R., BEGUM, N., MURAMATSU, M., NAKATA, M. & HONJO, T. 2004. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci U S A,* 101**,** 1975-80.

JAEHNING, J. A. 2010. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta,* 1799**,** 379-88.

JANEWAY, C. & NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (U.S.) 2001. Immunobiology the immune system in health and disease. *NCBI bookshelf.* 5th ed. New York: Garland Pub.

JANSEN, J. G., FOUSTERI, M. I. & DE WIND, N. 2007. Send in the clamps: control of DNA translesion synthesis in eukaryotes. *Mol Cell,* 28**,** 522-9.

JEEVAN-RAJ, B. P., ROBERT, I., HEYER, V., PAGE, A., WANG, J. H., CAMMAS, F., ALT, F. W., LOSSON, R. & REINA-SAN-MARTIN, B. 2011. Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. *J Exp Med,* 208**,** 1649-60.

JIRICNY, J. 2006. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol,* 7**,** 335-46.

JONSSON, Z. O. & HUBSCHER, U. 1997. Proliferating cell nuclear antigen: more than a clamp for DNA polymerases. *BioEssays : news and reviews in molecular, cellular and developmental biology,* 19**,** 967-75.

KANNOUCHE, P. L., WING, J. & LEHMANN, A. R. 2004. Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Mol Cell,* 14**,** 491-500.

KELLER, B. O., WANG, Z. & LI, L. 2002. Low-mass proteome analysis based on liquid chromatography fractionation, nanoliter protein concentration/digestion, and microspot matrix-assisted laser desorption ionization mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci,* 782**,** 317-29.

KIM, J., GUERMAH, M., MCGINTY, R. K., LEE, J. S., TANG, Z., MILNE, T. A., SHILATIFARD, A., MUIR, T. W. & ROEDER, R. G. 2009. RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell,* 137**,** 459-71.

KIM, J., GUERMAH, M. & ROEDER, R. G. 2010. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell,* 140**,** 491-503.

KIM, S., DAVIS, M., SINN, E., PATTEN, P. & HOOD, L. 1981. Antibody diversity: somatic hypermutation of rearranged VH genes. *Cell,* 27**,** 573-81.

KIM, S., HUMPHRIES, E. H., TJOELKER, L., CARLSON, L. & THOMPSON, C. B. 1990. Ongoing diversification of the rearranged immunoglobulin light-chain gene in a bursal lymphoma cell line. *Mol Cell Biol,* 10**,** 3224-31.

KIM, T. H., BARRERA, L. O., ZHENG, M., QU, C., SINGER, M. A., RICHMOND, T. A., WU, Y., GREEN, R. D. & REN, B. 2005. A high-resolution map of active promoters in the human genome. *Nature,* 436**,** 876-80.

KIM, Y. & TIAN, M. 2009. NF-kappaB family of transcription factor facilitates gene conversion in chicken B cells. *Mol Immunol,* 46**,** 3283-91.

KLEIN, I. A., RESCH, W., JANKOVIC, M., OLIVEIRA, T., YAMANE, A., NAKAHASHI, H., DI VIRGILIO, M., BOTHMER, A., NUSSENZWEIG, A., ROBBIANI, D. F., CASELLAS, R. & NUSSENZWEIG, M. C. 2011. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell,* 147**,** 95-106.

KLEIN, U. & DALLA-FAVERA, R. 2008. Germinal centres: role in B-cell physiology and malignancy. *Nat Rev Immunol,* 8**,** 22-33.

KLEMM, L., DUY, C., IACOBUCCI, I., KUCHEN, S., VON LEVETZOW, G., FELDHAHN, N., HENKE, N., LI, Z., HOFFMANN, T. K., KIM, Y. M., HOFMANN, W. K., JUMAA, H., GROFFEN, J., HEISTERKAMP, N., MARTINELLI, G., LIEBER, M. R., CASELLAS, R. & MUSCHEN, M. 2009. The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer Cell,* 16**,** 232-45.

KLIX, N., JOLLY, C. J., DAVIES, S. L., BRUGGEMANN, M., WILLIAMS, G. T. & NEUBERGER, M. S. 1998. Multiple sequences from downstream of the J kappa cluster can combine to recruit somatic hypermutation to a heterologous, upstream mutation domain. *Eur J Immunol,* 28**,** 317-26.

KLOTZ, E. L., HACKETT, J., JR. & STORB, U. 1998. Somatic hypermutation of an artificial test substrate within an Ig kappa transgene. *J Immunol,* 161**,** 782-90.

KOTHAPALLI, N. R., NORTON, D. D. & FUGMANN, S. D. 2011. Classical Mus musculus Igkappa enhancers support transcription but not high level somatic hypermutation from a V-lambda promoter in chicken DT40 cells. *PLoS One,* 6**,** e18955.

KRISHNA, T. S., KONG, X. P., GARY, S., BURGERS, P. M. & KURIYAN, J. 1994. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell,* 79**,** 1233-43.

KROGAN, N. J., KIM, M., TONG, A., GOLSHANI, A., CAGNEY, G., CANADIEN, V., RICHARDS, D. P., BEATTIE, B. K., EMILI, A., BOONE, C., SHILATIFARD, A., BURATOWSKI, S. & GREENBLATT, J. 2003. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol,* 23**,** 4207-18.

KRUMM, A., MEULIA, T., BRUNVAND, M. & GROUDINE, M. 1992. The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. *Genes Dev,* 6**,** 2201-13.

KUANG, F. L., LUO, Z. & SCHARFF, M. D. 2009. H3 trimethyl K9 and H3 acetyl K9 chromatin modifications are associated with class switch recombination. *Proc Natl Acad Sci U S A,* 106**,** 5288-93.

KUNKEL, T. A. & ERIE, D. A. 2005. DNA mismatch repair. *Annu Rev Biochem,* 74**,** 681-710.

KUPPERS, R. & DALLA-FAVERA, R. 2001. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene,* 20**,** 5580-94.

LANGERAK, P., KRIJGER, P. H., HEIDEMAN, M. R., VAN DEN BERK, P. C. & JACOBS, H. 2009. Somatic hypermutation of immunoglobulin genes: lessons

from proliferating cell nuclear antigenK164R mutant mice. *Philos Trans R Soc Lond B Biol Sci,* 364**,** 621-9.

LANGERAK, P., NYGREN, A. O., KRIJGER, P. H., VAN DEN BERK, P. C. & JACOBS, H. 2007. A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J Exp Med,* 204**,** 1989-98.

LARIJANI, M., FRIEDER, D., BASIT, W. & MARTIN, A. 2005. The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in ung(-/-)msh2(-/-) mice. *Immunogenetics,* 56**,** 840-5.

LEACH, C. A. & MICHAEL, W. M. 2005. Ubiquitin/SUMO modification of PCNA promotes replication fork progression in Xenopus laevis egg extracts. *J Cell Biol,* 171**,** 947-54.

LEE, C. H., MELCHERS, M., WANG, H., TORREY, T. A., SLOTA, R., QI, C. F., KIM, J. Y., LUGAR, P., KONG, H. J., FARRINGTON, L., VAN DER ZOUWEN, B., ZHOU, J. X., LOUGARIS, V., LIPSKY, P. E., GRAMMER, A. C. & MORSE, H. C., 3RD 2006. Regulation of the germinal center gene program by interferon (IFN) regulatory factor 8/IFN consensus sequence-binding protein. *J Exp Med,* 203**,** 63-72.

LEHMANN, A. R. & FUCHS, R. P. 2006. Gaps and forks in DNA replication: Rediscovering old models. *DNA repair,* 5**,** 1495-8.

LI, X., LI, J., HARRINGTON, J., LIEBER, M. R. & BURGERS, P. M. 1995. Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. *J Biol Chem,* 270**,** 22109-12.

LIAO, W., HONG, S. H., CHAN, B. H., RUDOLPH, F. B., CLARK, S. C. & CHAN, L. 1999. APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochem Biophys Res Commun,* 260**,** 398-404.

LINDAHL, T. 1993. Instability and decay of the primary structure of DNA. *Nature,* 362**,** 709-15.

LINDAHL, T. & WOOD, R. D. 1999. Quality control by DNA repair. *Science,* 286**,** 1897-905.

LIS, J. & WU, C. 1993. Protein traffic on the heat shock promoter: parking, stalling, and trucking along. *Cell,* 74**,** 1-4.

LIS, J. T. 2007. Imaging Drosophila gene activation and polymerase pausing in vivo. *Nature,* 450**,** 198-202.

LIU, M., DUKE, J. L., RICHTER, D. J., VINUESA, C. G., GOODNOW, C. C., KLEINSTEIN, S. H. & SCHATZ, D. G. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature,* 451**,** 841-5.

LIU, M. & SCHATZ, D. G. 2009. Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol,* 30**,** 173-181.

LONGERICH, S., TANAKA, A., BOZEK, G., NICOLAE, D. & STORB, U. 2005. The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. *J Exp Med,* 202**,** 1443-54.

LUMSDEN, J. M., MCCARTY, T., PETINIOT, L. K., SHEN, R., BARLOW, C., WYNN, T. A., MORSE, H. C., 3RD, GEARHART, P. J., WYNSHAW-BORIS, A., MAX, E. E. & HODES, R. J. 2004. Immunoglobulin class switch recombination is impaired in Atm-deficient mice. *J Exp Med,* 200**,** 1111-21.

LYST, M. J., NAN, X. & STANCHEVA, I. 2006. Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. *EMBO J,* 25**,** 5317-28.
MACDUFF, D. A., NEUBERGER, M. S. & HARRIS, R. S. 2006. MDM2 can interact with the C-terminus of AID but it is inessential for antibody diversification in DT40 B cells. *Mol Immunol,* 43**,** 1099-108.
MAEDA, K., SINGH, S. K., EDA, K., KITABATAKE, M., PHAM, P., GOODMAN, M. F. & SAKAGUCHI, N. 2010. GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. *J Biol Chem,* 285**,** 23945-53.
MAJKA, J. & BURGERS, P. M. 2004. The PCNA-RFC families of DNA clamps and clamp loaders. *Progress in nucleic acid research and molecular biology,* 78**,** 227-60.
MANIS, J. P., MORALES, J. C., XIA, Z., KUTOK, J. L., ALT, F. W. & CARPENTER, P. B. 2004. 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. *Nat Immunol,* 5**,** 481-7.
MARTIN, A., BARDWELL, P. D., WOO, C. J., FAN, M., SHULMAN, M. J. & SCHARFF, M. D. 2002. Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature,* 415**,** 802-6.
MARTOMO, S. A., YANG, W. W. & GEARHART, P. J. 2004. A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. *J Exp Med,* 200**,** 61-8.
MARUSAWA, H., TAKAI, A. & CHIBA, T. 2011. Role of activation-induced cytidine deaminase in inflammation-associated cancer development. *Adv Immunol,* 111**,** 109-41.
MATSUMOTO, Y., MARUSAWA, H., KINOSHITA, K., ENDO, Y., KOU, T., MORISAWA, T., AZUMA, T., OKAZAKI, I. M., HONJO, T. & CHIBA, T. 2007. Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med,* 13**,** 470-6.
MAYER, A., LIDSCHREIBER, M., SIEBERT, M., LEIKE, K., SODING, J. & CRAMER, P. 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol,* 17**,** 1272-8.
MCBRIDE, K. M., BARRETO, V., RAMIRO, A. R., STAVROPOULOS, P. & NUSSENZWEIG, M. C. 2004. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J Exp Med,* 199**,** 1235-44.
MCBRIDE, K. M., GAZUMYAN, A., WOO, E. M., SCHWICKERT, T. A., CHAIT, B. T. & NUSSENZWEIG, M. C. 2008. Regulation of class switch recombination and somatic mutation by AID phosphorylation. *J Exp Med,* 205**,** 2585-94.
MCCORMACK, W. T. & THOMPSON, C. B. 1990. Chicken IgL variable region gene conversions display pseudogene donor preference and 5' to 3' polarity. *Genes Dev,* 4**,** 548-58.
MCKINNON, P. J. & CALDECOTT, K. W. 2007. DNA strand break repair and human genetic disease. *Annu Rev Genomics Hum Genet,* 8**,** 37-55.
MEDVEDOVIC, J., EBERT, A., TAGOH, H. & BUSSLINGER, M. 2011. Pax5 a master regulator of B cell development and leukemogenesis. *Adv Immunol,* 111**,** 179-206.

MELLO, J. A. & ALMOUZNI, G. 2001. The ins and outs of nucleosome assembly. *Curr Opin Genet Dev,* 11**,** 136-41.

MENCIA, M. & DE LORENZO, V. 2004. Functional transplantation of the sumoylation machinery into Escherichia coli. *Protein Expr Purif,* 37**,** 409-18.

METZNER, M., SCHUH, W., ROTH, E., JACK, H. M. & WABL, M. 2010. Two forms of activation-induced cytidine deaminase differing in their ability to bind agarose. *PloS one,* 5**,** e8883.

MICHAEL, N., SHEN, H. M., LONGERICH, S., KIM, N., LONGACRE, A. & STORB, U. 2003. The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity,* 19**,** 235-42.

MIRCHANDANI, K. D., MCCAFFREY, R. M. & D'ANDREA, A. D. 2008. The Fanconi anemia core complex is required for efficient point mutagenesis and Rev1 foci assembly. *DNA repair,* 7**,** 902-11.

MIROUX, B. & WALKER, J. E. 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol,* 260**,** 289-98.

MIURA, M., DOMON, M., SASAKI, T. & TAKASAKI, Y. 1992. Induction of proliferating cell nuclear antigen (PCNA) complex formation in quiescent fibroblasts from a xeroderma pigmentosum patient. *J Cell Physiol,* 150**,** 370-6.

MOLDOVAN, G. L., PFANDER, B. & JENTSCH, S. 2007. PCNA, the maestro of the replication fork. *Cell,* 129**,** 665-79.

MORGAN, H. D., DEAN, W., COKER, H. A., REIK, W. & PETERSEN-MAHRT, S. K. 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem,* 279**,** 52353-60.

MORRISON, A. M., JAGER, U., CHOTT, A., SCHEBESTA, M., HAAS, O. A. & BUSSLINGER, M. 1998. Deregulated PAX-5 transcription from a translocated IgH promoter in marginal zone lymphoma. *Blood,* 92**,** 3865-78.

MOTEGI, A., LIAW, H. J., LEE, K. Y., ROEST, H. P., MAAS, A., WU, X., MOINOVA, H., MARKOWITZ, S. D., DING, H., HOEIJMAKERS, J. H. & MYUNG, K. 2008. Polyubiquitination of proliferating cell nuclear antigen by HLTF and SHPRH prevents genomic instability from stalled replication forks. *Proc Natl Acad Sci U S A,* 105**,** 12411-6.

MOTEGI, A., SOOD, R., MOINOVA, H., MARKOWITZ, S. D., LIU, P. P. & MYUNG, K. 2006. Human SHPRH suppresses genomic instability through proliferating cell nuclear antigen polyubiquitination. *J Cell Biol,* 175**,** 703-8.

MURAMATSU, M., KINOSHITA, K., FAGARASAN, S., YAMADA, S., SHINKAI, Y. & HONJO, T. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell,* 102**,** 553-63.

MURAMATSU, M., SANKARANAND, V. S., ANANT, S., SUGAI, M., KINOSHITA, K., DAVIDSON, N. O. & HONJO, T. 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem,* 274**,** 18470-6.

MUSCHEN, M., RE, D., JUNGNICKEL, B., DIEHL, V., RAJEWSKY, K. & KUPPERS, R. 2000. Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J Exp Med,* 192**,** 1833-40.

NAGAOKA, H., MURAMATSU, M., YAMAMURA, N., KINOSHITA, K. & HONJO, T. 2002. Activation-induced deaminase (AID)-directed hypermutation in the immunoglobulin Smu region: implication of AID involvement in a common step of class switch recombination and somatic hypermutation. *J Exp Med,* 195**,** 529-34.

NAKAMURA, M., KONDO, S., SUGAI, M., NAZAREA, M., IMAMURA, S. & HONJO, T. 1996. High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. *Int Immunol,* 8**,** 193-201.

NAMBU, Y., SUGAI, M., GONDA, H., LEE, C. G., KATAKAI, T., AGATA, Y., YOKOTA, Y. & SHIMIZU, A. 2003. Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science,* 302**,** 2137-40.

NAVARATNAM, N., MORRISON, J. R., BHATTACHARYA, S., PATEL, D., FUNAHASHI, T., GIANNONI, F., TENG, B. B., DAVIDSON, N. O. & SCOTT, J. 1993. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J Biol Chem,* 268**,** 20709-12.

NAVIAUX, R. K., COSTANZI, E., HAAS, M. & VERMA, I. M. 1996. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. *J Virol,* 70**,** 5701-5.

NEUBERGER, M. S., HARRIS, R. S., DI NOIA, J. & PETERSEN-MAHRT, S. K. 2003. Immunity through DNA deamination. *Trends Biochem Sci,* 28**,** 305-12.

NEWEY, P. J., BOWL, M. R. & THAKKER, R. V. 2009. Parafibromin--functional insights. *J Intern Med,* 266**,** 84-98.

NEWMAN, E. N., HOLMES, R. K., CRAIG, H. M., KLEIN, K. C., LINGAPPA, J. R., MALIM, M. H. & SHEEHY, A. M. 2005. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol,* 15**,** 166-70.

NIKOLAISHVILI-FEINBERG, N., JENKINS, G. S., NEVIS, K. R., STAUS, D. P., SCARLETT, C. O., UNSAL-KACMAZ, K., KAUFMANN, W. K. & CORDEIRO-STONE, M. 2008. Ubiquitylation of proliferating cell nuclear antigen and recruitment of human DNA polymerase eta. *Biochemistry,* 47**,** 4141-50.

NONAKA, T., DOI, T., TOYOSHIMA, T., MURAMATSU, M., HONJO, T. & KINOSHITA, K. 2009. Carboxy-terminal domain of AID required for its mRNA complex formation in vivo. *Proc Natl Acad Sci U S A,* 106**,** 2747-51.

ODEGARD, V. H., KIM, S. T., ANDERSON, S. M., SHLOMCHIK, M. J. & SCHATZ, D. G. 2005. Histone modifications associated with somatic hypermutation. *Immunity,* 23**,** 101-10.

ODEGARD, V. H. & SCHATZ, D. G. 2006. Targeting of somatic hypermutation. *Nat Rev Immunol,* 6**,** 573-83.

OH, C. E., MCMAHON, R., BENZER, S. & TANOUYE, M. A. 1994. bendless, a Drosophila gene affecting neuronal connectivity, encodes a ubiquitin-conjugating enzyme homolog. *J Neurosci,* 14**,** 3166-79.

OKAZAKI, I. M., HIAI, H., KAKAZU, N., YAMADA, S., MURAMATSU, M., KINOSHITA, K. & HONJO, T. 2003. Constitutive expression of AID leads to tumorigenesis. *J Exp Med,* 197**,** 1173-81.

OKAZAKI, I. M., KINOSHITA, K., MURAMATSU, M., YOSHIKAWA, K. & HONJO, T. 2002. The AID enzyme induces class switch recombination in fibroblasts. *Nature,* 416**,** 340-5.

OKAZAKI, I. M., OKAWA, K., KOBAYASHI, M., YOSHIKAWA, K., KAWAMOTO, S., NAGAOKA, H., SHINKURA, R., KITAWAKI, Y., TANIGUCHI, H., NATSUME, T., IEMURA, S. & HONJO, T. 2011. Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. *Proc Natl Acad Sci U S A,* 108**,** 7920-5.

OSBORNE, C. S., CHAKALOVA, L., MITCHELL, J. A., HORTON, A., WOOD, A. L., BOLLAND, D. J., CORCORAN, A. E. & FRASER, P. 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol,* 5**,** e192.

PAPOULI, E., CHEN, S., DAVIES, A. A., HUTTNER, D., KREJCI, L., SUNG, P. & ULRICH, H. D. 2005. Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. *Mol Cell,* 19**,** 123-33.

PARKER, J. L., BUCCERI, A., DAVIES, A. A., HEIDRICH, K., WINDECKER, H. & ULRICH, H. D. 2008. SUMO modification of PCNA is controlled by DNA. *EMBO J,* 27**,** 2422-31.

PASQUALUCCI, L., GUGLIELMINO, R., HOULDSWORTH, J., MOHR, J., AOUFOUCHI, S., POLAKIEWICZ, R., CHAGANTI, R. S. & DALLA-FAVERA, R. 2004. Expression of the AID protein in normal and neoplastic B cells. *Blood,* 104**,** 3318-25.

PASQUALUCCI, L., KITAURA, Y., GU, H. & DALLA-FAVERA, R. 2006. PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc Natl Acad Sci U S A,* 103**,** 395-400.

PASQUALUCCI, L., MIGLIAZZA, A., FRACCHIOLLA, N., WILLIAM, C., NERI, A., BALDINI, L., CHAGANTI, R. S., KLEIN, U., KUPPERS, R., RAJEWSKY, K. & DALLA-FAVERA, R. 1998. BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc Natl Acad Sci U S A,* 95**,** 11816-21.

PASQUALUCCI, L., NEUMEISTER, P., GOOSSENS, T., NANJANGUD, G., CHAGANTI, R. S., KUPPERS, R. & DALLA-FAVERA, R. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature,* 412**,** 341-6.

PATENAUDE, A. M., ORTHWEIN, A., HU, Y., CAMPO, V. A., KAVLI, B., BUSCHIAZZO, A. & DI NOIA, J. M. 2009. Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol*.

PAUKLIN, S. & PETERSEN-MAHRT, S. K. 2009. Progesterone inhibits activation-induced deaminase by binding to the promoter. *J Immunol,* 183**,** 1238-44.

PAUKLIN, S., SERNANDEZ, I. V., BACHMANN, G., RAMIRO, A. R. & PETERSEN-MAHRT, S. K. 2009. Estrogen directly activates AID transcription and function. *J Exp Med,* 206**,** 99-111.

PAVRI, R., GAZUMYAN, A., JANKOVIC, M., DI VIRGILIO, M., KLEIN, I., ANSARAH-SOBRINHO, C., RESCH, W., YAMANE, A., REINA SAN-MARTIN, B., BARRETO, V., NIELAND, T. J., ROOT, D. E., CASELLAS, R. & NUSSENZWEIG, M. C. 2010. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell,* 143**,** 122-33.

PAVRI, R. & NUSSENZWEIG, M. C. 2011. AID targeting in antibody diversity. *Adv Immunol,* 110**,** 1-26.

PAVRI, R., ZHU, B., LI, G., TROJER, P., MANDAL, S., SHILATIFARD, A. & REINBERG, D. 2006. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell,* 125**,** 703-17.

PEITZ, M., PFANNKUCHE, K., RAJEWSKY, K. & EDENHOFER, F. 2002. Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. *Proc Natl Acad Sci U S A,* 99**,** 4489-94.

PELED, J. U., KUANG, F. L., IGLESIAS-USSEL, M. D., ROA, S., KALIS, S. L., GOODMAN, M. F. & SCHARFF, M. D. 2008. The biochemistry of somatic hypermutation. *Annu Rev Immunol,* 26**,** 481-511.

PERKINS, D. N., PAPPIN, D. J., CREASY, D. M. & COTTRELL, J. S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis,* 20**,** 3551-67.

PERLOT, T., LI, G. & ALT, F. W. 2008. Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. *Proc Natl Acad Sci U S A,* 105**,** 3843-8.

PETERS, A. & STORB, U. 1996. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity,* 4**,** 57-65.

PETERSEN, S., CASELLAS, R., REINA-SAN-MARTIN, B., CHEN, H. T., DIFILIPPANTONIO, M. J., WILSON, P. C., HANITSCH, L., CELESTE, A., MURAMATSU, M., PILCH, D. R., REDON, C., RIED, T., BONNER, W. M., HONJO, T., NUSSENZWEIG, M. C. & NUSSENZWEIG, A. 2001. AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature,* 414**,** 660-5.

PETERSEN-MAHRT, S. 2005. DNA deamination in immunity. *Immunol Rev,* 203**,** 80-97.

PETERSEN-MAHRT, S. K., COKER, H. A. & PAUKLIN, S. 2009. DNA deaminases: AIDing hormones in immunity and cancer. *J Mol Med (Berl),* 87**,** 893-7.

PETERSEN-MAHRT, S. K., HARRIS, R. S. & NEUBERGER, M. S. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature,* 418**,** 99-103.

PETERSEN-MAHRT, S. K. & NEUBERGER, M. S. 2003. In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). *J Biol Chem,* 278**,** 19583-6.

PFANDER, B., MOLDOVAN, G. L., SACHER, M., HOEGE, C. & JENTSCH, S. 2005. SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature,* 436**,** 428-33.

PHAM, P., SMOLKA, M. B., CALABRESE, P., LANDOLPH, A., ZHANG, K., ZHOU, H. & GOODMAN, M. F. 2008. Impact of phosphorylation and phosphorylation-null mutants on the activity and deamination specificity of activation-induced cytidine deaminase. *J Biol Chem,* 283**,** 17428-39.

PHUNG, Q. H., WINTER, D. B., ALREFAI, R. & GEARHART, P. J. 1999. Hypermutation in Ig V genes from mice deficient in the MLH1 mismatch repair protein. *J Immunol,* 162**,** 3121-4.

PHUNG, Q. H., WINTER, D. B., CRANSTON, A., TARONE, R. E., BOHR, V. A., FISHEL, R. & GEARHART, P. J. 1998. Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J Exp Med,* 187**,** 1745-51.

POLTORATSKY, V., WOO, C. J., TIPPIN, B., MARTIN, A., GOODMAN, M. F. & SCHARFF, M. D. 2001. Expression of error-prone polymerases in BL2 cells activated for Ig somatic hypermutation. *Proc Natl Acad Sci U S A,* 98**,** 7976-81.

POPP, C., DEAN, W., FENG, S., COKUS, S. J., ANDREWS, S., PELLEGRINI, M., JACOBSEN, S. E. & REIK, W. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature,* 463**,** 1101-5.

PRELICH, G., TAN, C. K., KOSTURA, M., MATHEWS, M. B., SO, A. G., DOWNEY, K. M. & STILLMAN, B. 1987. Functional identity of proliferating cell nuclear antigen and a DNA polymerase-delta auxiliary protein. *Nature,* 326**,** 517-20.

RABBITTS, T. H., FORSTER, A., HAMLYN, P. & BAER, R. 1984. Effect of somatic mutation within translocated c-myc genes in Burkitt's lymphoma. *Nature,* 309**,** 592-7.

RADA, C., DI NOIA, J. M. & NEUBERGER, M. S. 2004. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell,* 16**,** 163-71.

RADA, C., EHRENSTEIN, M. R., NEUBERGER, M. S. & MILSTEIN, C. 1998. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity,* 9**,** 135-41.

RADA, C., JARVIS, J. M. & MILSTEIN, C. 2002. AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proc Natl Acad Sci U S A,* 99**,** 7003-8.

RADA, C. & MILSTEIN, C. 2001. The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. *The EMBO journal,* 20**,** 4570-6.

RADA, C., YELAMOS, J., DEAN, W. & MILSTEIN, C. 1997. The 5' hypermutation boundary of kappa chains is independent of local and neighbouring sequences and related to the distance from the initiation of transcription. *European journal of immunology,* 27**,** 3115-20.

RAHL, P. B., LIN, C. Y., SEILA, A. C., FLYNN, R. A., MCCUINE, S., BURGE, C. B., SHARP, P. A. & YOUNG, R. A. 2010. c-Myc regulates transcriptional pause release. *Cell,* 141**,** 432-45.

RAI, K., HUGGINS, I. J., JAMES, S. R., KARPF, A. R., JONES, D. A. & CAIRNS, B. R. 2008. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell,* 135**,** 1201-12.

RAJAGOPAL, D., MAUL, R. W., GHOSH, A., CHAKRABORTY, T., KHAMLICHI, A. A., SEN, R. & GEARHART, P. J. 2009. Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *J Exp Med,* 206**,** 1237-44.

RAMIRO, A. R., JANKOVIC, M., EISENREICH, T., DIFILIPPANTONIO, S., CHEN-KIANG, S., MURAMATSU, M., HONJO, T., NUSSENZWEIG, A. & NUSSENZWEIG, M. C. 2004. AID is required for c-myc/IgH chromosome translocations in vivo. *Cell,* 118**,** 431-8.

RAMIRO, A. R., STAVROPOULOS, P., JANKOVIC, M. & NUSSENZWEIG, M. C. 2003. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat Immunol,* 4**,** 452-6.

RASCHKE, E. E., ALBERT, T. & EICK, D. 1999. Transcriptional regulation of the Ig kappa gene by promoter-proximal pausing of RNA polymerase II. *J Immunol,* 163**,** 4375-82.

REIK, W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature,* 447**,** 425-32.

REINA-SAN-MARTIN, B., CHEN, H. T., NUSSENZWEIG, A. & NUSSENZWEIG, M. C. 2004. ATM is required for efficient recombination between immunoglobulin switch regions. *J Exp Med,* 200**,** 1103-10.

REINA-SAN-MARTIN, B., DIFILIPPANTONIO, S., HANITSCH, L., MASILAMANI, R. F., NUSSENZWEIG, A. & NUSSENZWEIG, M. C. 2003. H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *J Exp Med,* 197**,** 1767-78.

REVY, P., MUTO, T., LEVY, Y., GEISSMANN, F., PLEBANI, A., SANAL, O., CATALAN, N., FORVEILLE, M., DUFOURCQ-LABELOUSE, R., GENNERY, A., TEZCAN, I., ERSOY, F., KAYSERILI, H., UGAZIO, A. G., BROUSSE, N., MURAMATSU, M., NOTARANGELO, L. D., KINOSHITA, K., HONJO, T., FISCHER, A. & DURANDY, A. 2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell,* 102**,** 565-75.

REYNAUD, C. A., ANQUEZ, V., DAHAN, A. & WEILL, J. C. 1985. A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell,* 40**,** 283-91.

REYNAUD, C. A., ANQUEZ, V., GRIMAL, H. & WEILL, J. C. 1987. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell,* 48**,** 379-88.

REYNAUD, C. A., AOUFOUCHI, S., FAILI, A. & WEILL, J. C. 2003. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nat Immunol,* 4**,** 631-8.

REYNAUD, C. A., DELBOS, F., FAILI, A., GUERANGER, Q., AOUFOUCHI, S. & WEILL, J. C. 2009. Competitive repair pathways in immunoglobulin gene hypermutation. *Philos Trans R Soc Lond B Biol Sci,* 364**,** 613-9.

ROA, S., AVDIEVICH, E., PELED, J. U., MACCARTHY, T., WERLING, U., KUANG, F. L., KAN, R., ZHAO, C., BERGMAN, A., COHEN, P. E., EDELMANN, W. & SCHARFF, M. D. 2008. Ubiquitylated PCNA plays a role in somatic

hypermutation and class-switch recombination and is required for meiotic progression. *Proc Natl Acad Sci U S A,* 105**,** 16248-53.

ROBBIANI, D. F., BUNTING, S., FELDHAHN, N., BOTHMER, A., CAMPS, J., DEROUBAIX, S., MCBRIDE, K. M., KLEIN, I. A., STONE, G., EISENREICH, T. R., RIED, T., NUSSENZWEIG, A. & NUSSENZWEIG, M. C. 2009. AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol Cell,* 36**,** 631-41.

ROGOZIN, I. B., BASU, M. K., JORDAN, I. K., PAVLOV, Y. I. & KOONIN, E. V. 2005. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle,* 4**,** 1281-5.

ROGOZIN, I. B. & KOLCHANOV, N. A. 1992. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim Biophys Acta,* 1171**,** 11-8.

ROSS, A. L. & SALE, J. E. 2006. The catalytic activity of REV1 is employed during immunoglobulin gene diversification in DT40. *Mol Immunol,* 43**,** 1587-94.

ROZENBLATT-ROSEN, O., HUGHES, C. M., NANNEPAGA, S. J., SHANMUGAM, K. S., COPELAND, T. D., GUSZCZYNSKI, T., RESAU, J. H. & MEYERSON, M. 2005. The parafibromin tumor suppressor protein is part of a human Paf1 complex. *Mol Cell Biol,* 25**,** 612-20.

SABERI, A., NAKAHARA, M., SALE, J. E., KIKUCHI, K., ARAKAWA, H., BUERSTEDDE, J. M., YAMAMOTO, K., TAKEDA, S. & SONODA, E. 2008. The 9-1-1 DNA clamp is required for immunoglobulin gene conversion. *Mol Cell Biol,* 28**,** 6113-22.

SAEKI, H. & SVEJSTRUP, J. Q. 2009. Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. *Mol Cell,* 35**,** 191-205.

SALE, J. E. 2004. Immunoglobulin diversification in DT40: a model for vertebrate DNA damage tolerance. *DNA Repair (Amst),* 3**,** 693-702.

SALE, J. E., BATTERS, C., EDMUNDS, C. E., PHILLIPS, L. G., SIMPSON, L. J. & SZUTS, D. 2009. Timing matters: error-prone gap filling and translesion synthesis in immunoglobulin gene hypermutation. *Philos Trans R Soc Lond B Biol Sci,* 364**,** 595-603.

SALE, J. E., CALANDRINI, D. M., TAKATA, M., TAKEDA, S. & NEUBERGER, M. S. 2001. Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. *Nature,* 412**,** 921-6.

SATO, Y., PROBST, H. C., TATSUMI, R., IKEUCHI, Y., NEUBERGER, M. S. & RADA, C. 2010. Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *J Biol Chem,* 285**,** 7111-8.

SAUNDERS, A., CORE, L. J. & LIS, J. T. 2006. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol,* 7**,** 557-67.

SAYEGH, C. E., QUONG, M. W., AGATA, Y. & MURRE, C. 2003. E-proteins directly regulate expression of activation-induced deaminase in mature B cells. *Nat Immunol,* 4**,** 586-93.

SCHANZ, S., CASTOR, D., FISCHER, F. & JIRICNY, J. 2009. Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs

may play a key role in somatic hypermutation. *Proc Natl Acad Sci U S A,* 106**,** 5593-8.

SCHAR, P. & FRITSCH, O. 2011. DNA repair and the control of DNA methylation. *Prog Drug Res,* 67**,** 51-68.

SCHATZ, D. G. & JI, Y. 2011. Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol,* 11**,** 251-63.

SCHOTTA, G., SENGUPTA, R., KUBICEK, S., MALIN, S., KAUER, M., CALLEN, E., CELESTE, A., PAGANI, M., OPRAVIL, S., DE LA ROSA-VELAZQUEZ, I. A., ESPEJO, A., BEDFORD, M. T., NUSSENZWEIG, A., BUSSLINGER, M. & JENUWEIN, T. 2008. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev,* 22**,** 2048-61.

SELTH, L. A., SIGURDSSON, S. & SVEJSTRUP, J. Q. 2010. Transcript Elongation by RNA Polymerase II. *Annu Rev Biochem,* 79**,** 271-93.

SHAPIRO, G. S. & WYSOCKI, L. J. 2002. DNA target motifs of somatic mutagenesis in antibody genes. *Crit Rev Immunol,* 22**,** 183-200.

SHAPIRO-SHELEF, M., LIN, K. I., SAVITSKY, D., LIAO, J. & CALAME, K. 2005. Blimp-1 is required for maintenance of long-lived plasma cells in the bone marrow. *J Exp Med,* 202**,** 1471-6.

SHEEHY, A. M., GADDIS, N. C., CHOI, J. D. & MALIM, M. H. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature,* 418**,** 646-50.

SHEN, H. M., PETERS, A., BARON, B., ZHU, X. & STORB, U. 1998. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science,* 280**,** 1750-2.

SHILATIFARD, A. 2008. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr Opin Cell Biol,* 20**,** 341-8.

SHINKURA, R., ITO, S., BEGUM, N. A., NAGAOKA, H., MURAMATSU, M., KINOSHITA, K., SAKAKIBARA, Y., HIJIKATA, H. & HONJO, T. 2004. Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat Immunol,* 5**,** 707-12.

SHINKURA, R., TIAN, M., SMITH, M., CHUA, K., FUJIWARA, Y. & ALT, F. W. 2003. The influence of transcriptional orientation on endogenous switch region function. *Nat Immunol,* 4**,** 435-41.

SIMPSON, L. J., ROSS, A. L., SZUTS, D., ALVIANI, C. A., OESTERGAARD, V. H., PATEL, K. J. & SALE, J. E. 2006. RAD18-independent ubiquitination of proliferating-cell nuclear antigen in the avian cell line DT40. *EMBO Rep,* 7**,** 927-32.

SIMPSON, L. J. & SALE, J. E. 2003. Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line. *EMBO J,* 22**,** 1654-64.

SOHAIL, A., KLAPACZ, J., SAMARANAYAKE, M., ULLAH, A. & BHAGWAT, A. S. 2003. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res,* 31**,** 2990-4.

STANLIE, A., AIDA, M., MURAMATSU, M., HONJO, T. & BEGUM, N. A. 2010. Histone3 lysine4 trimethylation regulated by the facilitates chromatin

transcription complex is critical for DNA cleavage in class switch recombination. *Proc Natl Acad Sci U S A,* 107**,** 22190-5.
STAUDT, L. M. & LENARDO, M. J. 1991. Immunoglobulin gene transcription. *Annu Rev Immunol,* 9**,** 373-98.
STAVNEZER-NORDGREN, J. & SIRLIN, S. 1986. Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. *EMBO J,* 5**,** 95-102.
STELTER, P. & ULRICH, H. D. 2003. Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature,* 425**,** 188-91.
STORB, U., SHEN, H. M., MICHAEL, N. & KIM, N. 2001. Somatic hypermutation of immunoglobulin and non-immunoglobulin genes. *Philos Trans R Soc Lond B Biol Sci,* 356**,** 13-9.
STROBL, L. J. & EICK, D. 1992. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. *EMBO J,* 11**,** 3307-14.
SWANSON, M. S., MALONE, E. A. & WINSTON, F. 1991. SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Mol Cell Biol,* 11**,** 3009-19.
SZUTS, D. & SALE, J. E. 2006. Subnuclear immunofluorescence. *Subcell Biochem,* 40**,** 395-8.
TA, V. T., NAGAOKA, H., CATALAN, N., DURANDY, A., FISCHER, A., IMAI, K., NONOYAMA, S., TASHIRO, J., IKEGAWA, M., ITO, S., KINOSHITA, K., MURAMATSU, M. & HONJO, T. 2003. AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat Immunol,* 4**,** 843-8.
TAN, C. K., CASTILLO, C., SO, A. G. & DOWNEY, K. M. 1986. An auxiliary protein for DNA polymerase-delta from fetal calf thymus. *J Biol Chem,* 261**,** 12310-6.
TANAKA, A., SHEN, H. M., RATNAM, S., KODGIRE, P. & STORB, U. 2010. Attracting AID to targets of somatic hypermutation. *J Exp Med,* 207**,** 405-15.
TASHIRO, J., KINOSHITA, K. & HONJO, T. 2001. Palindromic but not G-rich sequences are targets of class switch recombination. *Int Immunol,* 13**,** 495-505.
TENG, B., BURANT, C. F. & DAVIDSON, N. O. 1993. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science,* 260**,** 1816-9.
TENG, G., HAKIMPOUR, P., LANDGRAF, P., RICE, A., TUSCHL, T., CASELLAS, R. & PAPAVASILIOU, F. N. 2008. MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity,* 28**,** 621-9.
TERAI, K., ABBAS, T., JAZAERI, A. A. & DUTTA, A. 2010. CRL4(Cdt2) E3 ubiquitin ligase monoubiquitinates PCNA to promote translesion DNA synthesis. *Molecular cell,* 37**,** 143-9.
TOM, S., HENRICKSEN, L. A. & BAMBARA, R. A. 2000. Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1. *J Biol Chem,* 275**,** 10498-505.
TOM, S., HENRICKSEN, L. A., PARK, M. S. & BAMBARA, R. A. 2001. DNA ligase I and proliferating cell nuclear antigen form a functional complex. *J Biol Chem,* 276**,** 24817-25.
TORRES-RAMOS, C. A., PRAKASH, S. & PRAKASH, L. 2002. Requirement of RAD5 and MMS2 for postreplication repair of UV-damaged DNA in Saccharomyces cerevisiae. *Molecular and cellular biology,* 22**,** 2419-26.

TSURIMOTO, T. & STILLMAN, B. 1989. Purification of a cellular replication factor, RF-C, that is required for coordinated synthesis of leading and lagging strands during simian virus 40 DNA replication in vitro. *Mol Cell Biol,* 9**,** 609-19.

TSURIMOTO, T. & STILLMAN, B. 1990. Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. *Proc Natl Acad Sci U S A,* 87**,** 1023-7.

TSURIMOTO, T. & STILLMAN, B. 1991. Replication factors required for SV40 DNA replication in vitro. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins. *J Biol Chem,* 266**,** 1950-60.

TUMAS-BRUNDAGE, K. & MANSER, T. 1997. The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. *J Exp Med,* 185**,** 239-50.

UCHIMURA, Y., BARTON, L. F., RADA, C. & NEUBERGER, M. S. 2011. REG-gamma associates with and modulates the abundance of nuclear activation-induced deaminase. *J Exp Med*.

ULRICH, H. D. 2009. Regulating post-translational modifications of the eukaryotic replication clamp PCNA. *DNA Repair (Amst),* 8**,** 461-9.

UNK, I., HARACSKA, L., GOMES, X. V., BURGERS, P. M., PRAKASH, L. & PRAKASH, S. 2002. Stimulation of 3'-->5' exonuclease and 3'-phosphodiesterase activities of yeast apn2 by proliferating cell nuclear antigen. *Mol Cell Biol,* 22**,** 6480-6.

VAIRAPANDI, M., LIEBERMANN, D. A., HOFFMAN, B. & DUKER, N. J. 2000. Human DNA-demethylating activity: a glycosylase associated with RNA and PCNA. *J Cell Biochem,* 79**,** 249-60.

VANOLI, F., FUMASONI, M., SZAKAL, B., MALOISEL, L. & BRANZEI, D. 2010. Replication and recombination factors contributing to recombination-dependent bypass of DNA lesions by template switch. *PLoS Genet,* 6**,** e1001205.

VIRET, C. & JANEWAY, C. A., JR. 1999. MHC and T cell development. *Rev Immunogenet,* 1**,** 91-104.

WADA, T., TAKAGI, T., YAMAGUCHI, Y., FERDOUS, A., IMAI, T., HIROSE, S., SUGIMOTO, S., YANO, K., HARTZOG, G. A., WINSTON, F., BURATOWSKI, S. & HANDA, H. 1998. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev,* 12**,** 343-56.

WAGA, S. & STILLMAN, B. 1998a. Cyclin-dependent kinase inhibitor p21 modulates the DNA primer-template recognition complex. *Mol Cell Biol,* 18**,** 4177-87.

WAGA, S. & STILLMAN, B. 1998b. The DNA replication fork in eukaryotic cells. *Annu Rev Biochem,* 67**,** 721-51.

WANG, L., WHANG, N., WUERFFEL, R. & KENTER, A. L. 2006. AID-dependent histone acetylation is detected in immunoglobulin S regions. *J Exp Med,* 203**,** 215-26.

WANG, L., WUERFFEL, R., FELDMAN, S., KHAMLICHI, A. A. & KENTER, A. L. 2009. S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J Exp Med,* 206**,** 1817-30.

WANG, P., BOWL, M. R., BENDER, S., PENG, J., FARBER, L., CHEN, J., ALI, A., ZHANG, Z., ALBERTS, A. S., THAKKER, R. V., SHILATIFARD, A., WILLIAMS, B. O. & TEH, B. T. 2008. Parafibromin, a component of the human PAF complex, regulates growth factors and is required for embryonic development and survival in adult mice. *Mol Cell Biol,* 28**,** 2930-40.

WARD, I. M., REINA-SAN-MARTIN, B., OLARU, A., MINN, K., TAMADA, K., LAU, J. S., CASCALHO, M., CHEN, L., NUSSENZWEIG, A., LIVAK, F., NUSSENZWEIG, M. C. & CHEN, J. 2004. 53BP1 is required for class switch recombination. *J Cell Biol,* 165**,** 459-64.

WEIGERT, M. G., CESARI, I. M., YONKOVICH, S. J. & COHN, M. 1970. Variability in the lambda light chain sequences of mouse antibody. *Nature,* 228**,** 1045-7.

WEILL, J. C. & REYNAUD, C. A. 2008. DNA polymerases in adaptive immunity. *Nat Rev Immunol,* 8**,** 302-12.

WIESENDANGER, M., KNEITZ, B., EDELMANN, W. & SCHARFF, M. D. 2000. Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J Exp Med,* 191**,** 579-84.

WILSON, D. M., 3RD & BARSKY, D. 2001. The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. *Mutat Res,* 485**,** 283-307.

WILSON, T. M., VAISMAN, A., MARTOMO, S. A., SULLIVAN, P., LAN, L., HANAOKA, F., YASUI, A., WOODGATE, R. & GEARHART, P. J. 2005. MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. *J Exp Med,* 201**,** 637-45.

WIT, N., KRIJGER, P. H., VAN DEN BERK, P. C. & JACOBS, H. 2011. Lysine residue 185 of Rad1 is a topological but not a functional counterpart of lysine residue 164 of PCNA. *PLoS One,* 6**,** e16669.

WU, X., GERALDES, P., PLATT, J. L. & CASCALHO, M. 2005. The double-edged sword of activation-induced cytidine deaminase. *J Immunol,* 174**,** 934-41.

WU, X., LI, J., LI, X., HSIEH, C. L., BURGERS, P. M. & LIEBER, M. R. 1996. Processing of branched DNA intermediates by a complex of human FEN-1 and PCNA. *Nucleic Acids Res,* 24**,** 2036-43.

XIAO, W., LIN, S. L., BROOMFIELD, S., CHOW, B. L. & WEI, Y. F. 1998. The products of the yeast MMS2 and two human homologs (hMMS2 and CROC-1) define a structurally and functionally conserved Ubc-like protein family. *Nucleic Acids Res,* 26**,** 3908-14.

XU, H., ZHANG, P., LIU, L. & LEE, M. Y. 2001. A novel PCNA-binding motif identified by the panning of a random peptide display library. *Biochemistry,* 40**,** 4512-20.

XU, Z., PONE, E. J., AL-QAHTANI, A., PARK, S. R., ZAN, H. & CASALI, P. 2007. Regulation of aicda expression and AID activity: relevance to somatic hypermutation and class switch DNA recombination. *Crit Rev Immunol,* 27**,** 367-97.

XUE, K., RADA, C. & NEUBERGER, M. S. 2006. The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice. *J Exp Med,* 203**,** 2085-94.

YAMADA, T., YAMAGUCHI, Y., INUKAI, N., OKAMOTO, S., MURA, T. & HANDA, H. 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell,* 21**,** 227-37.

YAMAGUCHI, Y., TAKAGI, T., WADA, T., YANO, K., FURUYA, A., SUGIMOTO, S., HASEGAWA, J. & HANDA, H. 1999a. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell,* 97**,** 41-51.

YAMAGUCHI, Y., WADA, T., WATANABE, D., TAKAGI, T., HASEGAWA, J. & HANDA, H. 1999b. Structure and function of the human transcription elongation factor DSIF. *J Biol Chem,* 274**,** 8085-92.

YAMANE, A., RESCH, W., KUO, N., KUCHEN, S., LI, Z., SUN, H. W., ROBBIANI, D. F., MCBRIDE, K., NUSSENZWEIG, M. C. & CASELLAS, R. 2011. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol,* 12**,** 62-9.

YELAMOS, J., KLIX, N., GOYENECHEA, B., LOZANO, F., CHUI, Y. L., GONZALEZ FERNANDEZ, A., PANNELL, R., NEUBERGER, M. S. & MILSTEIN, C. 1995. Targeting of non-Ig sequences in place of the V segment by somatic hypermutation. *Nature,* 376**,** 225-9.

YOSHIKAWA, K., OKAZAKI, I. M., ETO, T., KINOSHITA, K., MURAMATSU, M., NAGAOKA, H. & HONJO, T. 2002. AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science,* 296**,** 2033-6.

YU, K., CHEDIN, F., HSIEH, C. L., WILSON, T. E. & LIEBER, M. R. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol,* 4**,** 442-51.

YU, K., HUANG, F. T. & LIEBER, M. R. 2004. DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J Biol Chem,* 279**,** 6496-500.

ZEITLINGER, J., STARK, A., KELLIS, M., HONG, J. W., NECHAEV, S., ADELMAN, K., LEVINE, M. & YOUNG, R. A. 2007. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet,* 39**,** 1512-6.

ZHU, B., MANDAL, S. S., PHAM, A. D., ZHENG, Y., ERDJUMENT-BROMAGE, H., BATRA, S. K., TEMPST, P. & REINBERG, D. 2005. The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev,* 19**,** 1668-73.