

Nuclear exclusion of AID limits off target activity by enforcing productive targeting.

Stephen P. Methot

Faculty of Medicine, Division of Experimental Medicine

McGill University, Montreal.

August 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctoral of Philosophy in Experimental Medicine.

© Stephen P Methot, 2017

Dedicated to Ludi, for everything...

Abstract

Activation induced deaminase (AID) is an enzyme that deaminates deoxycytidine to deoxyuridine on single stranded DNA. AID expression is mostly restricted to antigen activated B-lymphocytes, wherein its activity is critical for both somatic hypermutation (SHM) and class-switch recombination (CSR), which underpin secondary antibody diversification. This is due to the targeted activity of AID at the immunoglobulin (Ig) genes, which encode for the antibody genes. DNA deamination catalyzed by AID initiates the generation of single point mutations for SHM, and double stranded breaks for CSR, both potentially harmful lesions. Importantly, multiple mechanisms regulate AID in order to limit its off-target activity, outside the Ig loci, which has been found to be potentially oncogenic and cytotoxic. For my thesis, I used structure-function studies of AID in order to better understand post-translational regulation of its activity.

In a first chapter, evolutionary comparison and structure-function analyses revealed that a structural conformation of a C-terminal domain of AID was required for its association to eEF1A1, which mediated the retention of AID in the cytoplasm. Mutations that disrupted AID cytoplasmic retention, or pharmacological inhibition of eEF1A1, caused AID to accumulate in the nucleus and to generate more mutations on- and off- target. This demonstrated that eEF1A1-dependent retention of AID in the cytoplasm was necessary to restrict its activity.

In a second chapter, a structure-function study of AID identified a domain dispensable for its catalytic activity, but critical for SHM and CSR activity in B cells.

Mutants of this domain were able to functionally distinguish the recruitment of AID to the chromatin from its mutagenic activity in B cells. This revealed that at the chromatin AID requires a licensing step in order to associate with the elongating transcription machinery, and become mutagenic. Interestingly, this licensing is bypassed when AID is forced to accumulate in the nucleus.

In a third chapter, studies of the C-terminal region of AID further clarified its role during CSR. This study identified a structural requirement of the C-terminus during CSR, downstream from DNA deamination. This study supports a role for AID, via the C-terminus, to recruit other factors during CSR, likely to mediate proper DNA repair.

All together, these studies highlight the intricate regulation of AID activity in B cells. They also suggest that preventing accumulation of AID in the nucleus limits off-target activity by enforcing a regulatory licensing step at the chromatin.

Keywords: Activation induced deaminase (AID); Antibody diversification; Somatic hypermutation (SHM); Class-switch recombination (CSR); Protein structure; Cytoplasmic retention; Eukaryotic translation elongation factor 1-alpha-1 (eEF1A1); Off-target activity; Transcription elongation.

Résumé

Désaminase induite par l'activation (AID) est une enzyme qui désamine le désoxycytidine au désoxyuridine sur l'ADN simple brin. Après avoir reconnu spécifiquement des antigènes, les lymphocytes B vont exprimer AID pour mettre en place l'hypermutation somatique (SHM) et la commutation de classe (CSR), deux mécanismes essentiels pour la diversification des anticorps. L'activité d'AID est alors restreinte aux régions des immunoglobulines (Ig) du génome, qui codent pour les gènes d'anticorps. La désamination par AID initie la génération de mutations simples pour le SHM, et des cassures d'ADN double brins pour le CSR, les deux comprenant des lésions potentiellement néfastes. Plusieurs mécanismes régulent AID avant et après sa transcription afin de limiter son activité en dehors des régions Ig, qui peut être potentiellement oncogénique et cytotoxique. Pendant ma thèse, j'ai effectué des études sur la relation structure-fonction d'AID pour mieux comprendre la régulation de son activité au niveau protéique.

Dans un premier chapitre, une comparaison évolutionnaire et une analyse de structure-fonction révèlent que la conformation structurale de la région C-terminal d'AID est essentielle à sa rétention au cytoplasme, ainsi que son association avec la protéine eEF1A1. Des mutations simples qui empêchent la rétention cytoplasmique d'AID ou des inhibiteurs de eEF1A1, entraînent l'accumulation d'AID dans le noyau et augmentent sa capacité à générer des mutations ciblées et non-ciblées. Ceci démontre ainsi que la rétention cytoplasmique d'AID, dépendant de la protéine eEF1A1, est nécessaire pour restreindre son activité mutagénique.

Dans un deuxième chapitre, une analyse de structure-fonction d'AID a révélé qu'une de ces régions n'est pas nécessaire pour son activité catalytique, mais est critique pour induire le SHM et le CSR dans les cellules B. Des mutants de cette région d'AID sont toujours recrutés à la chromatine mais n'exercent plus d'activité mutagénique. Cela démontre qu'au niveau de la chromatine, AID a besoin d'une étape supplémentaire pour initier son activité mutagénique aux sites du génome qui y sont spécifiquement sensibles. De plus, cette étape d'activation peut être contournée si AID s'accumule dans le noyau.

Dans un troisième chapitre, l'étude de la région C-terminal d'AID a permis également de clarifier son rôle pendant le CSR. En effet, la région C-terminal d'AID joue un rôle primordial dans son activité mutagénique lors du CSR, *via* le recrutement de facteurs, qui sont probablement impliqués dans la réparation de l'ADN.

L'ensemble de ces études confirme la régulation complexe de l'activité d'AID dans les cellules B en particulier au niveau de la chromatine et souligne l'importance d'empêcher son accumulation dans le noyau pour limiter son activité mutagénique non-ciblée.

Mots-clés: Désaminase induite par l'activation (AID); Diversification des anticorps; Hypermutation somatique (SHM); Commutation de classe (CSR); Structure protéique; Rétention cytoplasmique; Facteur d'élongation eucaryote 1 alpha 1 (eEF1A1); Activité non-ciblée; Élongation de la transcription.

Table of Contents

Abstract.....	ii
Résumé.....	iii
Acknowledgements	xi
Preface and Contribution to Knowledge.....	xiii
Contribution of Authors.....	xv
List of Figures.....	xvii
List of Tables.....	xix
Abbreviations	xx
CHAPTER 1: INTRODUCTION	1
1.1. Stepwise antibody diversification during B cell development.....	2
1.1.1 Humoral immunity	2
1.1.2 Antibody structure	3
1.1.3 Primary antibody diversification in the bone marrow	6
1.1.4 The peripheral B cell response.....	8
1.2 Activation induced deaminase	12
1.2.1 The AID/APOBEC family.....	12
1.2.2 AID expression.....	15
1.2.3 Deamination of deoxycytidine on ssDNA	18
1.3 Antibody diversification by DNA deamination	20
1.3.1 Somatic hypermutation.....	22

1.3.2 Immunoglobulin gene conversion.....	24
1.3.3 Class switch recombination.....	26
1.4 The clinical relevance of AID	28
1.4.1 AID in immunity	28
1.4.2 AID and autoimmunity.....	29
1.4.3 AID is oncogenic	30
1.4.4 DNA demethylation	32
1.4.5 Gene editing.....	33
1.5 Regulation of AID.....	35
1.5.1 Structural determinants of AID activity	35
1.5.2 AID subcellular localization	39
1.5.3 AID protein stability	41
1.5.4 AID regulation by DNA damage and cell cycle.....	42
1.5.5 AID phosphorylation.....	43
1.6 AID targeting.....	45
1.6.1 Transcription association dictates AID targeting	45
1.6.2 Genetic determinants of CSR.....	47
1.6.3 Genetic determinants of SHM	48
1.6.4 Recruiting AID to transcription.....	49
1.6.5 Chromatin marks that guide AID	51
1.6.6 RNA processing and AID targeting	52
1.7 Rationale	55

CHAPTER 2: CONSECUTIVE INTERACTIONS WITH HSP90 AND EEF1A1	
UNDERLIE A FUNCTIONAL MATURATION AND STORAGE PATHWAY OF AID IN	
THE CYTOPLASM.....	56
2.1. Abstract	57
2.2. Introduction	58
2.3. Results.....	61
2.3.1. AID cytoplasmic retention is evolutionarily conserved.....	61
2.3.2. Cytoplasmic retention requires a specific AID conformation	65
2.3.3. Mutually exclusive interaction of AID with eEF1A and HSP90	70
2.3.4. eEF1A is necessary for AID cytoplasmic retention.....	72
2.3.5. eEF1A limits isotype switching and chromosomal translocations.....	77
2.3.6. Functional integration of mechanisms regulating cytoplasmic AID.....	80
2.4. Discussion	83
2.5. Materials and methods	88
2.6. Acknowledgements.....	95
CHAPTER 3: A LICENSING STEP LINKS AID TO TRANSCRIPTION ELONGATION	
FOR B CELL MUTAGENESIS	96
3.1. Abstract	97
3.2. Introduction	98
3.3. Results.....	101
3.3.1. Three Arginine residues in AID α 6 define a critical functional domain.....	101
3.3.2. AID R-mutants access the nucleus but lack off-target activity.....	104
3.3.3. The RR domain is dispensable for AID association to nuclear complexes.	109

3.3.4. Chromatin-associated AID is tethered by Spt5, RNAPII and RNA.	112
3.3.5. R-mutants can be rescued by fusion with catalytically inactive AID	115
3.3.6. AID R-mutants lose specific interactions in vivo	115
3.3.7. AID R-mutants fail to progress with elongating RNAPII.....	118
3.3.8. Constitutively nuclear AID bypasses promoter licensing	120
3.4. Discussion	121
3.5. Materials and methods	125
3.6. Acknowledgements.....	136
3.7. Supplementary figures.....	137
CHAPTER 4: STRUCTURAL CONFORMATION OF THE AID C-TERMINAL DOMAIN	
IS NECESSARY FOR EFFICIENT CLASS SWITCH RECOMBINATION.....	146
4.1. Abstract	147
4.2. Introduction	148
4.3. Results.....	150
4.3.1. AID lacking the E5 region is still dependant on S38-phosphorylation	150
4.3.2. A role for the E5 in regulating enzymatic activity	150
4.3.3. Identification of D188 as an important functional residue.....	152
4.3.4. Fusions to AID C-terminus affect enzymatic activity	155
4.3.5. AID C-terminal integrity is critical for CSR.....	157
4.4. Discussion	159
4.5. Materials and Methods	162
4.6. Acknowledgements.....	165
CHAPTER 5: DISCUSSION.....	166

5.1. Stepwise regulation of AID function	167
5.2. Nuclear export and cytoplasmic retention exclude different pools of AID from the nucleus	170
5.3. Cytoplasmic retention enforces targeting by limiting AID nuclear accumulation	172
5.4. The E5 conformation regulates AID function	174
5.5. Mechanistic link between CR and licensing	176
5.6. The transcription environment favouring licensing.....	179
5.7. Pharmacological manipulation of AID	181
CHAPTER 6: CONCLUDING REMARKS	184
CHAPTER 7: BIBLIOGRAPHY	186

Acknowledgements

A PhD is a long, and arduous journey. I've been fortunate to have amazing support throughout, and would not be here without a lot of help.

First, I have to thank my supervisor, Dr. Javier M. Di Noia. I don't know if I can express the extent of my gratitude for your patience, confidence, guidance and support. You constantly challenged me to improve, and provided me with the tools to be a better scientist and person. I will miss our never-ending discussions.

Thank you to the Di Noia lab, past, present and future. You're like family, and I've relied on all of you for support! Especially thank you to Ludivine, who has been involved in almost every aspect of my research, and life (I'll get back to that), and to Anne-Marie and Astrid, who have both supported in immeasurable ways from start to finish. Thank you also to Alex, for his special kind of support... and also to Anil, Ganesh, Damien, Shiva, Debashree, Therence and Julian for your help, conversations, and for generally making the lab a fun place to be.

Science is collaborative and requires different skills and points of view to be successful. I must thank my scientific committee, Dr. Jerry Pelletier, Dr. Jacques Archambault and Dr. Ursula Stochaj, who helped guide my research, and always challenged my ideas, for the better. I am also grateful for my advisor, Dr. Jean-Francois Côté, who always made himself available when I needed advice. I am grateful to Dr. Patricia Gearhart, Dr. Anastasia Nijnik, Dr. Marc-André Langlois, Dr. Nicole Francis and Dr. Woong-Kyung Suh for taking the time evaluate my thesis. And thank you to my external collaborators, Dr. Ramiro Verdun, Dr. Francis Robert, Dr. Mani Larijani, Dr. Felipe Trajtenberg, Dr. Alejandro Buschiazzi and Dr. Brad Magor, because this work would have been much less interesting without your help.

The IRCM has been a great place to work because of the people. I have received countless advice, support and reagents from professors, post-docs and students at the IRCM so thank you all. An extra big thank you to our next-door neighbours in the Möroy lab, a rowdy bunch, but always generous with reagents and

knowledge. The core facilities at the IRCM are also world class, in particular because amazing people are involved. Thank you to Dominic Fillon, Eric Massicotte, Julie Lord, Philippe St-Onge, Julien Leconte, Mylène Cawthorne, Ève-Lyne Thivierge, Simone Terouz, Odile Neyret, Agnès Dumont, Myriam Rondeau, Denis Faubert, Jean-Francois Lauzon and Richard Cimon for all your help.

I am indebted to Jennifer Frazszak, Poorani Ganesh Subramani and Ludivine Litzler for critically reading my thesis, and to April Youn for help formatting the thesis.

During my studies I've made amazing friends with whom I've discussed science, but mostly explored the natural beauty surrounding Montreal. Thank you for the great times Jen, Damien, Manisha, Alex O., Vas, Marine, Julie, Max, April, Maxime, Eszter, Carine, Anais and everyone else that I cannot fit here. Thank you to the Primers, Julien, Jen, Elliot, Aurèle, Noumeira, Steve, Jin, Sarah, Melanie and Thomas for all the entertainment. And an extra thank you to my roommates Alex A. and Joe for being there when I really needed you.

Une immense merci a ma famille Québécoise. Vous m'avez accueilli avec amour et je me suis senti comme chez moi à cause de vous. Nos visite m'on toujours changez d'idée et sont parmi mes meilleurs souvenir du Doctorat. Je vous aimes!

Thank you to my family and friends in Vancouver. Especially my parents, who managed to always put things in perspective despite the 3,680 km separating us. To my brothers, Paul and Andre, and all my aunts, uncles and cousins, who constantly pushed me to finish and move back home. To my Yiayia and Papou who inspired my work ethic by setting an unbelievable example. And to my friends, Cailen, Matt, Elena, Sasha, Jeff, Jimmy, Hesh, Jesse, Derek, Dave, Cathy, and many more.

Finally, thank you to Ludi. For sharing this adventure, and making it the best time of my life. You have inspired me, supported me, and shared amazing adventures. You remind me everyday why life is so amazing. Je t'aime. Also thank you to all the Litzlers, for your support now, and to come.

Preface and Contribution to Knowledge

This thesis is presented in manuscript format, and was written with input from my supervisor, Dr. Javier M. Di Noia. I have included a general introduction, Chapter 1, a discussion of my work, Chapter 5, and a final conclusion, Chapter 6. My original research work focussed on using structure-function studies to understand AID regulation, revealing mechanistic understanding of AID retention in the cytoplasm, its functional targeting in B cells, and the structural role for the C-terminal region. This work is presented as three distinct manuscripts in Chapters 2, 3 and 4. These are published, submitted or being prepared for publication in peer review journals, respectively:

1. **Methot SP**, Litzler LC, Trajtenberg F, Zahn A, Robert F, Pelletier J, Buschiazio A, Magor BG, Di Noia JM. (2015) *Consecutive Interactions with HSP90 and eEF1A1 Underlie a Functional Maturation and Storage Pathway of AID in the Cytoplasm*. *J. Exp. Med.* 212(4): 581-96.
2. **Methot SP**, Litzler LC, Subramani PG, Eranki AK, Fifield H, Patenaude A-M, Gilmore JC, Bagci H, Santiago GE, Cote J-F, Verdun RE, Larijani M, Di Noia JM. *A Licensing Step Links AID to Transcription Elongation for B cell Mutagenesis*. Manuscript submitted.
3. **Methot SP**, Zahn A, Eranki AK, Gilmore JC, Di Noia JM. *Structural Conformation of the AID C-terminal Domain is Necessary for Efficient Class Switch Recombination*. Manuscript in preparation.

During my doctoral studies I have also contributed to the following publications:

- Zahn A, Eranki AK, Patenaude AM, **Methot SP**, Fifield H, Cortizas EC, Foster P, Imai K, Durandy A, Larijani M, Verdun RE, Di Noia JM. (2014) *Activation Induced Deaminase Prevents End Resection and Promotes End Joining During Class Switch Recombination*. *PNAS* 111(11): E988-97.

- Hu Y, Ericsson I, Torseth K, **Methot SP**, Sundheim O, Liabakk NB, Slupphaug G, Di Noia JM, Krokan HE and Kavli B. (2013) *A Combined Nuclear and Nucleolar Localization Motif in Activation-Induced Cytidine Deaminase (AID) Controls Immunoglobulin Class Switching*. J. Mol. Biol. 425(2): 424-43.
- Orthwein A, Zahn A, **Methot SP**, Godin D, Conticello SG, Terada K, Di Noia JM. (2011) *Optimal functional levels of activation-induced deaminase specifically require the Hsp40 DnaJa1*. EMBO J. 31(3): 679-91.

Contribution of Authors

Chapter 2: Stephen P. Methot designed, performed and analyzed most of the experiments, generated the figures and helped write the manuscript. Ludivine C. Litzler performed colPs with AID variants (Fig 2.3A). Dr. Felipe Trajtenberg generated AID structure models (Fig 2.2E,H). Dr. Astrid Zahn performed IgH-cMyc PCR reactions (Fig 2.6H). Dr. Francis Robert helped to perform sucrose gradient fractionations (Fig 2.5A). Dr. Brad G. Magor transfected and treated catfish B cells (Fig 2.1D). Dr. Jerry Pelletier and Dr. Alejandro Buschiazzi provided mentorship. Dr. Javier M. Di Noia conceived the project and wrote the manuscript. All authors discussed and interpreted data and critically read the manuscript.

Chapter 3: Stephen P. Methot designed, performed and analyzed most of the experiments, generated the figures and helped write the manuscript. Ludivine C. Litzler generated many AID variants, performed many functional assays and helped with ChIP experiments (Fig 3.1b-d, supplementary Fig 3.1b-d, supplementary Fig 3.2, Fig 3.3a). Poorani Ganesh Subramani analyzed the BioID data (Fig 3.6d-e, supplementary Fig 3.5c-f). Anil K. Eranki helped conceive the project and performed initial functional assays (supplementary Fig 3.1b-d). Heather Fifield performed *in vitro* deaminase activity and DNA binding assays (Fig 3.1e-f). Anne-Marie Patenaude helped generate constructs and helped with different assays. Julian C. Gilmore helped generate AID variants and helped with functional assays (Fig 3.3d-g, Fig 3.6a, Fig 3.7c-e). Halil Bagci performed IPs and generated peptide lists for BioID. Dr.

Ramiro E. Verdun and Gabriel E. Santiago performed ChIP IPs and analysis (Fig 3.3a, Fig 3.7a,b). Dr. Jean-Francois Coté, Dr. Mani Larijani and Dr. Ramiro E. Verdun provided mentorship. Dr. Javier M. Di Noia conceived the project and wrote the manuscript. All authors discussed and interpreted data and critically read the manuscript.

Chapter 4: Stephen P. Methot designed, performed and analyzed many of the experiments, generated the figures and wrote the manuscript. Dr. Astrid Zahn, helped conceive the project, designed the fusion constructs and help generate the fusion constructs (Fig. 4.3). Julian C. Gilmore, helped generate the fusion constructs and performed some functional assays (Fig 4.3). Anil K. Eranki generated and performed functional assays for certain AID variants (Fig 4.2). Dr. Javier M. Di Noia conceived the project and helped write the manuscript. All authors discussed and interpreted data.

List of Figures

Chapter 1:

Figure 1.1. Antibody structure and function.....	4
Figure 1.2. IgH locus diversification.....	8
Figure 1.3. The germinal center reaction.....	11
Figure 1.4. The AID/APOBEC family.....	13
Figure 1.5. The DNA deaminase model for antibody diversification.....	21
Figure 1.6. Downstream DNA repair mediating SHM and CSR.	23
Figure 1.7. The AID/APOBEC deaminase domain is highly conserved.	36
Figure 1.8. AID structural elements and regulatory features.	38
Figure 1.9. Regulation of AID subcellular localization.	40
Figure 1.10. Targeting AID to the IgV and S-regions.	50

Chapter 2:

Figure 2.1. Cytoplasmic retention of AID is conserved.....	63
Figure 2.2. A specific conformation of AID is necessary for cytoplasmic retention.....	66
Figure 2.3. Cytoplasmic AID is present in at least two mutually exclusive complexes. .	71
Figure 2.4. eEF1A1 inhibition causes nuclear accumulation of AID	74
Figure 2.5. eEF1A is necessary for cytoplasmic retention of AID.....	75
Figure 2.6. AID cytoplasmic retention limits CSR and chromosomal translocations.	78
Figure 2.7. eEF1A, CRM1 and Hsp90 associate with functionally distinct AID fractions.	81

Chapter 3:

Figure 3.1. Identification of functionally inactive AID variants.....	102
Figure 3.2. Normal nucleocytoplasmic shuttling of AID R-mutants.....	105
Figure 3.3. The RR domain is necessary for AID association to Igh and for genome-wide DNA damage.	108
Figure 3.4. The RR domain is dispensable for AID association with tightly held nuclear complexes.....	111
Figure 3.5. Requirements for chromatin association of AID.	113
Figure 3.6. Protein-protein interactions of AID and R-mutants.	116
Figure 3.7. AID R-mutants are uncoupled from transcription elongation.....	119
Supplementary Figure 3.1. The alpha helix 6 of AID is required for biological function.	138
Supplementary Figure 3.2. Arginines 171, 174 and 178 play different roles.	139
Supplementary Figure 3.3. Reconstitution of AID-deficient CH12 B cells	140
Supplementary Figure 3.4. AID chromatin association controls	142
Supplementary Figure 3.5. BioID controls.....	144

Chapter 4:

Figure 4.1. AID Δ E5 hyperactivity requires S38-phosphorylation in B cells.....	151
Figure 4.2. The E5 domain regulates AID catalytic and functional activities	154
Figure 4.3. AID E5 structural integrity is necessary for CSR.....	156

Chapter 5:

Figure 5.1. Stepwise model of AID activity.....	168
Figure 5.2. Structural modeling of AID regulation by E5.....	172
Figure 5.3. Mechanisms linking cytoplasmic retention and licensing.	177

List of Tables

Chapter 3:

Supplementary Table 3.1. AID and R-mutants BiID comparison table.....	145
---	-----

Chapter 4:

Table 4.1. Compilation of known AID C-terminal mutations.....	153
Table 4.2. Comparison of AID fusion constructs.....	158

Abbreviations

3'RR: 3' regulatory region

5mC: 5-Methylcytosine

A1/A2/A3: APOBEC1/2/3

A-EJ: Alternative end-joining

ActD: Actinomycin D

AID: Activation induced deaminase

APOBEC: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like

APE: apurinic/pyrimidinic endonuclease

ASC: Antibody secreting cell

ATM: Ataxia-telangiectasia mutated

BCR: B-cell receptor

BER: Base excision repair

CDR: Complementarity determining region

ChIP: Chromatin immunoprecipitation

CHX: Cycloheximide

C-NHEJ: classical non-homologous end joining

CR motif: Cytoplasmic retention motif

CRM1: Exportin 1

CSR: Class switch recombination

CytA: Cytotrienin A

dCMP: Deoxycytidine monophosphate

DidB: Didemnin B

DIVAC: Diversification activator

DSB: Double stranded DNA breaks

DVL3: Dishevelled 3

DZ: Germinal center dark zone

E5: The C-terminal 17 amino acids of AID

E μ : IgM enhancer

eEF1A1: Eukaryotic translation elongation factor 1 alpha 1

FACT complex: Spt16 and ssrp1

Fab: Fragment, antigen binding

Fc: Fragment, crystallizable

G4: G-quadruplex

GANP: Germinal center-associated nuclear protein

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

GC: Germinal center

GFP: Green fluorescent protein

HIGM: Hyper-IgM syndrome

hnRNP: Heterogeneous nuclear ribonucleoprotein

HR: Homologous recombination

HSP90: Heat-shock protein 90 kDa

IF: Immunofluorescence

IG: Immunoglobulin

IGC: Immunoglobulin gene conversion

IgH/IgL: Immunoglobulin heavy chain and light chain

IgV: Immunoglobulin variable region

IL-4: Interleukin 4
IP: Immunoprecipitation
iPSC: induced pluripotent stem cell
LacO/LacR: Lac operon/repressor
LMB: Leptomycin B
LZ: Germinal center light zone
LPS: Lipopolysaccharide
MBD4: Methyl-CpG binding domain 4
MFI: Mean fluorescence intensity
mIR: microRNA
MMR: Mismatch repair
MNase: Micrococcal nuclease
MRN: MRE11-RAD50-NBS1
MutS α : MSH2/MSH6
MutL α : MLH1/PMS2
NES: Nuclear export signal
PCNA: Proliferating cell nuclear antigen
Pol β / η / ζ / ι : DNA polymerase beta / eta / zeta / iota
PTBP2: Polypyrimidine tract binding protein 2
PKA: Protein kinase A
RAG: Recombination activating gene

R-loop: RNA-DNA hybrids formed in the context of dsDNA
RNAi: RNA interference
RNAPII: RNA polymerase II
RNP: Ribonucleoprotein
RPA: Replication protein A
S-region: Switch region
SD: Standard deviation
SEM: Standard error of the mean
SHM: Somatic hypermutation
S-S synapse: Physical association between switch regions
TET: Ten-eleven translocation
TDG: Thymine-DNA glycosylase
TdT: Terminal deoxynucleotide transferase
TLR: Toll-like receptor
TSS: Transcription start site
UNG: Uracil DNA glycosylase
VLR: Variable lymphocyte receptor
XRCC: x-ray cross-complementing group

CHAPTER 1: INTRODUCTION

1.1. Stepwise antibody diversification during B cell development

1.1.1 Humoral immunity

The immune system is divided into two main branches, innate and adaptive immunity. Innate immunity comprises the programmed ability to recognize invariant antigens, while adaptive immunity comprises the ability to generate and maintain specific recognition of novel antigens. Together, the immune system protects the body from toxins, infections and oncogenic transformations (Murphy et al., 2012; Parkin and Cohen, 2001). Adaptive immunity can be further divided into humoral immunity, characterized by the production of antibodies by B-lymphocytes, and cell-mediated immunity, driven by T-lymphocytes (Boehm and Swann, 2014).

The hallmarks of the adaptive immune system, the ability to recognize an antigen specifically and to maintain memory of this recognition – are dictated by the antigen receptors of B and T lymphocytes. This is possible because antigen specificity is hardwired into B and T cells by genetic alteration of the genes encoding for the antigen receptors. The antigen receptor in B cells (or BCR) is in fact a plasma membrane-bound form of the expressed antibody; therefore, B cell antigen specificity is always identical with respect to antigen recognition and antibody secretion (Gold, 2002).

Generating antigen-specific, high-affinity antibodies, is a multi-step process with two main stages. The first stage occurs during B cell development and centers

on generating a diverse antibody repertoire independently of antigen exposure. The second, a differentiation stage, sequentially drives affinity maturation and functional specialization in B cells that are exposed to antigen. The end goal of these processes is the generation of B cells that can secrete massive amounts of high affinity antibody in order to clear the invading pathogen. These stages are spatially segregated in the bone marrow and peripheral lymphoid organs (spleen, lymph nodes, Peyer's patches and appendix), respectively (Hardy and Hayakawa, 2001).

These developmental pathways allow B cells to generate an almost limitless antibody repertoire, capable of binding antigen of almost any chemical composition and configuration. This capacity is necessary to recognize and adapt to constantly changing pathogens (Perelson and Oster, 1979). Antibody binding contributes to immunity in four ways (Figure 1.1): neutralization, which blocks the entry of viruses, bacteria or toxins into cells; opsonisation, which promotes phagocytosis of pathogens by phagocytes; complement activation, whereby antibody binding promotes pathogen lysis by activating complement at its cell surface; and antibody-dependent cellular cytotoxicity, in which antibodies activate cytotoxic lymphocytes (Davies and Metzger, 1983; Scornik et al., 1974).

1.1.2 Antibody structure

Antibodies are generated by the association of four polypeptides, comprising two identical heavy chains (IgH) and two identical light chains (IgL), linked by disulphide bonds (Harris et al., 1992). They can be divided in two parts based on their structure and function (Figure 1.1).

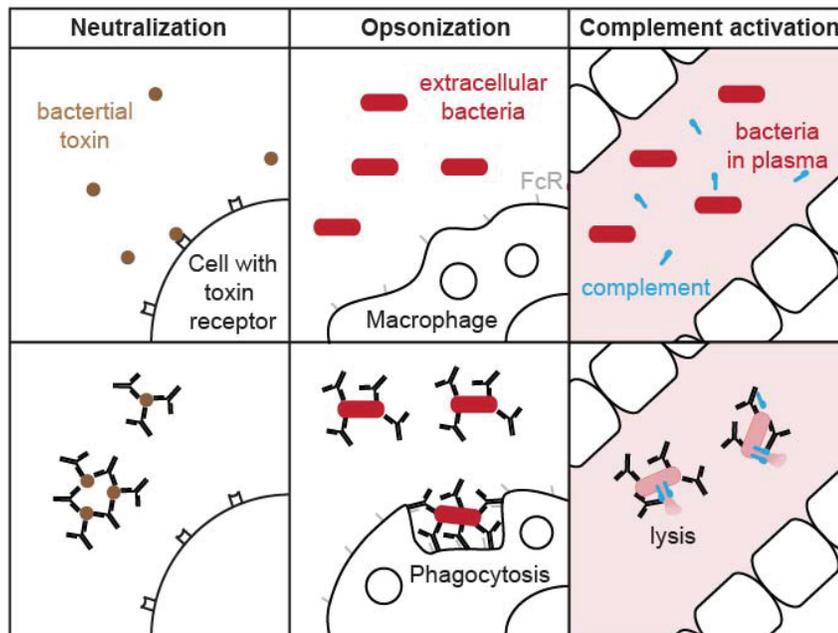
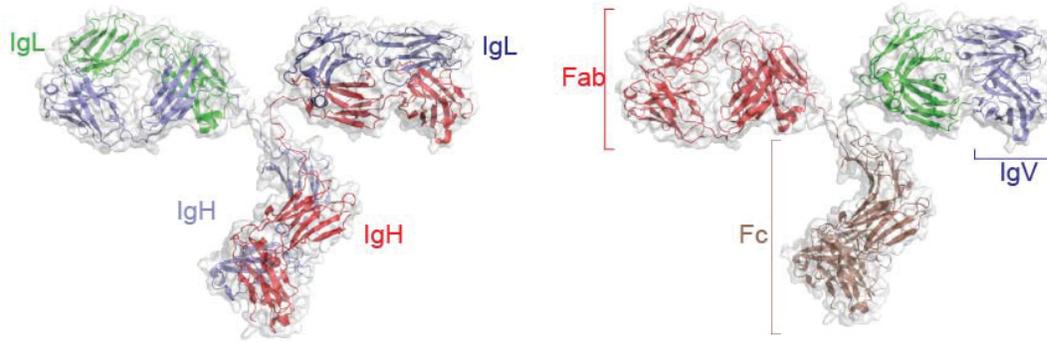


Figure 1.1. Antibody structure and function.

(Top) An antibody crystal structure (PDB: 1IGT) with different antibody chains and domains highlighted (Harris et al., 1997). IgL (light chain), IgH (heavy chain), Fab (fragment antigen binding), Fc (fragment crystallizable), IgV (variable region). (Bottom) The various mechanisms by which antibodies protect the host from toxins or pathogens. Neutralization occurs by blocking the association of toxins or pathogens to their target receptors on cells. Opsonization, involves the phagocytosis of antigen-coated pathogens, due to association of the antibody Fc-domain to the Fc-receptor on the phagocyte. Antibodies can also activate complement on the surface of pathogens and thereby induce lysis. Adapted from (Murphy et al., 2012).

Antigen binding is mediated by two identical antigen-binding domains (Fabs). Each Fab is formed by the N-terminal regions of both an IgH and IgL, forming two

beta barrel domains. The N-terminal beta barrel comprises the variable region, and has six distinct hypervariable loops called complementarity-determining regions (CDRs), three from IgH and IgL respectively, which are responsible for most contact sites with the antigen.

Each Fab is connected to the C-terminal region of the IgH. The association of two IgH chains, linked by disulphide bonds, generates the effector domain (Fc), which mediates downstream functions of the antibody (Davies and Metzger, 1983). While the Fabs can assume an almost limitless number of conformations with distinct amino acid sequences, the constant region can assume only one of five main forms, referred to as the antibody isotypes (Davies and Metzger, 1983).

The antibody isotypes are broken down into five classes in humans according to the identity of the heavy chain: μ (IgM), δ (IgD), γ (IgG1, 2, 3, 4), ϵ (IgE) and α (IgA1, 2). In mice only one IgA subtype exists, and IgG are referred differently (IgG1, 2a (or IgG2c in C57BL/6 (Martin et al., 1998)), 2b and 3). These isotypes confer distinct physical and functional properties to the antibody (Davies and Metzger, 1983). The Fc domain has three main roles during the antibody response. First, it can associate with Fc receptors present on other immune cells, generating antigen specific receptors to promote phagocytosis, or the release of cytokines (Ravetch and Kinet, 1991). Second, the Fc is necessary for the activation of complement by the antibody (Klaus et al., 1979). Third, the Fc can permit antibody access to specific compartments of the body by binding to specific receptors. This allows IgA to be secreted in mucous, tears and milk, and IgG to cross the placenta during pregnancy (Hay et al., 1971; Phalipon et al., 2002).

1.1.3 Primary antibody diversification in the bone marrow

Primary antibody diversification is the process of genetically assembling a unique, functional antibody gene during B cell development in the bone marrow. The intimate details of this process, underpinned by V-D-J recombination (Roth, 2014), are beyond the scope of this thesis.

Each B cell that leaves the bone marrow has a unique BCR on its cell surface, based on differences of the variable region. In adult humans, the potential diversity can theoretically reach $\sim 10^{15}$ distinct BCR affinities (Schroeder, 2006), although only $\sim 10^{11}$ B cells exist at any one time and are constantly turning over (Apostaei and Trabalka, 2010). This vast diversity is critical for a functional humoral immune system, as this permits the potential for recognizing with some affinity virtually any physico-chemical configuration possible. Achieving such a repertoire involves an elegant mix of genetic inheritance and somatic diversification, whereby genetically encoded gene segments are randomly rearranged in individual, developing B cells, in order to generate a functional, unique antibody. By consequence, only mature B cells leaving the bone marrow possess a functional antibody gene (Hozumi and Tonegawa, 1976).

The genetic rearrangements that generate the functional antibody occur sequentially at the *IgH* and *IgL* loci. In humans, the *IgH*, encoded on chromosome 14, contains 40 V_H , 25 D_H and 6 J_H gene segments along with the individual constant regions for each antibody isotype (Matsuda and Honjo, 1996). The RAG recombinase, made of the RAG1 and RAG2 subunits, randomly recombines a single

V_H , D_H and J_H segment in order to generate a functional variable region for the IgH (Figure 1.2) (Early et al., 1980; McBlane et al., 1995). This is done sequentially, starting with D_H to J_H recombination, followed by addition of a V_H segment (Alt et al., 1987). If successful, this process is then repeated at one of two IgL loci, λ (located on chromosome 22) or κ (located on chromosome 2). Unlike the IgH , the IgL contain only V and J segments (Barbie and Lefranc, 1998; Pallares et al., 1998).

Based solely on all possible recombinations, these processes can generate approximately 9.6×10^6 potential antibody specificities (Schroeder, 2006); however, further diversity is achieved by random modification of the DNA sequence at the gene segment junctions, during recombination (Alt et al., 1987). For example, terminal deoxynucleotidyl transferase can incorporate random nucleotides at the junctions, predominantly during recombination of the IgH locus (Gilfillan et al., 1993; Komori et al., 1993). Successful V-D-J or V-J recombination generates the variable region (IgV) for each respective locus.

Besides the limited number of B cells at any given time, the potential B cell repertoire will also be limited by selection steps that occur after successful rearrangement at both the IgH and IgL . The cells first undergo positive selection, which tests their ability to express a stable, functional BCR on the cell surface. Positive selection ensures that a functional B cell repertoire is available in the periphery (Rolink et al., 2001). The cells then undergo negative selection to test their ability to recognize self-antigen. B cells that recognize self-antigens either undergo receptor editing, i.e. further rearrangement to eliminate self-reactivity, or apoptose. Negative selection is critical to eliminate potentially auto-reactive B cells that could

initiate autoimmune responses; this process is also referred to as central B cell tolerance (Nemazee, 2017). B cells that successfully pass these selection steps emerge from the bone marrow as mature B cells.

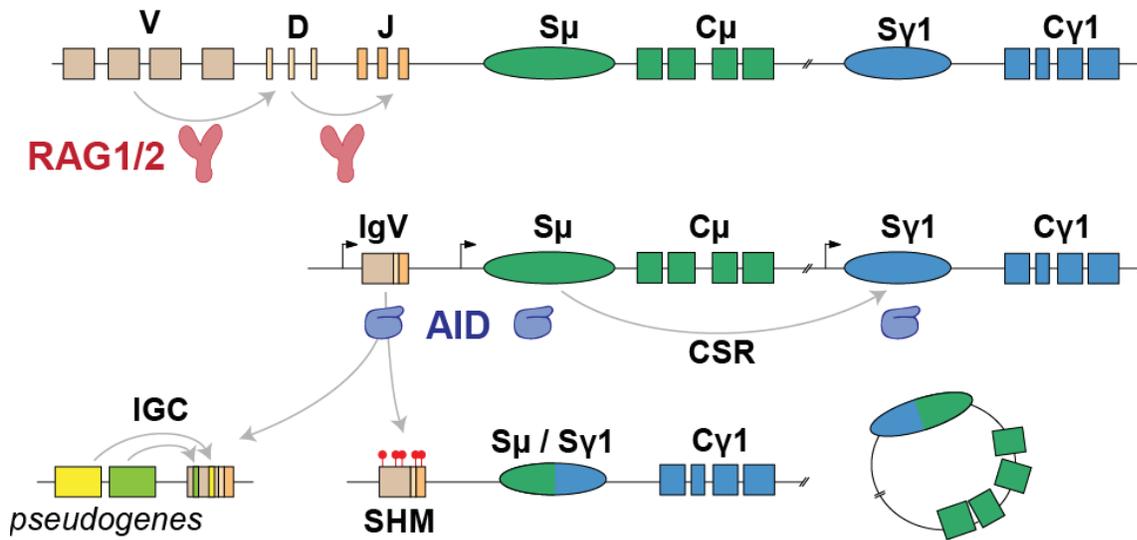


Figure 1.2. IgH locus diversification.

The IgH locus is composed of several V, D and J segments, followed by the constant regions that encode for the various antibody isotypes. RAG1 and RAG2 mediate the rearrangement of the V-D-J segments into a functional IgV region during B cell development in the bone marrow. In the periphery, activated B cells will express the enzyme AID, which can target and deaminate the IgV, as well as the S-regions that precede each constant region. AID deamination at the IgV can introduce single point mutations to generate diversity, this is called somatic hypermutation. In certain farm animals and birds, IgV deamination results in gene conversion, whereby regions from nearby pseudogenes are copied to the IgV. At the switch regions, processing of AID induced lesions can induce DSBs. Recombination of DSBs between different switch regions, (here S μ to S γ 1) will alter the constant region that is proximal to the IgV, and release the intervening DNA as a switch circle. This process is called class switch recombination and alters the antibody isotype. Adapted from (Di Noia and Neuberger, 2007).

1.1.4 The peripheral B cell response

After emerging from the bone marrow, B cells migrate to the secondary lymphoid organs where they accumulate in distinct follicles; these B cells are referred to as follicular B cells, and comprise the majority (~75%) of B cells in the body

(Westermann and Pabst, 1992). Follicular B cells are mobile, continuously circulating within the B cell follicle and between lymphoid organs to search for potential antigen. BCR recognition of certain antigens can directly stimulate proliferation and differentiation of B cells into antibody secreting cells. These antigens usually involved direct BCR stimulation with extensive cross-linking by repetitive antigens or by stimulation of Toll-like receptors (TLR). These responses are considered T-cell independent, as they result in rapid generation of antibody secreting cells (ASCs) without the need for T cell dependent signals (Mond et al., 1995).

For most protein antigens, a T cell dependent response is generally initiated by a low affinity recognition of antigen *via* the BCR, resulting in BCR crosslinking (Treanor, 2012) and initiating a signalling cascade (Woyach et al., 2012). These “activated B cells” migrate to the T cell border, and internalize the antigen-associated BCRs in order to process and present the antigen as peptides on their cell surface, in the context of MHCII (Yuseff et al., 2013). Association with CD4+ T cells by TCR recognition of the peptide-MHCII, leads to additional positive signals to the B cell, in the form of cell-cell contact through surface receptors and secreted cytokines (Parker, 1993). The combination of positive signalling from the BCR and T cell help induces rapid proliferation, generating cells that differentiate into plasmablasts, which will secrete low affinity antibodies, or that will initiate a germinal center (GC) reaction.

GCs are micro-anatomical regions within the B cell follicles, which can be both spatially, and functionally segregated into two distinct zones, the dark zone (DZ) and light zone (LZ) (Nieuwenhuis and Opstelten, 1984). A proportion of GC B cells transition between the DZ and LZ compartments. DZ B cells undergo rapid

proliferation while expressing high levels of the enzyme activation induced deaminase (AID) (Cattoretti et al., 2006). AID is necessary to initiate two distinct mechanisms, somatic hypermutation (SHM) and class-switch recombination (CSR), which genetically alter the antigen recognition domain and Fc region isotype, respectively. LZ B cells proliferate very little, but are in contact with GC specific T cells, called T_{FH} , and a network of specialized mesenchymal cells, called follicular dendritic cells, which act as reservoirs of antigen. LZ B cells undergo selection, as they acquire antigen and compete amongst each other for stimulating interactions from a limited number of T_{FH} cells (Victora et al., 2010). Positive interactions with T_{FH} cells drives B cells to re-enter the DZ and proliferate, whereas B cells that fail to interact with T_{FH} cells undergo apoptosis. Thus, cells that enter the LZ with higher antigen affinity will outcompete cells for T cell help, and will undergo additional rounds of DZ proliferation and antibody diversification. This creates a microenvironment that promotes progressively higher antibody affinity as the B cells undergo additional rounds of DZ, LZ transitions (Victora and Nussenzweig, 2012) (Figure 1.3). Once a GC B cell reaches a certain threshold of T cell stimulation, it can exit the GC in order to differentiate.

Exiting the GC, B cells have three possible fates, to become a memory B cell, a plasmablast, or a plasma cell (Benson et al., 2007). Memory cells do not secrete antibodies, and instead remain in the circulation for an extended period, poised to re-initiate a GC reaction or become ASCs if they encounter the same antigen (Phan and Tangye, 2017). Plasmablasts are short-lived ASCs, remaining near the GC from which they emerged, and secreting large amounts of antibody into the circulation in

order to rapidly clear the infection. Plasma cells on the other hand, can be long-lived ASCs, yet they home to the bone marrow and generate a continuous supply of antibody to generate a protective barrier to re-infection (Nutt et al., 2015).

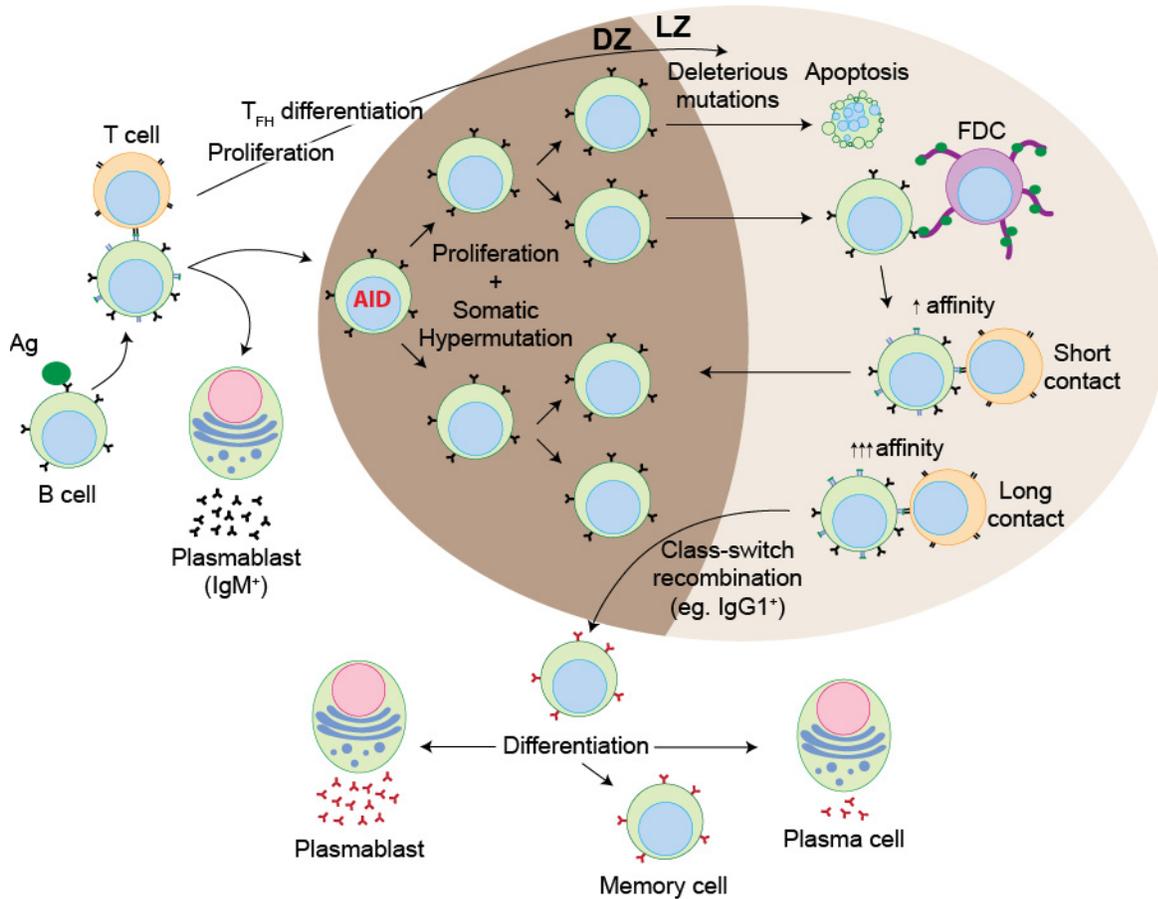


Figure 1.3. The germinal center reaction.

Antigen recognition in the periphery and subsequent interaction with antigen specific T-helper cells, stimulates B cells to undergo a burst of proliferation and to differentiate into plasmablasts, or initiate a germinal center reaction. In the GC, B cells within the dark zone proliferate and express AID to promote SHM. DZ B cells migrate to the light zone, where they can acquire antigen from FDCs and then present to T_{FH} cells. LZ B cells with improved antigen affinity will acquire more antigen and receive more signal from T_{FH} cells. Positive signals from T_{FH} drive B cells back to the DZ to undergo additional rounds of proliferation and diversification. B cells with reduced affinity are outcompeted and undergo apoptosis. After a threshold of signalling from T_{FH} cells is reached, B cells will undergo CSR, and can subsequently differentiate into plasmablasts, memory B cells, or plasma cells. T-helper cells that interact with B cells are also activated, and generate the GC T_{FH} cells.

1.2 Activation induced deaminase

1.2.1 The AID/APOBEC family

AID is part of a family of cytidine deaminases called the AID/APOBEC family, which represents a branch of the zinc-dependent deaminase superfamily (Conticello et al., 2005), defined by the specific catalytic zinc-coordination motif found in all members. The appearance of AID/APOBEC enzymes during evolution coincides with the appearance of adaptive immune systems in vertebrates, with AID believed to be the ancestral member of the family (Conticello et al., 2008; Rogozin et al., 2007).

Along with AID, the family comprises APOBEC1 (A1), APOBEC2 (A2), APOBEC3 (A3), APOBEC4 (A4) and APOBEC5 (A5) (Figure 1.4). Most of the AID/APOBECs can deaminate cytosine in DNA.

A1 has acquired additional RNA editing activity in mammals, but is a DNA deaminase in reptiles and birds, likely working in viral restriction (Severi et al., 2011). In mammals, the principal function of A1 seems to be mRNA editing, with the best example being the editing of apolipoprotein B (ApoB) pre-mRNA at position C6666, this generates a pre-mature stop codon leading to expression of truncated Apolipoprotein B protein (Navaratnam et al., 1993; Teng et al., 1993). In line with its activity, A1 is important for regulating lipoprotein metabolism, and in humans its expression is mostly restricted to the small intestine (Hadjigapiou et al., 1994; Morrison et al., 1996). Additional A1 mRNA editing targets have been identified, though the function of these editing events is unclear (Rosenberg et al., 2011; Skuse et al., 1996; Yamanaka et al., 1997).

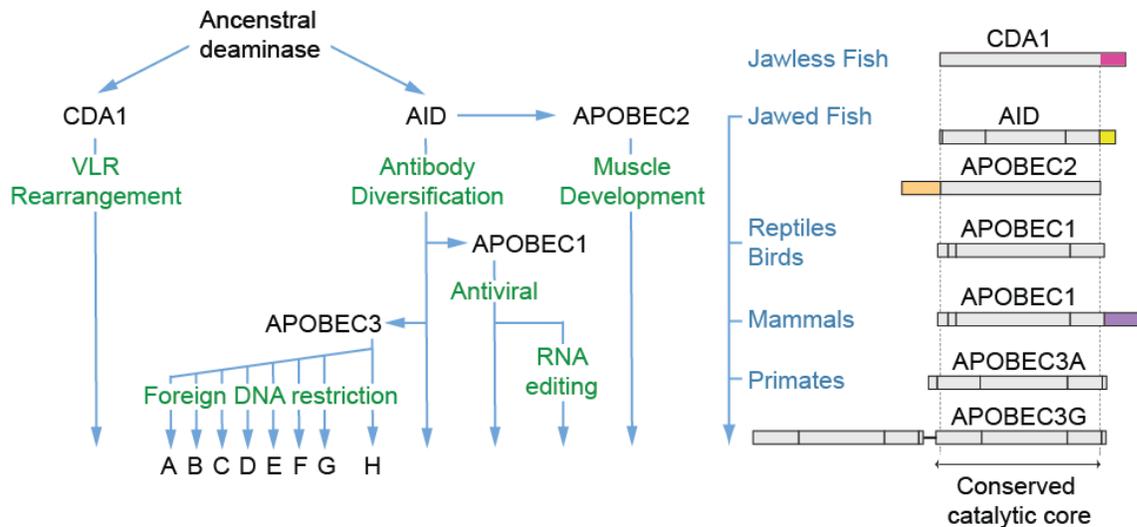


Figure 1.4. The AID/APOBEC family.

The AID/APOBEC family of deaminases emerged as adaptive immunity developed in vertebrates. In jawless fish, such as lampreys, the enzyme CDA1 has similar activity to AID, but is necessary for VLR rearrangement. AID is likely the ancestral member of the family in jawed vertebrates, though A2 emerged around the same period. Whether A2 has deaminase activity is unclear, but it seems to have a role in muscle development, at least in mice. A1 and A3 are the direct result of AID gene duplications. A1 originally emerged in reptiles as an antiviral factor, but in mammals acquired a unique C-terminal domain that promotes RNA editing activity. The A3 family, involved in foreign DNA restriction, emerged in mammals, and underwent a series of duplications in primates. The different A3s have distinct activity and structural organization, such as single- (A3A) or double- (A3G) deaminase domain configurations. Adapted from (Methot et al., 2015).

The activity of A2, A4 and A5 is still unknown. A2 seems to have a role in muscle development (Sato et al., 2010), A4 can enhance HIV replication (Marino et al., 2016) and A5 is a marsupial specific form of A1, lacking ApoB mRNA-editing function (Severi et al., 2011).

A3 has undergone massive genetic radiation, expanding from one A3 gene in mice to seven functionally and structurally distinct genes in humans, all clustered in tandem on chromosome 22. This expansion is believed to be a result of the important activity of the A3s as retrovirus and retrotransposon restriction factors (Harris and

Dudley, 2015). Expression of the A3s is found throughout the hematopoietic system and in various tissues, and can be induced by interferon in a protective manner; furthermore, T cells, the main target of HIV infection induce expression of A3s after activation (Koning et al., 2009; Refsland et al., 2010).

AID is highly conserved in all the jawed vertebrates (Conticello et al., 2005), while the jawless vertebrate, lamprey, encode a homologue of AID, called CDA1, despite the fact that they do not produce antibodies (Rogozin et al., 2007). Instead, lampreys produce distinct, highly variable genes, called variable lymphocyte receptors (VLRs) that are also secreted and act as antigen receptors (Guo et al., 2009; Pancer et al., 2004). Further, the reassembling of repetitive cassettes in order to generate VLR diversity is dependent on CDA1 (Rogozin et al., 2007). The exact nature of the AID and CDA1 ancestral functions are unclear as they emerged concomitantly with antibodies and VLR receptors, respectively; however, they share a common evolutionary origin and catalytic characteristics (Boehm et al., 2012; Quinlan et al., 2017). Despite strong conservation of AID among the jawed vertebrates, there are important evolutionary differences in the processes leading to antibody diversification. For example, bony fish do not undergo CSR, yet bony fish AID is capable of reconstituting such activity in AID deficient B cells (Barreto et al., 2005).

Despite their functional differences, the AID/APOBECs share significant sequence and a common structural fold, containing the catalytic core (Pham et al., 2016) (see below 1.3.1) (Figures 1.4 and 1.7). AID, A1 and A2 all have a single catalytic domain, with AID and mammalian A1 having unique C-terminal sequences important for their respective functions (Barreto et al., 2003; Severi et al., 2011;

Shinkura et al., 2004; Zahn et al., 2014). The A3s on the other hand are a mixture of single and double domain enzymes, with most double domain proteins having one catalytically active and one catalytically inactive domain; the exception A3B, has two active domains (Harris and Dudley, 2015).

The DNA deamination activity of AID, A1 and the A3s have also been correlated with mutations observed in various human tumours (Supek and Lehner, 2017). This observation demonstrates the risk associated with the mutagenic activity of these enzymes, and highlights the importance of regulatory mechanisms that have evolved to limit their activity.

1.2.2 AID expression

AID was initially discovered in a screening for genes that were specifically upregulated after B cell stimulation to induce CSR in the CH12F3 cell line (Muramatsu et al., 1999). AID transcripts were also found in secondary lymphoid organs, with high expression specifically in stimulated B cells *ex vivo*, and in GC B cells after immunization *in vivo* (Muramatsu et al., 1999). The fact that antibody diversification defects were the only obvious phenotypes in both AID-deficient mice and patients, strongly supported that its role was mostly restricted to antibody diversification in activated B cells (Muramatsu et al., 2000; Revy et al., 2000). In human tonsils, AID protein expression was further shown to be highest in DZ B cells of the GC, where SHM and CSR are believed to occur (Cattoretti et al., 2006). A transgenic mouse model, carrying a BAC containing the AID gene locus, *Aicda*, with an integrated GFP tag fused to the last exon, confirmed that expression of AID

transcripts and protein was almost completely restricted to activated B cells (Crouch et al., 2007). Importantly, AID expression is selective, with an almost complete absence of AID in naïve cells, preceding the GC reaction, and in plasma cells and memory cells, following the GC reaction (Crouch et al., 2007).

In B cells, AID expression is stimulated at the transcriptional level by signals emanating from cytokine receptors, antigen-dependent crosslinking of the BCR, as well as stimulatory interactions with T cells, such as CD40/CD154 (Zan and Casali, 2013). These stimuli cause rapid induction of *Aicda* mRNA, and AID protein. The tight regulation of AID expression is mediated by regulatory elements at the *Aicda* locus, separated into 4 regulatory regions (Zan and Casali, 2013). The details of *Aicda* transcription are not relevant for this work, but it's important to note that although AID expression is tightly regulated, it can be induced by multiple stimuli, acting *via* different transcription factors, such as NF- κ B, HoxC4 and E2A (Zan and Casali, 2013). The levels of AID expression are critical to its function, as heterozygous *Aicda*^{+/-} mice show reduced levels of SHM and CSR (Sernández et al., 2008; Takizawa et al., 2008). There is also evidence that AID is limiting for CSR, as cells overexpressing transgenic AID have increased isotype switching (Robbiani et al., 2009).

Fine-tuning of AID expression is also achieved post-transcriptionally. The micro RNAs mir181b and mir155 have been shown to associate with *Aicda* mRNA, leading to translational silencing and mRNA degradation (de Yébenes et al., 2008; Dorsett et al., 2008; Teng et al., 2008). Interestingly, mir155 is downregulated by Bcl6, the master transcription regulator of the GC reaction (Basso and Dalla-Favera, 2010),

thereby contributing to GC-specific expression of AID (Basso et al., 2012). Further regulation of AID at the protein level exists and will be discussed in section 1.5.

Outside of the activated or GC B cell compartments, AID has also been found to be expressed in developing bone marrow B cells, as well as in plasmablasts (Crouch et al., 2007). A physiological role for AID in plasmablasts is not known but may represent residual AID expression from the GC reaction or AID expression in extrafollicular plasmablasts, which undergo CSR (Di Niro et al., 2015). AID expression in the bone marrow during B cell development on the other hand is necessary for efficient central B cell tolerance (Kuraoka et al., 2011; Meyers et al., 2011).

AID expression is also regulated by the hormones estrogen and progesterone (Pauklin and Petersen-Mahrt, 2009; Pauklin et al., 2009). Estrogen was found to increase AID mRNA and protein expression in B cells as well as ovarian and breast tissue, while progesterone inhibited AID expression in B cells (Pauklin and Petersen-Mahrt, 2009; Pauklin et al., 2009). Hormonal regulation may be related to AID expression in the ovaries (Morgan et al., 2004), with a potential role in epigenetic reprogramming (Popp et al., 2010).

Expression in other tissues, such as the stomach and liver has been linked to inflammation, and though the physiological role of such expression is unclear, the oncogenic consequences are evident (Shimizu et al., 2012). These pathological roles of AID, beyond antibody diversification, will be further described in Section 1.4.

1.2.3 Deamination of deoxycytidine on ssDNA

The relevance of AID for antibody diversification was first demonstrated studying B cells from mice lacking AID, as they were incapable of undergoing either SHM or CSR (Muramatsu et al., 2000). At the same time, inactivating mutations of AID were found in humans with a rare immunodeficiency syndrome, called hyper-IgM syndrome type 2 (HIGM2), which is characterized by a deficiency for both SHM and CSR (Revy et al., 2000). The homology of AID with APOBEC1, the only other known family member at the time, led to speculation that AID was an mRNA-editing deaminase (Muramatsu et al., 2000). This model was quickly challenged by the fact that AID expression led to uracils in the DNA, as demonstrated by the accumulation of C to T mutations in E coli lacking the Uracil DNA glycosylase (UNG) (Petersen-Mahrt et al., 2002) and a predictable change to the pattern of SHM after inhibiting UNG in chicken DT40 B cells (Di Noia and Neuberger, 2002). Furthermore, despite strong RNA binding affinity, AID cytidine deaminase activity *in vitro* was shown to be specific for ssDNA (Bransteitter et al., 2003; Dickerson et al., 2003).

Further confirmation of direct AID deaminase activity at the *Ig* came from the fact that AID preferentially targets WRC (W = A/T, R = A/G) motifs when deaminating ssDNA *in vitro*. This correlates perfectly with the WRCY hotspots (Y = C/T) that are preferentially mutated during SHM (Beale et al., 2004; Pham et al., 2003). The hotspot preference of AID is actually conferred by a substrate recognition loop, spanning residues 113-123 of AID, with similar loops in the A3s providing differential target sequence preference (Kohli et al., 2009; Langlois et al., 2005). Specificity seems to be dictated by the size and shape of the loop, with the larger loop in AID

being permissive to a large purine residue preceding the C (Pham et al., 2016). Further proof of DNA deamination during antibody diversification was demonstrated by the fact that swapping AID and A3 recognition loops, generating an AID with altered mutagenic site preference, results in an altered spectrum of mutations during SHM in B cells (Wang et al., 2010). Definite proof was demonstrated by direct measurement of uracil at the Ig loci of B cells expressing AID, but not in AID deficient B cells (Maul et al., 2011).

Despite AID activity being focused at the Ig loci, aberrant SHM at other loci has been described in B cell lymphomas (Pasqualucci et al., 2001) and in normal cells (Liu et al., 2008; Shen et al., 1998), suggesting that AID can produce off-target deamination with some frequency. It was immediately appreciated that an enzyme capable of generating point mutations and double strand breaks in the genome would require significant regulation (de Yebenes and Ramiro, 2006). The relevance and mechanisms for this regulation will be discussed in sections 1.4 to 1.6.

1.3 Antibody diversification by DNA deamination

The DNA deaminase model stipulates that uracil in the DNA, generated by AID, can act as the starting point for SHM and CSR, along with Ig gene conversion (IGC), see below (Petersen-Mahrt et al., 2002) (Figures 1.2 and 1.5). DNA deamination by AID, transforms a C:G pair into a U:G mismatch, which can then remain unprocessed, be processed by UNG as part of the uracil base excision repair (BER), or be recognized by MutS α (comprising MSH2 and MSH6) a member of the mismatch repair (MMR) pathway (Petersen-Mahrt et al., 2002; Rada et al., 2004; Rada et al., 1998; Rada et al., 2002b). Non-canonical processing of AID-induced uracils can generate the spectrum of mutations observed during SHM, as well as the dsDNA breaks (DSBs) necessary for IGC and CSR (Methot and Di Noia, 2017).

UNG is able to detect uracils in either ss- or ds-DNA, and removes the uracil base from DNA, leaving an abasic site (Krokan and Bjoras, 2013). During normal BER, the abasic site is cleaved by an apurinic/pyrimidinic endonuclease (APE), generating a ssDNA nick. The scaffold XRCC1 then coordinates recruitment of the high-fidelity DNA polymerase β (Pol β) and a ligase, to faithfully replace the residue and seal the nick (Robertson et al., 2009).

MutS α can detect uracil in dsDNA only in the context of a U:G mismatch (Jiricny, 2013; Larson et al., 2008). During normal MMR, MutS α recruits MutL α (composed of MLH1 and PMS2) which has endonuclease activity (*via* PMS2) and can generate a ssDNA nick 5' of the mismatch. The exonuclease EXO1 is recruited to the nick in order to degrade a patch of DNA 5'-3', passing over the mismatch to remove

the erroneous base. PCNA and a high-fidelity DNA polymerase are recruited to fill the gap and a ligase will finalize the repair (Jiricny, 2013).

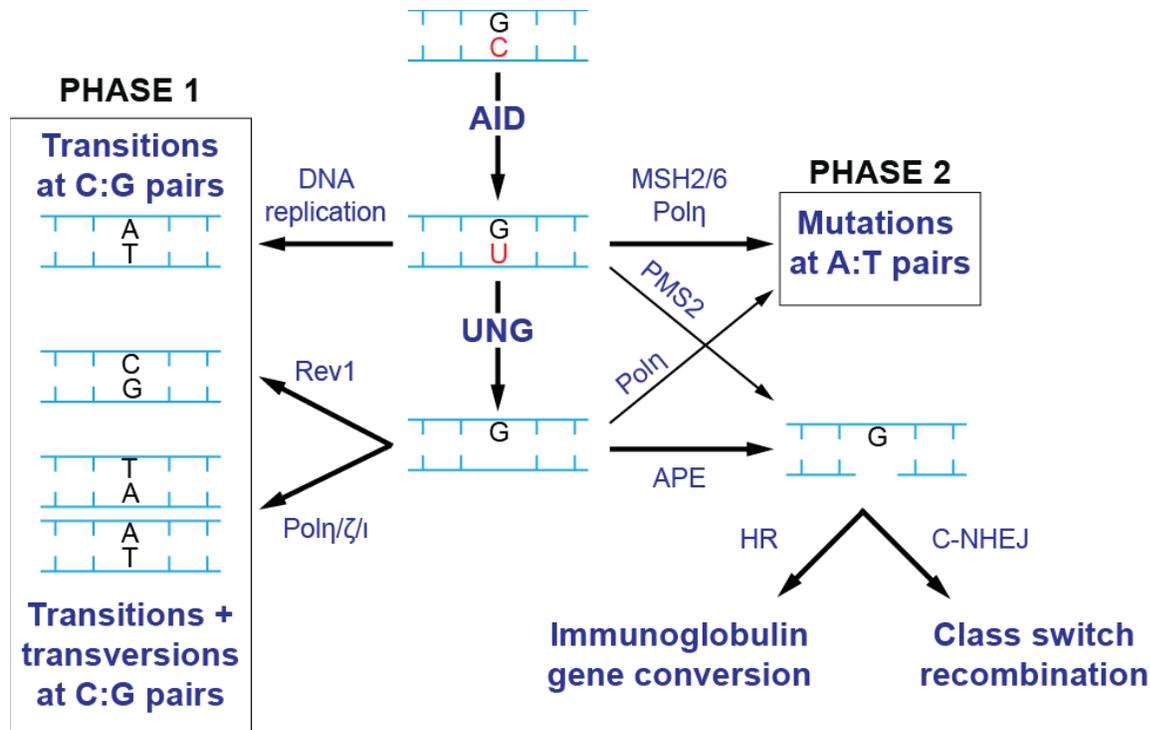


Figure 1.5. The DNA deaminase model for antibody diversification.

The DNA deaminase model, proposes that the AID dependent deamination of cytosine to uracil in DNA, can generate SHM, IGC and CSR if processed properly. Adapted from (Petersen-Mahrt et al., 2002). Phase 1 mutations, at the target C:G pair, occur either by DNA replication over the uracil, or downstream of UNG glycosylase activity. Phase 2 mutations, at A:T pairs, occur mostly downstream of the MMR machinery (MSH2 and MSH6), and require the low-fidelity DNA polymerase η . There is evidence that some phase 2 mutations do occur downstream of UNG. APE generates ssDNA nicks, at UNG dependent abasic sites. If two ssDNA nicks are generated on opposite strands, and in proximity, they can generate DSBs, which are necessary for IGC and CSR. DSBs can also be generated by the endonuclease PMS2, recruited by MSH2/MSH6; however, this is less efficient for generating DSBs during CSR, and does not contribute to IGC.

Interestingly, haploinsufficiency for XRCC1 or Pol β results in increased mutation levels during SHM, suggesting that canonical repair does occur with some frequency at the *Ig* (Saribasak et al., 2011; Wu and Stavnezer, 2007). Nonetheless,

perturbation of these repair mechanisms must occur during SHM, IGC and CSR, as will be discussed below (Figure 1.6).

1.3.1 Somatic hypermutation

SHM is the process of introducing single point mutations within the IgV of both the *IgL* and *IgH* loci (Figure 1.2). Within the GC, mutations that increase antibody affinity are selected by T_{FH} help, while mutations that reduce affinity or alter BCR stability and/or expression lead to apoptosis. These mutations are not randomly distributed, with a typical GC reaction generating around 10-20 mutations, clustered within the CDR loops in order to alter the antibody-antigen binding affinity (Tiller et al., 2007). Many of the mutations occur at WRCY AID hotspots, which also seem to be enriched at the IgV regions and specifically within the CDRs (Wagner et al., 1995).

In the DNA deaminase model for antibody diversification, SHM can occur in two phases. The first phase generating mutations at the deaminated cytosine residues, and the second spreading mutations to A:T base pairs, which are not AID targets (Petersen-Mahrt et al., 2002).

The first phase can occur in two ways. Either the U:G mismatch remains unprocessed, resulting in C:G to T:A transition mutations after DNA replication, as demonstrated in cells that lack both UNG and MSH2 (Rada et al., 2004). Or, if UNG generates an abasic site, DNA replication will result in recruitment of translesion polymerases that will insert residues based on their preference, which can generate either transition or transversion mutations. REV1 for example can only insert dCMP into

DNA, and will therefore create C:G to G:C transversions (Jansen et al., 2006; Simpson and Sale, 2003).

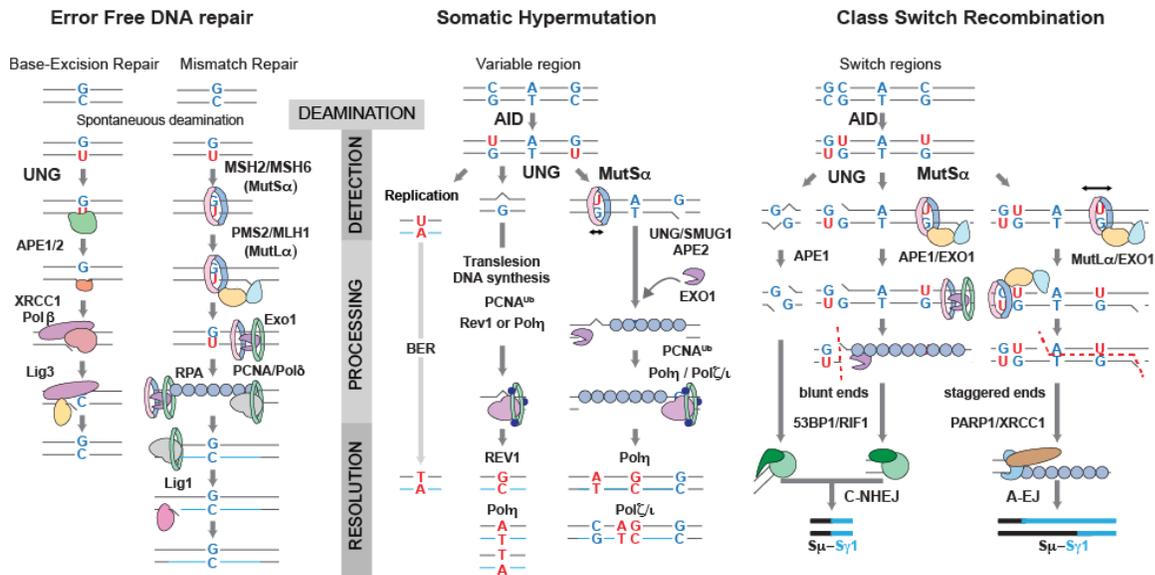


Figure 1.6. Downstream DNA repair mediating SHM and CSR.

(Left) Canonical pathways for BER, left, and MMR, right. Canonical uracil excision repair is initiated by UNG, creating an abasic site that is recognized by APE1 or APE2. APE1/2 nick the DNA 5' of the abasic site. XRCC1 is recruited to the nick and recruits Polβ. Polβ can remove the remaining 5' deoxyribose and insert a single nucleotide, followed by ligation. In MMR, MutSα recognizes the U:G mispair and recruits MutLα, which nicks the DNA 5' of the mismatch via PMS2. The PCNA-associated EXO1 5'-3' exonuclease activity creates an extended patch of ssDNA from the nick going past the mismatch site. PCNA recruits Polδ to replicate over the patch and Ligase 1 finalizes repair. (Right) During SHM, uracil can act as a template for replication leading to a C-T transition mutation. Alternatively, non-canonical BER leads to transition and transversion mutations. MutSα initiates mutagenic repair affecting A:T pairs. Both pathways recruit low fidelity polymerases (Polη,ζ,ι) through PCNA ubiquitination (PCNA-Ub). For CSR, non-canonical BER can lead to DSBs when two uracils in opposite strands are closely spaced. MMR can process distantly spaced uracils, leading to staggered DSBs. DSBs can also be produced by the collaboration of BER and MMR. Blunt DSBs are joined by C-NHEJ during CSR, whereas staggered breaks are preferentially repaired by A-EJ. Adapted from (Methot and Di Noia, 2017).

The second phase of mutations requires MutSα, as evidenced by the almost complete lack of A:T mutations in MSH2 or MSH6 deficient mice (Rada et al., 1998;

Wiesendanger et al., 2000). Generation of mutations requires recruitment of an error prone DNA polymerase during gap filling. Pol η , which has a very high error rate of 3.5×10^{-2} bases compared to 10^{-6} to 10^{-4} for replicative DNA polymerases (Matsuda et al., 2000), is critical for phase two mutations. A:T mutations are almost completely absent in Pol η -deficient humans and mice (Delbos et al., 2005; Wilson et al., 2005; Zeng et al., 2001). In the absence of Pol η , the polymerases Pol ζ and Pol ι are able to generate some A:T mutations (Maul et al., 2016; Saribasak et al., 2012). UNG also seems to have a small role in phase two mutations, as *msh2*^{-/-} mice maintain some A:T mutations (Rada, 2004). This likely involves a version of long-patch BER, dependent on Pol η (Delbos et al., 2007; Robertson et al., 2009; Shen et al., 2006).

It is still unclear exactly why BER and MMR recruit error prone DNA polymerases during SHM; however, one hypothesis is that the localized, high density of uracils generated by AID could force BER and MMR to work in proximity, interfering with one another. The details of this collaboration are beyond the scope of this thesis, but were discussed in a review I wrote recently (Figure 1.6) (Methot and Di Noia, 2017).

1.3.2 Immunoglobulin gene conversion

In many farm animals, and notably in chickens, primary antibody diversification only generates a limited repertoire, due to a small repertoire of functional V, D, and J segments as well as a low incidence of junction diversification (Ratcliffe, 2006). In order to expand the antibody repertoire, a second diversification event, Immunoglobulin gene conversion (IGC), takes place during B cell development in the Bursa of Fabricius, prior to antigen encounter, as well as in the B cell follicles and

GC after antigen encounter (Figure 1.2) (Weinstein et al., 1994). This process, initiated in a similar fashion to SHM, requires AID deaminase activity at the Ig V-regions (Arakawa et al., 2002). However, during IGC uracils are processed into DSBs and V-region diversity is then generated using the homologous recombination (HR) DNA repair machinery, and a series of pre-rearranged V-region pseudogenes, which act as templates during the repair process. This mechanism thereby incorporates stretches of pseudogene sequence into the V-region in order to create diversity. Interestingly, during the GC reaction, chicken B cells use a mix of IGC and SHM to further improve antibody affinity, generating large changes (*via* IGC) and fine-tuning affinity (*via* SHM) (Tang and Martin, 2007).

The process of IGC has been extensively studied using the DT40 chicken B cell line, which is highly amenable to genetic manipulations (Arakawa and Buerstedde, 2006). Using DT40 cells, the role and the molecular mechanisms of HR during IGC was demonstrated, for example, HR inhibition by XRCC2 ablation converts IGC into SHM in DT40 cells (Sale et al., 2001). Similarly, SHM instead of IGC occurs in cells deficient for UNG (Di Noia and Neuberger, 2002; Saribasak et al., 2006) or in cells in which the donor pseudogenes cluster is deleted (Arakawa et al., 2004). The molecular mechanism for IGC has been well characterised (Tang and Martin, 2007). In the lab, we use DT40 cells deficient in XRCC2 or with a complete deletion of the pseudogene repertoire to study SHM. Reconstitution of AID-deficient versions of these cell lines with wt AID or AID variants permits structure function studies during SHM (Arakawa et al., 2004).

1.3.3 Class switch recombination

Within the IgH locus, the constant region for each isotype is preceded by an independent promoter that drives transcription of repetitive regions, called switch regions (S-regions) (Figure 1.2). CSR specificity to a particular isotype is driven by transcription of the respective S-region, with each responding differentially to the B cell microenvironment; thus, appropriate antibody isotypes are generated in the appropriate setting (Stavnezer and Schrader, 2014). For example, CSR in the Peyer's patches is primarily to IgA, as these antibodies will be secreted into the gut to regulate the microbiota. As AID activity is dependent on transcription, specific transcription induction directs AID deaminase activity to the appropriate S-region, in order to initiate the CSR mechanism.

CSR occurs via DSBs, so unlike SHM, it cannot proceed without recognition and processing of the AID generated uracil, as demonstrated in *ung^{-/-} msh2^{-/-}* mice (Rada et al., 2004). Nonetheless, UNG seems to have a more important role, as *ung^{-/-}* mice have an almost complete loss of CSR, while *msh2^{-/-}* mice retain ~50% of CSR activity (Ehrenstein and Neuberger, 1999; Rada et al., 2002b). During CSR, UNG and MutS α function may be spatially separated, with UNG favoring S-region uracils, where there is a high density of WRCY motifs, and MutS α favoring uracils upstream of the S-region, which is less dense in WRCY (Cortizas et al., 2013). This functional separation may explain the different importance for UNG and MSH2 during CSR. Furthermore, this distinction may be important for CSR as UNG can generate closely spaced nicks in order to generate a DSB, while MMR processing of two distant uracils can lead to a DSB. In any case, as in SHM, efficient CSR works best when BER and MMR work

together (Figure 1.6), as demonstrated by a shift in the break sites, becoming focused at hotspot motifs, in the absence of MSH2 (Ehrenstein and Neuberger, 1999).

Efficient CSR requires not only the generation of DSB, but also efficient repair *via* the classical non-homologous end joining (C-NHEJ) DNA repair pathway. The importance of C-NHEJ for CSR is demonstrated by the critical role for its components, such as Ku70, Ku80, DNA ligase IV and 53BP1, (Casellas et al., 1998; Manis et al., 1998; Manis et al., 2004; Pan-Hammarstrom et al., 2005; Ward et al., 2004). 53BP1 is particularly important for CSR, as it may also enhance formation of the synapse between S-regions and coordinate the orientation of the breaks during repair to ensure productive recombination (Dong et al., 2015; Reina-San-Martin et al., 2007). In the absence of 53BP1 or essential members of the C-NHEJ pathway, CSR still occurs but uses an alternative NHEJ (A-EJ) mechanism (Bothmer et al., 2010; Yan et al., 2007). A-EJ is dependent on sequence microhomology to promote joining, and is dependent on the factors XRCC1, CtIP and ATM (Bothmer et al., 2010; Han et al., 2012; Lee-Theilen et al., 2011; Saribasak et al., 2011). Interestingly, the breaks leading to A-EJ seem to be generated by the MMR machinery, and may be both spatially and temporally distinct from the breaks that result in C-NHEJ (Cortizas et al., 2013; Eccleston et al., 2011). Though A-EJ can compensate for C-NHEJ, balance between these repair mechanisms is important, as A-EJ promotes an increased frequency of chromosomal translocations (Iliakis et al., 2015).

1.4 The clinical relevance of AID

AID is a double-edged sword. It is absolutely required for an efficient humoral immune response, and also seems to have roles in preventing autoimmunity, as an antiviral factor, and in DNA demethylation. On the other hand, AID expression is linked with severity of autoimmune responses, and can promote oncogenic transformation and progression. Thus, regulation of AID is critical in order to permit its important functions while limiting its potential side effects.

1.4.1 AID in immunity

The effects of *Aicda* mutations in HIGM2 patients demonstrate the critical role of AID for the humoral immune response. In general, HIGM patients are characterized by normal or elevated serum IgM levels, with an absence of IgG, IgA and IgE (Notarangelo et al., 1992). This defect is often detected in children, as patients present with recurrent infections, and a deficiency to respond to immunizations. Prior to the identification of *Aicda* mutations in HIGM2, the defining characteristic of HIGM2 was the absence of typical CD40 signalling defects observed in B cells of HIGM patients (Durandy et al., 1997).

More generally, the importance of AID for immune responses can be demonstrated in mice, by the reduction of AID expression levels with age, which correlates with reduced CSR (Frasca et al., 2004). A potential consequence of reduced AID expression with aging in humans has been demonstrated, as AID expression decreases with age, and correlates with the efficiency of the humoral response to H1N1 vaccination (Khurana et al., 2012).

AID may also have a role in innate immune defence against viruses. Though in mammals this is the role of A3s, it is possible that AID retains this activity as a functional relic from the ancestral DNA deaminase. For example, like the A3s AID can restrict the movement of retrotransposable elements, though this is independent of its deaminase activity (MacDuff et al., 2009). Furthermore, AID expression can be induced by proinflammatory cytokines in hepatocytes, and restricts hepatitis B virus infection by deaminating the viral genome (Qiao et al., 2016; Watashi et al., 2013).

1.4.2 AID and autoimmunity

An additional phenotype in HIGM2 patients is an enhanced predisposition to autoimmune conditions (Quartier et al., 2004). In fact, AID is key to efficient central and peripheral B cell tolerance, as auto-reactive B cells are not efficiently removed during negative selection in AID deficient mice and humans (Kuraoka et al., 2011; Meyers et al., 2011). The role of AID in tolerance can explain the low level of AID expression observed in the bone marrow, as only auto-reactive B cells induce sufficient AID expression. In these cells, AID seems to generate cytotoxicity, and promotes apoptosis of the auto-reactive B cells (Cantaert et al., 2015).

Autoantibodies contribute to many autoimmune pathologies, such as rheumatoid arthritis, Sjögren's syndrome and systemic lupus erythematosus (Suurmond and Diamond, 2015). In patients, AID is highly expressed in ectopic GCs linked to autoimmune reactions (Bombardieri et al., 2007), and AID expression can be correlated with disease severity (Xu et al., 2009). In mice, increased AID expression has been linked to the development of autoreactive antibodies (Hsu et al.,

2007), and in a mouse model of lupus, autoantibodies show high levels of SHM (Zan et al., 2009). Accordingly, AID deficiency results in disease improvement, indicating that AID-dependent SHM and CSR activities could exacerbate established autoimmune reactions (Hsu et al., 2011; Jiang et al., 2007).

Furthermore, women are more prone to develop autoimmunity and would be expected to express more AID, based on the regulation of AID expression by estrogen and progesterone (Pauklin and Petersen-Mahrt, 2009; Pauklin et al., 2009). This has not been experimentally analyzed, but if this was the case, AID expression could influence this sex-bias in autoimmunity.

1.4.3 AID is oncogenic

AID is the only known eukaryotic enzyme that functions by mutating the genome, and this genome-wide targeting is not without consequences. The activity of AID is likely one reason why B cell malignancies are significantly over-represented compared to T cell malignancies. Many lines of evidence have demonstrated AID oncogenic activity both in animal models and human patients.

In mice, ubiquitous overexpression of AID leads to tumorigenesis, with most tumours being T cell lymphomas (Okazaki et al., 2003). Specific overexpression of AID in B cells can only promote spontaneous transformation if p53 is concomitantly deleted, suggesting that B cells undergo distinct regulation to prevent transformation (Robbiani et al., 2009). Nonetheless, mouse models of B cell malignancies demonstrate a clear role for AID in both the development and progression of B cell lymphomas and

leukemias (Gruber et al., 2010; Kotani et al., 2007; Montamat-Sicotte et al., 2015; Pasqualucci et al., 2008; Robbiani and Nussenzweig, 2013; Swaminathan et al., 2015).

AID is also expressed in a number of human B cell neoplasms, correlating with poor prognosis (Feldhahn et al., 2007; Greeve et al., 2003; Leuenberger et al., 2010; McCarthy et al., 2003; Swaminathan et al., 2015).

The role of AID in promoting malignancy is primarily due to its ability to target and deaminate genes outside the Ig locus (Casellas et al., 2016), which occurs with some frequency in normal mouse B cells undergoing antibody diversification (Hakim et al., 2012; Klein et al., 2011; Liu et al., 2008; Meng et al., 2014; Qian et al., 2014; Staszewski et al., 2011b). Off-target SHM by AID can be enhanced in several loci in the absence of UNG and MSH2, which indicates that canonical repair occurs predominantly at these non-Ig sites (Liu et al., 2008). AID off-target activity can trigger oncogenic events that are frequently found in human B cell neoplasms, such as the IgH-cMyc translocation that causes Burkitt's lymphoma (Ramiro et al., 2004; Robbiani et al., 2008) and the cMyc-mir142 translocation which can likely cause aggressive B cell leukemia (Gauwerky et al., 1989; Robbiani et al., 2009). In human patients, hypermutation also occurs at proto-oncogenes such as *BCL-6*, *PAX5* and *PIM1*, with mutation patterns similar to AID SHM at the Ig loci (Liso et al., 2006; Pasqualucci et al., 2001; Shen et al., 1998). Similar mutations can be detected in mouse models of lymphoma, and are dependent on AID expression (Pasqualucci et al., 2008). AID can also promote resistance to chemotherapy, as AID mutations of the BCR-ABL1 oncogene can

accelerate Imatinib resistance in a mouse model of chronic myelogenous leukemia (Klemm et al., 2009).

Chronic infections can also favour oncogenic AID activity in B cells, with direct links for AID induction and oncogenic activity during Epstein bar virus and malaria infections (Epeldegui et al., 2007; He et al., 2003; Robbiani et al., 2015).

In other tissues, inflammation can induce aberrant AID expression, potentially via NF- κ B, and has been linked to the development of various cancers (Endo et al., 2007; Shimizu et al., 2012). Notably, *H. pylori* infection has been linked to stomach inflammation and subsequent induction of AID expression, thereby promoting the development of stomach cancer (Endo et al., 2008; Matsumoto et al., 2007).

1.4.4 DNA demethylation

AID deaminase activity has been demonstrated *in vitro* on 5-Methylcytosine (5mC), an abundant epigenetic mark involved in gene silencing (Morgan et al., 2004). This, along with the expression of AID in mouse ovaries correlating with pluripotency genes, led to the hypothesis that AID may have a role in epigenetic reprogramming. Demethylation by AID would entail, deamination of 5mC by AID to create a C to T transition. Recognition of the ensuing T:G mismatch by a glycosylase, such as MBD4 or TDG, would promote repair, and insertion of a new, unmethylated cytosine. A functional role for AID in active demethylation has been demonstrated in zebrafish embryos (Rai et al., 2008), and in mammalian cells during epigenetic reprogramming (Popp et al., 2010), during iPSC generation (Bhutani et al., 2010), and may underpin the observed need for AID in epithelial-mesenchymal transition in mammary cell lines

(Muñoz et al., 2013). In B cells, *in vitro* activated *Aicda*^{-/-} cells have an identical transcriptome and methylation patterns compared with wildtype cells (Fritz et al., 2013); yet, *Aicda*^{-/-} GC B cells show differences in methylation patterns, specifically at genes involved in the GC program (Dominguez et al., 2015). Interestingly, AID-dependent demethylation in GC B cells may overlap with off-target mutagenic activity of AID, suggesting that AID off-target mutations may correlate with a physiological role in DNA demethylation (Kenter et al., 2016).

More thorough study of AID activity on methylated cytosines led to the observation that steric hindrance from base modifications significantly disrupts the efficiency of AID dependent deamination, with deamination of 5mC being 10 fold lower than cytosine (Nabel et al., 2012). The physiological importance of AID in DNA demethylation has also been questioned, as AID deficient mice seem to develop normally, which would be unlikely if additional demethylation mechanisms did not exist (Kohli and Zhang, 2013). In the end, AID likely has a niche role in DNA demethylation, as recent work has identified a clear and important role for the TET family of proteins, which can oxidize 5mC in order to generate products that are recognized and removed by the glycosylase TDG (Wu and Zhang, 2017).

1.4.5 Gene editing

CRISPR based genome-editing is a burgeoning field, with particular utility for generating gene insertions and gene silencing (Komor et al., 2017). In order to bypass the need for DSB intermediates, recent work has demonstrated the utility of fusing AID or APOBECs to catalytically inactive Cas9 (the catalytic subunit in

CRISPR technology) in order to generate targeted deamination tools. Targeting these fusion proteins using CRISPR guide RNAs and boosting mutagenic activity by inhibition of UNG, has generated potent and site-specific deamination activity. The technology has already gone through a number of iterations, but is still limited by off-target activity, possibly due to the intrinsic targeting of AID (Komor et al., 2016; Nishida et al., 2016; Shimatani et al., 2017; Yang et al., 2016). In any case, this technology is an exciting use for AID deamination activity, and demonstrates an additional relevance to understanding the mechanisms that regulate AID targeting and activity.

1.5 Regulation of AID

1.5.1 Structural determinants of AID activity

AID is a small, 198 amino acid protein, generating a globular catalytic core that is highly conserved with other APOBEC family members. The C-terminal domain, comprising 17 amino acids encoded by the *Aicda* exon 5 (E5), are unique to AID. The first 29 residues, which are sufficient to induce nuclear import of GFP (Ito et al., 2004), are often referred to as the N-terminal domain.

A 3D structure is essential to understand the physical constraints of an enzyme. The first crystal structure of an APOBEC enzyme was generated 10 years ago, and permitted the modelling of AID based on homology (Prochnow et al., 2007). Subsequently, other APOBECs were crystalized, either whole or truncated, revealing striking similarity despite significant sequence differences (Bohn et al., 2013; Chen et al., 2008; Holden et al., 2008; Kitamura et al., 2012). Based on these APOBEC structures, AID has been modelled with high confidence giving insight into its biochemistry and regulation (King et al., 2015; Patenaude et al., 2009).

Recently, two AID variants have been crystalized (Pham et al., 2016; Qiao et al., 2017); however, the biologically important N and C terminal domains were modified, or removed entirely, to allow purification and crystallization of the protein. These structures have revealed that like the APOBECs, the AID core domain is formed by a series of α -helices and β -sheets that together support the zinc-binding catalytic pocket (Figure 1.7) (Pham et al., 2016). This places the key catalytic residues in a conformation to accept deoxycytidine in the appropriate orientation to allow deamination. The structure of AID

bound to dCMP also demonstrates that DNA specific contacts in the catalytic site are likely to account for its inability to deaminate RNA (Qiao et al., 2017). Furthermore, AID seems to have two distinct DNA binding grooves, which favour deamination activity on branched and G-quadruplex (G4) DNA, which may have a particularly important role during CSR (Qiao et al., 2017).

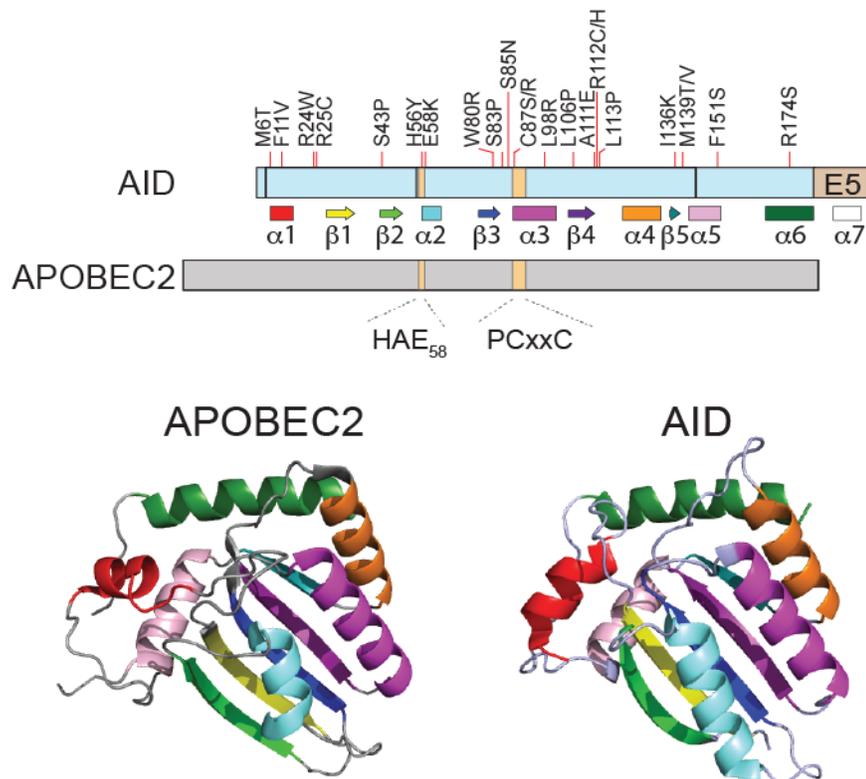


Figure 1.7. The AID/APOBEC deaminase domain is highly conserved.

(Top) Schematic of AID and APOBEC2 structural organization, with the conserved catalytic site residues indicated (orange and under the scheme). Individual AID exons are marked by black lines, with exon 5 labelled. Single point mutations from HIGM2 patients are indicated (above), and highlight how changes throughout the deaminase domain can affect its function (Mahdaviyani et al., 2012). Secondary structure elements are indicated and colour coded. (Bottom) Experimental structures of AID (PDB 5JJ4) and APOBEC2 (PDB 2NYT), coloured according to the scheme above, clearly demonstrate the structural conservation of the deaminase domain (Pham et al., 2016; Prochnow et al., 2007).

Figure 1.8. AID structural elements and regulatory features.

Mechanisms that positively (green) or negatively (red) affect AID function are indicated, along with the protein factors that mediate them (black). Drugs that can indirectly affect AID function by inhibiting or activating these factors are indicated (grey) connected to their target. AID structures are shown in the middle: Top, experimental structure of the AID catalytic core (PDB 5JJ4) (Pham et al., 2016). The catalytic site (teal), specificity loop (blue), catalytic pocket (grey surface) and Zn⁺⁺ ion (orange ball) are highlighted. Note that the N-terminal domain (pink) was altered to permit solubility and the C-terminus was truncated. Bottom, a model of AID, based on the APOBEC3C structure, including the E5 domain (pink) and native N-terminus. Residues D187 and D188 (blue) necessary for cytoplasmic retention and the hydrophobic nuclear export sequence (NES – dark green), are highlighted. Phosphorylation sites are indicated with balls. Adapted from (Methot and Di Noia, 2017).

Many AID mutations have been identified in HIGM2 patients, revealing residues critical for AID structure and catalytic pocket integrity (Figure 1.7) (Mahdavian et al., 2012). These have helped to understand AID function and regulation, for example, patients with heterozygous mutations in AID leading to a C-terminal truncation lack CSR activity despite intact SHM (Imai et al., 2005). *In vitro* studies have formally demonstrated that the C-terminal truncation disrupts CSR activity of AID, and further acts in a dominant negative fashion to block CSR activity in cells expressing wt AID (Ucher et al., 2014; Zahn et al., 2014). On the other hand, certain N-terminal residues seem to be dispensable for CSR but absolutely required for SHM (Shinkura et al., 2004). The mechanism by which these N-terminal residues mediate SHM is still unclear, yet they highlight differential requirements for AID activity in SHM and CSR. In my thesis, I further study the requirements for the E5 region of AID, revealing that E5 structural determinants can affect AID enzymatic activity and somehow promote CSR, downstream of deamination (Chapter 4).

1.5.2 AID subcellular localization

Though SHM, IGC and CSR are initiated by AID deaminating genomic DNA in the nucleus, it was initially puzzling to observe AID localization predominantly in the cytoplasm of B cells (Rada et al., 2002a). Further studies have now demonstrated that in B cells approximately 10% of AID localizes to the nucleus in steady state (Orthwein and Di Noia, 2012). The subcellular distribution of AID is determined by an intricate balance of nuclear-cytoplasmic shuttling, retention in the cytoplasm and differential stability in the cytoplasm relative to the nucleus (Figure 1.9). Though it seems counter-intuitive to restrict AID access to the nucleus, limiting its access to the genome is necessary to prevent deleterious off-target activity (Methot and Di Noia, 2017). In this thesis, I report the elucidation of the molecular mechanism for AID retention in the cytoplasm and propose that limiting nuclear accumulation is necessary to restrict AID activity (Chapter 2).

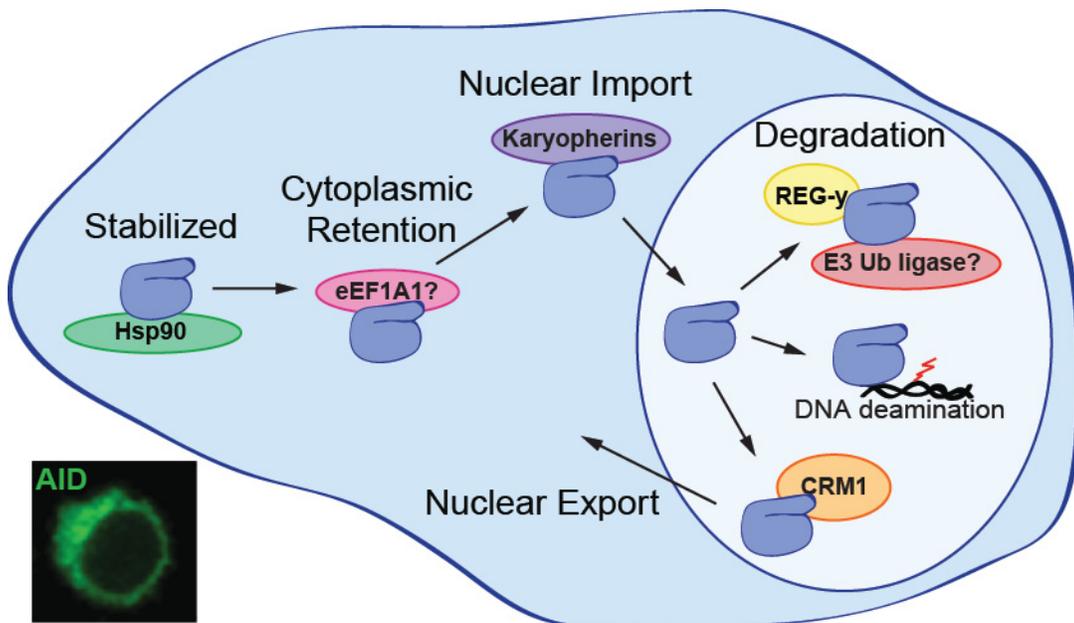


Figure 1.9. Regulation of AID subcellular localization.

Schematic of the various mechanisms regulating AID subcellular localization. As in the microscopy image, in B cells most AID is localized to the cytoplasm at steady state. AID stability favours cytoplasmic localization, as in the cytoplasm it is stabilized by HSP90, but it is rapidly degraded in the nucleus. Nuclear instability is due to ubiquitination-dependent and Reg- γ -dependent proteasomal degradation. AID is also actively imported into the nucleus, possibly by various karyopherins. Nonetheless, nuclear accumulation is limited by active nuclear export, by CRM1, and by cytoplasmic retention. During my studies, cytoplasmic retention was proposed, but was not formally demonstrated to be mediated by eEF1A1.

AID is a small protein of 198 amino acids. Nonetheless it must be actively imported into the nucleus *via* the karyopherin family of nuclear importins, dependent on a structural motif, involving positively charged residues spanning almost the entire protein (Hu et al., 2013; Patenaude et al., 2009). In DT40 B cells, AID can associate with the karyopherin, CTNNBL1, which is important for AID activity. As CTNNBL1 binding required residues involved in nuclear import (Conticello et al., 2008), and CTNNLB1 is structurally similar to karyopherins (Ganesh et al., 2011), it was believed to mediate AID nuclear import. Further studies in mice and mouse cell lines revealed that CTNNBL1 is dispensable for AID activity, suggesting that there is either redundancy amongst the karyopherins that mediate AID nuclear import or that CTNNBL1 has a different function (Chandra et al., 2013; Han et al., 2010). Interestingly, nuclear AID can co-localize with CTNNBL1 in nucleoli (Hu et al., 2013). This localization is possibly linked to nuclear import but the functional relevance is unknown.

The E5 of AID mediates two distinct functions to restrict nuclear accumulation, nuclear export *via* a canonical leucine rich nuclear export signal (NES) (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004) and cytoplasmic retention, which requires at least two adjacent aspartic acid residues (Patenaude et al., 2009). Accordingly, in the

absence of E5, AID becomes fully nuclear, as nuclear import becomes the only force acting on AID localization (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004). Nuclear export of AID is mediated by the Ran dependent nuclear exportin CRM1 (McBride et al., 2004), and cytoplasmic retention was suggested to require eEF1A1, which binds cytoplasmic AID, dependent on the retention motif (Hasler et al., 2011), but prior to this thesis the mechanism was unknown.

Prior to this work, the interplay between nuclear export and cytoplasmic retention was also unclear. Both mechanisms require adjacent residues in the E5, and would seem redundant in their role to exclude AID from the nucleus. Furthermore, studies of AID localization using GFP fusions, found that pharmacological inhibition of CRM1 led to rapid nuclear accumulation of AID, and that NES mutants were completely nuclear at steady state, suggesting that export was the major force restricting AID nuclear access (Ito et al., 2004; McBride et al., 2004).

1.5.3 AID protein stability

AID protein stability also influences its localization, as AID half-life in the nucleus is significantly shorter than in the cytoplasm (Figure 1.9) (Aoufouchi et al., 2008). One reason for this is that cytoplasmic AID associates with the chaperone HSP90 and co-chaperone DnaJa1, which protect it from ubiquitin dependent degradation (Orthwein et al., 2010; Orthwein et al., 2012). Nuclear AID on the other hand, undergoes an ubiquitin dependent degradation for which the exact molecular mechanism remains unclear (Aoufouchi et al., 2008), and associates with the protein Reg- γ , which actively promotes degradation in an ubiquitin independent but proteasomal dependent manner (Uchimura

et al., 2011). It is not clear whether these different nuclear degradation mechanisms act on distinct pools of AID; however, nuclear AID is stable enough to access and mutate the DNA during antibody diversification.

The mechanistic complexity involved in regulating AID shuttling and protein stability has a clear role in protecting the genome, permitting just the right amount of AID to access the genome at any given moment. Reducing AID stability with pharmacological inhibition of HSP90 or genetic DnaJa1 deficiency, results in reduced AID expression and limited antibody diversification (Orthwein et al., 2010; Orthwein et al., 2012). On the contrary, forcing AID to accumulate in the nucleus by removing its C-terminal domain or knocking out Reg- γ (Barreto et al., 2003; Shinkura et al., 2004; Uchimura et al., 2011; Zahn et al., 2014) results in excessive mutagenic activity, promoting off-target mutations, oncogenic translocations and cytotoxicity.

1.5.4 AID regulation by DNA damage and cell cycle

Less defined mechanisms may also be involved in regulating AID stability and localization. Excessive DNA damage can cause AID to accumulate in the nucleus (Brar et al., 2004; Lambert et al., 2013). This accumulation seems to be dependent on PARP signalling, but its functional role remains unclear (Tepper et al., 2016).

Nuclear AID also seems to be more stable during G1 phase of the cell cycle compared to S/G2 (Le and Maizels, 2015), and AID can accumulate in the nucleus during cytokinesis, remaining nuclear during early G1 (Wang et al., 2017). Fittingly, Ig locus DNA damage and uracils are detected in G1 (Petersen et al., 2001; Wang et al., 2017), and UNG activity is specifically required in G1 for efficient SHM and CSR

(Sharbeen et al., 2012). The exact purpose for G1 restriction is unclear, but it may protect the genome by limiting AID deamination potential during DNA replication in S phase, or may enforce appropriate repair as c-NHEJ works in G1/S/G2-phases but HR only works during S/G2-phases (Rothkamm et al., 2003).

1.5.5 AID phosphorylation

AID function can be modified by phosphorylation at various sites (Figure 1.7). The most studied phosphorylation site is serine 38, which is carried out by protein kinase A (PKA) and specifically modulates AID activity at the chromatin. Phosphorylation of S38 does not affect AID catalytic activity on ssDNA, but greatly enhances SHM, IGC and CSR activities (Basu et al., 2005; Chatterji et al., 2007; McBride et al., 2006; Pasqualucci et al., 2006). Mechanistically, S38 phosphorylation promotes AID activity on transcribed dsDNA by enhancing its association with the ssDNA binding protein RPA (Basu et al., 2005; Chaudhuri et al., 2004).

Modifying PKA activity in B cells can directly affect CSR activity of AID (Pasqualucci et al., 2006), while mutating S38 *in vivo* greatly reduces SHM and CSR activities (Cheng et al., 2009; McBride et al., 2008). Interestingly, AID from bony fish lack S38, yet they maintain binding to RPA *via* a phospho-mimetic aspartic acid at position 44 (Basu et al., 2008). A possible evolutionary explanation for why bony fish lack S38 phosphorylation, is that they do not undergo CSR (Stavnezer and Amemiya, 2004), suggesting that S38 phosphorylation would regulate the formation of DSBs during CSR. In line with this, DNA damage can activate PKA *via* the DNA damage response kinase ATM, thereby promoting S38 phosphorylation, creating a positive

feedback loop that specifically boosts AID activity at the S-regions during CSR (Vuong et al., 2013).

Other, less studied phosphorylation sites for AID exist. These include residues Serine 3 (S3), which may negatively regulate AID activity (Gazumyan et al., 2011), Threonine 27 (T27), which may affect AID enzymatic activity (Basu et al., 2005), and Tyrosine 184 (Y184), which has no known effects on AID activity (Basu et al., 2005; Patenaude et al., 2009; Zahn et al., 2014). AID residue Threonine 140 (T140) is also phosphorylated. Since the point mutation, AID T140A, only modestly affects CSR activity but dramatically reduces SHM, it is believed that T140 phosphorylation can fine-tune AID activity (McBride et al., 2008).

Similar to AID localization and stability, AID phosphorylation seems to regulate when and where AID deamination occurs, permitting appropriate activity while limiting off-target potential.

1.6 AID targeting

AID activity at the Ig locus is 100-1000 fold higher than at off-target sites (Liu et al., 2008). The exact basis for this preferential targeting has been of great interest, with recent work trying to establish the locus contributions, sequence-intrinsic determinants or trans-acting factors involved.

1.6.1 Transcription association dictates AID targeting

For antibody diversification, AID needs to access ssDNA at the Ig loci in the nucleus. This is not trivial, as the genome is very large, is protected by chromatin, and rarely exposes ssDNA in order to prevent genomic instability (Richard et al., 2009). In bacteria and in yeast, efficient mutagenic activity of AID was found to require active transcription of the target genes (Chaudhuri et al., 2003; Gomez-Gonzalez and Aguilera, 2007; Ramiro et al., 2003). This was in line with the observation that AID prefers a small transcription-bubble-like substrate *in vitro* (Bransteitter et al., 2003; Larijani and Martin, 2007), and that both SHM and CSR activity require active transcription at the *Ig* loci in B cells (Lee et al., 2001; Peters and Storb, 1996; Stavnezer, 1996). The functional link between AID activity and transcription is now well established (Casellas et al., 2016; Storb, 2014).

In B cells, AID associates with actively transcribed regions in the Ig locus, and is associated with the RNA polymerase II (RNAPII) complex (Nambu et al., 2003). The ability for AID to deaminate is believed to require RNAPII stalling, as this could allow sufficient time for AID deaminase activity, given its poor enzymatic kinetics. This link has been formally demonstrated by engineering an *IgV* that contains

sequences to promote stalling (Kodgire et al., 2013). Furthermore, SPT5, a transcription elongation factor involved in stabilizing stalled RNAPII, is also necessary for the association of AID with the transcription complex (Pavri et al., 2010). Thus, SPT5 can help promote AID interaction with stalled transcription and can thereby promote efficient AID activity (Wang et al., 2014).

AID also interacts with the RNA exosome, a multisubunit complex that can degrade antisense RNA transcripts, which has been found to promote AID activity by exposing ssDNA. The exosome may also degrade R-loops, RNA-DNA hybrids formed when the transcript remains annealed to the template, promoting AID activity at both strands of the S-regions during CSR (Basu et al., 2011; Pefanis et al., 2014).

Interestingly, AID has been found to associate with thousands of distinct sites within the B cell genome by chromatin immunoprecipitation (Yamane et al., 2011); yet, only about 300 sites show consistent AID-dependent DNA damage (Casellas et al., 2016). The majority of AID off-target sites associate with super enhancers and have convergent and/or divergent transcription, demonstrating that the transcriptional environment, perhaps *via* increased transcription stalling or ssDNA availability, plays a role in targeting AID activity (Casellas et al., 2016; Meng et al., 2014; Pefanis et al., 2014; Pefanis et al., 2015; Qian et al., 2014). In my thesis, I uncover a licensing step that dissociates the recruitment of AID to the chromatin from its ability to associate with elongating RNAPII, and to mutate a particular gene. This work helps to further define the requirements for AID mutagenic targeting (Chapter 3).

1.6.2 Genetic determinants of CSR

For CSR, the organization of the *IgH* locus has a clear role, with each S-region having its own promoter/enhancer element that only drives transcription under the appropriate cytokine stimulation (Matthews et al., 2014b). Topology is also important, as the S-regions must physically associate (S-S synapse) to promote recombination during CSR. This is mediated by the 3' regulatory region (3'RR), a composite enhancer that physically recruits the E μ along with the enhancer of the S-region undergoing CSR (Wuerffel et al., 2007). The formation of this looping conformation seems to be intrinsic to the *IgH* locus, as even the I-Sce-I endonuclease can induce CSR in cells where the S-regions are replaced with I-Sce-I sites (Zarrin et al., 2007).

This chromatin re-structuring is regulated at least in part by the cohesin, mediator and INO80 complexes (Kracker et al., 2015; Thomas-Claudepierre et al., 2013; Thomas-Claudepierre et al., 2016). 53BP1, which is critical to c-NHEJ repair, also seems to promote formation of the S-S synapse (Wuerffel et al., 2007), to affect timing and directionality of the repair, preventing inverted recombination that would inactivate the gene (Dong et al., 2015; Feldman et al., 2017; Rocha et al., 2016).

The S-region sequences themselves are also important for CSR, promoting AID mutagenesis (Luby et al., 2001; Zarrin et al., 2004). The S-regions are GC rich and highly repetitive, containing multiple AGCT repeats (Zarrin et al., 2004). The AGCT sequence is important, containing two overlapping AID hotspots, so the simultaneous deamination of both cytosines on opposite strands can increase the chances of DSBs (Han et al., 2011; Zarrin et al., 2004). The GC richness of the S-regions also makes

them particularly prone to forming R-loops (Yu et al., 2003). S-regions from *Xenopus laevis* do not form R-loops, but they can still promote CSR (Zarrin et al., 2004), so R-loops are not absolutely required. Nonetheless, R-loops can promote transcription stalling and thereby increase AID activity, so they likely increase CSR efficiency (Rajagopal et al., 2009). The S-regions also contain repeated stretches of Gs that can form G4 structures *in vitro*, favouring AID binding and deamination (Qiao et al., 2017).

1.6.3 Genetic determinants of SHM

Demonstrating whether the *IgV* sequence favours AID activity has been more difficult. The CDRs are the main SHM targets in the *IgV*, and they contain a higher proportion of AID hotspots, which can favour mutagenic activity (Wagner et al., 1995). Nonetheless, *IgV* regions undergo selection *in vivo*, and are not efficiently mutated in cultured B cells, making it difficult to separate intrinsic mutability from antigen selection. This problem can be solved using a passenger *IgV* allele system (Betz et al., 1993), which can dissociate mutagenesis from selection. This approach initially helped to determine the intrinsic mutation pattern of AID at the *IgV* and the existence of hotspots (Di Noia and Neuberger, 2007). Recently, this technique was adapted by knocking-in unrelated genes at the *IgV* locus, demonstrating definitively that independent of *IgV* sequence, the genetic context of the *IgV*, i.e. promoters, enhancers, etc., is critical for promoting AID mutagenic activity (Yeap et al., 2015). Thus, the high levels of mutation achieved at the *IgV* are not necessarily related to specific sequence elements but are driven by its genetic context.

Many of the genetic elements within the Ig loci that drive AID activity are known. Enhancers, 3' of the *IgV* are well studied, and help drive SHM beyond their role in transcription. Of these, the DIVAC, a 3' enhancer element from the chicken $V\lambda$ locus, is the best studied. The DIVAC is conserved in mammalian *IgH* and *Igk* loci, and contains several transcription factor-binding elements that can enhance AID targeting to unrelated, transcribed genes if placed in close proximity (Blagodatski et al., 2009; Buerstedde et al., 2014; Kohler et al., 2012; Kothapalli et al., 2008). In addition to the DIVAC, the 3'RR, also seems to drive efficient SHM (Dunnick et al., 2009; Rouaud et al., 2013). Similar to the DIVAC, the 3'RR seems to both enhance *IgV* transcription, and promote AID recruitment to the *IgV* (Rouaud et al., 2013). Whether these enhancers recruit AID *via* similar mechanisms, and whether they are related to super enhancers or convergent transcription remains unknown.

1.6.4 Recruiting AID to transcription

Genome-wide association of AID with RNAPII (Pavri et al., 2010; Yamane et al., 2011) does not correlate with its deaminase activity, measured by generation of DSBs or off-target SHM (Hakim et al., 2012; Keim et al., 2013; Liu et al., 2008; Meng et al., 2014; Qian et al., 2014; Staszewski et al., 2011b). This clearly demonstrates that additional factors are involved in dictating AID activity in B cells (Figure 1.10).

Recruitment of AID during SHM and CSR may not be entirely equivalent. For example, transient knockdown of SPT5 reduces CSR (Pavri et al., 2010; Stanlie et al., 2012) but actually increases SHM (Wang et al., 2014). This disparity may reflect the fact that SPT5, along with SPT4, is involved in both stabilizing stalled RNAPII as well as

promoting transcription processivity (Wada et al., 1998). To that effect, S-region R-loops promote polymerase stalling, while the *IgV* does not form R-loops; thus, reduced SPT5 levels may destabilize stalled RNAPII in the S-regions, while promoting transcription stalling in the *IgV* (Wang et al., 2014). On the other hand, the exosome is equally important for CSR and SHM activity (Basu et al., 2011; Wang et al., 2014) highlighting the need for AID to target both DNA strands during both processes.

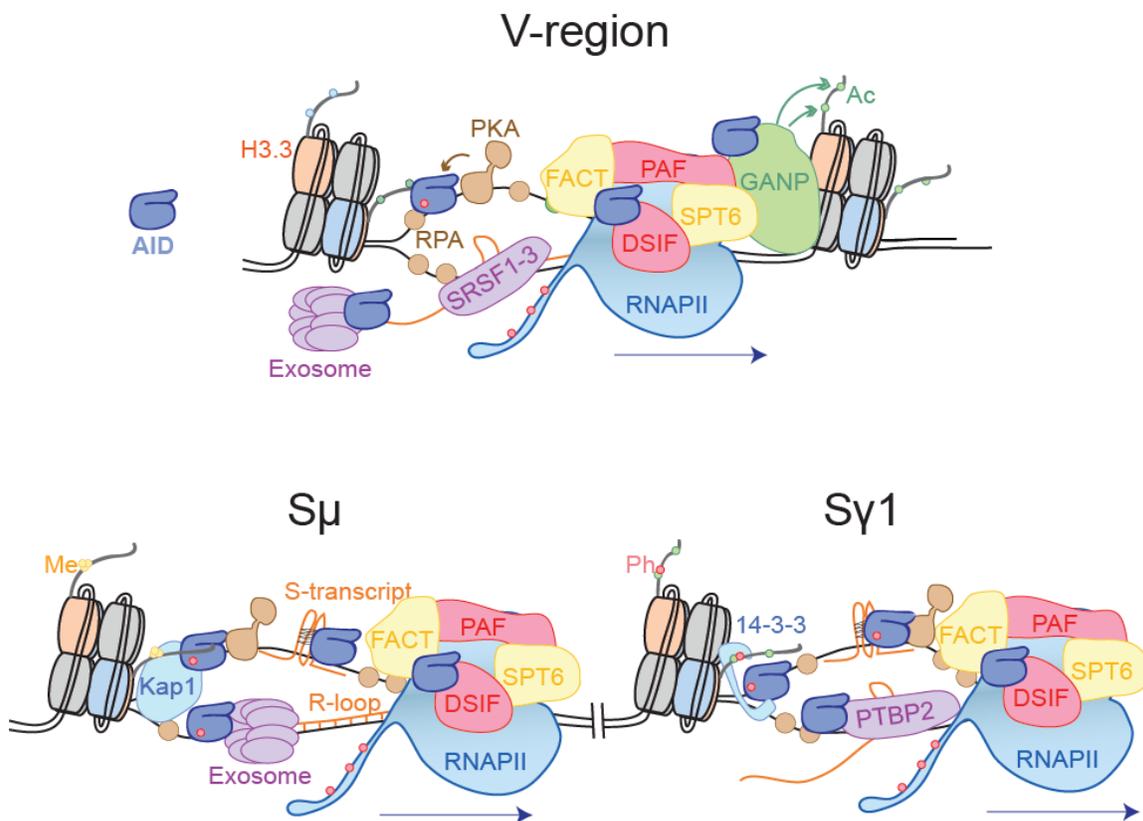


Figure 1.10. Targeting AID to the *IgV* and S-regions.

Schematic representation of different factors that contribute to targeting AID activity during SHM or CSR at the *IgV* or S-regions, respectively (see text for details). Factors are color coded according to their function: histone chaperones (yellow), transcription elongation factors (red), RNA-processing factors (purple), RNA molecules (orange), factors promoting AID catalytic activity (brown) and factors promoting AID targeting (green). Adapted from (Methot and Di Noia, 2017).

Depletion of either SPT6 or the FACT complex, which are involved in transcription elongation and nucleosome shuttling, results in reduction of both CSR and SHM (Aida et al., 2013; Begum et al., 2012; Okazaki et al., 2011; Stanlie et al., 2010). Interestingly, both SPT6 and FACT are important for maintaining H3K4me3 marks at the *Ig* loci, which may somehow affect AID activity (Begum et al., 2012; Stanlie et al., 2010). FACT levels also correlate with H3.3 deposition at the *IgV* and S-regions, and H3.3 levels in turn seem to correlate with AID mutagenic activity (Aida et al., 2013). Consistently, H3.3 bound DNA has an increased propensity of being single stranded, and H3.3-deficient DT40 cells are severely compromised for both SHM and IGC (Romanello et al., 2016). The PAF complex, a scaffold that coordinates transcription elongation and promotes stable recruitment of FACT and SPT6 to the transcription machinery, is found at the *Ig* locus and also seems to promote AID recruitment (Willmann et al., 2012).

1.6.5 Chromatin marks that guide AID

Along with these transcriptional partners, certain histone modifications can also recruit factors that can in turn recruit AID. For instance, the combination of H3K9ac and H3S10ph is specifically enriched at the transcriptionally active target S-regions, and can recruit the 14-3-3 adaptors as well as AID (Li et al., 2013). The 14-3-3 adaptors can also associate with AGCT repeat sequences, as found in the S-regions, which may confer specificity of their recruitment to the S-regions (Xu et al., 2010).

It has also been found that H3K9me3, specifically at the S_{μ} , recruits the KAP1 and HP1 chromatin factors (Jeevan-Raj et al., 2011). It remains unclear how H3K9me3

and H3K9ac can simultaneously promote AID activity during CSR (Methot and Di Noia, 2017); however, KAP1 seems to specifically recruit AID to the S μ region during CSR, and may therefore help ensure that the S μ is broken prior to the acceptor S-region (Jeevan-Raj et al., 2011; Rocha et al., 2016). These histone marks may also work in concert with H3K14ac and H3K4me3 in order to recruit other mediators necessary to coordinate the downstream DNA repair events in CSR (Stanlie et al., 2010; Wang et al., 2009a).

GANP, which functions as part of the RNA export machinery (Wickramasinghe et al., 2010), is important for regulating SHM. *Ganp*^{-/-} mice have severely reduced SHM, whereas transgenic mice overexpressing GANP have increased SHM (Kuwahara et al., 2004; Sakaguchi et al., 2005). During SHM, the role of GANP seems to be predominantly as a histone acetyltransferase, as GANP levels at the *IgV* are associated with H3K9Ac and H3K27Ac, activating histone marks that enhance AID recruitment and mutagenesis by promoting DNA accessibility, *IgV* transcription and possibly RNAPII stalling (Singh et al., 2013). As GANP may also promote AID nuclear import, the exact mechanism by which it promotes *IgV* targeting remains unclear (Maeda et al., 2010).

1.6.6 RNA processing and AID targeting

Splicing has long been known to affect CSR activity, as blocking splicing of the S μ transcript prevents CSR (Hein et al., 1998; Lorenz et al., 1995). The role of splicing during CSR is multifactorial, as both the splicing machinery and spliced transcripts are implicated in recruiting AID to the Ig locus.

The spliced S-region transcript itself has recently been shown to bind selectively to the corresponding S-region, and to recruit AID (Zheng et al., 2015). AID binding to the S-region transcripts requires G4 structures, as observed for the S-region DNA (Qiao et al., 2017). These are only produced in the S-region transcript after splicing (Zheng et al., 2015). Blocking splicing inhibits CSR, but this can be rescued by direct expression of spliced S-region transcripts (Zheng et al., 2015). AID interacts with polypyrimidine tract binding protein 2 (PTBP2), an RNA binding protein involved in alternative splicing (Nowak et al., 2011), which may recruit AID to the S-regions *via* the spliced S-region transcript. Intriguingly, PTBP2 knockdown reallocates AID from the S-regions to the IgV, suggesting that these regions may compete for AID recruitment (Matthews et al., 2014a). AID association with G4 may not be unique to the S-region transcripts, as AID is recruited to and can deaminate the telomeres (Cortizas et al., 2016) and telomeric transcripts can also form G4 structures (Rhodes and Giraldo, 1995).

During SHM, AID targeting may also depend on splicing events and the splicing machinery to efficiently induce mutations. In DT40 cells, a specific isoform of the splicing factor SRSF1 (SRSF1-3) is necessary for SHM, though it does not seem to promote AID recruitment. Rather, SRSF1-3 may somehow facilitate AID access to ssDNA by regulating the *Ig* transcript splicing (Kanehiro et al., 2012), or by promoting AID nuclear accumulation (Kawaguchi et al., 2017).

Various heterogeneous nuclear ribonucleoproteins (hnRNPs) that are also implicated in alternative splicing have been found to associate with AID in B cells (Hu et al., 2015; Mondal et al., 2016). A few of these factors are actually implicated in regulating AID activity, such as hnRNPK, which is necessary for both SHM and CSR

activity, and hnRNPL, which is dispensable for mutagenic activity but is nonetheless required for CSR in CH12F3 cells (Hu et al., 2015). Nonetheless, AID is part of large RNP complexes, but more work is necessary to dissect their exact role in regulating AID activity (Mondal et al., 2016).

Altogether, it is clear that different mRNPs and/or splicing machinery can differentially regulate AID targeting to either the IgV (SRSF1-3, GANP) or S-regions (PTBP2, spliced S-transcripts) perhaps depending on the timing during the GC reaction. As splicing and transcription are tightly coordinated (Kornblihtt et al., 2013), they could synchronize AID targeting and access to ssDNA in order to specifically direct its activity.

1.7 Rationale

For such a small protein, AID associates with many factors, and is regulated by a complex network of mechanisms. The relevance of such regulation is evidenced by the pathological consequences of deregulated AID activity, with immunodeficiency (Revy et al., 2000) and oncogenic transformation (Shimizu et al., 2012) representing opposite ends of the spectrum. In light of this, ascertaining how AID regulatory mechanisms cooperate and compete is important to understand how deregulated AID activity can occur. Therefore, the goal of my doctoral studies has been to use structure function approaches to determine the molecular mechanisms and functional relevance of different regulatory mechanisms of AID.

Prior to my arrival, the lab had determined that AID was actively retained in the cytoplasm (Patenaude et al., 2009). The mechanism was unknown, but two residues in the E5 region were critical for retention. We therefore set out to further define the motif necessary for cytoplasmic retention, to identify its molecular mechanism, and to determine its functional relevance (Chapter 2). Concurrently, we had a project attempting to dissociate AID enzymatic activity from functional activity in B cells. This led to the identification of an arginine-rich motif that was necessary for SHM and CSR but dispensable for AID catalytic activity, for which we attempted to determine the molecular mechanism (Chapter 3). During my studies we also published a manuscript that described how the E5 region of AID influences downstream DNA repair during CSR (Zahn et al., 2014). Using structure-function studies, we have tried to further define the role of the E5 region in regulating AID activity (Chapter 4).

**CHAPTER 2: CONSECUTIVE INTERACTIONS WITH
HSP90 AND eEF1A UNDERLIE A FUNCTIONAL
MATURATION AND STORAGE PATHWAY OF AID IN
THE CYTOPLASM**

Stephen P. Methot, Ludivine C. Litzler, Felipe Trajtenberg, Astrid Zahn, Francis Robert, Jerry Pelletier, Alejandro Buschiazso, Brad G. Magor, Javier M Di Noia

This article is published in the *Journal of Experimental Medicine* (2015), 212(4): 581-

2.1. Abstract

Activation-induced deaminase (AID) initiates mutagenic pathways to diversify the antibody genes during immune responses. The access of AID to the nucleus is limited by CRM1-mediated nuclear export and by an uncharacterized mechanism of cytoplasmic retention. Here, we define a conformational motif in AID that dictates its cytoplasmic retention and demonstrate that the translation elongation factor eEF1A is necessary for AID cytoplasmic sequestering. The mechanism is independent of protein synthesis but dependent on a tRNA-free form of eEF1A. Inhibiting eEF1A prevents the interaction with AID, which accumulates in the nucleus and increases class switch recombination as well as chromosomal translocation byproducts. Most AID is associated to unspecified cytoplasmic complexes. We find that the interaction of AID with eEF1A and HSP90 are inversely correlated. Despite both interactions stabilizing AID, the nature of the AID fractions associated with HSP90 or eEF1A are different, defining two complexes that sequentially produce and store functional AID in the cytoplasm. In addition, nuclear export and cytoplasmic retention cooperate to exclude AID from the nucleus but might not be functionally equivalent. Our results elucidate the molecular basis of AID cytoplasmic retention, define its functional relevance and distinguish it from other mechanisms regulating AID.

2.2. Introduction

During immune responses, B cells producing high affinity antibodies of the IgG, IgA and IgE classes are generated from the low affinity, IgM⁺ B cells that initially recognize the invading antigens. The molecular mechanisms underpinning the affinity maturation and change of class of the antibody response are somatic hypermutation (SHM) and class switch recombination (CSR), respectively; mutagenic processes that modify the antibody genes. The enzyme Activation Induced Deaminase (AID) initiates SHM and CSR by converting deoxycytidine to deoxyuridine in DNA. Repair enzymes that recognize uracil in DNA trigger further mutagenic processing to generate the full spectrum of SHM or the DNA breaks that are necessary for CSR. Human patients lacking AID have no antibody affinity maturation or class switching and are immunodeficient (Revy et al., 2000). On the other hand, excessive or deregulated AID activity can be cytotoxic (Zahn et al., 2014), contribute to autoimmunity (Diaz, 2013) or predispose to B cell lymphomas (Robbiani et al., 2009). The delicate balance between the physiological and pathological effects of AID is enforced by multiple levels of AID regulation (Keim et al., 2013; Vuong and Chaudhuri, 2012).

Subcellular localization and protein stability are major points of AID regulation (Orthwein and Di Noia, 2012; Vuong and Chaudhuri, 2012). AID is a nuclear-cytoplasmic shuttling protein (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004) and its stability is intricately related to its compartmentalization. AID is stabilized in the cytoplasm by an HSP90 molecular chaperoning pathway that requires the DnaJa1 HSP40 (Orthwein et al., 2010; Orthwein et al., 2012) and destabilized in the nucleus by ubiquitin-dependent and -independent pathways (Aoufouchi et al., 2008; Uchimura et

al., 2011). The small size of AID (24 kDa) should allow it to diffuse through the nuclear pores; however, it requires active import to enter the nucleus (Patenaude et al., 2009) and 90% of AID is localized to the cytoplasm under steady state conditions (Pasqualucci et al., 2004; Rada et al., 2002a). Two mechanisms that exclude AID from the nucleus have been identified. AID is exported from the nucleus by CRM1, which recognizes a Leucine-rich nuclear export signal (NES) within positions 188-198 of AID (McBride et al., 2004). AID is also retained in the cytoplasm by a still ill-defined mechanism that requires residues Asp 187 and 188 in human AID, which overlap with the NES (Patenaude et al., 2009). The relative contribution of CRM1-mediated nuclear export and cytoplasmic retention to nuclear exclusion and functional regulation of endogenous AID is also unknown because of the lack of reagents to block each mechanism without resorting to AID mutants that might affect both processes.

HSP90 and DnaJa1 bind cytoplasmic AID but do not mediate its retention (Orthwein et al., 2010; Orthwein et al., 2012). Cytoplasmic AID also interacts with the translation elongation factor eEF1A in human, mouse and chicken B cells (Hasler et al., 2011). This factor delivers aminoacyl-tRNA to the elongating ribosomes (Andersen et al., 2003) but it has other functions that are unrelated to protein synthesis (Mateyak and Kinzy, 2010). Since mutations in AID residues Asp187 and Asp188 disrupt cytoplasmic retention (Patenaude et al., 2009), as well as the interaction with eEF1A (Hasler et al., 2011), it is possible that eEF1A is part of a complex retaining AID in the cytoplasm (Hasler et al., 2012). However, mutating Asp187/188 could also disrupt the interaction of AID with other factors. Moreover, the N-terminal region of AID is required for the interaction with eEF1A (Hasler et al., 2011) but not for cytoplasmic retention (Patenaude

et al., 2009). In addition, eEF1A and AID form part of an uncharacterized ~350 kDa cytoplasmic complex (Hasler et al., 2011), which could contain redundant or additional factors. Thus, the mechanism of cytoplasmic retention and the involvement of eEF1A remain to be elucidated. In addition, whether AID forms distinct complexes with HSP90 and eEF1A and the possibility that AID shuttles between these two factors, as well as the comparative functional relevance of each of these interactions also remain to be determined.

Here, we demonstrate that the mechanism of AID cytoplasmic retention depends on a specific AID conformation and on eEF1A and that it limits AID function. We also identify differences between eEF1A-, HSP90- and CRM1-mediated AID regulation that suggest a model to explain the transit and functional maturation of AID in the cytoplasm, which impacts on the efficiency of antibody gene diversification and its collateral damage.

2.3. Results

2.3.1. AID cytoplasmic retention is evolutionarily conserved

The structure of the AID/APOBEC family proteins along evolution shows that many of these deaminases have acquired unique extensions at either end of the common catalytic domain, as epitomized by the C-terminal tail that APOBEC1 acquired in mammals (Severi et al., 2011) (Fig. 2.1A). These extensions are often encoded by a single exon of the corresponding gene, such as exon 1 of APOBEC2 or exon 5 of mammalian APOBEC1 (Fig. 2.1A). Similarly, AID has a specific C-terminal extension of 17 residues encoded by exon 5 (henceforth E5). E5 is critical for AID nuclear exclusion, as it harbors both the NES and a poorly defined cytoplasmic retention motif (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004). The NES is conserved between AID homologs (Fig. 2.1B). Two acidic residues important for cytoplasmic retention are also conserved, albeit in poikilotherms Glu replaces Asp187 and the region preceding it is very different (Fig. 2.1B). To determine whether both nuclear exclusion mechanisms of AID were functional throughout evolution, we analyzed the subcellular localization of multiple AID orthologs. AID-GFP fusions from all tested homeotherms were excluded from the nucleus of HeLa cells while AID-GFP from poikilotherms was not (Fig. 2.1C). These differences in localization did not reflect proteolytic generation of free GFP, as controlled for by Western blot, and were also observed in HEK293 and mouse NIH3T3 cells (data not shown). Treatment with the CRM1 inhibitor leptomycin B (LMB) led to increased nuclear abundance of every AID ortholog in all cell lines tested (Fig. 2.1C and data not shown) indicating that the NES of all the orthologs was functional. Thus, the homogenous distribution of AID from poikilotherms in mammalian cells could reflect

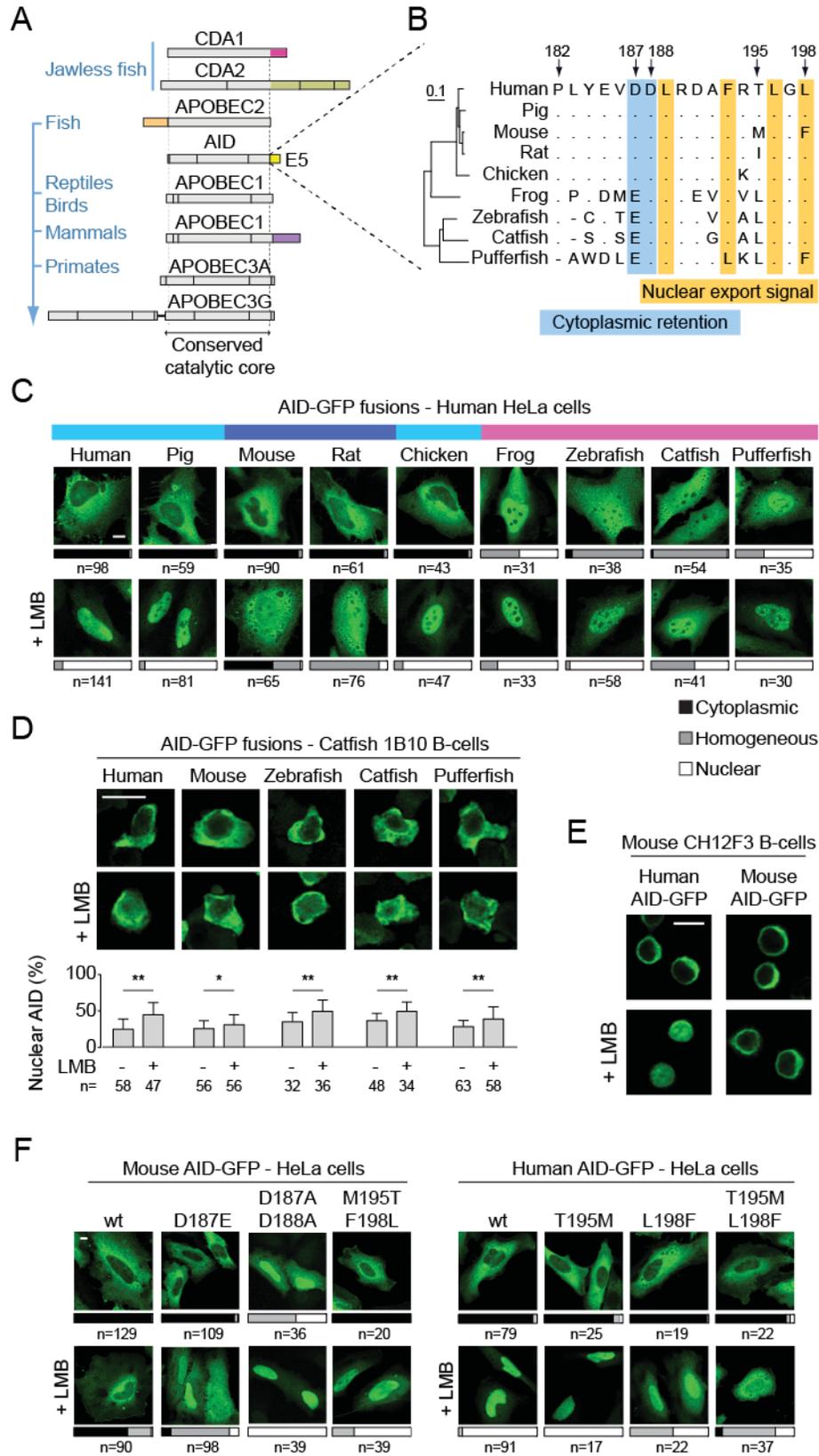


Figure 2.1. Cytoplasmic retention of AID is conserved

(A) In-scale representation of AID/APOBEC family members, indicating protein regions encoded by different exons and their phylogenetic distribution. Domains unique to specific members of the family are colored. (B) Amino acid alignment of the E5 region from AID homologs. Arrows indicate positions according to, and dots denote identities with, human AID. The NES and a motif required for cytoplasmic retention are highlighted. A N-J tree built on the full-length proteins' alignment (left) illustrates the phylogenetic distances between homologs. (C) Representative confocal microscopy images of HeLa cells transiently expressing AID-GFP homologs under steady state conditions or after inhibition of nuclear export with LMB (50 ng/ml, 2 h). Bars represent the proportion of n cells from ≥ 3 experiments showing the indicated types of subcellular distribution (cytoplasmic, black; homogenous, grey; nuclear, white). The color bars at the top identify AID-GFP homologs that show similar subcellular distributions in human cells (human-like, teal; stronger retention, blue; weaker retention, pink). (D) Representative confocal microscopy images of AID-GFP homologs transiently expressed in catfish 1B10 B cells and treated as in (C). Nuclear GFP signals were measured by line intensity analysis for n cells from ≥ 2 experiments and the mean + SEM, plotted (graph below). Differences were assessed by Student's t -test ($P < 0.05 = *$; $P < 0.001 = **$). (E) Representative confocal microscopy images of mouse CH12F3 B cells overexpressing human or mouse AID-GFP and stimulated with CIT (1 μ g/ml anti-CD40, 1 ng/mL TGF- β , and 10 ng/mL IL-4) for 24 h prior to treatment with either vehicle or 10 ng/mL LMB for 2 h. Images are representative of 2 independent experiments. (F) Representative confocal microscopy images of HeLa cells transiently expressing mouse and human AID-GFP and indicated mutants. Cells were treated as in (C) and subcellular localization scored from 3 independent, pooled experiments. Magnification: 400X (A, B) or 630X (E, F), scale bars, 10 μ m throughout.

defective cytoplasmic retention. However, the sequence divergence of the E5 domain during evolution could prevent the interaction of fish and frog AID with the human/mouse retention factors. We therefore expressed AID from poikilotherms species in the catfish B cell line 1B10. We found that all fish AID were excluded from the nucleus in 1B10 cells and that this exclusion was partially relieved by LMB treatment (Fig. 2.1D). These results indicate that nuclear export was functional in 1B10 cells and that fish AID were retained in the cytoplasm in the proper cellular context (Fig. 2.1D).

We discovered the mechanism of cytoplasmic retention using human AID (Patenaude et al., 2009). However, for reasons that are described below, the C-terminal

GFP fusion weakens the retention of human AID-GFP so that this mechanism is not readily apparent in HeLa cells (Fig. 2.1C). This is probably also the case for pig and chicken AID-GFP that have E5 regions identical to the human's (Fig. 2.1B). In contrast, the mouse and rat orthologs accumulated only partially in the nuclei of HeLa cells after LMB treatment (Fig. 2.1C). Even in mouse CH12F3 B cells stimulated for CSR, mouse AID-GFP remained largely cytoplasmic after LMB treatment while human AID-GFP accumulated in the nucleus (Fig. 2.1E). Given that the mouse AID NES is functional (McBride et al., 2004) (Fig. 2.1B,C), our results suggested that there are amino acid differences between mouse and human AID that could modulate the strength of cytoplasmic retention in the context of AID-GFP fusions. We confirmed that cytoplasmic retention was preventing mouse AID-GFP nuclear accumulation in LMB-treated cells by mutating Asp187 and 188. Substituting Asp187 with the Glu residue present in fish was sufficient to render mouse AID-GFP responsive to LMB (Fig. 2.1F), which could partially explain the lack of retention of fish AID-GFP in HeLa cells. Mutating both positions to Ala eliminated its nuclear exclusion (Fig. 2.1F). We then focused on the only two positions within E5 that are different between human and mouse AID (Fig. 2.1B). Mutating mouse AID Met195 and Phe198 to the human residues (Thr and Leu, respectively) allowed nuclear accumulation of mouse AID-GFP when nuclear export was inhibited (Fig. 2.1F), suggesting reduced cytoplasmic retention. Conversely, mutations T195M and L198F in human AID-GFP synergistically reduced its response to LMB (Fig. 2.1F), thus suggesting increased retention.

We conclude that AID cytoplasmic retention was acquired early on during evolution and therefore likely plays an important function in AID biology. We also find

that amino acid sequence differences between AID homologs could dictate cytoplasmic retention efficiency, at least in the context of AID-GFP fusions.

2.3.2. Cytoplasmic retention requires a specific AID conformation

We then asked whether the difference in cytoplasmic retention efficiency between human and mouse AID-GFP held true for the endogenous enzymes in B cells. In the human Ramos B cell line, in which transfected human AID-GFP completely relocalized to the nucleus after LMB treatment (Fig. 2.2A), endogenous AID detected by immunofluorescence (IF) responded significantly less (Fig. 2.2B). The same modest response to LMB was observed for endogenous AID in two other human B cell lines (Fig. 2.2B), as well as for the endogenous mouse AID in stimulated CH12F3 cells (Fig. 2.2C). Thus, in contrast to their C-terminal GFP fusion counterparts, endogenous human and mouse AID show similar cytoplasmic retention, which plays a large role in their nuclear exclusion (Fig. 2.2B,C).

The effect of the GFP fusion on the human AID C-terminus was still informative. The different behavior of human AID-GFP and endogenous AID was not an artifact of overexpression. Untagged human AID also accumulated much less than AID-GFP in the nucleus of HeLa cells after LMB treatment (Fig. 2.2D), suggesting that the C-terminal GFP fusion weakened cytoplasmic retention. We also observed that a natural human AID variant bearing an insertion of 34 amino acids between exons 4 and 5 (Ito et al., 2004) was not excluded from the nucleus despite having an intact NES (Fig. 2.2D).

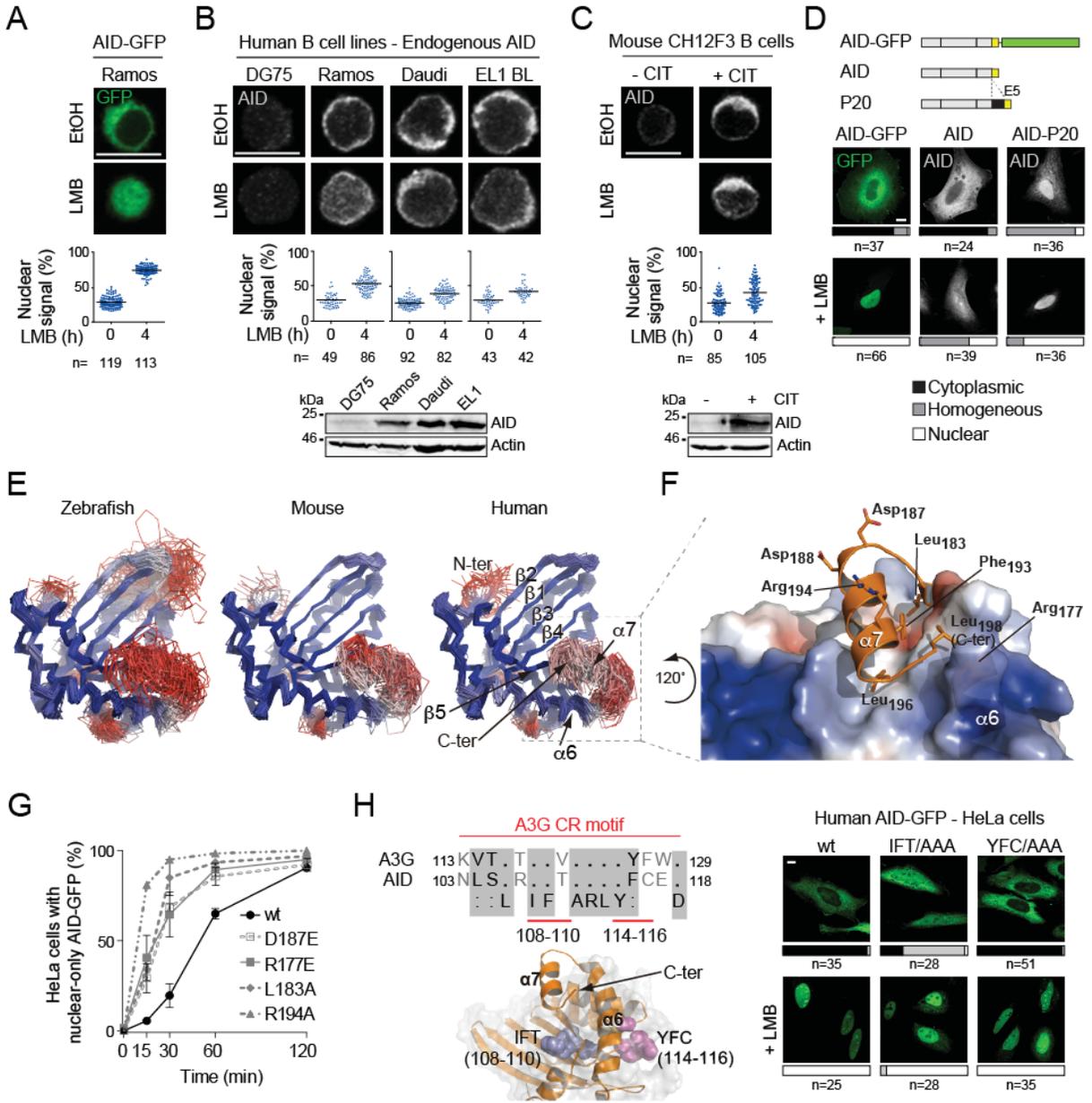


Figure 2.2. A specific conformation of AID is necessary for cytoplasmic retention.

(A) Representative confocal microscopy images of AID-GFP in human Ramos B cell stable transfectants treated for 4 h with either vehicle (EtOH) or 10 ng/mL LMB. The proportion of nuclear AID signal was calculated as the ratio between the AID signal overlapping with propidium iodide signal to the total AID signal, and was plotted for *n* individual cells (dots). The mean for each group is indicated by a horizontal line (graph below). (B) Representative confocal microscopy images of endogenous AID detected by IF in the indicated human B cell lymphoma lines treated, and quantified as in (A) (graph below). DG75 cells were used as negative control. A Western blot of AID and actin expression in total cell extracts for each cell line is shown (lower panel). (C) Representative confocal microscopy images of endogenous AID

detected by IF in mouse CH12F3 B cells stimulated with CIT for 24 h prior to LMB treatment and quantification as in (A) (graph below). Non-stimulated cells were used as negative control. A Western blot as in (B) is shown (lower panel). (A-C) Representative data from one out of 2 independent experiments. (D) Representative confocal microscopy images of HeLa cells transiently expressing human AID-GFP or untagged AID wt or variant P20 with 34 amino acids inserted upstream of E5. Localization was assessed by detecting GFP, or by IF for overexpressed untagged AID. Cells were treated with either vehicle or 50 ng/mL LMB for 2 h. The aligned schemes of the proteins are shown on the top panel and quantification of subcellular localization in *n* cells from 2 independent, pooled experiments is shown below each image (cytoplasmic, black; homogenous, grey; nuclear, white). (E) Ensemble of top scoring models from the best represented cluster were derived from the conformational sampling of the E5 C-terminal tail for zebrafish, mouse and human AID in which the catalytic core was modeled on the APOBEC3C structure. Models were colored according to the root-mean-square deviation (rmsd) differences for each position with a ramp from blue (low) to red (high). The lower the rmsd, the better the coincidence between multiple models at any given position. (F) Top scoring model of human AID from the best represented cluster. The electrostatic potential (negative in red, positive in blue) was mapped onto the molecular surface of the core domain, while the C-terminal tail is represented as a cartoon. Selected residues are represented as sticks. This model is rotated 120° counter-clockwise relative to the models in (E). (G) Nuclear localization kinetics of the indicated human AID-GFP variants were assessed in HeLa cells. Cells were treated with 50 ng/mL LMB and fixed at the indicated times. The mean ± SEM proportion of cells with full nuclear localization were plotted for each time point from 2 independent experiments, after scoring >40 cells per time point, per experiment. (H) Sequence alignment of an APOBEC3G (A3G) cytoplasmic retention motif with its homologous AID region. The structural model of AID shows the relative position of residues 108-110 (blue spheres) and 114-116 (magenta spheres) (left panel). Representative confocal images of HeLa cells transiently expressing human AID-GFP variants in which either one of these residue stretches were mutated to Ala. Cells were treated with either vehicle or 50 ng/mL LMB for 2 h, and scored as in (D), from 2 independent, pooled experiments (right panels). Magnification 630X (A,B,C) or 400X (D,H). Scale bars, 10 μm, throughout.

Together, our results suggested that cytoplasmic retention requires that the E5 domain adopt a certain conformation with respect to the enzyme core. This conformation would be variably affected by a GFP fusion depending on the sequence of E5 (Fig. 2.1F).

Since AID has not been crystallized we resorted to molecular modeling to ask whether an interaction between the E5 region and the catalytic core of AID was thermodynamically possible. We modeled the AID catalytic core using as template

APOBEC3C (Kitamura et al., 2012), which among the AID paralogs of known 3D structure displays the highest sequence similarity (42%). The C-terminal tail of AID is absent in the APOBECs (Fig. 2.1A) and cannot be modeled by comparison. However, secondary structure predictions for AID of all species consistently showed an amphipathic helix within the E5 (helix $\alpha 7$), in which the hydrophobic residues of the NES and the Asp187 and 188 are on opposite sides (not depicted). We used the Rosetta suite (Das and Baker, 2008) to perform a conformational sampling of the AID E5 from human, mouse and zebrafish, all of which yielded a similar result (Fig. 2.2E). In all the best energy clusters (see Methods), E5 sat within the same circumscribed region interacting with the helix $\alpha 6$ and the strand $\beta 5$ from the catalytic core. This convergence of all three models despite significant sequence differences (zebrafish and human AID are 59% similar) strongly suggests that E5 spontaneously adopts this conformation.

The predicted binding of the helix $\alpha 7$ buries the hydrophobic residues of the NES (Phe193, Leu196 and Leu 198 in human AID) in a groove on the catalytic core conformed by residues Leu126, Ala137, Ile138, Leu176, Leu180 and Leu181, thus avoiding their exposure to the solvent (Fig 2.2F). The model shows several contacts between NES residues and the AID core, such as a salt bridge between the side chain of Arg177 and the C-terminal carboxylate in Leu198 and multiple van der Waals contacts (Leu183:Arg127, Leu183:His130, Arg177:Phe193, Phe109:Leu196, etc). Such a conformation predicts that mutating the NES would not only eliminate nuclear export but also compromise cytoplasmic retention by affecting the position of E5. Indeed, we have shown that mutating the NES caused more efficient AID nuclear accumulation than inhibiting CRM1 (Patenaude et al., 2009). Moreover, the reduced retention of AID-

GFP (compare Fig. 2.2A to 2.2B) and of the P20 variant (Fig. 2.2D), were also consistent with our model, as the GFP fusion or the insertion separating E5 from the core would interfere with this conformation. To further validate the model we mutated other residues, outside the NES, that would participate in the interaction between E5 and the enzyme core. Mutations L183A and R177E compromised cytoplasmic retention of AID, as judged by faster nuclear import kinetics (Fig. 2.2G). Charged residues Asp187, Asp 188, Arg194, were predicted to be on the solvent-exposed face of helix $\alpha 7$ (Fig. 2.2F) and could mediate the interaction with retention factors. Consistently, mutants R194A and D187E had faster nuclear localization kinetics than wt AID (Fig. 2.2G).

The AID paralogs APOBEC1, APOBEC3A and APOBEC3G have mechanisms for nuclear exclusion (Bennett et al., 2006; Chester et al., 2003; Land et al., 2013; Stenglein et al., 2008), which might have evolved from those of AID. Interestingly, the AID model predicted interactions between E5 and an internal region corresponding to residues 103-118, which is homologous to a cytoplasmic retention determinant identified in APOBEC3G (Bennett et al., 2006; Stenglein et al., 2008) (Fig. 2.2H). Mutating AID Ile108-Phe109-Thr110 to Ala, in which Phe109 would form van der Waals interactions with Leu196 in E5, abrogated AID nuclear exclusion (Fig. 2.2H). On the other hand, mutating residues 114 to 116, which do not contact E5 in the 3D model, did not affect AID localization (Fig. 2.2H).

We conclude that cytoplasmic retention of AID relies on a specific conformation in which E5 interacts with the enzymatic core through the NES, leaving residues around Asp187 poised to interact with the retention factor(s). The measurable effect of the

conservative D187E substitution in human and mouse AID (Fig. 2.2G and 2.1F) suggested that even subtle changes on the surface could affect this interaction.

2.3.3. Mutually exclusive interaction of AID with eEF1A and HSP90

The factors mediating AID retention in the cytoplasm are unknown. eEF1A is a likely candidate because its interaction with cytoplasmic AID is prevented by AID mutations D187A/D188A (Hasler et al., 2011), but whether it is necessary for AID retention is unknown. Moreover, eEF1A is in a large cytoplasmic complex with AID and whether HSP90 is part of the same or a distinct AID complex is unknown (Hasler et al., 2011). The graded cytoplasmic retention efficiency of AID-GFP proteins from different species and of human AID-GFP with amino acid substitutions in E5 provided us with unique tools to probe the interaction of AID with eEF1A and HSP90. Co-immunoprecipitation (co-IP) experiments showed that the binding efficiency of each tested AID homolog or point mutant to eEF1A in human cells directly correlated with its relative cytoplasmic retention efficiency and was inversely correlated to its interaction with HSP90 (Fig. 2.3A). Thus, the fish AID-GFP homologs, which were not retained, barely interacted with eEF1A, while mouse AID-GFP, of stronger retention, bound more efficiently than human AID-GFP. The reciprocal replacement of the two amino acid differences between human and mouse AID E5 affected their eEF1A binding, exactly mirroring the effect of each replacement on cytoplasmic retention. Human AID-GFP with the Ile108-Phe109-Thr110 motif mutated to Ala failed to interact with eEF1A, in line with its loss of nuclear exclusion. Finally, substituting human AID Asp187 with Glu or Ala

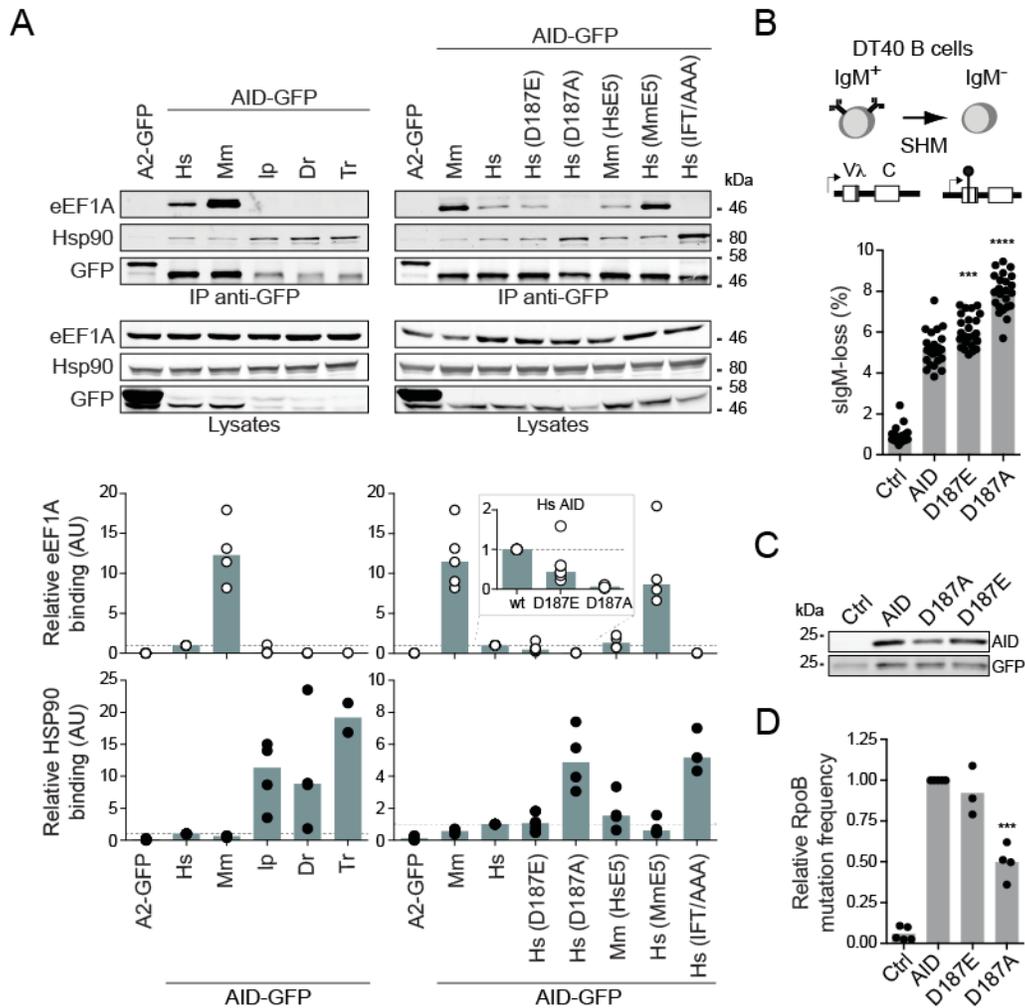


Figure 2.3. Cytoplasmic AID is present in at least two mutually exclusive complexes.

(A) Total lysates from HEK293 cells transfected with GFP-tagged APOBEC2 (A2), AID orthologs or AID mutants were immunoprecipitated with anti-GFP microbeads and analyzed by Western blot as indicated (top panels). One of ≥ 3 independent experiments is shown. Hs, human; Mm, mouse; Ip, catfish; Dr, zebrafish; Tr, pufferfish. Quantification of Western blot signals by densitometry from ≥ 3 co-IP for each AID variant were plotted to compare their binding to eEF1A (white circles) and HSP90 (black circles) (bottom graphs). Values were normalized to that of human AID in each experiment. Grey columns indicate medians. (B) SHM assay in IgM⁺ DT40 $\Delta\psi$ VL *Aicda*^{-/-} B cells complemented with AID variants in pMXs-(AID)-ires-GFP or empty vector (Ctrl). SHM frequency is proportional to the median accumulation of surface IgM-negative cells in multiple populations over time (top panel). Medians IgM-loss (grey bars) for 24 individual populations (black dots) per variant were plotted (lower graph). One representative out of 2 experiments is shown. (C) Western blots for AID and GFP expression in total cell lysates from (C) are

*shown. One representative out of 2 repeats. (D) Relative enzymatic activity of AID variants was estimated from the frequency of rpoB mutation they induce in E. coli. Means (grey bars) of the relative medians (black dots) from ≥ 3 independent experiments normalized to AID are shown. (B, D) Differences relative to wt AID were tested by Anova with Dunnetts post-test (***, $P < 0.0004$; ****, $P < 0.0001$).*

showed a progressive decrease in their interaction with eEF1A that was proportional to their relative effect on AID localization. As mentioned, the relative interaction of each AID variant with HSP90 was inversely correlated to their eEF1A binding. Since the effect of mutating Asp187 on cytoplasmic retention of human AID does not depend on the GFP tag (Hasler et al., 2011), we compared the relative ability of untagged human AID wt, D187E and D187A to perform SHM. We used an engineered AID-deficient DT40 B cell line, in which the SHM capacity of transduced AID variants is directly proportional to the fraction of cells losing surface IgM expression over time (Arakawa et al., 2004; Zahn et al., 2014) (Fig. 2.3B). In line with their relative effect on retention, AID D187E and D187A displayed increasingly higher SHM capacity (Fig. 2.3B) despite their expression levels and catalytic activity were similar to or lower than wt AID (Fig. 2.3C,D). These results strongly suggest that eEF1A is part of the AID cytoplasmic retention complex and that this interaction limits SHM, with the caveat that mutations in AID E5 could affect other AID interactions that we cannot control for. We also conclude that eEF1A and HSP90 form distinct complexes with AID.

2.3.4. eEF1A is necessary for AID cytoplasmic retention

The fact that eEF1A is extremely abundant and an essential cellular factor (Trachsel, 1996) precludes genetic ablation approaches to test whether it is necessary to retain AID in the cytoplasm. To circumvent this limitation, we resorted to didemnin B (DidB), a natural cyclic depsipeptide that binds eEF1A to inhibit its function and block

protein synthesis (Crews et al., 1996; Marco et al., 2004). DidB treatment rapidly increased the nuclear abundance of both AID-GFP and untagged AID in HeLa cells (Fig. 2.4A). This effect was specific; APOBEC3G-GFP and GFP-APOBEC1, which do not interact with eEF1A (Hasler et al., 2011), as well as endogenous GAPDH and HSP90 remained largely cytoplasmic after treatment (Fig. 2.4A). We also found that DidB prevented the co-IP of eEF1A with AID (Fig. 2.4B). To exclude the possibility that these effects were due to off-target consequences of DidB, we used two other inhibitors of translation elongation. These were cytotrienin A (CytA), an ansamycin compound chemically distinct from DidB (Fig. 2.4C) that we previously showed inhibits translation elongation specifically by interfering with eEF1A function (Lindqvist et al., 2010) and cycloheximide (CHX), an elongation inhibitor that targets the ribosome (Robert et al., 2009). CytA had similar effects to DidB in inducing nuclear accumulation of AID-GFP, but not GFP-APOBEC1, in HeLa cells, while CHX did not affect either protein (Fig. 2.4D).

To confirm these results in B cells, we first used Ramos B cells stably expressing human AID-GFP (2-fold over endogenous AID, not depicted). As expected, the three drugs inhibited translation elongation in Ramos to similar extents, as inferred from polysome stabilization profiles (Fig. 2.5A). None of the treatments changed the relative distribution of AID, HSP90 or eEF1A in sucrose gradients, according to Western blots of the fractions (not shown). However, DidB and CytA significantly increased nuclear AID accumulation in a dose-dependent and saturable manner, while CHX had little effect (Fig. 2.5B). Anticipating toxicity issues (see below), we tested very low doses of DidB that were tolerated for extended periods of time. Even at 1 nM DidB we could detect a

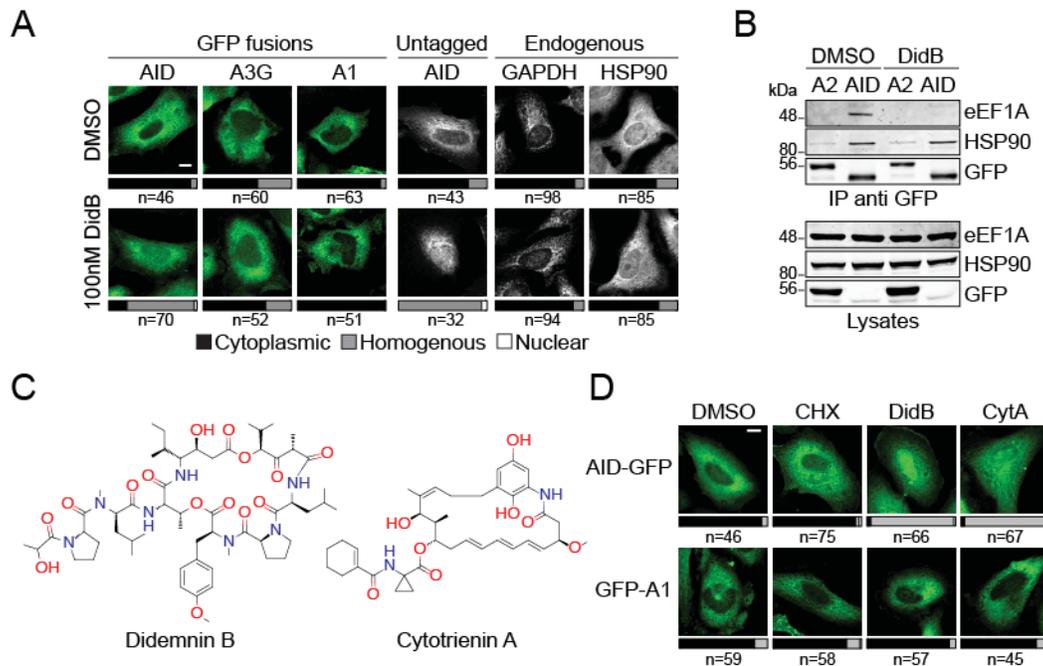


Figure 2.4. eEF1A1 inhibition causes nuclear accumulation of AID

(A) Representative confocal microscopy images of HeLa cells transiently expressing GFP-tagged AID, APOBEC3G (A3G) or APOBEC1 (A1) or untagged AID are shown. Localization was assessed by detecting GFP, or by IF for overexpressed untagged AID and endogenous GAPDH and HSP90. Cells were treated with either vehicle or 100 nM Didemnin B (DidB) for 2 h prior to fixation. Bars represent the proportion of *n* cells from 2 independent, pooled experiments showing each subcellular distribution (cytoplasmic, black; homogenous, grey; nuclear, white). (B) Total cell extracts from HEK293 cells transiently transfected with GFP-tagged APOBEC2 (A2) or AID and treated for 4 h with either vehicle or 100 nM DidB were immunoprecipitated with anti-GFP microbeads and analyzed by Western blot as indicated. One representative out of 3 independent experiments is shown. (C) Chemical structure of Didemnin B and Cytotrienin A. (D) Representative confocal microscopy images of HeLa cells transiently transfected with GFP-A1 or AID-GFP are shown. Cells were treated with DMSO, 2.5 μ M CHX, 100 nM DidB or 1 μ M CytA for 2 h and scored as in (A) from 2 independent, pooled experiments. Magnification 400X (A,D). Scale bars, 10 μ m, throughout.

significant increase in nuclear AID-GFP in Ramos cells by 24 h (Fig. 2.5C). Since nuclear AID is actively destabilized (Aoufouchi et al., 2008), the reduction in cellular AID-GFP levels in DidB-treated cells quantified by flow cytometry (Fig. 2.5C), was also

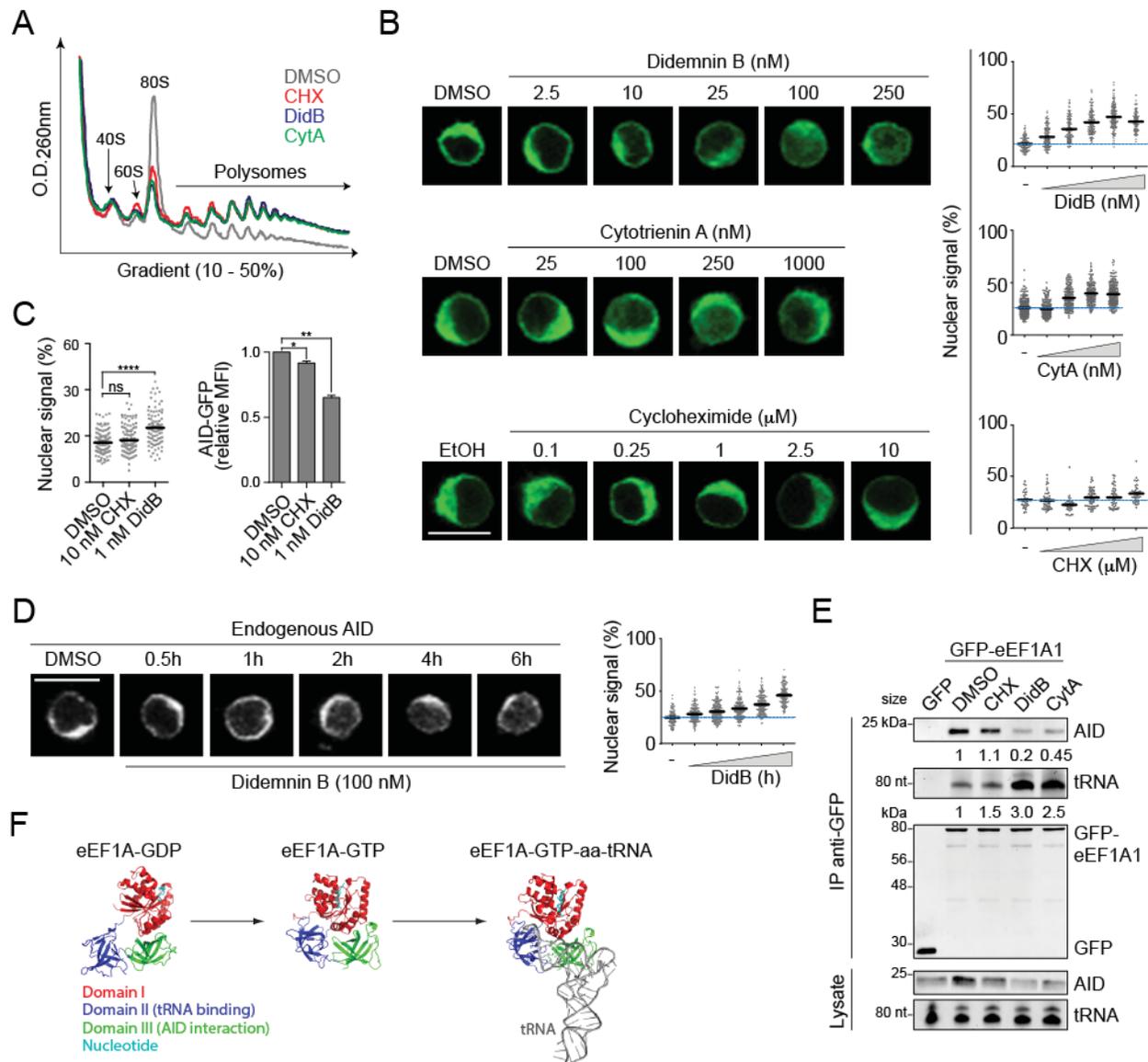


Figure 2.5. eEF1A is necessary for cytoplasmic retention of AID.

(A) UV absorbance profile of total extracts from Ramos B cells treated for 2 h with either vehicle, 2.5 μ M CHX, 100 nM DidB or 1 μ M CytA, and fractionated by sucrose gradient. One representative out of 3 independent experiments is shown. (B) Representative confocal microscopy images (left) and nuclear signal quantitation (right) of Ramos B cells stably expressing AID-GFP treated for 2 h with either vehicle or increasing concentrations of DidB (top), CytA (middle) or CHX (bottom), as indicated. Plots (right graphs) indicate the quantitated nuclear signal in >40 individual cells per point, from 1 out of 2 independent experiments. Mean values are shown as black horizontal bars, and the mean value of untreated cells is shown as a blue line across the plot. (C) Ramos B cells stably expressing AID-GFP were treated for 24 h with either DMSO, 10 nM CHX, or 1 nM DidB and the nuclear GFP signal was quantified as in (B) (Left panel). One out of 2 independent experiments is shown. The mean fluorescence

intensity (MFI) of AID-GFP measured by flow cytometry and normalized to DMSO + SD was plotted (Right panel) for 3 independent experiments. Differences relative to DMSO were analysed by Anova with Dunnett's post-test (*, $P < 0.05$; **, $P < 0.005$; ****, $P < 0.0001$). (D) Representative confocal microscopy images are shown (left) of endogenous AID (detected by IF) in Ramos B cells after DMSO or 100 nM DidB treatment for the indicated times. Plots show nuclear signal quantified as in (B) from one out of 2 independent experiments (right panel). (E) Total cell extracts from HeLa cells transiently expressing AID and GFP or GFP-eEF1A1 and treated for 4 h with either vehicle, 2.5 μ M CHX, 100 nM DidB or 1 μ M CytA, were immunoprecipitated with anti-GFP microbeads and analyzed by Western blot or by Urea-polyacrylamide gel electrophoresis, stained with SYBR Gold to detect tRNA. One representative out of 3 independent experiments is shown. The mean binding of GFP-eEF1A1 to AID or tRNA normalized to DMSO (measured by densitometry) is indicated under the relevant blots. (F) Structure of the eEF1A homolog EF-Tu (33% identical to eEF1A1) in the GDP- (PDB:1TUI) and GTP-bound forms (PDB:1TTT, drawn with and without the tRNA molecule (grey)). The domains of EF-Tu (I, red; II, blue; III, green), and bound nucleotide (teal) are indicated. Magnification 630X (B, D). Scale bars, 10 μ m throughout.

consistent with an increased proportion of nuclear AID. CHX did not increase nuclear AID-GFP and had a much smaller effect on its levels (Fig. 2.5C), indicating that inhibiting translation *per se* was not at the root of our observations with DidB and CytA. Finally, we confirmed that endogenous AID in Ramos B cells also relocalized to the nucleus in response to DidB treatment, although with slower kinetics than AID-GFP (Fig. 2.5D), in keeping with the weaker cytoplasmic retention of AID-GFP compared to untagged AID (Fig. 2.2D).

Mechanistically, DidB and CytA, but not CHX, impaired the interaction of eEF1A with AID (Fig. 2.5E). There are two distinct conformers of eEF1A depending on whether it is bound to GTP or GDP (Andersen et al., 2003) (Fig. 2.5F). It is unknown if AID binds specifically to one form or the other, but the GTP-bound form of eEF1A binds to the aminoacyl-tRNA, which largely overlaps with the eEF1A domain III that is necessary for the interaction with AID (Hasler et al., 2011) (Fig. 2.5F). It is thought that DidB freezes eEF1A in the GTP-bound conformation (Marco et al., 2004). Accordingly, we found that

in cells treated with DidB or CytA the association of eEF1A with tRNA was increased at the same time that the interaction with AID was decreased; while neither interaction was affected by CHX (Fig. 2.5E). This suggests that the interaction of AID and tRNA with eEF1A are mutually exclusive. We conclude that cytoplasmic retention of AID is mediated by a mechanism dependent on the tRNA-free conformer of eEF1A but independent of eEF1A's role in translation.

2.3.5. eEF1A limits isotype switching and chromosomal translocations

eEF1A could be expected to limit AID activity, through cytoplasmic retention, and/or promote it, by stabilizing AID (Fig. 2.5C). To measure the biological relevance of the AID cytoplasmic retention mechanism by eEF1A we induced isotype switching in mouse splenic B cells, which also induces the expression of eEF1A (Fig. 2.6A), and 24 h later treated with DidB. We used DidB at around 1 nM, which had low toxicity over several days and still caused a measurable accumulation of AID in the nucleus (Fig. 2.5C). We observed a dose dependent increase in overall CSR efficiency to IgG1 in B cells treated with DidB, with the 1 nM dose increasing CSR by ~70% and to IgG3 by ~50% (Fig. 2.6B,C). Even at these low concentrations DidB still reduced B cell proliferation, as measured by CFSE dilution (Fig. 2.6B,C), so the results likely underestimate the positive effect of DidB on CSR. Indeed, when we compared cells that had undergone the same number of cell divisions, DidB increased CSR to IgG1 by >100% (Fig. 2.6D). CytA treatment caused a similar increase in CSR per cell division as DidB (Fig. 6E), while CHX caused only a marginal increase in CSR (Fig. 2.6F). Next, we

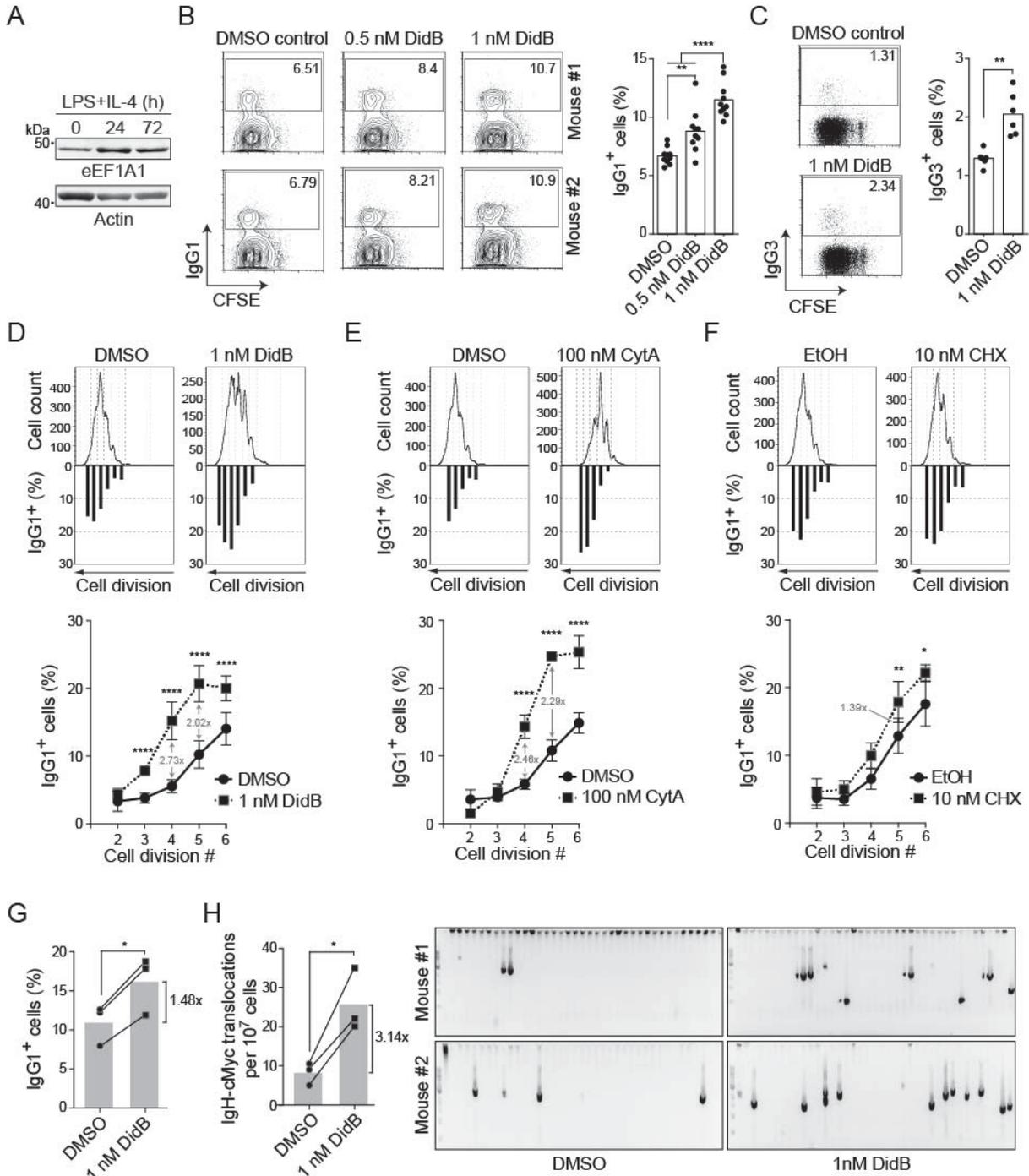


Figure 2.6. AID cytoplasmic retention limits CSR and chromosomal translocations.

(A) Western blot analysis was performed with anti-eEF1A and anti-Actin antibodies on 40 μ g of total lysate from mouse splenic B cells at 0, 24 and 72 h after stimulation with 5 μ g/mL LPS and 5 ng/mL IL-4.

One representative experiment out of two is shown ($n=1$ mouse/experiment). **(B-F)** CSR assays were performed on purified mouse naïve splenic B cells loaded with CFSE and stimulated with 5 μ g/ml LPS and 5 ng/ml IL-4 to induce switching to IgG₁, or with 15 μ g/ml LPS for IgG₃. The indicated inhibitors were added to the medium 24 h after stimulation and CSR assayed 3 days post-stimulation by flow cytometry using biotinylated anti-IgG1 or anti-IgG3 antibodies, followed by anti-biotin APC. **(B)** Representative flow cytometry plots (left) of B cells induced for class switching to IgG₁, and treated with vehicle or DidB as indicated. The graph (right) compiles the proportion of IgG₁⁺ B cells from 10 mice in 6 independent experiments. Means were compared by one-way Anova (**, $P=0.0029$; ****, $P<0.0001$). **(C)** Representative flow cytometry (left) and compilation plots (right) as in (B) for class switching assays to IgG₃ of 6 mice in 3 independent experiments ($n=2$ mice/experiment). Means were compared by Student's *t*-test (**, $P=0.0047$). **(D-F)** Representative graphs of the proportion of IgG₁⁺ splenic B cells per cell division were plotted below their corresponding CFSE plots as shown (top panels). Graphs in the bottom panels show compiled data from 6 mice in 4 independent experiments in which B cells were treated with DidB **(D)** or CytA **(E)**, and from B cells which were treated with CHX ($n=4$ mice; 2 experiments) **(F)**. Differences for each point were analyzed by two-way Anova (*, $P<0.05$; **, $P<0.005$; ****, $P<0.0001$). **(G)** CSR to IgG₁ in B cells from three TP53^{-/-} mice was performed as in (B) in 2 independent experiments, compared by Student's *t*-test (*, $P=0.0151$). **(H)** Igh-cMyc chr12 derivative translocations were measured in B cells from (G) (left graph). Each dot indicates the frequency of translocations for an individual mouse and bars show the means, compared by Student's *t*-test (*, $P=0.0387$). The gels show representative ethidium bromide staining of PCR bands for Igh-cMyc fusions from 2 of the mice (right panels). Translocations were confirmed by sequencing (data not shown).

asked whether cytoplasmic retention would limit the potential of AID to induce the chromosomal translocations that occur as side effects of CSR (Ramiro et al., 2004). We used B cells from TP53^{-/-} mice because they facilitate the detection of cells with translocations (Ramiro et al., 2006). As observed for wt B cells, 1 nM DidB increased CSR by ~50% in activated TP53^{-/-} B cells (Fig. 2.6G) and concomitantly increased the frequency of translocations by ~3-fold (Fig. 2.6H). We conclude that cytoplasmic retention of AID mediated by eEF1A limits CSR as well as oncogenic translocations.

2.3.6. Functional integration of mechanisms regulating cytoplasmic AID

The existence of two mechanisms to exclude AID from the nucleus (i.e. nuclear export and cytoplasmic retention) begs the question as to their relative roles in regulating AID. We analyzed the localization of endogenous AID in Ramos B cells treated with the eEF1A inhibitors and/or LMB. Incubation with 100 nM DidB, 1 μ M CytA or 10 ng/mL LMB each modestly increased nuclear signal of AID after 2 h (Fig. 2.7A). However, the combination of DidB or CytA with LMB had a synergistic effect, resulting in full nuclear accumulation of AID (Fig. 2.7A). CHX showed little effect alone and no interaction with LMB (Fig. 2.7A). An increase in the proportion of nuclear AID would be reflected in reduced AID stability (Aoufouchi et al., 2008). Indeed, DidB and CytA reduced AID stability in Ramos B cells, as shown by Western blot of endogenous AID, and by the kinetics of AID-GFP decay (Fig. 2.7B,C). CHX had a very small effect on AID stability, indicating that the effect of DidB and CytA was not due to inhibiting translation. Note that DidB and CHX have equal effects on translation at the doses used (Robert et al., 2009). DidB and CytA synergized with LMB to destabilize AID, while CHX did not (Fig. 2.7B,C). However, we noted that LMB alone did not destabilize endogenous AID to the same extent as DidB or CytA (Fig. 2.7B). We then asked whether nuclear export and cytoplasmic retention were also complementary in limiting AID function. Treating activated mouse splenic B cells with LMB was toxic, but low doses that significantly increased the proportion of AID-GFP in the nucleus of Ramos B cells over 24 h (Fig. 2.7D) were tolerated, albeit they reduced cell proliferation (Fig. 2.7E). Nonetheless, LMB failed to increase CSR while DidB increased CSR in the same cells (Fig. 2.7E).

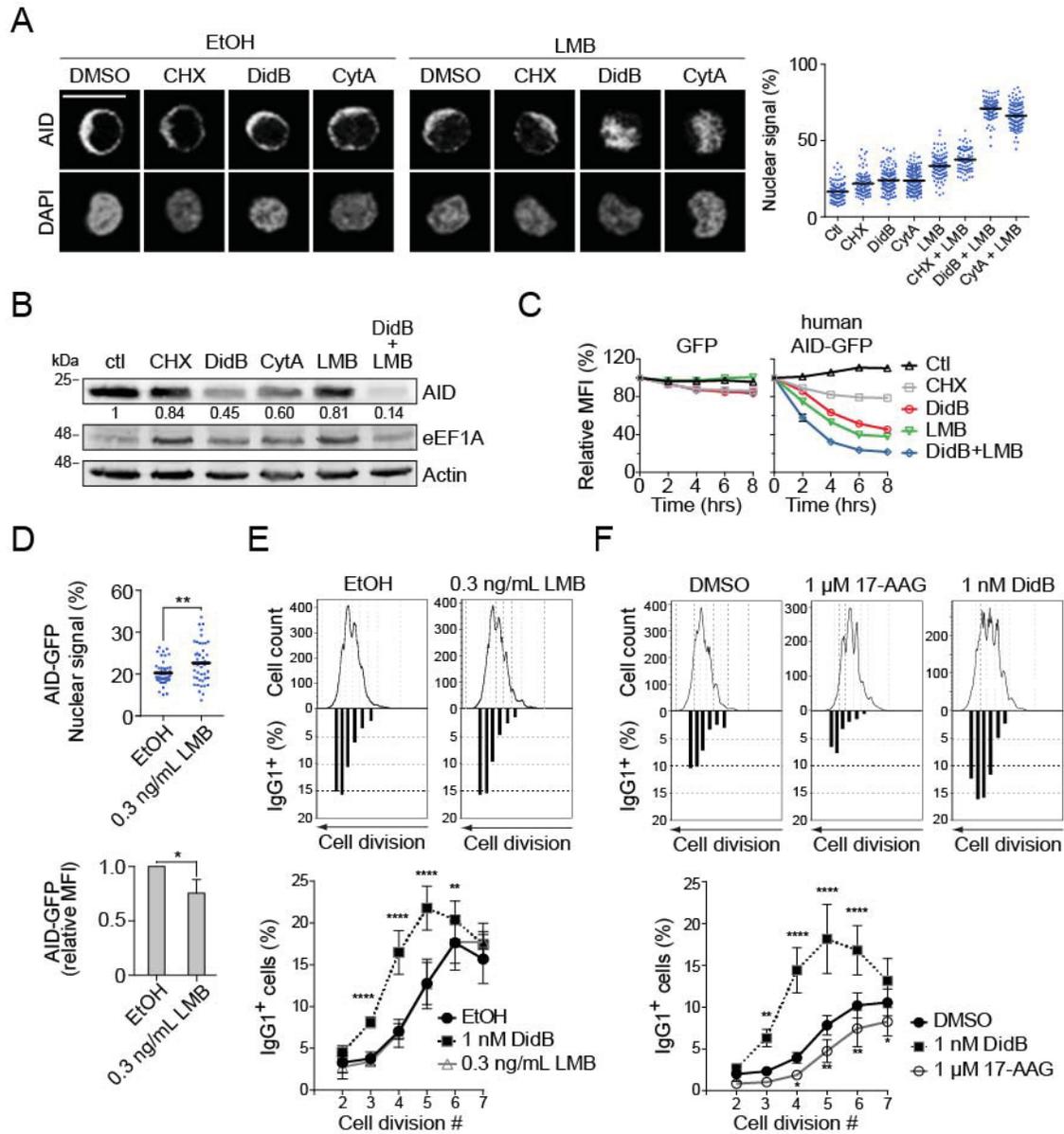


Figure 2.7. eEF1A, CRM1 and Hsp90 associate with functionally distinct AID fractions.

(A) Representative confocal microscopy images are shown of endogenous AID (detected by IF) in Ramos B cells (left panels). Cells were treated for 2 h with DMSO, 2.5 μ M CHX, 100 nM DidB or 1 μ M CytA in combination with either vehicle or 10 ng/mL LMB. Magnification, 630X. Scale bar, 10 μ m. The proportion of nuclear AID signal for ≥ 100 cells per point was assessed as the ratio of AID signal overlapping DAPI signal to total AID signal $\times 100$. The nuclear AID signal of each cell (dots), and the mean for each group (black bars) are shown for one out of 2 independent experiments, (right graph). (B) Ramos B cells were

treated as in (A) for 6 h and expression of AID, actin and eEF1A were analyzed by Western blot. One representative out of 2, independent experiments is shown. The mean ratio of AID to actin measured by densitometry, normalized to DMSO is indicated under the relevant blots. (C) Ramos B cells stably expressing either GFP or human AID-GFP were treated in duplicate as in (A). The MFI \pm SEM of GFP was measured by flow cytometry at different times post-treatment and normalized to $t_0=100\%$, and was plotted over time for one out of ≥ 2 representative, independent experiments. (D) Ramos B cells stably expressing AID-GFP were treated for 24 h as indicated, and the nuclear GFP signal was quantified as in (A) (top panel). The MFI + SD of AID-GFP was measured by flow cytometry and normalized to vehicle, and was plotted from 4 independent experiments (bottom panel). Means were compared by Student's *t*-test (*, $P < 0.05$; **, $P < 0.005$). (E, F) Representative plots of the proportion of IgG₁⁺ cells per division are shown below the corresponding CFSE profiles in wt mouse naïve splenic B cells stimulated with 5 μ g/ml LPS and 5 ng/ml IL-4 in the presence of the indicated inhibitors 24 h later. The proportions of IgG₁⁺ cells were determined by flow cytometry 3 days after stimulation (shown in bottom graphs). Data were compiled from 4 mice from 2 experiments, ($n=2$ mice/experiment). Differences relative to vehicle were analyzed by two-way Anova (*, $P < 0.05$; **, $P < 0.005$; ***, $P < 0.0004$; ****, $P < 0.0001$).

Together, these results show that cytoplasmic retention is complementary to nuclear export in excluding AID from the nucleus. However, the data suggest that these two mechanisms might not be functionally analogous.

We then compared the functional roles of the interaction of AID with HSP90 and eEF1A. Despite the HSP90 inhibitor 17-AAG affecting cell proliferation and AID stability (Orthwein et al., 2010) similarly to DidB, testing both drugs in parallel showed a clear difference, with DidB increasing and 17-AAG decreasing CSR per cell division (Fig. 2.7F). We conclude that, unlike HSP90-associated AID, the eEF1A complex stores functionally competent AID in the cytoplasm. Altogether our results suggest a model in which shuttling between different complexes underpins the production and storage of CSR-competent AID in the cytoplasm.

2.4. Discussion

We identify a mechanism that retains AID in the cytoplasm through eEF1A, thereby excluding functional AID from the nucleus of B cells and limiting CSR per cell division and oncogenic chromosomal translocations. We also compare cytoplasmic retention to nuclear export and cytoplasmic stabilization of endogenous AID by HSP90 in B cells to propose a minimal model for the transit of AID through distinct cytoplasmic complexes to acquire functional competence.

The stoichiometric interaction between AID and eEF1A in the cytoplasm, which required AID Asp187 and 188, suggested a role for eEF1A in cytoplasmic retention of AID (Hasler et al., 2011). This observation was not conclusive because the same mutations prevent the interaction of AID with other factors, at least with PABPC1 and hnRNPA (SPM and JMDN unpublished results). Moreover, it is unknown whether AID and eEF1A interact directly or not (Hasler et al., 2012). Thus, those observations were insufficient to tell whether eEF1A was necessary for AID cytoplasmic retention or just part of a larger complex containing both proteins (Hasler et al., 2011). We demonstrate that eEF1A is a necessary component of the complex that retains AID in the cytoplasm. First, the cytoplasmic retention efficiency of different AID-GFP variants correlates perfectly with their relative ability to interact with eEF1A. Second, two translation elongation inhibitors that target eEF1A, DidB and CytA, disrupt its interaction with AID, increase the AID nuclear accumulation and boost CSR and *Igh-cMyc* fusions. The very different chemical nature and structures of DidB and CytA make it extremely unlikely that non-specific (i.e. eEF1A-independent) effects can explain their identical effects on AID biology at so many levels. In addition, eEF1A is by far the most abundant target of

DidB (Crews et al., 1994; Crews et al., 1996), which has a ~20 nM IC₅₀ for translation inhibition (Robert et al., 2009). The only other known target of DidB, a lysosomal palmitoyl thioesterase, is inhibited at 5 μM IC₅₀ (Meng et al., 1998). Thus, at the ~1 nM doses we used in functional assays, eEF1A is in all likelihood the main target of DidB. The mechanism whereby eEF1A retains AID in the cytoplasm is neither related to its role in mRNA translation, as shown by the lack of effect of CHX, nor to the transcription-dependent nuclear export pathway in which eEF1A participates. The latter pathway requires a sequence signal that is absent in AID and is inhibited by Actinomycin D (Khacho et al., 2008), which does not affect AID localization (data not shown). Rather, our data indicates a mechanism in which DidB and CytA favor a specific conformer of eEF1A that is unable to interact, directly or indirectly, with AID. Although DidB does not inhibit the GTPase activity of eEF1A, it stabilizes a GTP-bound-like conformer by bridging domains I and II of eEF1A and preventing the conformational change after GTP hydrolysis (Marco et al., 2004). Our results are consistent with this mechanism, as DidB and CytA both favor the association of eEF1A to tRNA (Fig. 2.5E). The tRNA binds to eEF1A domain II but it is large enough to sterically interfere with AID (or an adaptor) binding to domain III (Fig. 2.5F). Thus, AID is retained in the cytoplasm in a complex with the eEF1A-GDP conformer. The complex does not contain polyribosomes or ribosomal units, as judged from sucrose gradient fractionations, either under basal conditions or after DidB treatment (data not shown), and is distinct from the complex that AID forms with HSP90. Future work will determine the composition of this complex.

Optimal binding of AID to eEF1A, and thus cytoplasmic retention, is conformation-dependent. We present a structural model in which the C-terminal E5

domain is folded onto the catalytic core of AID, interacting largely through the hydrophobic NES residues. We obtained the same model of AID from phylogenetically distant species simply based on thermodynamic restrictions. Additionally, the model is consistent with several experimental results: i) mutations in the NES are more efficient than LMB at disrupting nuclear exclusion of AID (Geisberger et al., 2009; Ito et al., 2004; McBride et al., 2004; Patenaude et al., 2009); ii) the C-terminal GFP fusion, or an insertion separating the E5 from the core, interferes with AID retention (Fig. 2.2D); iii) several model guided mutation of non-NES residues have the predicted effect on nuclear import kinetics (Fig. 2.2G,H); iv) Asp187 and 188, which mediate cytoplasmic retention and the eEF1A interaction (Hasler et al., 2011; Patenaude et al., 2009), are exposed on the surface.

The need of a specific AID conformation for interacting with eEF1A is consistent with AID forming a distinct complex with HSP90 (Fig. 2.3). Since HSP90 does not assist the initial folding of proteins but stabilizes metastable conformations (Young et al., 2004), a straightforward interpretation of our data is that this chaperoning pathway facilitates an eEF1A-compatible AID conformation (Orthwein et al., 2010; Orthwein et al., 2012). Accordingly, inhibiting HSP90 releases mostly unstable AID that is degraded in the cytoplasm (Orthwein et al., 2010) thus placing the interaction of AID with HSP90 before its association to eEF1A during the cellular life cycle of AID. This would also explain the slow kinetics of degradation of AID after HSP90 inhibition compared to other HSP90 client proteins in B cells (Orthwein et al., 2010). Indeed, the pool of AID in complex with eEF1A is insensitive to HSP90 inhibitors and stable in the cytoplasm, thereby delaying the decay of cellular AID levels observed after HSP90 inhibition.

Our results definitely demonstrate that nuclear export and cytoplasmic retention of AID are different mechanisms, since they are mediated by different factors and inhibited by different drugs. The AID 3D model suggests how this is possible despite being mediated by the same very small region of AID. The NES is hidden to expose the cytoplasmic retention motif. In turn, the E5 region must at some stage adopt an extended conformation for interacting with CRM1 (Sun et al., 2013b). The combined use of DidB or CytA and LMB indicates that cytoplasmic retention and nuclear export are complementary in excluding endogenous AID from the nucleus. However, the data suggest that they are not functionally equivalent because the eEF1A inhibitors increase CSR while the CRM1 inhibitor does not, despite both similarly augmenting the nuclear abundance of AID (Fig. 2.7E and compare 2.5C to 2.7D). Thus AID in complex with eEF1A is competent for CSR, as evidenced by the increased CSR upon its release by DidB. It has been reported that DidB can increase the titer of antigen-specific antibodies after immunization (Montgomery et al., 1987), suggesting it could work similarly *in vivo*. It is also conceivable that boosting the activity of AID through DidB or CytA could be used to sensitize cells expressing AID to DNA repair inhibitors (Lamont et al., 2013). In contrast, the AID pool that is exported by CRM1 seems unable to perform CSR. We cannot rule out that LMB prevents CSR by some other means but it is intriguing that LMB has a smaller effect on endogenous AID stability than DidB or CytA (Fig. 2.7B). We speculate that AID may enter the nucleus in either CSR-proficient or not-proficient states. The transit of AID through the eEF1A complex may promote the association of AID with some factor(s) that masks the NES and/or are necessary for its function. AID imported without those associated factors would be recognized by CRM1 and exported,

rather than degraded in the nucleus. Other possibilities exist, but our results provide a working model to test in future work.

Finally, our findings have implications for the well established observation that the probability of activated B cells to undergo CSR increases with the number of cell divisions (Hodgkin et al., 1996). Since AID levels also increase with subsequent B cell divisions, AID expression probably needs to reach a certain threshold, contributing to division-linked CSR (Rush et al., 2005). Cytoplasmic retention of AID is another mechanism limiting CSR per cell division. Releasing AID from eEF1A increases the probability of CSR by increasing the effective concentration of functional AID in the nucleus, thereby reducing the threshold level of total AID required. This release comes associated to a large increase in chromosomal translocations, indicating that cytoplasmic retention is important in moderating the oncogenic potential of AID.

2.5. Materials and methods

2.5.1. Mice

C57BL6/J wt and *TP53*^{-/-} mice (JAX, Bar Harbor, ME) were kept under SPF conditions at the IRCM animal house. All animal experimentation was approved by the IRCM Animal Protection Committee according to the guidelines from the Canadian Council of Animal Care.

2.5.2. DNA constructs

GFP fusions of APOBEC2, rat APOBEC1 and APOBEC3G have been described previously (Orthwein et al., 2010). *Xenopus tropicalis* AID, a gift from Dr Andrea Bottaro (University of Rochester, NY), was PCR amplified using oligonucleotides OJ492 and OJ493 and cloned into pEGFP-N3 as a BamHI-EcoRI fragment. All other AID orthologs were a gift from Dr Reuben S. Harris (University of Minnesota, MN) and were cloned in pEGFP-N3 as HindIII-Sall fragments. Retroviral expression vectors were made by subcloning HindIII-NotI fragments from pEGFP-N3 into pMXs. Point mutations were introduced by Quickchange (Stratagene). Oligonucleotide sequences are available upon request. Human eEF1A1 was PCR amplified from cDNA using oligonucleotides OJ641 and OJ642 and cloned into pEGFP-C1 as an EcoRI-Sall fragment.

2.5.3. Reagents and Antibodies

Stock aliquots of 50 µg/mL LMB (LC Laboratories) and 100 mM CHX (Sigma) were made in Ethanol. 17-AAG (LC Laboratories) and DidB (NSC 325319, generously provided by the Natural Products branch, NCI, Bethesda, MD, USA), were prepared as 20 mM stock solutions in DMSO. CytA (Cayman Chemical, MI, USA) was prepared as a

1 mM stock solution in DMSO. Drugs were kept in the dark at -20°C, and diluted fresh before each use. Antibodies (dilutions) for IF were as follows: MAb anti-GAPDH (1:500, SC Biotechnology, H-12), MAb anti-HSP90 (1:500, BD biotechnology 68/HSP90), to detect human AID we used mouse MAb anti-AID (1:500, Invitrogen ZA001), rat MAb anti-AID (1:500, eBioscience, mAID-2 or Cell Signaling, EK2 5G9), to detect mouse AID we used mAID-2 (1:250, eBioscience). Secondary antibodies (all 1:500, Invitrogen) were anti-mouse AlexaFluor 488, anti-Rat AlexaFluor 680, anti-Rabbit 546, anti-Rabbit 680. For Western blot we used rat MAb anti-AID (1:1000, Cell Signaling), rabbit anti-GFP (1:2000, Invitrogen), mouse MAb anti-PCNA (1:3,000, Abcam, PC-10), rabbit anti-eEF1A (1:2,000, Abcam), mouse MAb anti-HSP90 (1:3,000, BD bioscience), rabbit MAb anti-Actin (1:3,000, Sigma).

2.5.4. Microscopy and immunofluorescence

HeLa, HEK293 and NIH3T3 cells were plated on coverslips and transfected using TransIT-LT1 (Mirus) according to manufacturers instructions. Cells were treated as indicated 36-48 h post-transfection before fixation. Human Ramos B cells stably expressing AID-GFP were described previously (Patenaude et al., 2009). Mouse CH12F3 B cells were retrovirally transduced using culture medium of HEK293 cells co-transfected with pMXs vectors encoding the protein of interest and vectors encoding VSV-G and GAG-Pol. Virus was attached to 6-well plates coated with 0.2 ug/mL Retronectin (Takara) by centrifugation (32°C, 30 min, 2000 x g), cells were layered on top by centrifugation (32°C, 30 min, 600 x g) and cultured for 2 days. Infected cells were sorted for similar levels of GFP expression. For GFP visualization, cells were fixed with 3.7% (w/v) formaldehyde for 10 min, washed 3X with PBS and nuclei stained with

propidium iodide (PI). Catfish 1B10 B cells (Miller et al., 1994) were cultured at 28°C, cells were electroporated as described (Ellestad and Magor, 2005) and 48 h later treated for 2 h with EtOH or 50 ng/mL LMB, then fixed as above. Fixed cells were cytospun onto microscopy slides and nuclei were stained with PI before mounting for imaging. EL1 BL, Daudi and DG75 B cell lymphoma lines were a generous gift from Dr M. Neuberger (Medical Research Council, UK). B cells were kept in suspension at 10⁶ cells/mL during treatments, after which they were attached to poly-L lysine (Sigma) coated coverslips and processed for IF. Cells were fixed in either 3.7% formaldehyde (for AID) or -20°C methanol (for GAPDH and HSP90), before permeabilization and incubation for 30 min in blocking solution (PBS, 0.5% (v/v) Triton-X100, 1% (w/v) BSA, 5% (v/v) goat serum). Cells were then stained overnight at 4°C in blocking solution containing primary antibodies followed by 3x washes with 0.01% Triton X-100 PBS (PBS-T) then a 1 h incubation with secondary antibody in blocking solution and 3 x PBS-T washes. Specific AID signal in IF was confirmed using DG75 and unstimulated CH12F3 cells. The ability of anti-AID antibodies to detect nuclear AID was confirmed using LMB-treated AID-GFP Ramos B cells. The results were identical with the three different anti-AID antibodies listed above. After nuclear staining with either PI or Dapi (300 nM in PBS), coverslips were washed with ddH₂O and mounted on slides using Lerner Aqua-Mount (Thermo Scientific). Images were acquired at room temperature using ZEN 2010 on either a Zeiss LSM 510, or LSM 700 confocal microscope with excitation lasers at 405 nM (Dapi), 488 nM (EGFP and Alexa488), 543 nM (PI and Alexa546), and 633 nM (Alexa633, Alexa680), using either 40x/1.3 or 63x/1.4 oil immersion objectives, and collected with a Hamamatsu PMT. Subcellular localization of

fluorescent signal in HeLa, 293T and NIH3T3 cells was scored using ImageJ (Rasband, WS, NIH, Bethesda, MD, USA). In Image J we quantified the proportion of nuclear signal intensity using the profile tool to calculate the mean GFP fluorescence signal across the nucleus (i.e. overlapping with PI signal) over the total mean fluorescence signal along the line for multiple cells. In Volocity (Perkin Elmer), for each individual cell, we made masks for nuclear and total GFP or IF signal. The proportion of nuclear signal was calculated as the ratio of nuclear signal / total signal X 100. For each experiment, multiple fields were analyzed for subcellular localization. Cells showing saturated signal, abnormal DNA structure or mitotic figures were excluded. For making figures, images were transferred to Photoshop for adjusting contrast throughout the whole image when necessary to enhance visibility and for cropping.

2.5.5. Structural modelling

We generated a structural model of the core domain of AID using the Rosetta suite (version: rosetta_2013wk51) and APOBEC3C (PDB ID 3VOW) as template by comparative homology modeling. In order to systematically sample suitable positions of the C-terminal tail of AID, we generated 20,000 models using the core domain as a rigid seed and the E5 was sampled with fragments of 9, 3 and 1 residues. All models were evaluated by the Rosetta energy score and the top-ranking ones were clustered using VMD (Humphrey et al., 1996), eventually selecting the most populated ones. While human and mouse models resulted in one major cluster, zebrafish displayed three isoenergetic groups, one of them being equivalent to the major cluster found in the other two species.

2.5.6. Co-Immunoprecipitations

For AID-GFP co-IP, HEK293 cells were transiently transfected with APOBEC/AID-GFP constructs using CaPO₄ precipitation 48 h before lysis. Cells were lysed for 30 min on ice in lysis buffer [20 mM Hepes pH 7.5, 150 mM NaCl, 10% glycerol, 0.2% Triton X-100, 1x Complete protease inhibitor cocktail (Roche) and 2x HALT protease and phosphatase inhibitor cocktail (Thermo)]. For GFP-eEF1A1 co-IP, HeLa cells were transiently transfected with AID and GFP or GFP-eEF1A1 using TransIT-LT1 48 h before lysis. Cells were lysed for 10 min on ice in lysis buffer [5 mM Tris pH 7.5, 1.5 mM KCl, 2.5 mM MgCl₂, 0.1% Triton X-100, 0.1% Na deoxycholate, 1mM DTT, 40 U/mL RNaseIN (Promega) and 2X Complete]. After lysate clarification, GFP immunoprecipitation was carried out using the μ MACS GFP isolation kit according to manufacturers instructions (Miltenyi Biotech), with the exception of washing. For AID-GFP co-IP, washing was done 3x with 200 μ L of wash buffer [20 mM Hepes (pH 7.5), 150 mM NaCl, 0.1% Triton X-100, and 1x Complete]. For GFP-eEF1A1 co-IP, washing was done 5x with 200 μ L of wash buffer [50 mM Tris pH 7.5, 300 mM KCl, 12 mM MgCl₂, 1% Triton X-100, 1mM DTT, and 1X Complete]. When indicated, cells were treated with DMSO, 2.5 μ M CHX, 100 nM DidB or 1 μ M CytA for 4 h prior to lysis. Elutions and total lysates were analysed by Western blot using the Li-Cor Odyssey infrared imaging system (Li-COR Biosciences).

2.5.7. Protein stability assays and biochemistry

For endogenous protein stability, Ramos cells were plated at 0.5×10^6 cells/mL density and treated with indicated drugs for 6 h. After drug incubation, cells were counted, and 2×10^6 cells were harvested, washed with PBS, and lysed directly in SDS-PAGE loading buffer. Protein levels were analyzed by Western blot. For GFP-fused

protein stability assays, the GFP signal in stably transfected Ramos B cells was measured by flow cytometry at specific times after the indicated drug treatments, as done previously (Orthwein et al., 2010). Protein expression in mouse splenic B cells was measured by Western blot. Protein lysates were collected from either resting B cells, or B cells stimulated for 24 or 72 h with 5 $\mu\text{g}/\text{mL}$ LPS and 5 ng/mL IL-4. At least 5×10^6 cells were pelleted, resuspended in PBS + 5X complete protease inhibitor cocktail and lysed in NP-40 lysis buffer [20 mM Tris-HCl pH.8, 137 mM NaCl, 10% glycerol, 2 mM EDTA and 1% NP-40] for 30 min on ice. Protein was quantified in clarified lysates using a BCA protein assay kit (Thermo Scientific), and 40 μg total protein loaded for each Western blot. Sucrose gradient fractionation was done as described (Robert et al., 2009) using 20×10^6 Ramos B cells. Cells were treated for 2 h with DMSO, 2.5 μM CHX, 100 nM DidB, or 1 μM CytA, washed with PBS, then lysed in hypotonic lysis buffer (5 mM Tris pH 7.5, 2.5 mM MgCl_2 , 1.5 mM KCl, 2 mM DTT, 1% Triton X-100 and 0.5% sodium deoxycholate) in the presence of the appropriate inhibitor. Cleared lysates were loaded onto a continuous 10-50% sucrose gradient column (10% to 50% sucrose, 20 mM HEPES pH 7.6, 100 mM KCl, 5 mM MgCl_2 and 1 mM DTT), and centrifuged at 35,000 rpm for 2 h at 4°C in a SW41 rotor. Polysomes were visualized and fractionated using a fraction collector connected to a UV254nm detector (ISCO), and 10-drop fractions were collected. Proteins from each fraction were then concentrated using StrataClean resin (Agilent technologies) and separated on SDS-PAGE gels for analysis by Western blot.

2.5.8. Monitoring AID activity

E. coli rpoB mutation assays were performed using the Δung BW310 strain expressing 6XHis-AID fusions, as described (Zahn et al., 2014). SHM activity was assessed in IgM⁺ DT40 $\Delta\psi VL Aicda^{-/-}$ cells (Arakawa et al., 2004) retrovirally complemented with AID variants. Fluctuation analysis of IgM phenotype was performed as described {Zahn, 2014 #528} using initial populations of 1000 GFP⁺ IgM⁺ cells FACS-sorted into 96-well plates and expanded for 15 d in 24 well plates before staining with anti-chicken IgM-RPE (Southern Biotech) and measuring the proportion of GFP⁺ IgM⁻ cells by flow cytometry. Class switch recombination was measured in resting B cells purified from spleens of C57BL6/J mice using MACS anti-CD43 microbeads depletion (Miltenyi Biotech), as described (Orthwein et al., 2010). Cells were stained with CFSE (Invitrogen) according to manufacturers instructions, plated at 10⁶ cells/well in 24 well plates, and stimulated with 5 μ g/mL LPS and 5 ng/mL IL-4 for IgG₁ or 15 μ g/mL LPS for IgG₃. 24 h post-stimulation, cells were treated as indicated and analyzed 48 h later. B cells were stained with biotinylated rat anti-mouse IgG1 or IgG3 antibody (BD Bioscience), followed by APC-conjugated anti-biotin antibody (Miltenyi Biotech). IgG₁ or IgG₃ as well as CFSE dilution were measured by flow cytometry on a FACS Calibur (BD Bioscience), and analyzed on FlowJo. Chromosomal translocations were detected by nested PCR amplification of the *Igh-cMyc* fusion, chr12 derivative as described (Ramiro et al., 2004; Zahn et al., 2014). PCR products were purified and sequenced at Macrogen (Seoul, Korea) using oligonucleotide OJ317.

2.6. Acknowledgements

We thank Andrea Bottaro, Reuben S. Harris, and Michael S. Neuberger for reagents. We are indebted to the Natural Products branch of the National Cancer Institute for providing Didemnin B. We thank A.K. Eranki, Dr A. Orthwein and Dr J.F. Côté for discussions and Dr C. Rada for critical reading of the manuscript. We thank the assistance of D. Fillion with microscopy, E. Massicotte with flow cytometry and Dr J. Fraszcsak for the *TP53*^{-/-} mice. This work was funded by the Canadian Institutes of Health Research operating grant MOP125991 (to JMDN), and the National Science and Engineering Research Council of Canada (to BGM). SPM was supported by fellowships from the Canadian Institutes of Health Research, Cole Foundation and Fonds de Recherche du Québec – Santé. JMDN holds a Canadian Research Chair tier 2.

The authors have no conflicting financial interests

CHAPTER 3: A LICENSING STEP LINKS AID TO TRANSCRIPTION ELONGATION FOR B CELL MUTAGENESIS

Stephen P. Methot, Ludivine C. Litzler, Poorani Ganesh Subramani, Anil K. Eranki,
Heather Fifield, Anne-Marie Patenaude, Julian C. Gilmore, Halil Bagci, Gabriel E.
Santiago, Jean-Francois Coté, Mani Larijani, Ramiro E. Verdun, Javier M. Di Noia

This article is prepared for submission to *Nature Communications*.

While studying cytoplasmic retention, we revealed a role for nuclear exclusion to limit off-target activity of AID. Based on this, we wanted to better understand the mechanisms that regulate AID targeting. Concurrently to this work, we had identified a motif in $\alpha 6$ helix of AID, which was somehow necessary for functional activity, but dispensable for mutagenesis in *E. coli*. This suggested that this motif was necessary for AID functional activity in B cells, so we set out to identify the mechanism for this defect.

3.1. Abstract

Activation-Induced deaminase (AID) mutates the immunoglobulin genes (*Ig*) to initiate somatic hypermutation (SHM) and class switch recombination (CSR) in B cells. AID further mutates a few hundred other loci, but most AID-associated genes are spared. The mechanisms underlying productive AID targeting are unclear. We identify three clustered arginine residues defining a functional AID domain. Specific replacements of any of these residues abrogate SHM, CSR and off-target activity in B cells, without affecting enzymatic activity or *E coli* mutagenesis. These AID mutants are broadly associated to chromatin in an Spt5-, RNAPII-, and RNA-dependent manner, like wt AID; and occupy the *Igh* S μ promoter. On the other hand, they fail to occupy the gene body and lose association to transcription elongation factors. Thus, AID mutagenic activity is determined not by occupancy but by a licensing mechanism, which couples AID to transcription elongation, and is enforced by limiting nuclear levels of AID.

3.2. Introduction

The enzyme Activation induced deaminase (AID) initiates genetic modifications at the immunoglobulin (Ig) genes in activated B cells (Methot and Di Noia, 2017; Peled et al., 2008). AID catalyzes the deamination of deoxycytidine to deoxyuridine on single stranded DNA (Peled et al., 2008). This change is mutagenic, but further processing of the deoxyuridines by DNA repair enzymes underpins somatic hypermutation (SHM) and class switch recombination (CSR), which are indispensable for efficient antibody responses (Methot and Di Noia, 2017; Muramatsu et al., 2000; Peled et al., 2008). As deleterious side effects of SHM and CSR, AID can mutate and induce DNA damage outside the *Ig* loci, in many cases triggering chromosomal translocations (Casellas et al., 2016).

DNA repair pathways limit off-target mutations and DNA damage by AID (Cortizas et al., 2016; Hasham et al., 2010; Liu et al., 2008). Nevertheless, several additional layers of regulation are necessary to control AID oncogenic and cytotoxic activity (Vaidyanathan et al., 2014). Regulation of AID protein levels and nuclear access restrains both on- and off-target activities, but do not seem to provide specificity (Methot and Di Noia, 2017). The preferential targeting of AID to the Ig genes and how AID mutates a small number of additional genomic loci while sparing most others, is an area of active research (Casellas et al., 2016; Robbiani and Nussenzweig, 2013). The *Ig* loci possess an intrinsic ability to attract AID activity (Yeap et al., 2015) likely conferred in part by specialized cis-acting elements combining transcriptional enhancers and multiple transcription factor binding sites, which can ectopically function to target SHM (Buerstedde et al., 2014; Dunnick et al., 2009). Similar elements have not been

identified in AID off-targets, but they share with the *Ig* the characteristic of showing convergent transcription and being associated with strong super-enhancers (Meng et al., 2014; Pefanis et al., 2014; Qian et al., 2014). Nonetheless, there are many highly transcribed genes with similar characteristics that are not mutated, so an additional layer of regulation must exist. The identity of the trans-acting factors targeting AID to the *Ig* loci is also elusive, though non-coding RNA and transcription factors likely play a role (Casellas et al., 2016). Genome wide studies have identified a few factors that correlate with AID occupancy and mutagenic activity, such as RNA polymerase II (RNAPII), its associated factor Spt5 and the RNA processing exosome (Basu et al., 2011; Nambu et al., 2003; Pavri et al., 2010). Again, these factors function at a much larger number of loci than are mutated by AID and fail to explain its specificity on their own.

Thus, there is a three tier-system of AID targeting, with the *Ig* loci being targeted much more frequently than any AID off targets, but the latter restricted to a few hundred sites. Beyond specific examples of loci occupied by AID but not mutated (Matthews et al., 2014a), the analysis of AID occupancy by ChIP-seq has suggested its association with ~6,000 genes in B cells, while AID-induced damage is limited to ~300 loci (Hakim et al., 2012; Liu et al., 2008; Meng et al., 2014; Qian et al., 2014; Staszewski et al., 2011a). This begs the question of why most sites bound by AID are spared from its activity.

Here, we report a new functional domain of AID dispensable for enzymatic activity but necessary for on- and off-target biological activity in B cells. Systematic analysis of the function and interactome of AID variants with mutations in this arginine-rich (RR) domain demonstrate they have a defect specifically in their associating

with the *Igh* gene body during transcription elongation, explaining their failure to mutate. Our results uncover a licensing mechanism that explains why occupancy is not sufficient to predict AID activity and suggest a new model for productive AID targeting.

3.3. Results

3.3.1. Three Arginine residues in AID $\alpha 6$ define a critical functional domain

In previous structure-function analyses, we used a set of chimeric proteins in which contiguous regions of AID are replaced by their homologous region from APOBEC2 (A2), (Conticello et al., 2008; Orthwein et al., 2010; Patenaude et al., 2009). Only one of these, AID-A2#5, could mutate the *E coli* genome (Supplementary Fig. 3.1a,b). AID-A2#5 replaces a large C-terminal portion of AID, starting from the loop preceding alpha-helix 6 ($\alpha 6$) and eliminating the C-terminal E5 domain, which is necessary for CSR (Barreto et al., 2003); however, this chimera lacked not only CSR activity but also IgV SHM when used to complement *Aicda*^{-/-} B cells (Supplementary Fig. 3.1c,d). Adding back E5 did not rescue CSR (Supplementary Fig. 3.1c). A smaller chimera, replacing only the $\alpha 6$ of AID with that of A2 (AID-A2 $\alpha 6$) had measurable activity in *E. coli* but not in B cells (Supplementary Fig. 3.1a-d). The functional defect of AID-A2 $\alpha 6$ could not be explained by differences in protein abundance or nuclear access (Supplementary Fig. 3.1b-e). These results suggested that the AID $\alpha 6$ contained residues required for SHM and CSR but dispensable to mutate *E. coli*.

We sought to identify single amino acid substitutions within AID $\alpha 6$ that could separate its ability to mutate *E coli* from its biological activity in B cells. Comparing a 3D molecular model of AID (Methot et al., 2015) to the A2 structure (Prochnow et al., 2007) showed many residue and charge differences in $\alpha 6$ between these paralogs (Fig. 3.1a). We independently mutated several of these AID residues to the corresponding A2

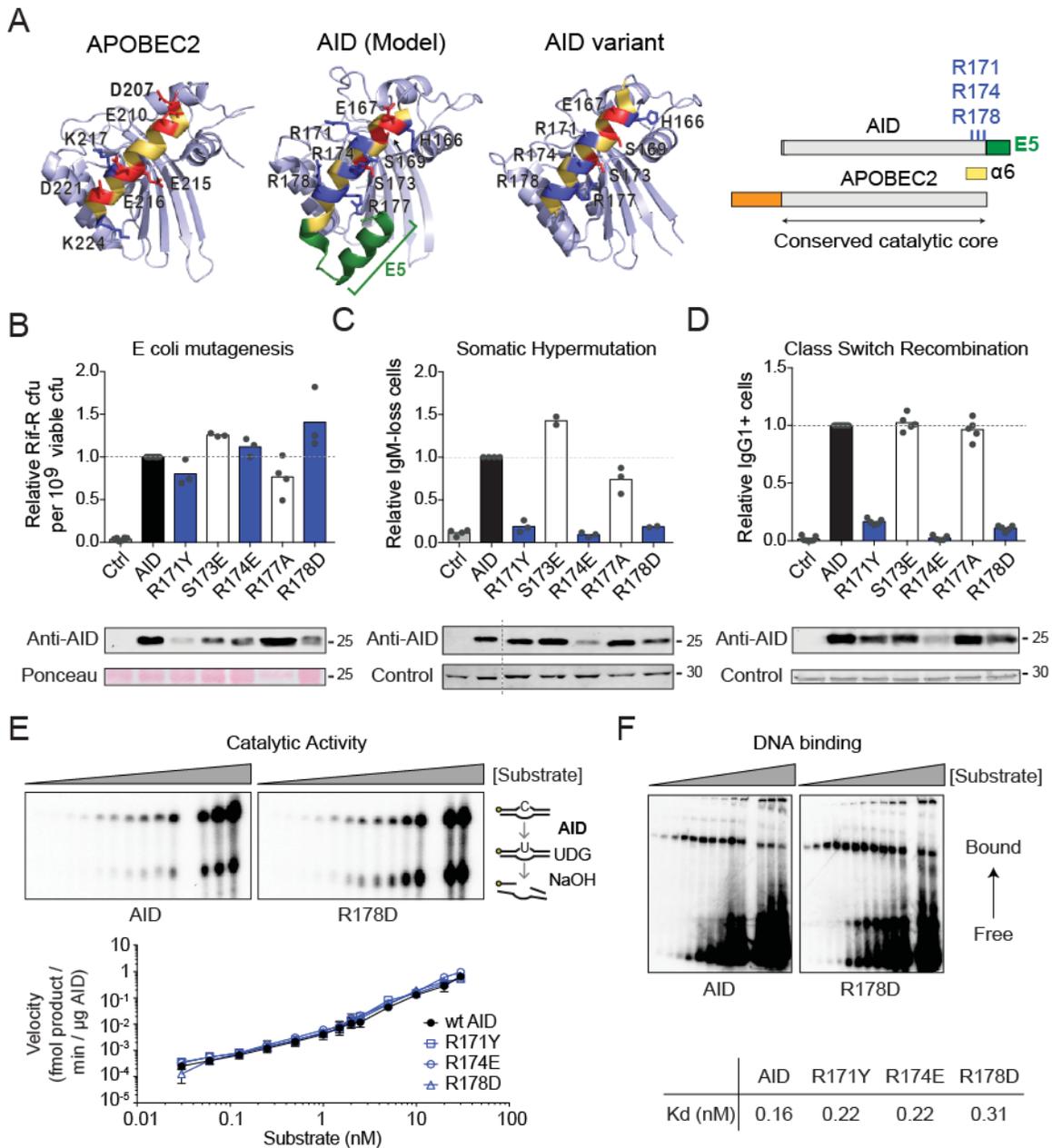


Figure 3.1. Identification of functionally inactive AID variants.

(a) (Left) Comparison of our 3D model of AID, the experimental structure of APOBEC2 (PDB 2NYT) and an AID variant (PDB 5JJ4) in the same orientation. Residues within helix $\alpha 6$ are labelled. Side chains for basic (blue) and acidic (red) residues are shown. (Right) Labeled schemes of AID and APOBEC2. (b) Mutagenic activity in *E. coli*, measured by the relative frequency of rifampicin (Rif) resistant colonies arising from cultures expressing AID variants or empty vector (Ctrl). Means (bars) of median values (dots) obtained from ≥ 3 independent experiments (5 cultures/experiment) are shown, normalized to AID. (c) Somatic hypermutation activity, assayed by the relative IgM-loss accumulation in cultures of DT40

Aicda^{-/-} ΔΨVλ B cells complemented with AID variants-ires-GFP or empty vector (Ctrl). Means (bars) of the median values (dots) obtained from ≥3 independent experiments (≥12 cultures/experiment) were normalized to the median value of AID. (d) Class switch recombination activity, monitored in naïve *Aicda*^{-/-} mouse primary B cells cultured with LPS and IL-4 and transduced with AID variants-ires-GFP or empty vector (Ctrl). The proportion of IgG1+ cells in the GFP+ population was determined 72 h after transduction. Means (bars) of ≥3 independent experiments (dots) are shown, normalized to AID. In (b-d) WB of cell extract probed with anti-AID antibody and loading control are shown on the bottom. (e) Catalytic kinetics of purified recombinant wt AID and R-mutants assayed by the standard alkaline cleavage assay for deamination. (Top) Representative alkaline cleavage assays used to measure specific activity of wt AID and R-mutants. (Bottom) Mean ± SD of 4 independent experiments were quantified. (f) DNA binding affinity of wt AID and the R-mutants assayed by EMSA. (Top) Representative EMSA gels are shown. (Bottom) Mean Kd was calculated from 4 independent experiments for each AID variant.

residue. Three of these recapitulated the results obtained with the chimeras. AID R171Y, R174E and R178D mutated *E. coli* with the same efficiency as AID but were inactive for SHM and CSR (Fig. 3.1b-d). In contrast, adjacent mutations AID R177A and S173E maintained all three activities (Fig. 3.1a-d). Notably, Arg 171, 174 and 178 are conserved in most jawed vertebrates but not in the APOBECs (Supplementary Fig. 3.1f) and form a contiguous surface in AID (Fig. 3.1a) (King et al., 2015; Methot et al., 2015; Pham et al., 2016). The natural AID variant R174S found in some immunodeficient HIGM2 patients (Quartier et al., 2004) conserves DNA binding and processivity but has substantially reduced catalytic activity (Mu et al., 2012). Furthermore, protein arginine residues are common contact points with nucleic acids (Nadassy et al., 1999). Nonetheless, the R-mutants have similar DNA binding kinetics and deaminate ssDNA within a bubble substrate *in vitro* with similar specific activities to wt AID (Fig. 3.1e,f).

To test the role of the positive charge contributed by these Arg residues, we substituted each of them for lysine (Supplementary Fig. 3.2). AID R171K had reduced *E. coli* mutation activity and a proportional decrease in SHM and CSR, indicating a

structural contribution to catalysis maintained by Tyr but not Lys. AID R174K mutated *E coli* with 50% efficiency compared to AID, and showed a proportional decrease in SHM but lacked CSR, indicating a structural contribution for R174 but also a specific role in CSR. AID R178K was indistinguishable from AID for all activities, indicating that in this case, the charge is sufficient for biological function.

We conclude that arginines 171, 174 and 178 contribute in charge but also structurally to creating a new functional domain of AID necessary for SHM and CSR. Since the substitutions to the A2 residues consistently retained enzymatic activity, we used those mutants for dissecting the role of this functional domain (hereafter, RR domain).

3.3.2. AID R-mutants access the nucleus but lack off-target activity

Nuclear access of AID is restricted, with ~10% of AID being nuclear in steady state, product of the integration of several mechanisms regulating AID nuclear import and protein stability (Orthwein and Di Noia, 2012). To exclude that the functional defect of the R-mutants was due to defective nuclear access, we analyzed their subcellular localization. We used AID-deficient CH12 B cells, in which wt AID reconstituted CSR activity but none of the R-mutants did, despite similar expression levels (Supplementary Fig. 3.3a-c). Similarly to wt AID, the R-mutants were cytoplasmic in steady state and accumulated in the nucleus after inhibiting AID nuclear export with Leptomycin B (LMB) and/or cytoplasmic retention with Didemnin B (DidB) (Fig. 3.2a). AID can be trapped

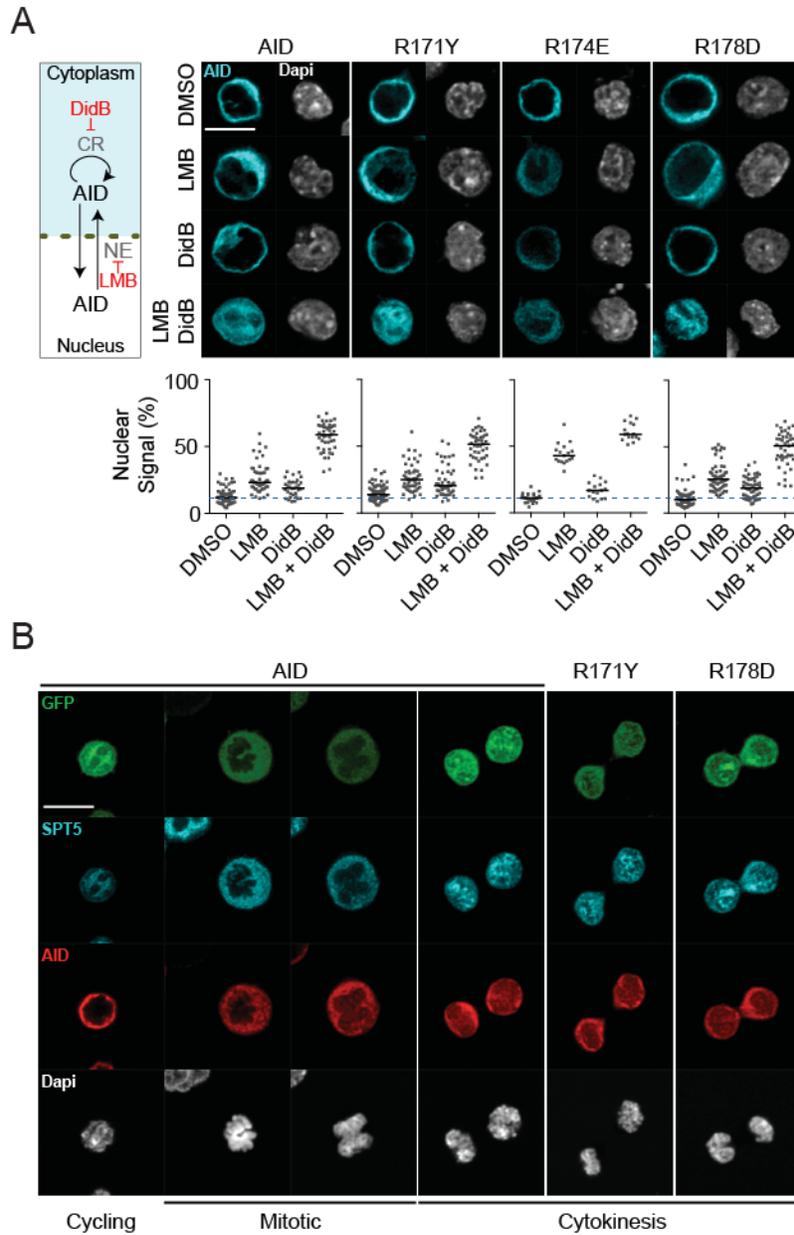


Figure 3.2. Normal nucleocytoplasmic shuttling of AID R-mutants.

(a) (Left) The mechanisms regulating AID nucleocytoplasmic shuttling. DidB inhibits cytoplasmic retention (CR); LMB inhibits nuclear export (NE). (Right) Representative confocal microscopy images of AID-deficient CH12 cells transduced with untagged AID variants and analyzed by immunofluorescence with anti-AID antibody. Cells were treated for 2 h with DMSO, or 10 ng/mL LMB, and/or 100 nM DidB, as indicated. (Bottom) Proportion of nuclear AID signal for individual cells (dots), with means indicated (black bars). A blue line across the plot indicates the mean value of untreated cells expressing wt AID. Compilation of 2 independent experiments. (b) Confocal microscopy images of CH12 cells analyzed

using anti-AID and anti-SPT5. Cells were determined to be cycling (G1/S/G2), mitotic or in cytokinesis based on DNA condensation and SPT5 access to the DNA. Images are representative of ≥ 10 observed events per construct from ≥ 2 independent experiments. (a, b) Magnification 630X. Scale bar, 10 μm .

inside the nucleus upon reforming of the nuclear envelope after mitosis (Wang et al., 2017). This was maintained by the R-mutants (Fig. 3.2b and Supplementary Fig. 3.3d). Thus, the mechanisms regulating AID nuclear access are functional in the R-mutants.

We then asked whether the R-mutants would associate to the S_{μ} region in the nucleus, the major AID target at the *Igh* locus during CSR (Xue et al., 2006). We used an antibody raised against the E5 domain of AID to perform ChIP from *Aicda*^{-/-} mouse B cells complemented with either wt AID or the R-mutants. The association of the R-mutants to S_{μ} was severely decreased compared to wt AID (Fig. 3.3a,b). As a control, we used the catalytically inactive AID E58A, which showed reduced occupancy, as expected (Vuong et al., 2013), but was still substantially higher than any of the R-mutants. Reduced association to the S_{μ} was also supported by sequencing from reconstituted *Aicda*^{-/-} *Ung*^{-/-} B cells, which showed lower mutation frequency for AID R171Y and R178D than for wt AID, proportionally to their chromatin association by ChIP (Fig 3.3a,b).

We next asked whether the apparent *Igh* targeting defect of the R-mutants extended to other loci. As a general assay for off-target mutagenesis, we measured the decrease in fitness that is associated with AID off-target DNA damage (Hasham et al., 2010; Zahn et al., 2014). AID expression was sufficient to compromise cell fitness in competitive cultures of CH12 cells, and this effect was enhanced by using a low dose of DidB, which increases the proportion of AID in the nucleus and enhances off-target activity (Methot et al., 2015) (Fig. 3.3c). Cells expressing the R-mutants showed no cell

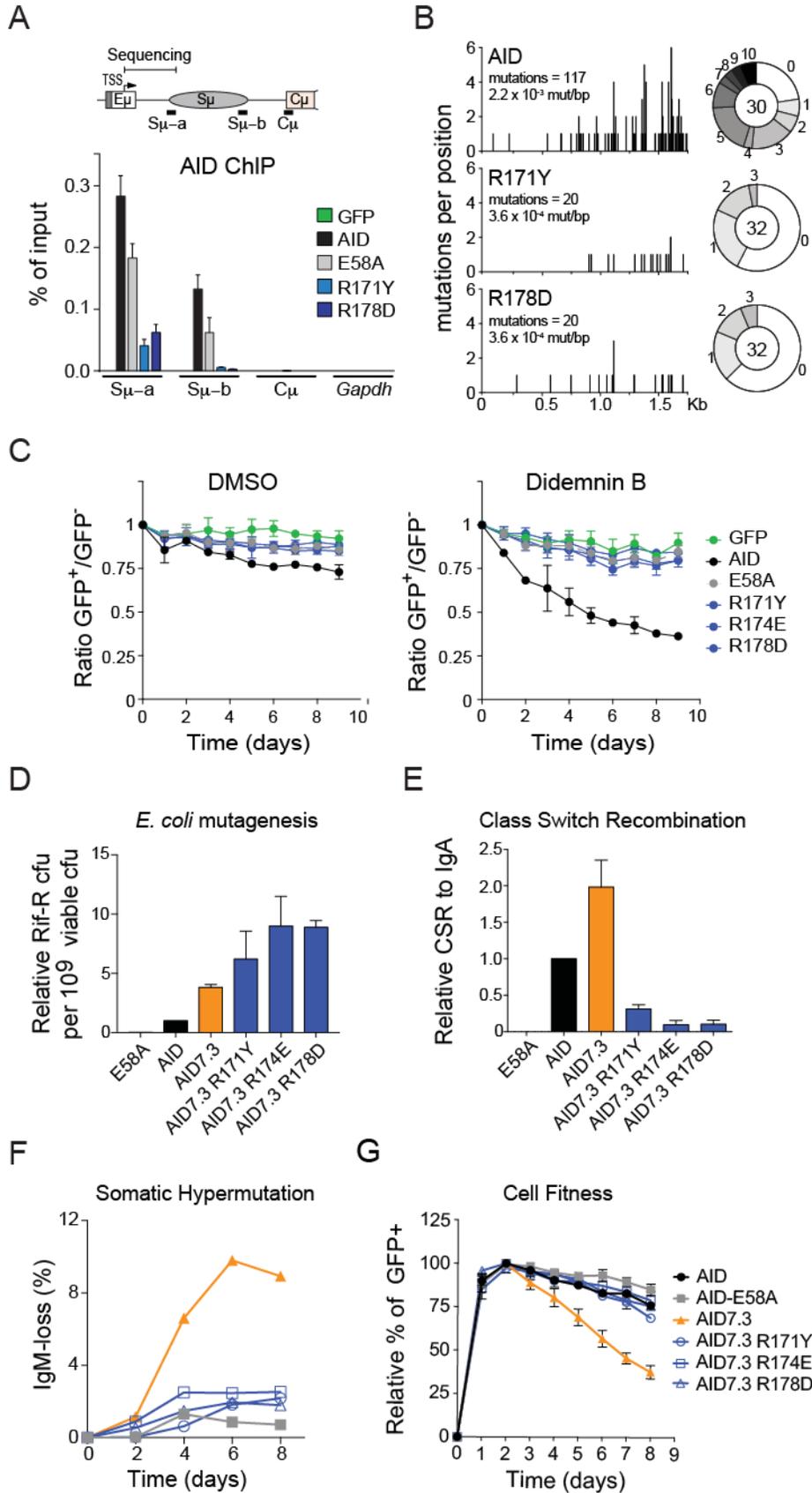


Figure 3.3. The RR domain is necessary for AID association to Igh and for genome-wide DNA damage.

(a) (Top) Schematic of the Igh S_μregion, highlighting ChIP amplicons. (Bottom) Real-time PCR ChIP analysis of AID occupancy at the Igh or control Gapdh in mouse Aicda^{-/-} B cells transduced with AID variants-ires-GFP or vector (GFP). Mean + SEM from 2 independent experiments. (b) (Left) Mutation profiles over the S_μregion (shown in a), from all analyzed sequences. The number and frequency (mut/bp) of mutations identified are indicated. (Middle) Pie charts of mutation load per sequence; slices represent proportion of sequences with indicated number of mutations. Number of sequences analyzed is indicated in the center. (Right) Proportion of mutations at C:G within WRC (W = A/T, R = A/G) motifs or not, or at A:T. (c) AID-deficient CH12 B cells complemented with AID variants-ires-GFP or empty vector (GFP), plated at a 1:1 ratio of GFP⁺ to GFP⁻, treated with either DMSO or 1nM DidB and the proportion of GFP⁺ cells monitored over time. Means ± SDs of 3 independent experiments, relative to day 0. (d) Mutagenic activity of AID7.3 variants in E. coli, measured by the relative frequency of Rif resistance. Means + SEMs of medians from 2 independent experiments (5 cultures/experiment) are shown, normalized to AID. (e) CSR activity of AID or AID7.3 variants-ires-GFP in AID-deficient CH12 B cells, as the proportion of IgA⁺ cells in GFP⁺ population (minus background) 72 h post-activation. Means + SDs of 3 independent experiments, normalized to AID. (f) SHM activity, assayed by IgM-loss over time in DT40 Aicda^{-/-} ΔΨVλ B cells complemented with AID variants-ires-GFP, 1 of 2 independent experiments is shown. (g) Effect of AID or AID7.3 variants-ires-GFP on the growth of transduced AID-deficient CH12 B cells. The proportion of GFP⁺ cells monitored over time, plotted as means ± SDs from 3 independent experiments, each normalized to max value.

fitness defect, just like AID-deficient cells or those expressing catalytically inactive AID E58A (Fig. 3.3c). To confirm these results, we introduced the R-mutants into the hyperactive AID7.3 variant, which bears three point mutations (outside α6) that increase enzymatic activity 3-fold, leading to proportionally higher SHM, CSR, chromosomal translocations and DNA damage in B cells (Wang et al., 2009b; Zahn et al., 2014). The AID7.3 R-mutants maintained hyperactivity in E. coli (Fig. 3.3d), yet they were still severely deficient for CSR, SHM and cytotoxicity in CH12 and DT40 B cells (Fig. 3.3f-h). These results show that the R-mutants enter the nucleus but have globally reduced

mutagenic activity in B-cells. The ChIP results at the *Igh* suggested this might reflect a defect in chromatin association.

3.3.3. The RR domain is dispensable for AID association to nuclear complexes

To determine whether the chromatin association of the R-mutants was different from wt AID, we combined a nuclear wash technique (Sawasdichai et al., 2010) with confocal microscopy. The nuclear fraction of AID is difficult to visualize in whole cells because of the signal coming from cytoplasmic AID. We first validated this in situ fractionation method on endogenous AID using CH12 cells, where eliminating the cytoplasmic signal allowed specific detection of AID in the nucleus (Fig. 3.4a). We then compared the R-mutants to wt AID in transduced AID-deficient CH12 cells (Fig. 3.4b). Washing away cytoplasm and nucleoplasm (as shown by the loss of the cytoplasmic AID and cell wide GFP signals), revealed a chromatin-associated pool of wt AID. AID E58A and the R-mutants showed the same distribution as wt AID (Fig. 3.4b).

To confirm the association of the R-mutants with chromatin, we used a biochemical fractionation protocol that uses an incomplete DNA digestion by Micrococcal nuclease (MNase) followed by sequential extractions with increasing salt concentrations (Henikoff et al., 2009) (Fig. 3.4c). As expected, the majority of AID was cytoplasmic, and the lack of cytoplasmic contamination in the isolated nuclei was confirmed by the *Gapdh* partition (Fig 3.4d). Nuclear fractionation showed some Spt5 but no RNAPII in the MNase fraction. All RNAPII and most SPT5 were found in the low (150 mM) and high (600 mM) salt fractions, representing loosely and tightly held

transcription complexes, respectively (Henikoff et al., 2009) (Fig 3.4d). The latter fraction also contained most of the chromatin, judging from nucleosome content.

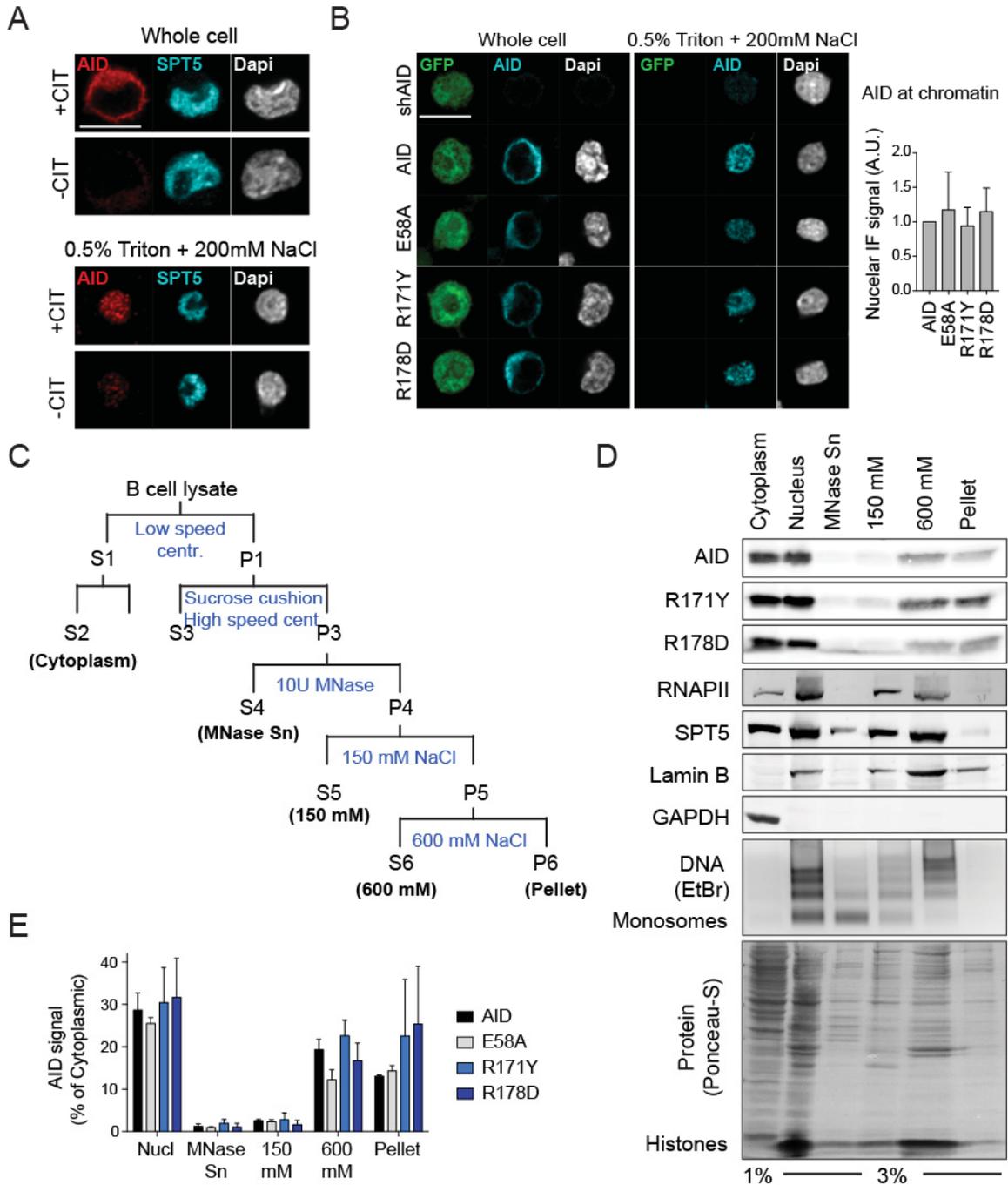


Figure 3.4. The RR domain is dispensable for AID association with tightly held nuclear complexes.

(a) Representative confocal microscopy images of CH12 B cells either fixed directly (Whole cell), or after nuclear washing (0.5% Triton + 200mM NaCl). Isolated nuclei were analyzed by IF to detect endogenous AID and SPT5 and DNA stained by Dapi. Cells were stimulated (+CIT) to induce AID expression or not (-CIT). (b) (Left) Representative confocal microscopy images of GFP, AID (detected by IF), and DNA (Dapi) in reconstituted AID-deficient CH12 B cells, imaged as in (a). AID and GFP expression were linked via IRES. (Right) Nuclear AID signal was calculated relative to whole cell intensity for each variant and normalized to the wt AID value for each of 3 independent experiments with ≥ 20 cells/cond/exp. Means +SD are plotted. (a, b) Magnification 630X. Scale bar, 10 μ m. Laser power and/or gain were increased for imaging after nuclear wash (see methods). (c) Scheme for biochemical fractionation of B cells. Fractions analysed are in bold. (d) Representative WBs on indicated fractions from reconstituted AID-deficient CH12 B cells. Antibodies against AID, RNAPII, SPT5, Lamin B, and GAPDH were used as indicated. Representative agarose gel, with purified DNA stained using ethidium bromide and Ponceau-S staining of total protein are also shown. (e) Quantification of AID signal in each lane from (d), normalized to the respective cytoplasmic AID. Means + SDs from 3 independent experiments.

Notably, ~60% of the nuclear AID was found in the 600 mM NaCl extract (Fig. 3.4d,e), indicating tight chromatin association, with little AID in the MNase or 150 mM NaCl fractions. The rest of AID was present in a remaining pellet that was largely devoid of RNAPII, Spt5 or chromatin but contained Lamin B, thus defining a second pool of nuclear AID that is not directly associated with transcription factors but is still part of non-soluble nuclear complexes. Importantly, the R-mutants had a similar distribution profile and proportions as wt AID in all fractions (Fig. 3.4d,e), indicating that they are normally associated with the chromatin in steady state. We conclude that at least two distinct pools of AID associate to the chromatin, independently of catalytic activity and biological function. The R-mutants dissociate the chromatin interaction from the biological activity of AID, suggesting that these are mechanistically distinct steps.

3.3.4. Chromatin-associated AID is tethered by Spt5, RNAPII and RNA.

We investigated the association of AID and the R-mutants to chromatin further. Spt5 is important for AID activity in B cells and correlates with AID occupancy genome wide (Pavri et al., 2010), but it is not known whether it is necessary or sufficient for chromatin interaction of AID. Spt5 knockdown in CH12 cells not only reduced CSR but also overall chromatin association of both endogenous and transduced AID (Supplementary Figs. 3.4a,b and 3.5a). Interestingly, the R-mutants were similarly dependent on Spt5 for chromatin association (Fig. 3.5a). We directly assessed whether Spt5 was sufficient to recruit AID to chromatin by using U2OS cells that contain a genomic Lac operon array (LacO), which is recognized by the Lac repressor (LacR) (Shanbhag et al., 2010) (Fig. 3.5b). In this system, a mCherry-LacR-Spt5 fusion recruited AID-GFP, as well as the R-mutants, to the LacO locus (Fig. 3.5b). APOBEC1-GFP acted as a negative control.

The association of AID to RNAPII depends on Spt5 (Pavri et al., 2010), but whether transcription or RNAPII itself are involved in retaining AID at the chromatin is unknown. To test this we treated CH12 cells with Actinomycin D (Act D) a DNA intercalator that disrupts transcription elongation (Bensaude, 2011). The eukaryotic RNA polymerases show unequal sensitivity to Act D, RNAPI > RNAPII > RNAPIII (Perry and Kelley, 1970). Act D at doses that inhibit RNAPI (0.04 μ M) or RNAPI and II (0.4 μ M), slightly reduced chromatin bound RNAPII and caused some AID redistribution to distinct nuclear sites but did not significantly change the amount of AID associated to the chromatin (Fig. 3.5c). Interestingly, at 4 μ M Act D, both RNAPII and AID were

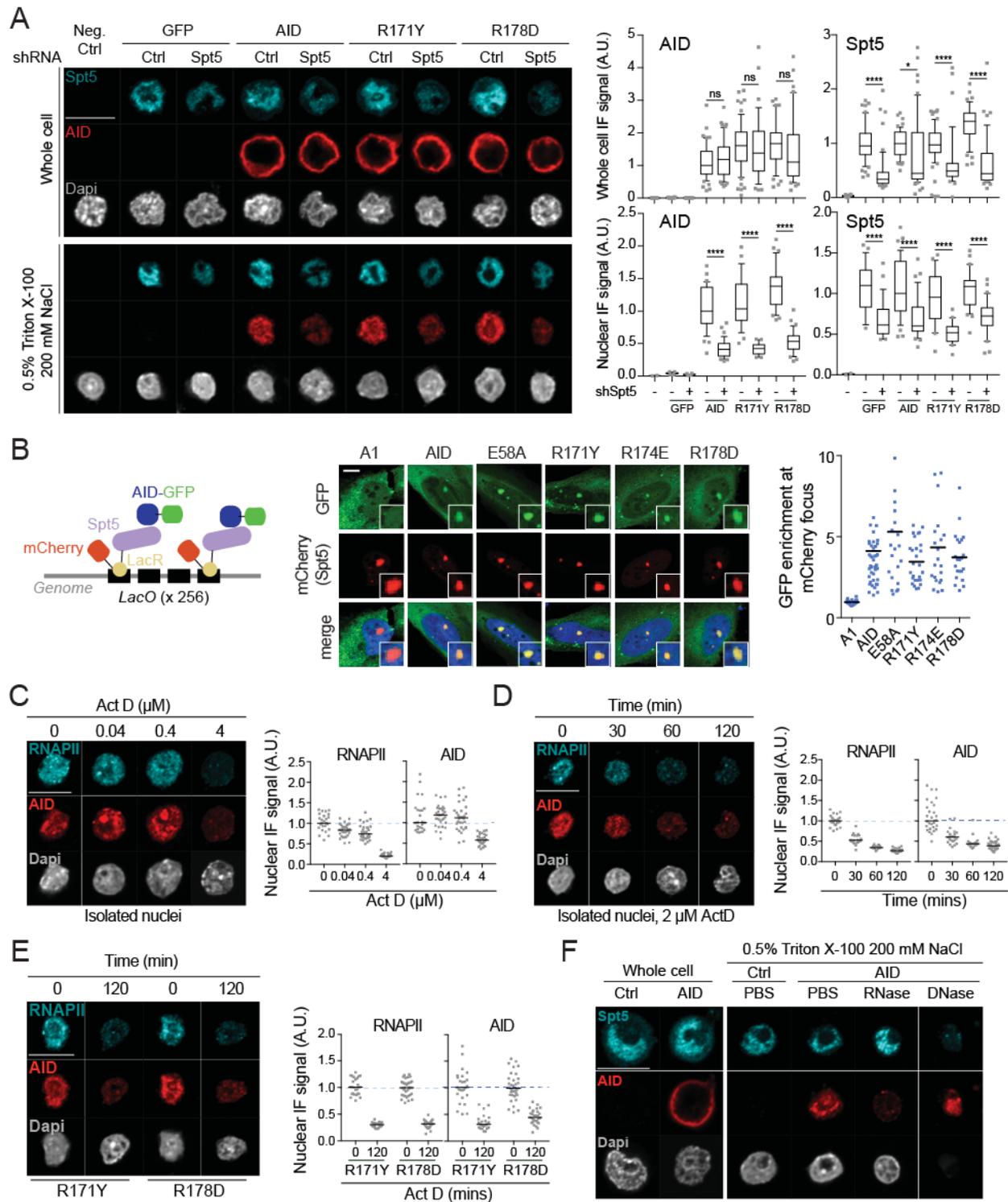


Figure 3.5. Requirements for chromatin association of AID.

(a,c-f) Representative confocal microscopy images of GFP, immunofluorescence for AID, Spt5 and RNAPII, with DNA staining (Dapi), in reconstituted AID-deficient CH12 B cells. Nuclear wash was carried

out as in Fig 4 unless otherwise stated. Magnification 630X. Scale bar, 10 μm . (a) (Left) CH12 cells were transduced with either a shRNA control or against Spt5 as indicated. (Right) Box plots represent the median, 25-75 percentiles (boxes), 10-90 percentiles (whiskers) of AID and Spt5 signal from individual whole cells or isolated nuclei, normalized to the median of wt AID. Representative of 2 independent experiments, (≥ 20 cells per condition). Differences were evaluated by unpaired, two tailed t-test (* < 0.05 , **** < 0.0001). (b) (Left) Schematic of the lacO/LacR system. (Middle) Representative confocal microscopy images of GFP or mCherry and an overlay including DNA (Dapi). Small boxes highlight typical LacO/LacR-mCherry foci. (Right) Quantification of GFP signal at each mCherry focus, normalized to non-focus nuclear GFP signal (dots). The mean for each construct (bars) is shown. Compilation of 2 independent experiments. Magnification 400X. Scale bar, 10 μm . (c) As in (a), but cells were treated for 60 min with DMSO (0) or the indicated dose of Actinomycin D (Act D) prior to plating and washing. (d) As in (a), but cells were treated with DMSO ($t=0$) or for various times with 2 μM Act D prior to plating and washing. (e) As in (a) for cells expressing AID R-mutants treated with either DMSO or 2 μM Act D for 120 min. (c-e) Average AID and RNAPII signal for individual nuclei (dots), with bars indicating population median, normalized to the median of untreated cells (dotted line). (f) As in (a) but nuclei were incubated at 37°C with PBS control, RNase or DNase during nuclear wash. Images are representative of at least 20 cells per condition and 2 independent experiments.

depleted from the chromatin (Fig. 3.5c). A time course at 2 μM Act D showed that RNAPII and AID were concomitantly depleted from the chromatin (Fig. 3.5d), suggesting a cause effect. Again, chromatin association of the R-mutants was similarly sensitive to Act D (Fig. 3.5e)

Surprisingly, AID was partially resistant to chromatin disruption by DNase treatment, which depletes nuclear Spt5 but leaves the nuclear envelope (evidenced by Lamin B staining) intact (Fig. 3.5f and Supplementary Fig 3.4c). In contrast, nuclear AID was largely depleted after RNase treatment, despite chromatin-associated Spt5 being resistant (Fig. 3.5f).

We conclude that the broad association of AID to chromatin requires RNA and the transcription machinery, but not transcriptional activity, with Spt5 being necessary and sufficient to tether AID to chromatin. The fact that chromatin-associated AID exists

in at least two distinct fractions, only one of which contains Spt5 and RNAPII (Fig. 3.4d), implies that AID is dynamically associated to these two fractions. The RR domain is dispensable for this dynamic interaction, yet necessary for function; suggesting it mediates productive targeting of AID.

3.3.5. R-mutants can be rescued by fusion with catalytically inactive AID

To obtain mechanistic insight into the defect of the AID R-mutants, we asked whether a wt AID domain could rescue their function by fusing AID E58A (with a wt RR domain) to either wt AID or the R-mutants. The control AID-AID E58A fusion protein had reduced mutagenic activity in *E coli*, compared to the AID monomer, but still produced substantial CSR (Fig. 3.6a). The analogous R-mutant fusions had the same activity as AID-AID E58A, not only in *E coli* but also for CSR (Fig. 3.6a). In this experiment, catalytic activity is derived from the R-mutant AID, while RR domain function is from AID E58A. This demonstrates that the two functions are modular, and confirms that the R-mutants are intrinsically capable of deaminating the Ig locus.

3.3.6. AID R-mutants lose specific interactions in vivo

Our results suggest that the RR domain may mediate an interaction important for targeting AID activity. To identify proteins that associate with AID in a manner that depends on the intact RR domain, we compared the interactome of AID and the R-mutants in live B cells. We used BiOLD, a proximity-based biotin labelling technique in which the bait is fused to a promiscuous BirA* biotin ligase that can label the protein environment of the bait in an ~10 nm radius (Roux et al., 2012). We generated an AID-

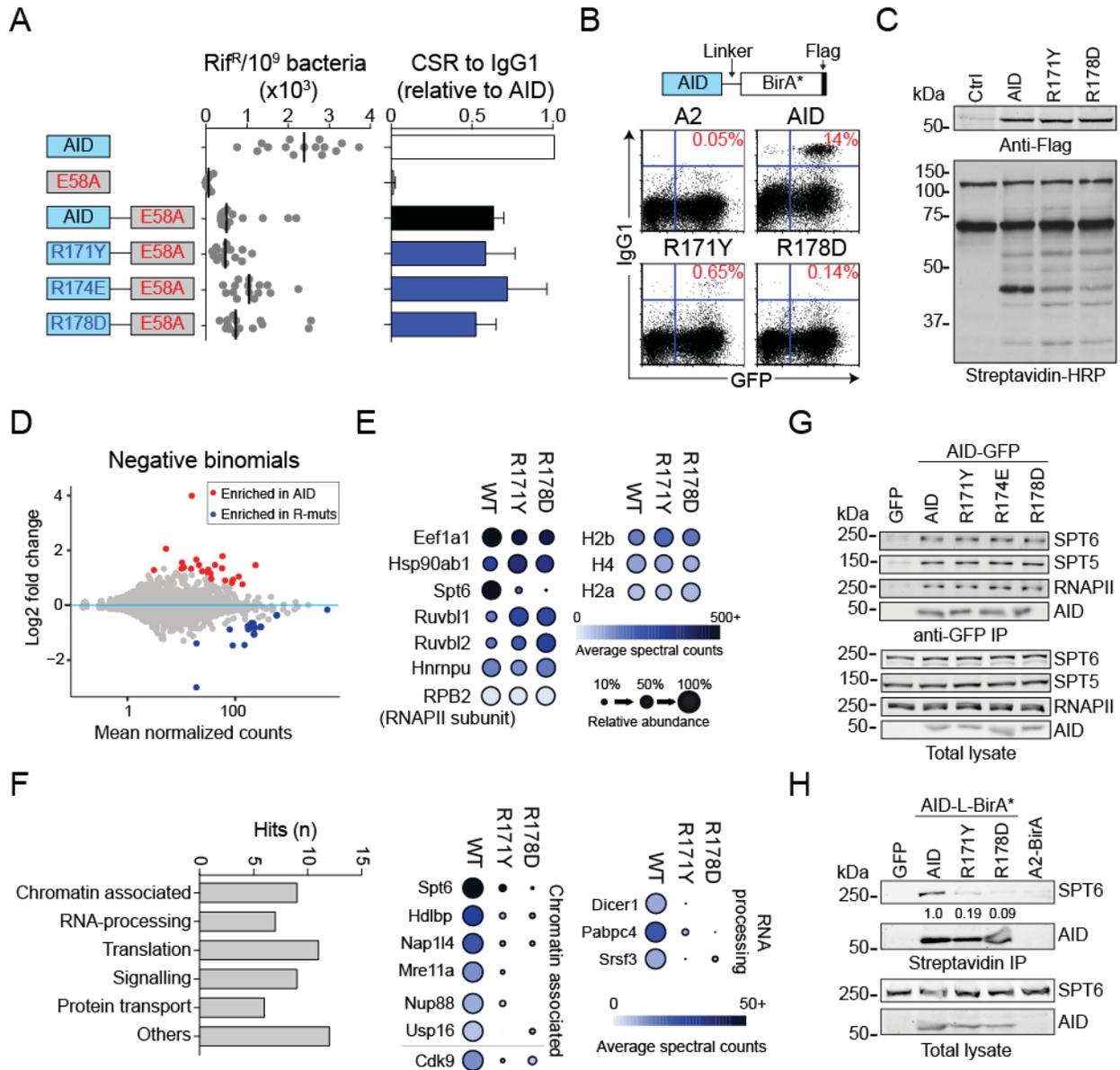


Figure 3.6. Protein-protein interactions of AID and R-mutants.

(a) (Left) AID fusions to AID-E58A and their relative *rpoB* mutation frequency in *E. coli*. Medians (bars) of 10 clones (dots) from 2 independent experiments are shown. (Right) Relative CSR capacity of the same proteins expressed in *Aicda*^{-/-} mouse B cells. Means + SDs of 6 mice, from 3 independent experiments are shown, normalized to AID. (b) Schematic of the AID fusions to BirA* and representative flow cytometry plots for CSR, the proportion of IgG1⁺ cells in the GFP⁺ population, in reconstituted *Aicda*^{-/-} mouse B cells. *Apobec2* (A2) was used as control. (c) Representative WB of the AID-BirA* fusions (anti-flag) and biotinylated targets (streptavidin), from reconstituted *Aicda*^{-/-} mouse B cells after incubation with biotin. Untransduced cells were used as control (Ctrl). (d) MA plot (fold change in sc as a function of

average sc of all hits) for wt AID versus R-mutants. Red and blue dots represent significant differences identified by DESeq2 software. (e) (Left) Circle plots of BioID signal for selected known AID interacting partners. Circle size indicates relative abundance normalized to the variant with the most sc. Circle color indicates actual sc (scales included). (f) (Left) Broad functional categories of BioID associations found reduced in the R-mutants compared to wt AID. (Right) Circle plots for hits in the indicated categories. (g) Whole cell lysates from AID-deficient CH12 B cells reconstituted with AID variant-GFP fusions subjected to anti-GFP pull down. Spt6, Spt5, RNAPII and AID were detected by WB from either immunoprecipitates or lysates. Representative of 2 independent experiments. (h) Streptavidin pull down of biotinylated proteins from whole cell lysates of AID-deficient CH12 B cells reconstituted with AID variant-BirA* fusions. Spt6 and AID were detected by WB from either immunoprecipitates or lysates. Ratios of Spt6 to AID signal normalized to AID-BirA* are indicated. Representative of 2 independent experiments.

BirA* fusion that was active for CSR in transduced *Aicda*^{-/-} B cells, and its R171Y or R178D derivatives (Fig. 3.6b). Addition of biotin to the cultures led to biotinylation of proteins in cells expressing BirA* fusions (Fig. 3.6c), proportionally to AID-BirA* expression levels (Supplementary Fig. 3.5a,b). The vast majority of AID interactions were conserved in the R-mutants, including many validated AID interactors (Fig. 3.6d,e and Supplementary Fig 3.5f), further confirming the structural integrity of the mutants. Using four different statistical methods we identified 54 proteins that interacted with AID but not with either one of the R mutants, 29 of these were significant by at least two methods (Fig. 3.6d and Supplementary Fig 3.5c-e, and Supplementary Table 3.1). Functional annotation of the BioID hits showed 9 chromatin associated factors, including the histone chaperones SPT6 and NAP1L4 as well as the P-TEFb component CDK9, which are functionally linked to transcription elongation (Jonkers and Lis, 2015) (Fig. 3.6f). We also found 7 factors linked to co-transcriptional RNA-processing; splicing and mRNA transport factors, as well as DICER which can bind to dsRNA resulting from convergent transcription induced by R-loops (Skourti-Stathaki et al., 2014). Of these, only Spt6 had been previously shown to interact with AID by co-immunoprecipitation.

Interestingly, standard pull down from cell extracts showed that GFP-tagged R-mutants still co-immunoprecipitated Spt6 (Fig. 3.6g). However, streptavidin pull down of proteins biotinylated in live CH12 B cells expressing AID- or R-mutants-BirA^{*}, confirmed the lack of interaction with the R-mutants (Fig. 3.6h), stressing the power of BioID to detect functional defects in live cells.

Thus, the RR domain mediates interactions necessary to reach factors that act during active transcription or post transcriptionally, such as Spt6, in live cells. We conclude that the defect in the R-mutants lies after chromatin association but prior to the transcriptional step in which Spt6 is recruited.

3.3.7. AID R-mutants fail to progress with elongating RNAPII

The R-mutants interacted with RNAPII (Fig. 3.6e,h), yet did not mutate B cells. As Spt5 is recruited to promoter-proximal paused RNAPII, and Spt6 only to active transcription (Andrulis et al., 2000), we hypothesized that the R-mutants might fail to progress from paused to elongating RNAPII. We therefore compared the occupancy of wt AID and the R-mutants near the transcription start site (TSS, -51 to 222 bp), where paused RNAPII is expected (Jonkers and Lis, 2015), to occupancy at the downstream S_μ-region. ChIP showed that wt AID was present in all amplicons but highest at the S_μ region (Fig. 3.7a), where elongating RNAPII was previously shown to stall (Rajagopal et al., 2009). This profile was not an artifact of AID overexpression, as endogenous AID had a similar distribution (Fig. 3.7b). Spt6 was also present in all amplicons (Fig 3.7b), as expected for a highly transcribed gene (Andrulis et al., 2000). In contrast, the R-mutants were enriched at the TSS, but depleted from the S_μ (Fig. 3.7a). Thus, the lack of association between the R-mutants and Spt6 most likely reflects the uncoupling of

the AID R-mutants from elongating and/or stalled RNAPII, which suggests the existence of a novel licensing step for productive targeting of AID.

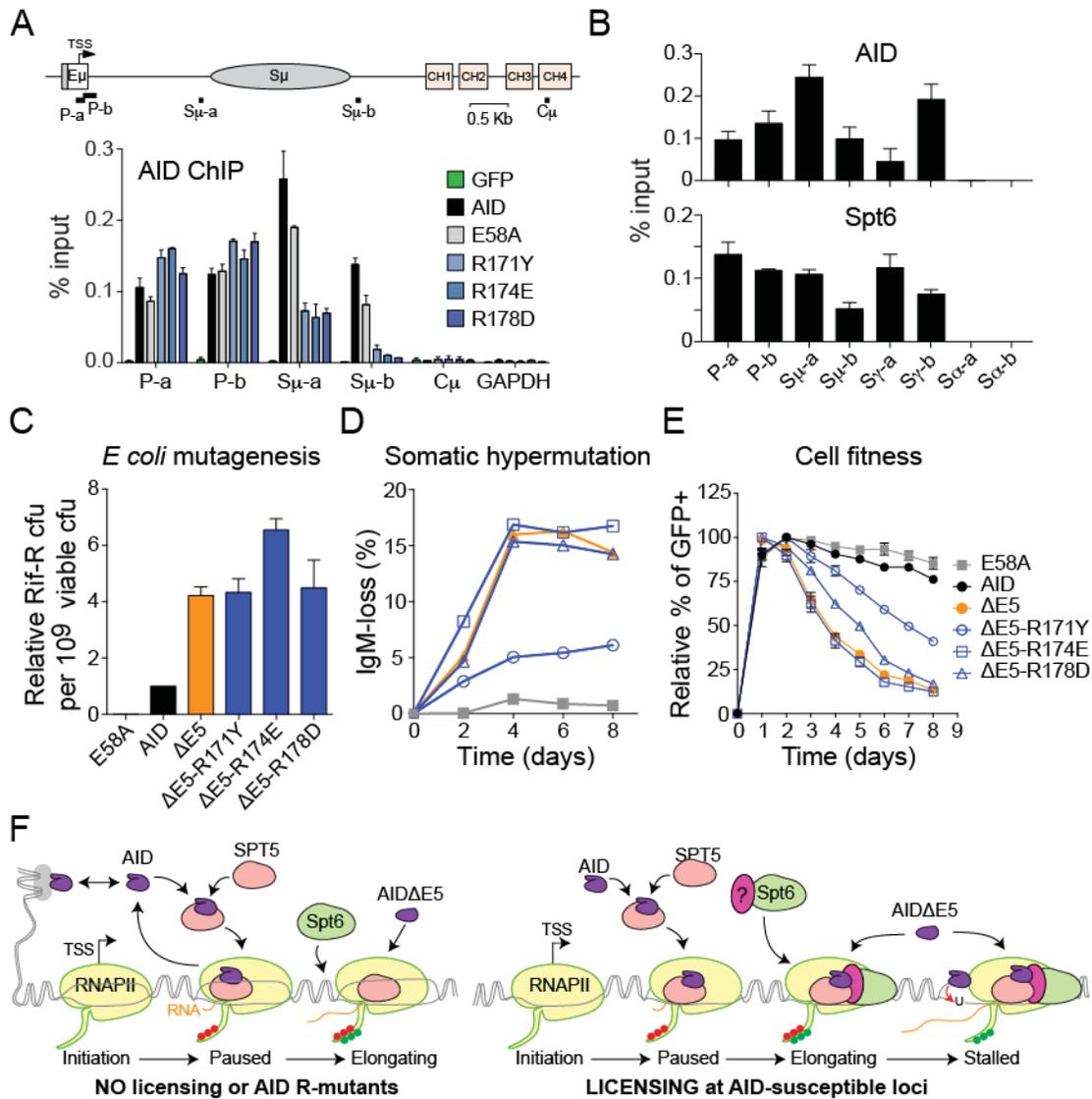


Figure 3.7. AID R-mutants are uncoupled from transcription elongation

(a) (Top) Scale schematic of the *Igh* S μ region, with ChIP qPCR amplicons indicated. (Bottom) ChIP qPCR for AID from reconstituted *Aicda*^{-/-} mouse B cells activated with LPS and IL-4. Means + SEMs from 2 independent experiments. (b) ChIP qPCR for endogenous AID and Spt6 from wt mouse B cells treated and using the same amplicons as in (a), as well as amplicons in the Sy1 and Sa regions. (c) Relative mutagenic activity of AID Δ E5 variants in *E. coli*, measured as the frequency of Rif resistant colonies.

Means + SEMs of medians from 2 independent experiments (5 cultures/experiment), normalized to AID. (d) Somatic Hypermutation capacity of AID Δ E5 variants-ires-GFP measured by IgM-loss over time in complemented DT40 *Aicda*^{-/-} $\Delta\Psi\forall\lambda$ B cells. One of 2 independent experiments is plotted. (e) Effect of AID or AID Δ E5 variants-ires-GFP on the growth of transduced AID-deficient CH12 B cells. Means \pm SEMs of the proportion of GFP+ cells monitored over time from 2 independent experiments is plotted, each normalized to maximal value. (f) Model for AID targeting (see discussion).

3.3.8. Constitutively nuclear AID bypasses promoter licensing

The RR domain was dispensable for mutagenesis in *E coli*, suggesting the licensing step can be bypassed. We therefore asked whether a constitutively nuclear AID might rescue the defect of the R-mutants by mimicking the *E coli* situation, where proximity to the genome is unregulated. We introduced the R-mutants into AID Δ E5, a nuclear AID variant (Patenaude et al., 2009; Zahn et al., 2014). The AID Δ E5 R-mutants were as active as AID Δ E5 in *E. coli* (Fig. 3.7c). AID R174E and R178D variants did not prevent SHM of AID Δ E5 in DT40 B cells, nor its effect on fitness in CH12 cells (Fig. 3.7d,e). AID Δ E5 R171Y produced substantial SHM and toxicity but less than AID Δ E5, suggesting an additional role of this residue for mutagenesis in B cells. We conclude that the licensing step is not necessary when nuclear AID levels are not limiting.

3.4. Discussion

The genes that are mutated by AID in B cells share a transcriptional landscape associated to super-enhancers with convergent transcription (Meng et al., 2014; Pefanis et al., 2014; Qian et al., 2014). This demonstrates that AID targeting is based on the local chromatin architecture, rather than fixed features of the target genes. The mechanism underlying the productive targeting of AID remains unclear, but cannot be explained by differential recruitment of AID, as AID occupancy is insufficient for mutation (Casellas et al., 2016; Hakim et al., 2012; Matthews et al., 2014a; Yamane et al., 2011). Here, we uncover an obligatory licensing step for AID that provides insight into this fundamental question.

We show that specific substitutions of either one of 3 arginine residues within AID helix $\alpha 6$ dissociates its ability to associate with the chromatin from its ability to induce mutations in B cells. This led to our characterization of the association of AID and the R-mutants to chromatin, where we show that Spt5 is necessary and sufficient. The physical presence of RNAPII may also be required, but transcription itself is dispensable. Nonetheless, a large proportion of chromatin-associated AID is in a fraction devoid of RNAPII or Spt5, as shown by salt fractionation. Furthermore, RNase treatment eliminates AID (but not Spt5) from the chromatin. Thus, our results suggest that AID association to chromatin depends not only on the transcription machinery but also on some distinct, RNA-dependent complex. Of note, AID and the R-mutants interact with hnRNP-U (Fig. 3.6f), an RNA binding protein that forms a chromatin-scaffold with newly transcribed RNA (Nozawa et al., 2017) and plays a role in CSR (Mondal et al., 2016). Our interpretation of these results is that the broad distribution of

AID at the chromatin is dynamic, allowing AID to sample multiple loci. Similar behavior has been observed as transcription factors search for their target DNA sequences (Chen et al., 2014); yet unlike transcription factors, AID does not recognize a specific DNA sequence (Peled et al., 2008) and therefore requires the special level of regulation provided by the licensing step to recognize its targets. It is likely that this mechanism functions within super-enhancers resulting in licensing at the off-targets in addition to the Ig genes.

The current model for AID targeting posits that AID mutates when RNAPII stalls, but this does not explain how AID associates with stalled polymerase, beyond invoking a role for Spt5 (Maul et al., 2014). As the R-mutants associate with Spt5, and are recruited to the IgH locus TSS, they uncover an additional level of regulation. In fact, our results imply that in physiological conditions AID is recruited to the TSS and must progress with transcription elongation until the polymerase becomes stalled, rather than being directly recruited to stalled polymerases.

The R-mutants abrogate SHM and DNA damage in the context of the full length AID7.3 but not of AID Δ E5, despite similar catalytic activities (Zahn et al., 2014). Though wt AID, and AID7.3 are excluded from the nucleus, AID Δ E5 is fully nuclear (Patenaude et al., 2009; Wang et al., 2009b), thus, high nuclear AID levels can bypass the licensing step, possibly by direct recruitment to stalled polymerases, resulting in a highly mutagenic enzyme. This would provide a rationale for the abundance of mechanisms that restrict AID nuclear access (Methot and Di Noia, 2017), which would serve to enforce licensing.

Recently, the RR domain residues have been implicated in CSR because they are necessary to deaminate the kind of structured, G-quadruplex, DNA formed by the S-regions (Qiao et al.). We did observe that AID R171Y reduced off-target activity of the nuclear AID Δ E5 variant, and seemed to also reduce the preference for deaminating WRCY hotspots at the S μ , which might be consistent with a defect for deaminating structured DNA. However, SHM does not require structured DNA, yet AID Δ E5 R171Y also had reduced SHM activity at the DT40 IgV, indicating some additional defect in this mutant. On the other hand, R174E and R178D did not affect the ability of AID Δ E5 to do SHM or off-target DNA damage in B cells, demonstrating that the AID RR domain is not intrinsically necessary for deaminating the genome. Thus, our results do not dispute the role of the RR domain for deaminating structured DNA and for CSR (Qiao et al.), but provide good evidence of an additional role by tethering AID to transcription elongation. It is tempting to speculate that the nuclear APOBEC3B and other AID paralogues with antiviral activity (Lackey et al., 2012), have lost the RR domain and licensing in order to prevent mutating the self-genome during transcription.

In conclusion, our data is consistent with a model (Fig. 3.7f) in which the broad but dynamic association to chromatin would allow AID to sample multiple loci. The low concentration of AID in the nucleus would favor its recruitment to many promoter-proximally paused RNAPII, probably via Spt5, consistent with previous findings (Pavri et al., 2010; Yamane et al., 2011). However, only those loci within a permissive transcriptional landscape would contain the factors that mediate the coupling of AID to transcription elongation, licensing AID to deaminate the downstream region. These licensing factors might be present in certain super-enhancers, likely more frequently or

more active at the *Ig* loci, compared to AID off-targets, but absent from those loci that are not mutated by AID.

3.5. Materials and methods

3.5.1. Animals

Aicda^{-/-} mice (a gift from Dr. T. Honjo, Kyoto University, Japan) and *Aicda*^{-/-} *Ung*^{-/-} mice (*Ung*^{-/-} mice were a gift of Dr H. Krokan, Norwegian University of Science and Technology, Norway) in C57BL6/J background were bred at the specific pathogens-free facility of IRCM. All animal work was approved by the IRCM animal protection committee in accordance to the guidelines of the Canadian Council for Animal Care.

3.5.2. DNA constructs

Retroviral vector pMXs human AID, AIDΔE5-ires-GFP, AID7.3-ires-GFP and mouse AID-GFP have been described (Methot et al., 2015; Patenaude et al., 2009; Zahn et al., 2014). Human AID fusions were assembled as BamHI-AID-EcoRI-Linker-HindIII-AID-E58A-XhoI cassettes into pTrcHisA (ThermoFisher) or pMX-ires-GFP (Cell Biolabs). The linker encoded a flexible (SGGGG)₃ peptide. Human SPT5 was PCR amplified and cloned into a gateway compatible mCherry-LacRep destination vector (Orthwein et al., 2015) (a gift from Dr. D. Durocher, University of Toronto, Canada). Mouse AID and Linker-BirA* were PCR amplified using gateway compatible primers and cloned into appropriate donor vectors to generate AID-BirA* fusions into a homemade gateway compatible pMX-ires-GFP (gateway cassette cloned BamHI-EcoRI) using Multisite gateway (Invitrogen). AID variants were generated by PCR amplification with *ad hoc* oligonucleotides or by quick-change site-directed mutagenesis using KOD1 DNA polymerase (Toyobo Inc.). Oligonucleotide sequences are available on request.

3.5.3. Cell culture and transduction

CH12 cells, Plat-E and primary B lymphocytes were cultured in RPMI 1640 media (Wisent) at 37 °C with 5% (vol/vol) CO₂. U2OS cells were cultured in McCoy's 5a media (Wisent). Media were supplemented with 10% FBS (Wisent), 1% penicillin/streptomycin (Wisent), and 0.1 mM 2-mercaptoethanol (Bioshop). DT40 cells were culture in RPMI 1640 supplemented as above plus 1% chicken serum (Wisent). CH12 cells stably expressing shRNA against AID have been described (Cortizas et al., 2013). Naïve splenic B cells from *Aicda*^{-/-} or *Aicda*^{-/-} *Ung*^{-/-} mice were isolated from total splenocytes depleted with anti-CD43 microbeads in an autoMACS cell separator (Miltenyi). For retroviral complementation of DT40 or CH12 cells, VSV-G, MLV gag-pol, and pMXs vectors (1:1:4 ratio, 2.5 µg DNA total) were transfected into HEK293 cells using Trans-IT LT-1 (Mirus Bio). Retrovirus for primary cell transduction was produced using Plat-E ecotropic packaging cells (Morita et al., 2000) transfected with pMXs vectors. For infections, 1 ml of HEK293 or 1.5 mL Plat-E supernatant at 48 h post-transfection was used to infect 10⁶ B cells in 24-well plates, in the presence of 8 µg/mL polybrene, by spinning at 600 x g for 90 min at 32°C. Medium was replaced 4 h later.

3.5.4. Reagents and antibodies

Stock aliquots: 50 µg/mL Leptomycin B in ethanol (LC Laboratories), 20 mM Didemnin B in DMSO (NSC 325319; provided by the Natural Products Branch, National Cancer Institute, Bethesda, MD) and 2 mM Actinomycin D in DMSO (Santa Cruz Biotechnology). Drugs were kept at -20°C in the dark and diluted fresh before each experiment. Antibodies used for IF: for human AID; rat mAb anti-AID (1:500, EK2 5G9, Cell Signaling), for mouse AID; rat mAb anti-AID (1:250, mAID-2, eBioscience), rabbit mAb anti-SPT5 (1:500, EPR5145(2), Abcam), rabbit anti-RNAPII (1:100, H-224, SC

Biotechnology), goat anti-LaminB (1:500, M-20, SC Biotechnology). With the exception of anti-rat Dylight 550 (1:500, SA5-10027, ThermoFisher Scientific) all other IF secondary Abs were AlexaFluor conjugates (1:500, invitrogen): anti-rat-680, anti-rabbit-546, anti-rabbit-680 and anti-goat-680. Antibodies use for Western blots: rat mAb anti-AID (for human AID; 1:1,000, EK2 5G9), mouse mAb anti-AID (for human AID; 1:5,000 with a 1:1 mixture of clones 52–1 and 39–1 specific for human AID N terminus, a gift from Dr M. Neuberger), rat mAb anti-AID (for mouse AID; 1:500, mAID-2), rabbit anti-Actin (1:3,000, A2066, Sigma), rabbit anti-GFP (1:2,000, 11122, Invitrogen), rabbit anti-GFP-HRP (1:5,000, 130-091-833, Miltenyi Biotec), mouse mAb anti-GAPDH (1:3,000, H-12, SC Biotechnology), mouse mAb anti-HSP90 (1:3,000, 68/Hsp90, BD), goat anti-LaminB (1:2,000, M-20), rabbit anti-SPT5 (1:500, H-300, SC Biotechnology), rabbit anti-RNAPII (1:500, H-224), rabbit anti-SPT6 (1:2,000, A300-801A, Bethyl Laboratories), Streptavidin-HRP (1:10,000, N100, Thermo Scientific). Secondary antibodies were anti-mouse-, anti-rat-, anti-goat- or anti-rabbit- AlexaFluor680 (1:10,000; Invitrogen), read using an Odyssey CLx apparatus (Li-COR), or the ChemiDoc XRS+ Imaging system (Biorad) for HRP conjugates developed by chemiluminescence (34080, Thermo Scientific).

3.5.5. Monitoring SHM and CSR

SHM was measured by fluctuation analysis of IgM-loss in DT40 *Aicda*^{-/-} $\Delta\Psi\forall\lambda$ cells (Arakawa et al., 2004) complemented by retroviral transduction with AID or mutants thereof, as described (Zahn et al., 2014). For AID7.3 and AID Δ E5, IgM-loss was measured in bulk ever 2-3 days to determine SHM kinetics. CSR to IgG1 in complemented mouse *Aicda*^{-/-} B cells was induced by adding 5 μ g/mL LPS before and

after the infection and 20 ng/mL mrlL-4 (PeproTech) 4 h post-infection. CSR efficiency in the infected (GFP⁺) subpopulations was measured by flow cytometry using biotinylated anti-IgG1 (BD) followed by anti-biotin-allophycocyanin (Miltenyi Biotech) and propidium iodide to exclude dead cells. CSR in CH12 B cells was induced using CIT [1 µg/mL rat-anti-CD40 (clone 1C10, eBioscience), 10 ng/mL IL-4 and 1 ng/mL TGF-β1 (R&D Systems)]. The proportion of IgA⁺ cells was measured 3 days later by flow cytometry using anti-IgA conjugated with R-phycoerythrin (SouthernBiotech). SHM at S_μ was analyzed in *Aicda*^{-/-} *Ung*^{-/-} mouse B cells transduced twice with pMX-AID variant-ires-GFP in order to get nearly 100% infection efficiency. Infected cells were cultured for 4 days with 10 µg/mL LPS and 25 ng/mL IL-4 before enriching for live cells using OptiPrep (Sigma) and purifying genomic DNA with DirectPCR lysis reagent (Viagen). A 1,749 bp fragment of the S_μ region was amplified using KOD1 DNA polymerase, cloned into pGEMT-easy (Promega), and individual clones sequenced at Macrogen (Seoul, Korea). The distribution and frequency of mutations were computed as described (Zahn et al., 2014).

3.5.6. Deaminase activity and DNA binding assays

E. coli mutation assays were performed using the *Δung* BW310 strain, as described (Zahn et al., 2014). AID variants were subcloned as BamHI-XhoI fragments into pTrcHisA (Invitrogen) to express His-AID fusions after Isopropyl β-D-1-thiogalactopyranoside induction. Mutation frequencies were calculated as the median number of cfu that survived rifampicin selection per 10⁹ ampicillin-resistant cells from at least 3 experiments with 5 independent cultures per construct. For biochemical assays, EcoRI fragments encoding the ORF of each AID were cloned into pGEX-5x-3 (GE

Healthcare) to generate and purify GST-AID as described (King et al., 2015). For each mutant and wt AID, 2-4 independent preparations were purified and tested. An end-labeled bubble substrate containing a 7-nt-long single-stranded region with the motif TGC, previously described to be an optimal AID substrate, was used in activity assays and EMSAs (Larijani et al., 2007). For alkaline cleavage, 0.03-4 nM substrate was incubated with 0.1 μ g AID, followed by addition of UDG, NaOH and heat, and electrophoresis on denaturing urea gels, as described (King et al., 2015; Larijani et al., 2007). For EMSA, 0.015-5 nM substrate was incubated with 0.1 μ g of GST-AID in binding buffer (50mM Tris, pH 7.5, 2 μ M MgCl₂, 50mM NaCl, and 1mM DTT) in a final volume of 10 μ l for 60 min at 37°C, followed by UV cross linking as previously described (Larijani et al., 2007). Samples were electrophoresed at 4°C on an 8% acrylamide native gel. Alkaline cleavage and EMSA gels were visualized using a PhosphorImager (Bio-Rad). Densitometry was performed using Quantity One 1-D Analysis Software (Bio-Rad). Data was graphed using GraphPad Prism to derive initial deamination velocity and K_d values.

3.5.7. AID shuttling and nuclear wash protocol

CH12 cells were treated for 2 h with DMSO, 10 ng/mL LMB (a CRM1 inhibitor), 100 nM DidB (an EEF1A inhibitor) or both drugs combined before harvesting. Cells were washed with PBS then plated on poly-L lysine coated coverslips and fixed in 3.7% (w/v) formaldehyde for 10 min, then washed 3x in PBS. The nuclear wash protocol was adapted from (Sawasdichai et al., 2010). Briefly, CH12 cells were plated on poly-L lysine coverslips and washed 1x with CSK buffer (10 mM PIPES, 300 mM sucrose, 200 mM NaCl, 3 mM MgCl₂, 1 mM EDTA and 1x fresh complete protease inhibitor (CPI,

Roche)). Cells were then either fixed directly in formaldehyde (whole cell) or washed to remove cytoplasm and loosely held nuclear proteins. Washing was done by sequentially incubating the coverslips on ice in: CSK buffer for 1 min, CSK + 0.1% triton X-100 for 1 min, CSK + 0.5 % triton X-100 for 20 min. For RNase and DNase, the last wash was 10 min on ice, then 10 min at 37°C in CSK buffer containing 10 mg/mL RNase or 10 mg/mL DNase. After washes, cells were fixed in formaldehyde. For all IF, cells were permeabilized and blocked for 1 h in blocking solution (PBS, 0.5% (v/v) Triton-X100, 1 mg/mL BSA, 5% (v/v) goat serum). For anti-LaminB IF, blocking buffer was 5% BSA to avoid cross-reactivity of anti-goat secondary. Cells were then incubated overnight at 4°C with primary antibodies in blocking solution, followed by 3x washes with PBS + 0.01% Triton X-100 (PBS-T) then a 1 hr incubation with secondary antibodies in blocking solution and 3x PBS-T washes. After nuclear staining with Dapi (300 nM in PBS), coverslips were washed with ddH₂O and mounted on slides using Lerner Aqua-Mount (Thermo Scientific).

3.5.8. LacR-LacO recruitment

U2OS cells with a 256 copy *lacO* array (Shanbhag et al., 2010) (a gift from Dr. R. Greenberg, University of Pennsylvania, USA) were plated on coverslips and co-transfected with mCherry-LacR-NLS-SPT5 along with GFP-tagged AID variants or APOBEC1 using TransIT-LT1 transfection reagent (Mirus). 30-40 h post-transfection, cells were fixed in 3.7% (w/v) formaldehyde for 10 min then washed 3x, stained with Dapi, washed and mounted as above.

3.5.9. Microscopy

Images were acquired at room temperature using ZEN 2010 on a Zeiss LSM 700 confocal microscope with excitation lasers at 405 nM (Dapi), 488 nM (GFP), 543 nM (Alexa546 and DyLight550) and 633 nM (Alexa680), using either 40x/1.3 or 63x/1.4 oil immersion objectives, and collected with a Hamamatsu PMT. Settings for nuclear wash: endogenous AID, whole cell (laser power 5, gain 550), nuclear wash (laser power 20, gain 650); overexpressed AID, whole cell (laser power 5, gain 475), nuclear wash (laser power 5, gain 625). Subcellular localization was scored in Volocity (Perkin Elmer). Masks were made for each individual cell for both nuclear and total IF signal. The proportion of nuclear signal was calculated as the ratio of nuclear signal/total signal x 100. For nuclear washes, whole cell IF signal was measured from a mask generated by GFP signal, whereas nuclear signal was measured from a mask generated by Dapi signal. For each experiment, multiple fields were analyzed, excluding cells showing saturated signal, abnormal DNA structure or mitotic figures. For making figures, images were transferred to Photoshop for adjusting contrast throughout the whole image when necessary to enhance visibility and for cropping.

3.5.10. Chromatin fractionation

Chromatin fractionation was adapted from (Henikoff et al., 2009). Briefly, $\sim 50 \times 10^6$ CH12 cells were collected and washed 1x with ice-cold PBS prior to re-suspension in 1 mL of Lysis buffer (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl₂, 0.34 M Sucrose, 10 % glycerol, 0.1 % Triton X-100, 1 mM DTT, 1x CPI). Cells were lysed for 8 min on ice then centrifuged @ 1,300 x g for 4 min. The supernatant was kept as the cytoplasm fraction. The pellet was resuspended in 200 μ L nuclear re-suspension buffer (10 mM Tris pH 8, 3 mM CaCl₂, 1 mM Mg Acetate, 0.34 M Sucrose, 0.5 % NP-40, 1

mM DTT, 1x CPI). Nuclei were then layered onto a sucrose cushion (10 mM Tris pH 8, 2 M Sucrose, 5 mM Mg Acetate, 0.1 mM EDTA, 1 mM DTT), and centrifuged at 20,000 RPM for 15 min. The nuclear pellet was resuspended in 500 μ L of nuclear resuspension buffer without NP-40 then centrifuged at 100 x g for 10 min. Nuclei were washed 1 x with nuclear wash buffer (10 mM Tris pH 7.4, 2 mM MgCl₂, 1 x CPI), then resuspended in 400 μ L of nuclear wash buffer and placed at 37°C for 5 min. CaCl₂ was added to 1 mM and DNA digested by adding Mircococcal nuclease (New England Biolabs) to 9.6 U/mL for 10 min. Digestion was stopped by adding EGTA to 2 mM final. An aliquot was saved as the total nuclear fraction. Nuclei were pelleted for 10 min at 100 x g, and the supernatant was saved as the MNase fraction. Nuclei were then re-suspended in 700 μ L of 150 mM extraction buffer (10 mM Tris pH 7.4, 140 mM NaCl, 1 mM MgCL₂, 2 mM EGTA, 0.1 % Triton X-100, 1 x CPI) and incubated for 2 h on a rocker at 4°C. After centrifugation for 10 min at 100 x g, the supernatant was saved as the 150 mM fraction. Nuclei were then re-suspended in 700 μ L of extraction buffer, with 590 mM NaCl, and incubated overnight on a rocker at 4°C. After centrifugation for 10 min at 500 x g, the supernatant was saved as the 600 mM fraction. Nuclei were finally resuspended in 10 mM Tris pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 x CPI, and kept as the pellet fraction. DNA was purified using a PCR purification kit (BioBasic) then run on an agarose gel in order to confirm efficient digestion and DNA extraction. Protein fractions were analysed by SDS-PAGE and for WBs.

3.5.11. CHIP assays

Naïve *Aicda*^{-/-} primary mouse B cells were stimulated with 12 μ g/mL LPS and 50 ng/mL IL-4 and retrovirally transduced 24 h. Cells were harvested at 18–20 h post-

infection when GFP⁺ cell proportions were 30–50%. ChIP procedures have been described in detail (Cortizas et al., 2013). Briefly, cells fixed with 1% formaldehyde for 30min, were lysed in RIPA buffer and sonicated to generate DNA fragments <500 bp. Lysate fractions (2 µg/µL) of 0.5 mg (for AID) or 1.25 mg (for SPT6) were precleared with G protein-Sepharose slurry before incubating overnight with 2–5 µg anti-AID antibody (328.8b, Active Motif) at 4°C. DNA was purified and used as template in real-time PCR reactions containing 1x SYBR Green Mix (Applied Biosystems), 1/10 fraction ChIP-enriched DNA, and 100 nM primers. Plates were read in an Applied Biosystems StepOnePlus instrument. Standard curves with different amounts of the input extracts were run in each plate for each individual amplicon and used to calculate input %. The input % of the IgG control immunoprecipitation was subtracted from each sample to calculate the values reported.

3.5.12. Co-Immunoprecipitations

For coIPs, AID-deficient CH12 cells were reconstituted with GFP tagged AID variants then lysed as in (Pavri et al., 2010). After lysate clarification, GFP immunoprecipitation was carried out using the µMACS GFP isolation kit according to manufacturer's instructions (Miltenyi Biotech). Elution and total lysate were analysed by SDS-PAGE and blotted as for other WBs.

3.5.13. Bioid and Mass spectrometry

Bioid samples were processed as described elsewhere with modifications (Couzens et al., 2013). For each construct, 18×10^6 *Aicda*^{-/-} mouse B cells were pre-cultured for 48 h with 0.5 µg/mL anti-CD180 (BD). Cells were then infected twice over consecutive days with pMX- AID variant -BirA*-Ires-GFP or pMX-A2-BirA*-Ires-GFP

retrovirus in the presence of 5 µg/mL LPS. After the second infection, media was supplemented with 5 µg/mL LPS + 25 ng/mL IL-4. The next day, 50 µM of biotin (Sigma) was added to the media. Cells were harvested 24h later (~40-50 x 10⁶ cells), washed 3x with PBS, then lysed in 1.5 mL of RIPA buffer and sonicated 30 secs at 30% amplitude (3 x 10 sec bursts with a 2 sec break in between). Benzoyl-DL-homoserine (250U, EMD Millipore) was added to the lysates during centrifuging, 30 min at 16,000 x g, 4°C. 40 µL aliquots of supernatant were kept to monitor expression and biotinylation, and the remaining lysate was incubated with 70 µL of pre-washed streptavidin-sepharose beads (GE Healthcare) for 3 h on a rotator at 4°C. Beads were then washed with 1 mL of RIPA buffer, transferred to a new tube, and washed again 2x with 1 mL of RIPA buffer and then 3x with 1 mL of 50 mM Ammonium Bicarbonate (ABC) (Biobasic). Beads were then resuspended in 100 µL of ABC with 1 µg of trypsin (Sigma) and incubated overnight at 37°C with rotation. The following day, 1 µg of trypsin was added for a further 2 h digestion. Samples were centrifuged 1 min at 2000 RPM, and the supernatant transferred to a new tube. Beads were rinsed twice with 100 µL of water and all supernatants pooled and taken to 5% formic acid. Samples were then centrifuged for 10 min at 16,000 x g for clarification. Trypsin-digested peptides in the supernatant were dried in a SpeedVac (Eppendorf) for 3 h at 30°C. Samples were resuspended in 15 µL of 5% formic acid and kept at -80°C for Mass Spectrometry analysis.

Samples were injected into an Orbitrap Fusion (Thermo Fisher). Protein identification and analysis was carried out as described elsewhere (Couzens et al., 2013). .RAW files were converted to .mzXML in Proteowizard (Kessner et al., 2008).

Peptide search and identification was processed using Human RefSeq Version 57 and the iProphet pipeline (Shteynberg et al., 2011) integrated in ProHits (Liu et al., 2010).

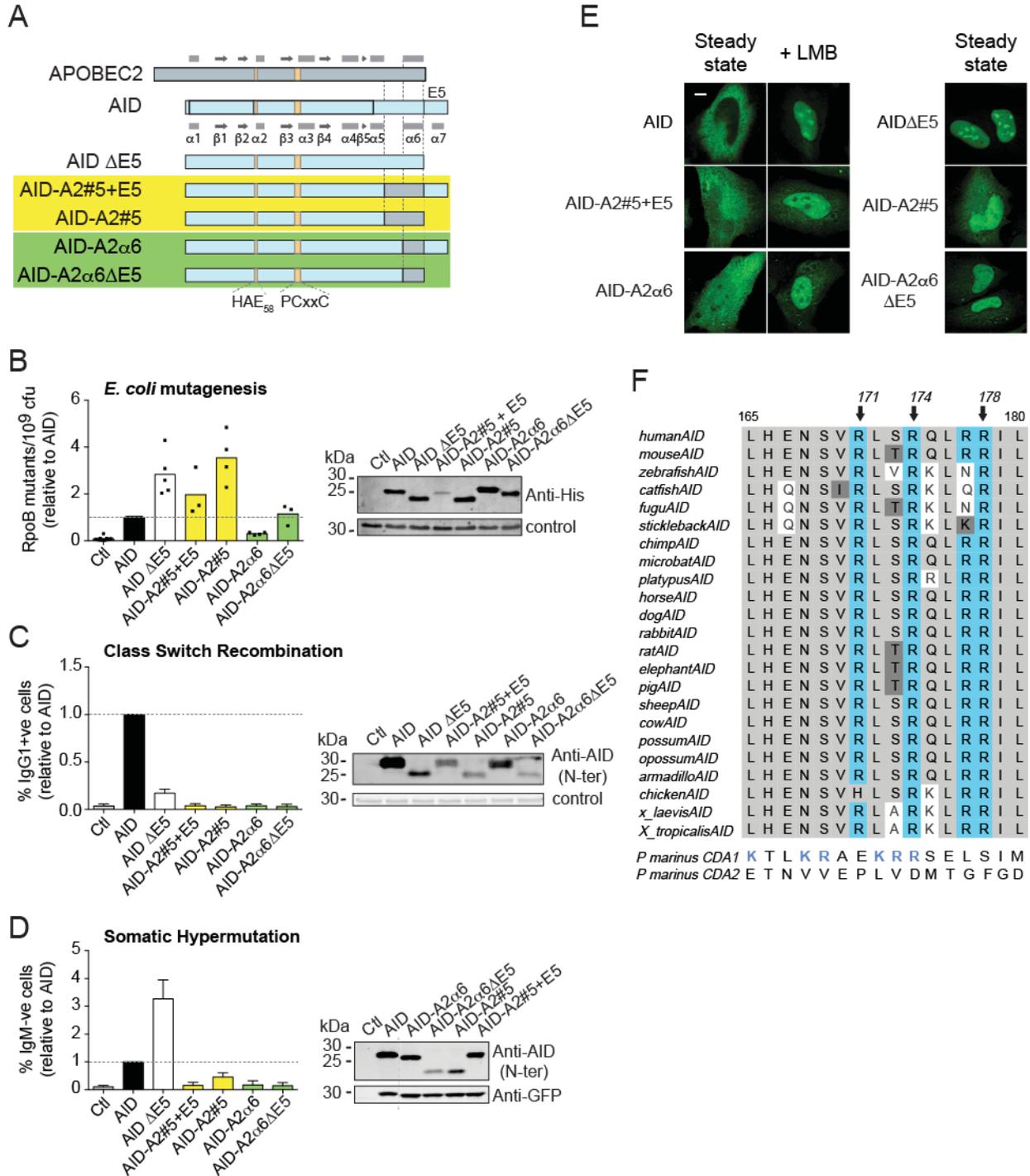
3.5.14. BioID

We used 4 different statistical methods to identify the most consistent wt-AID interactions that were significantly reduced in the R-mutants. All methods were implemented through ad-hoc scripts in R version 3.3.2. Tables were handled in R using the data.table package. Method 1 used fold-enrichment, calculated using mean spectral counts (SCs). First, SCs were normalized in all samples to their corresponding BirA* SC levels. Interactors with at least 2.5 fold enrichment in wt-AID over the R mutants or vice versa were identified as differential interactors. To eliminate interactions that were not AID-specific, only interactions with at least 5-fold enrichment over APOBEC2 were considered. Method 2 used log-transformed fold-enrichment values. Method 3 calculated local Z-scores using sliding windows in an R-I plot, as described (Quackenbush, 2002). In methods 2 and 3, hits with local Z-scores ≥ 2 or ≤ -2 (i.e., 2 SD from the mean) were considered as differential interactors. Method 4 used DESeq2 v.1.14.1 (Love et al., 2014), an R package that uses negative binomial generalized linear models to identify the differential interactions. This analysis was carried out using default parameters from the package and specifying the mutants as reference. Preys with multiple-testing adjusted p-values (Benjamini-Hochberg procedure) less than 0.1 were considered as differential interactors. All figures were plotted using ggplot2 package in R. Dot plots were generated using Prohits-Viz web tool (Knight et al., 2015).

3.6. Acknowledgements

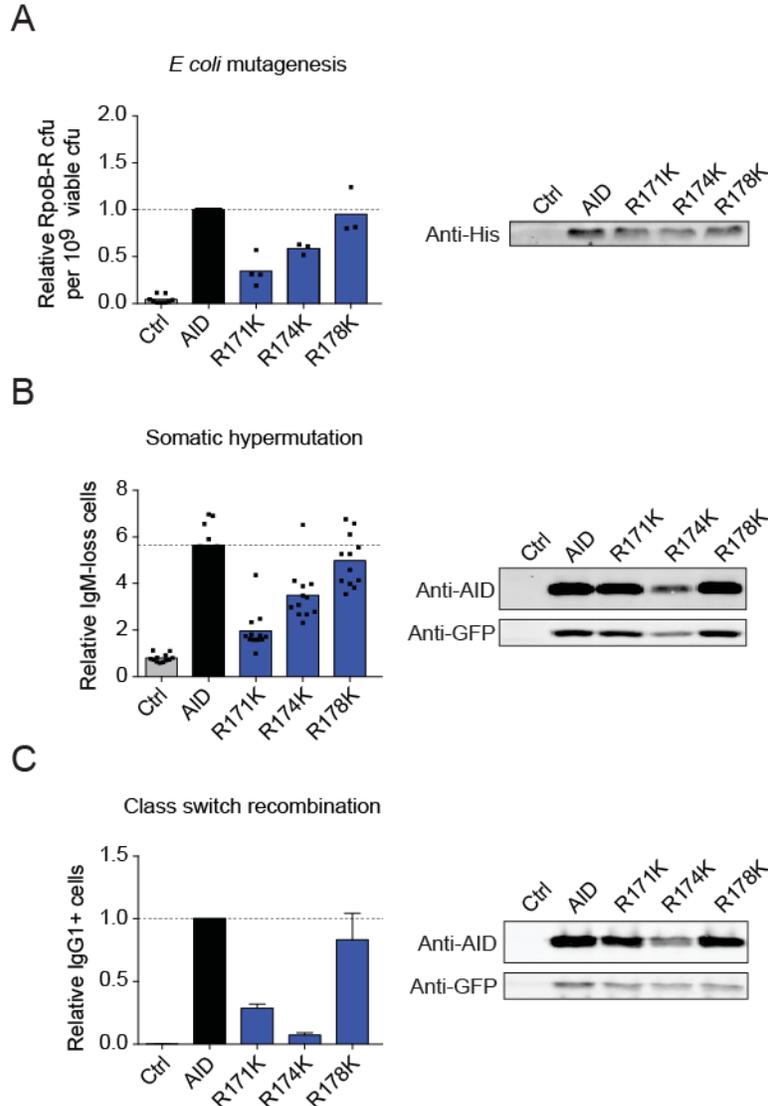
We thank Dr. N. Francis for critical reading and Drs. F. Robert, A. Orthwein and A. Zahn for discussions. We thank Jean-Felix Côté for technical help and the assistance of D. Fillion with microscopy, E. Massicotte, J. Lord, J Leconte and P St-Onge with flow cytometry and M. Cawthorne and E. Thivierge with animal care. We thank Drs. M. Neuberger, C. Rada, F Robert, A. Orthwein, R. Pavri and T. Honjo for sharing data and reagents. This work was funded by grants from the Cancer Research Society and Canadian institutes of Health research MOP130535 (to JMDN), MOP111132 (to ML), and a Discovery grant from The Natural Sciences and Engineering Research Council of Canada (to JFC). SPM and HB hold doctoral training awards from the Fonds de Recherche du Québec-Santé. SPM and LCL were supported by doctoral fellowships from the Cole Foundation. JFC holds the Transat Chair in Breast Cancer Research and is supported by an FRQS Senior investigator career award. JMDN holds a Canadian Research Chair tier 2. The authors have no conflicting financial interests.

3.7. Supplementary figures



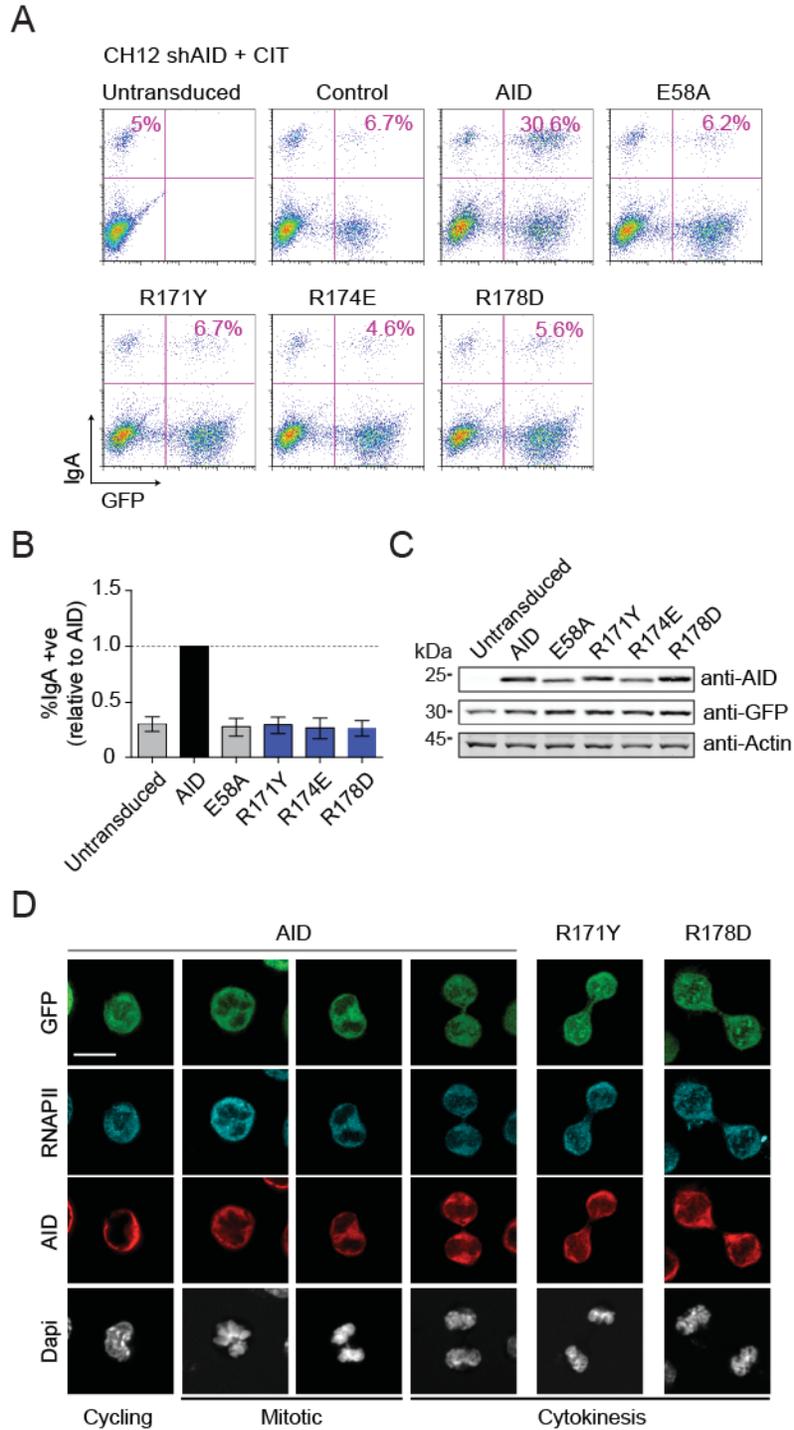
Supplementary Figure 3.1. The alpha helix 6 of AID is required for biological function.

(a) Schematics of APOBEC2 (A2), AID and AID-A2 chimeras. E5 denotes the region encoded by AICDA exon 5. Secondary structure elements, α -helices (light grey rectangles) and β -sheets (dark grey arrows) are identified. (b) Mutagenic activity in *E. coli* measured by the frequency of Rif resistant colonies arising from cultures expressing AID or the chimeras. Means (bars) of median values (dots) obtained from at least 3 independent experiments (5 cultures/experiment) are shown, normalized to AID. (c) Class switch recombination activity was assayed by transducing purified naïve *Aicda*^{-/-} mouse primary B cells with AID or the indicated chimeras -ires-GFP and culturing the cells with LPS and IL-4. The proportion of IgG1+ cells in the GFP+ population was assessed 72 h after transduction. Means \pm SDs from 3 independent experiments are shown, normalized to AID. (d) Somatic hypermutation activity was assayed by the relative IgM-loss accumulation in cultures of DT40 *Aicda*^{-/-} $\Delta\Psi V\lambda$ B cells complemented with AID or the indicated chimeras -ires-GFP. Means \pm SDs of the median values obtained from ≥ 3 independent experiments (≥ 12 cultures/experiment) were normalized to the mean value of AID. In (b-d) WB of cell extracts probed with anti-AID antibody and loading control are shown on the right. (e) Representative confocal microscopy images of HeLa cells transiently expressing AID and chimeras fused to GFP under steady state or after nuclear export inhibition with LMB (50 ng/mL, 2h). Representative of ≥ 2 independent experiments. Magnification 400X. Scale bar, 10 μ m. (f) Alignment of amino acid sequence of the region corresponding to the $\alpha 6$ helix of AID from multiple vertebrate species (top) or the $\alpha 6$ helix of AID and various APOBECs (bottom). Arg 171, 174 and 178 residues are indicated and basic residues at those positions are highlighted in blue.



Supplementary Figure 3.2. Arginines 171, 174 and 178 play different roles.

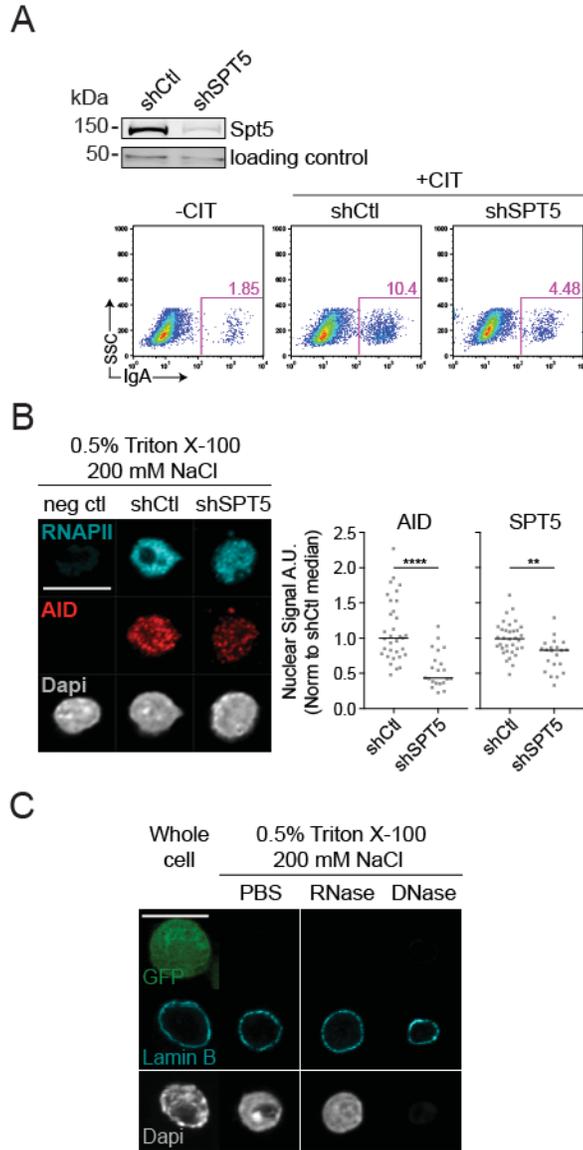
(a) Mutagenic activity in *E. coli* measured by the frequency of Rif resistant colonies arising from cultures expressing AID variants or empty vector (Ctrl). Means (bars) of median values (dots) obtained from at least 3 independent experiments (5 cultures/experiment) are shown, normalized to AID. (b) Somatic hypermutation activity was assayed by the relative IgM-loss accumulation in cultures of DT40 *Aicda*^{-/-} $\Delta\Psi\lambda$ B cells complemented with the indicated AID variants-ires-GFP or empty vector (Ctrl). Medians (bars) from 12 cultures/construct from 1 experiment are shown. (c) Class switch recombination activity was assayed by transducing purified naïve *Aicda*^{-/-} mouse primary B cells with the indicated AID variants-ires-GFP and culturing the cells with LPS and IL-4. The proportion of IgG1+ cells in the GFP+ population was assessed 72 h after transduction. Means \pm SDs from 3 independent experiments are shown, normalized to AID. In (b-d) WB of cell extracts probed with anti-AID antibody and loading control are shown on the bottom.



Supplementary Figure 3.3. Reconstitution of AID-deficient CH12 B cells

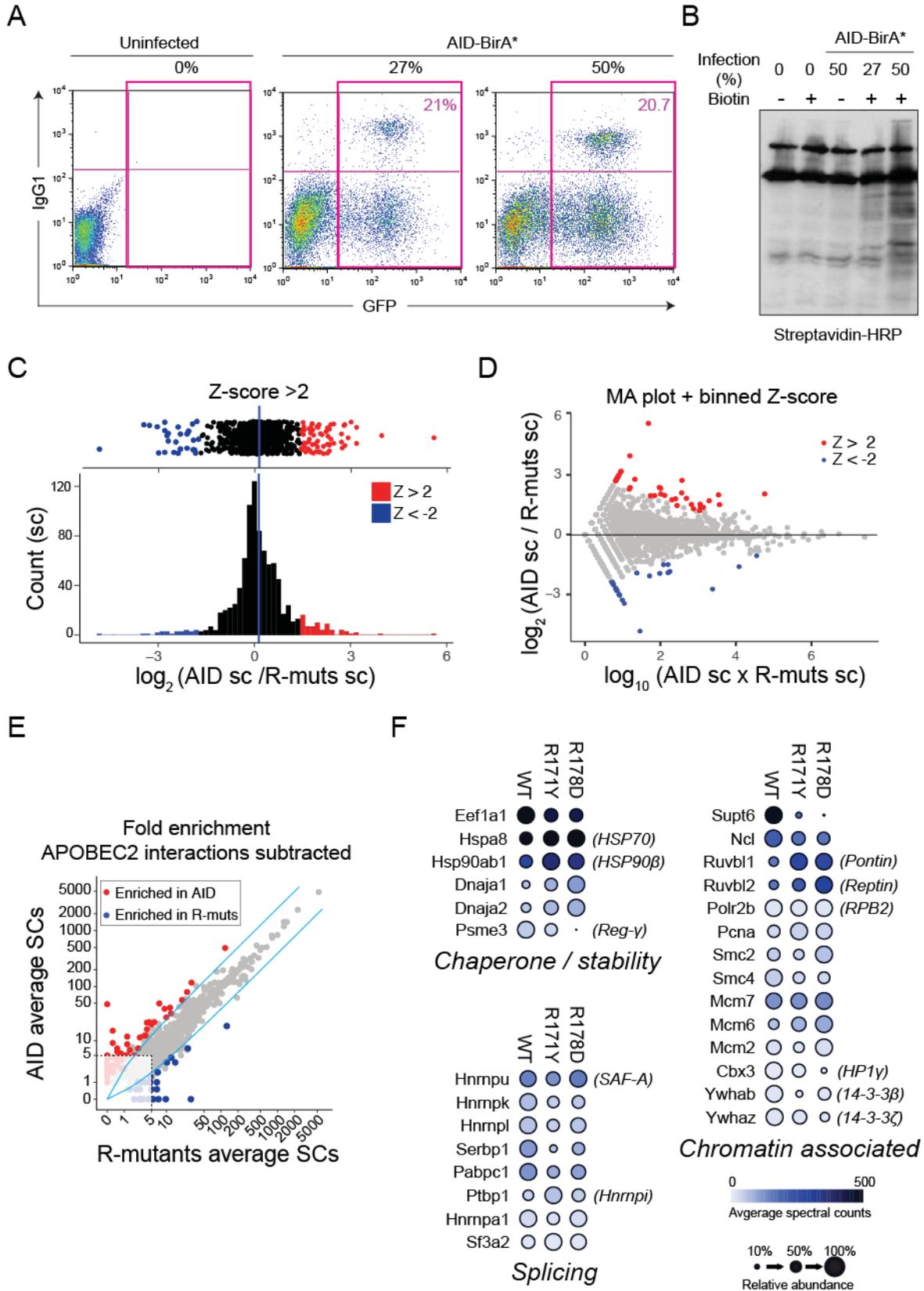
(a) CH12 B cells constitutively expressing an shRNA against AID were reconstituted with AID variants by transducing with pMX-AID variant-ires-GFP. Cultures were then stimulated with CIT for 72 h to induce CSR to IgA. Representative flow cytometry plots comparing GFP infection and IgA levels in CIT-

stimulated cells. The proportion of GFP+ cells that are IgA+ is indicated. (b) Means \pm SDs of the proportion of GFP+ cells that are IgA+ from 3 independent experiments, normalized to AID. (c) Western blots from reconstituted AID-deficient CH12 B cells. GFP is used as a control of reconstitution and actin as a loading control. (d) Confocal microscopy images of CH12 cells analyzed by anti-AID and anti-RNAPII IF. Cells were determined to be cycling (G1/S/G2), mitotic or in cytokinesis based on DNA condensation and RNAPII access to the DNA. Images are representative of at least 10 different events per construct from at least 2 independent experiments. Magnification 630X. Scale bar, 10 μ m.



Supplementary Figure 3.4. AID chromatin association controls

(a) (Top) Western blot for SPT5 and loading control (non-specific band) from wt CH12 B cells transduced with shRNA against luciferase (shCtl) or SPT5. (Bottom) Representative flow cytometry plots showing the proportion of IgA⁺ cells in unstimulated cells (-CIT) or stimulated cells (+CIT) expressing the indicated shRNAs. (b) CH12 cells expressing indicated shRNAs were stimulated with CIT prior to nuclear wash. (Left) Confocal microscopy images of isolated nuclei analyzed by anti-AID and anti-RNAPII (by IF) or DNA (dapi). (Right) Plot of mean AID or SPT5 signal for each nucleus (dots) and the population median (bars). Significant changes in AID or SPT5 signal by unpaired, two tailed t-tests are shown (** < 0.01, **** < 0.0001). (c) Representative confocal microscopy images of GFP, Lamin B (by IF) or DNA (dapi) on whole cells expressing GFP control or isolated nuclei thereof. During nuclear wash, nuclei were incubated at 37°C with PB control, RNase or DNase, as indicated. (b, c) Magnification 630X. Scale bar, 10 μm.



Supplementary Figure 3.5. BioID controls

(a) B cells from *Aicda*^{-/-} mice were transduced once or twice with pMX-AID-BirA*-ires-GFP. Representative flow cytometry plots showing the infection efficiency (above) and the relative proportions of IgG1+ in infected cells. (b) Representative western blot probed with streptavidin-HRP to detect biotinylated proteins 24h after adding biotin. Endogenous biotinylation occurs in the cells (lanes 0%), but BirA* dependent signal is only detected when the cells are cultured with biotin (+), and is proportional to the level of infection. (c) Z-score analysis, with positive Z-score values representing hits enriched for wt AID over R-mutants and negative values represent hits enriched for the R-mutants over wt AID. As indicated, positive hits were determined as ± 2 SDs away from the median. (d) Hits were distributed based on their overall spectral counts for AID and the R-mutants, and then binned in order to for run independent Z-score analysis based on overall association. Positive hits were again determined as ± 2 SDs from the median. (e) Comparison of average spectral counts (sc) for either wt AID or R-mutants interactions after subtracting A2 interactions. Teal lines delimit 2.5 fold changes in enrichment. Dashed box indicates hits with ≤ 5 sc, not considered in the analysis. Red and blue dots represent hits that were enriched 2.5x in AID over both R-mutants and both R-mutants over AID respectively. (f) Dot plot for multiple known AID interactions that were detected by BioID using the list from method 1 (i.e. subtracting those that also interact with A2). Proteins were divided into categories based on their main function. Alternate names for certain proteins are indicated (*italics*). Circle size indicates relative abundance normalized to the AID variant with the most spectral counts, and colour indicates actual spectral counts of identified factor (scales are included).

Supplementary Table 3.1. AID and R-mutants BioID comparison table

Geneid	Fold	Normz	Maz	DeSeq	Total	genename	Functional category	Primary function	Additional functions, comments	aid2	aid1	m1712	m1711	m1782	m1781	aid	m171	m178
108723	1	1	1	1	4	Card11	Signaling	BCR signaling		23	18	6	5	7	2	20.5	5.5	4.5
15163	1	1	1	1	4	Hcls1	Signaling	BCR signaling		116	119	27	41	21	29	117.5	34	25
17060	1	1	1	1	4	Blnk	Signaling	BCR signaling		16	15	1	0	0	0	15.5	0.5	0
192119	1	1	1	1	4	Dicer1	Translation	Translation silencing	May process RNA-DNA hybrids during transcription	9	9	0	1	0	0	9	0.5	0
209206	1	1	1	1	4	Supt16	Chromatin associated	Histone chaperone	Transcription elongation	484	482	142	197	71	64	488	169.5	67.5
223891	1	1	1	1	4	Eif3l	Translation	Translation initiation	eIF3 complex	37	42	9	7	7	12	39.5	9	9.5
227646	1	1	1	1	4	Sec16a	Other	Protein transport	ER-b-Goig vesicular transport system	20	23	3	7	4	4	22.5	5	4
23970	1	1	1	1	4	Pac-sin2	Signaling	Signaling	Transcription elongation	49	47	0	0	0	0	48	0	0
17955	1	1	1	1	3	Nap114	Chromatin associated	Histone chaperone	Transcription elongation	24	21	6	6	5	6	22.5	6	5.5
217337	1	1	1	1	3	Srp88	Other	Protein transport	mRNA transport	50	64	25	18	15	21	57	21.5	18
217869	1	1	1	1	3	Eif5	Translation	Translation initiation	mRNA transport	69	90	18	27	32	21	79.5	22.5	26.5
67154	1	1	1	1	3	Mtdh	Signaling	Signaling		44	39	11	7	15	14	41.5	9	14.5
13544	1	1	1	1	3	Dvl3	Signaling	Signaling		78	21	0	15	2	12	49.5	7.5	7
17535	1	1	1	1	3	Mre11a	Chromatin associated	DNA damage response		11	13	3	2	0	0	12	2.5	0
19069	1	1	1	1	3	Nup88	Chromatin associated	Nuclear pore	mRNA transport / Transcription silencing	7	10	2	3	0	0	8.5	2.5	0
230721	1	1	1	1	3	Pabpc4	Translation	mRNA transport		22	22	10	4	0	2	22	7	1
363258	1	1	1	1	3	Ley1	Other	Ribosome biogenesis	rRNA processing in nucleolus / cytoplasm	9	5	1	1	0	0	7	1	0
110811	1	1	1	1	2	Hdhbp	Chromatin associated	RNA-binding protein	heterochromatin formation	30	29	12	6	3	13	29.5	9	8
13669	1	1	1	1	2	Eif3a	Translation	Translation initiation	eIF3 complex	113	91	28	31	43	42	102	28.5	42.5
226594	1	1	1	1	2	Rsd1	Other	Actin remodeling	Interacts with Dishevelled	42	45	13	18	17	13	43.5	15.5	15
56347	1	1	1	1	2	Eif3c	Translation	Translation initiation	mRNA transport, eIF3 complex	57	62	29	30	12	21	59.5	28.5	16.5
75705	1	1	1	1	2	Eif4b	Translation	Translation initiation	mRNA transport	50	60	12	13	22	37	55	12.5	29.5
110355	1	1	1	1	2	Adrbk1	Signaling	Signaling		4	6	0	0	0	0	5	0	0
16341	1	1	1	1	2	Eif3e	Translation	Translation initiation	eIF3 complex / Histone H2B interactor	26	29	6	13	11	10	27.5	9.5	10.5
20383	1	1	1	1	2	Srsf3	Chromatin associated	Splicing	mRNA export	8	10	0	1	3	2	9	0.5	2.5
224742	1	1	1	1	2	Abcr1	Translation	Translation initiation	Histone H2B interactor	13	20	3	0	6	7	16.5	1.5	6.5
320528	1	1	1	1	2	Vps13c	Other	mitochondrial maintenance		6	9	2	0	0	0	6	1	0
58194	1	1	1	1	2	Ski3btp1	Signaling	Signaling	Interacts with BLNK	6	9	0	0	0	0	7.5	0	0
74112	1	1	1	1	2	Usp16	Chromatin associated	H2A de-ubiquitinase	Transcription elongation	2	9	0	0	2	1	5.5	0	1.5
17886	1	1	1	1	1	Myh9	Other	Cytoskeleton		142	135	66	69	62	78	138.5	67.5	70
208643	1	1	1	1	1	Eif4g1	Translation	Translation initiation	mRNA transport	139	118	50	72	69	84	128.5	61	76.5
226562	1	1	1	1	1	Prrc2c	Other	Unknown		200	169	97	91	69	79	184.5	94	74
27979	1	1	1	1	1	Eif3b	Translation	Translation initiation	RNA-binding	52	62	23	22	34	25	57	22.5	29.5
73158	1	1	1	1	1	Larp1	Translation	Translation initiation		107	91	45	50	41	62	99	47.5	51.5
103963	1	1	1	1	1	Rpn1	Other	N-glycosaccharyl transferase	Proteasome	14	9	0	3	5	3	11.5	1.5	4
107951	1	1	1	1	1	Cdk9	Chromatin associated	Transcription elongation	Cyclin dependent kinase	15	10	2	3	2	8	12.5	2.5	5
19192	1	1	1	1	1	Psmc3	Other	Proteasome	known to promote nuclear AID degradation	43	43	32	31	3	4	43	31.5	3.5
213988	1	1	1	1	1	Tnrc6b	Translation	Translation silencing		7	7	3	2	2	3	7	2.5	2.5
218873	1	1	1	1	1	Wdhd1	Chromatin associated	Replication initiation		12	17	2	0	0	11	14.5	1	5.5
236732	1	1	1	1	1	Rbm10	Chromatin associated	Splicing		5	5	0	0	0	1	5	0	0.5
27984	1	1	1	1	1	Efhcd2	Other	Calcium binding		9	3	0	0	0	2	6	0	1
319322	1	1	1	1	1	Sf3b2	Chromatin associated	Splicing		6	4	0	0	0	0	5	0	0
56440	1	1	1	1	1	Snx1	Other	Protein transport		20	32	4	0	7	19	26	2	13
59021	1	1	1	1	1	Rab2a	Other	Protein transport		7	6	0	2	2	2	6.5	1	2
66085	1	1	1	1	1	Eif3f	Translation	Translation initiation		45	55	26	11	30	20	50	18.5	25
66448	1	1	1	1	1	Mrp120	Other	Mitochondrial ribosome		6	4	4	0	0	0	5	2	0
67062	1	1	1	1	1	Sic25as53	Other	Mitochondrial transport		5	5	0	1	4	0	5	0.5	2
67166	1	1	1	1	1	Arl8b	Signaling	GTPase		6	4	3	1	2	2	5	2	2
68926	1	1	1	1	1	Pabpc6	Translation	mRNA transport	mouse specific	30	35	0	26	23	0	32.5	13	11.5
67543	1	1	1	1	1	Ubaip2	Other	Protein ubiquitination		14	11	1	9	4	0	12.5	5	2
69710	1	1	1	1	1	Arap1	Other	Protein transport		4	7	0	0	0	0	5.5	0	0
72567	1	1	1	1	1	Bclaf1	Chromatin associated	Transcriptional repressor	Can promote apoptosis	5	6	0	0	0	0	5.5	0	0
75786	1	1	1	1	1	Ckap5	Other	Cytoskeleton		6	8	0	0	0	3	7	0	1.5
76302	1	1	1	1	1	Ponp	Other	Cell cycle		6	6	2	2	0	0	6	6	2

**CHAPTER 4: STRUCTURAL CONFORMATION OF THE
AID C-TERMINAL DOMAIN IS NECESSARY FOR
EFFICIENT CLASS SWITCH RECOMBINATION**

Stephen P. Methot, Astrid Zahn, Julian C. Gilmore, Anil K. Eranki, Javier M. Di Noia

This article has been prepared for submission to the *European Journal of
Immunology*.

The regulatory mechanisms I had studied thus far, relating to AID cytoplasmic retention and promoter licensing, required residues in the C-terminal region of AID. During my studies we also published an article relating to a functional role for the E5 region during CSR (Zahn et al., 2014). Based on these works studying the requirements for the C-terminal region in AID regulatory mechanisms, we wanted to further define the requirements of the C-terminal residues for AID regulation, and in particular for downstream processing during CSR.

4.1. Abstract

Activation induced deaminase (AID) is critical to both somatic hypermutation (SHM) and class switch recombination (CSR) in B-lymphocytes. The E5 C-terminal domain of AID is essential for CSR, but dispensable for SHM. We previously reported that lesions generated by AID lacking this domain (AID Δ E5) are extensively resected and toxic. Moreover, AID Δ E5 is hyperactive. We now sought to further understand the role of E5 during CSR by performing a mutagenic screen of the AID E5 residues guided by a structural model. We unexpectedly find a role for E5 in modulating AID catalytic activity. Using modular fusion constructs, we find that the CSR defect is independent of AID Δ E5 hyperactivity and toxicity, but depends on the structural context of the E5 domain with respect to the enzyme core. These results support a model wherein the C-terminal domain of AID mediates efficient CSR by recruiting factor(s) downstream of its deamination activity.

4.2. Introduction

Secondary antibody diversification in peripheral B-lymphocytes involves two distinct processes, somatic hypermutation (SHM) and class switch recombination (CSR), which diversify antibody recognition and functional domains respectively (Di Noia and Neuberger, 2007; Stavnezer et al., 2008). Both these mechanisms require the enzyme Activation-induced deaminase (AID), which deaminates deoxycytidine to deoxyuridine at the immunoglobulin (Ig) loci (Di Noia and Neuberger, 2002; Muramatsu et al., 2000; Petersen-Mahrt et al., 2002).

Processing of the AID dependent uracils by different DNA repair pathways underpins CSR, as well as the varied spectrum of mutations observed in SHM (Methot and Di Noia, 2017). SHM, at the Ig variable regions, leads to the generation of single point mutations either directly by replication over the uracil, or after uracil excision, via the recruitment of error prone polymerases during the repair process. For CSR, dsDNA breaks (DSBs) are generated at the *IgH* switch regions (S-regions) by processing closely spaced uracils, introduced by AID on opposite strands, into ssDNA nicks. The c-NHEJ pathway of DSB repair then processes these lesions in order to re-join the break with a simultaneous break occurring at another S-region (Stavnezer et al., 2008).

Surprisingly, though SHM and CSR are similarly initiated by AID generated uracils, they can be functionally distinguished by single point mutations in distinct regions of the enzyme. SHM is more sensitive to specific N-terminal mutations, while CSR is specifically ablated by truncations and some mutations in a C-terminal domain of AID (Barreto et al., 2003; Shinkura et al., 2004; Ta et al., 2003).

The C-terminal region of AID that is exactly encoded by the exon 5 (E5) of the gene (*Aicda*), forms a discrete domain that is particularly important for AID regulation and function. AID lacking the E5 domain (AID Δ E5) is constitutively nuclear (Ito et al., 2004), as E5 contains both a nuclear export sequence (NES), necessary for CRM1 exportin binding (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004) and a structural motif necessary for its retention in the cytoplasm *via* the translation factor eEF1A1 (Hasler et al., 2011; Methot et al., 2015; Patenaude et al., 2009). AID Δ E5 is also hyperactive, producing ~3-fold more mutations than wt AID when expressed in *E. coli*, which translates into a proportional increase in SHM or Ig gene conversion activity in B cells (Barreto et al., 2003). Nonetheless, AID Δ E5 is not only defective for CSR (Barreto et al., 2003), but it can block CSR activity when co-expressed with wt AID (Ucher et al., 2014; Zahn et al., 2014). This dominant negative effect on CSR of AID Δ E5 likely explains the autosomal dominant hyper-IgM syndrome seen in patients carrying C-terminal AID truncations, which display normal levels of SHM (Imai et al., 2005; Kasahara et al., 2003).

Here, we performed a systematic analysis of E5 residues, confirming the importance of the NES residues (Doi et al., 2009; Geisberger et al., 2009), and identifying a role for aspartic acid 188 in AID functional activity, both for CSR and SHM. We also demonstrate that the structural context of the E5 is important for AID catalytic activity, and is essential to its role during CSR.

4.3. Results

4.3.1. AID lacking the E5 region is still dependant on S38-phosphorylation

As previously reported (Barreto et al., 2003; Zahn et al., 2014), AID Δ E5 has increased mutagenic activity relative to full-length AID both in *E. coli*, as well as at the IgV region in DT40 B cells (Fig. 4.1A,B). This hyperactivity also reduces cell fitness in DT40 B cells (Fig. 4.1C), most likely due to excessive off-target DNA damage shown by the production of visible γ H2AX foci (Zahn et al., 2014). AID needs phosphorylation of serine 38, which mediates its interaction with RPA, to be able to deaminate both strands of transcribed DNA from cell extracts, as well as for efficient SHM and CSR (Basu et al., 2005; Cheng et al., 2009; McBride et al., 2008). To ask whether the exacerbated activity of AID Δ E5 *in vivo* was still regulated by Ser38 phosphorylation, we mutated it to alanine to generate AID Δ E5-S38A. Interestingly, though AID Δ E5-S38A maintained hyperactivity in *E. coli*, it had substantially reduced SHM in DT40 B cells (Fig. 4.1A-C). This result shows that exacerbated SHM capacity of AID Δ E5 is still dependent on Ser38 phosphorylation, demonstrating that AID Δ E5 mutagenic activity is similarly regulated to wt AID. AID Δ E5-S38A also had a much weaker effect on DT40 B cell fitness than AID Δ E5, suggesting that Ser38 phosphorylation does not specifically regulate the activity of this hyperactive variant at the Ig loci, but extends genome-wide.

4.3.2. A role for the E5 in regulating enzymatic activity

The biochemistry regarding the role of E5 on AID activity remains unresolved. The *in vitro* activity of AID Δ E5 on ssDNA was found to be equivalent to full

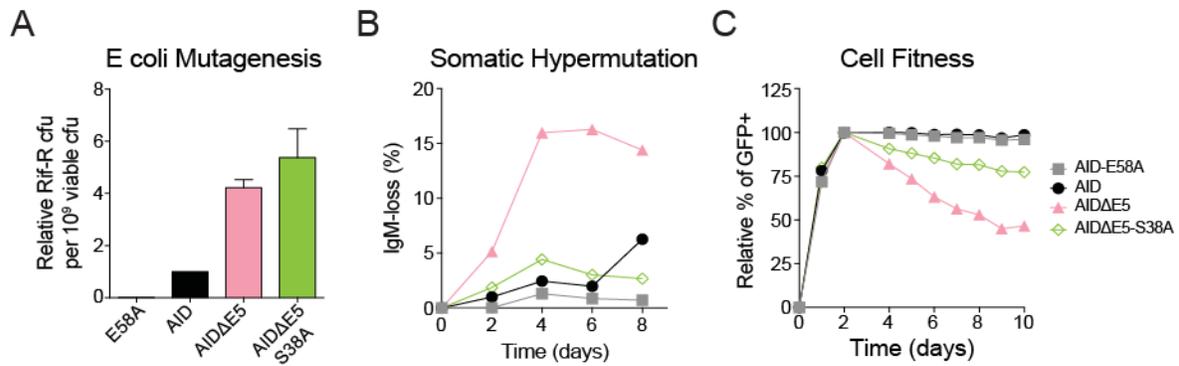


Figure 4.1. AIDΔE5 hyperactivity requires S38-phosphorylation in B cells

(A) Mutagenic activity in *E. coli*, measured by the relative frequency of rifampicin (Rif) resistant colonies arising from cultures expressing AID variants or empty vector (Ctrl). Means + SEM (bars) of median values obtained from 2 independent experiments (5 cultures/experiment), normalized to AID. (B) Somatic hypermutation activity, assayed by IgM-loss over time in DT40 *Aicda*^{-/-} ΔΨVλ B cells complemented with AID variants-ires-GFP. (C) Effect of AID or AIDΔE5 variants-ires-GFP on the growth of transduced AID-deficient DT40 B cells. The proportion of GFP+ cells monitored over time, normalized to max value.

length AID in two reports (Mu et al., 2012; Zahn et al., 2014) but 3-fold higher in another (Kohli et al., 2009). The comparison of purified AID mutants in deaminating oligonucleotides is likely influenced by the different expression systems used and the fact that recombinant AID is known to be prone to aggregation (Larijani and Martin, 2007; Larijani et al., 2007; Qiao et al., 2017). Interestingly, C-terminal truncations that include even a single residue from exon 4 almost completely abrogate enzymatic activity in vitro, demonstrating that the structural integrity of the region adjacent to E5 is important for activity (Mu et al., 2012). Partial truncations of E5 can also increase the mutagenic activity of the enzyme in *E. coli* (Zahn et al., 2014), suggesting that this domain does modulate the enzymatic activity of AID (Barreto et al., 2003; Zahn et al., 2014). Our structural modelling has suggested two possible conformations for E5, either one of which is unlikely to physically reach the catalytic residues or pocket of AID (King

et al., 2015; Methot et al., 2015) (Fig. 4.2A,B). Therefore, one possibility for the effect of removing E5 on catalytic activity, is that it modulates allosteric effects that either alter the catalytic pocket or promote substrate accessibility (King et al., 2015).

Previous screens for E5 residues important for AID activity have focussed on the last 10 amino acids, containing the NES (Table 4.1) (Doi et al., 2009; Geisberger et al., 2009). We decided to test additional E5 residues for their effect on catalytic activity, using *E. coli* mutagenesis as a proxy (Fig. 4.2C). Interestingly, mutating the initial E5 residue Proline 182 to Alanine (P182A), which would ease the rotational constraints induced by Pro on the polypeptide chain (Alber et al., 1988), reduced catalytic activity. AID-P182A also reduced SHM and CSR activity, proportionally to the reduction of catalytic activity (Fig. 4.2C-E). On the other hand, mutating Leucine 198, part of the NES, to serine, increased catalytic activity, as we have previously observed (Fig. 4.2C) (Zahn et al., 2014). There were other examples of disparate effects of mutations in E5 on catalytic activity, E185A reduced activity while L183A or R194A increased it (Fig. 4.2C).

These opposing effects of single point mutations on enzymatic activity strongly argue that allosteric effects of E5 can modulate AID catalytic activity. Such allosteric effects could in some way be related to E5 docking on the catalytic core, as we demonstrated previously that NES mutations could affect such a structure (Methot et al., 2015).

4.3.3. Identification of D188 as an important functional residue

Amino acid substitutions in E5 fell into three categories when assayed for SHM and CSR. First, those that reduced SHM and CSR proportionally to their effect on *E. coli*

Table 4.1. Compilation of known AID C-terminal mutations

		Mutation tested	E coli mut	SHM	CSR	Reference	
		wt	++	++	++		
Proline	182	P182A	+	+	+	here	
Leucine	183	L183A	+++	++	ND	here	
Tyrosine	184	Y184A	+++	++	++	(Basu et al., 2005; Zahn et al. 2014)	
Glutamate	185	E185A	+	++	++	Here	
Valine	186	-	ND	ND	ND		
Aspartate	187	D187A	+	+++	++	(Zahn et al., 2014; Methot et al., 2015)	
Aspartate	188	D188A	++	-	+/-	Here	
		D188E	++	+	+	Here	
NES	Leucine	189	L189A	ND	+++	+/-	(Doi et al., 2009; Geisberger et al., 2009)
	Arginine	190	R190A	ND	+/-	++	(Doi et al., 2009; Geisberger et al., 2009)
	Aspartate	191	D191A	ND	++	++	(Doi et al., 2009; Geisberger et al., 2009)
	Alanine	192	A192G	ND	++	++	(Doi et al., 2009)
	Phenylalanine	193	F193A	ND	+++	+/-	(Doi et al., 2009; Geisberger et al., 2009)
	Arginine	194	R194A	+++	++	++	Here and (Doi et al., 2009; Geisberger et al., 2009)
	Threonine	195	T195A	ND	++	++	(Doi et al., 2009; Geisberger et al., 2009)
	Leucine	196	L196A	ND	+++	+/-	(Doi et al., 2009; Geisberger et al., 2009)
	Glutamine	197	G197A	ND	+	+	(Doi et al., 2009; Geisberger et al., 2009)
	Leucine	198	L198S	+++	+++	+/-	Here and (McBride et al., 2004; Zahn et al. 2014)

mutagenesis, like P182A. Second, those that had no effect on functional activity, such as E185A, R194A and L183A (Fig. 4.2C-E). These mutations demonstrate that not all effects on catalytic activity result in functional defects. One possible explanation is that protein associations in B cells could stabilize the E5 structure and thereby normalize AID activity.

Third, and most informative were changes that preserved the capacity to mutate *E. coli*, but prevented SHM and/or CSR activity. As previously described, mutation of the NES residue L198, increased AID *E. coli* mutagenesis and SHM, but abrogated CSR activity (Zahn et al., 2014) (Fig. 4.2D,E). In addition, we found that Aspartic acid 188 (D188), was critical for AID functional activity, as AID D188A had little to no effect on AID *E. coli* mutagenesis, but completely compromised both SHM and CSR activity (Fig. 4.2C-E). As Aspartic acid is negatively charged, we mutated D188 to Glutamate, to investigate whether structural changes at this position affect activity. The D188E mutation reduced AID functional activity, though not to the extent of D188A, demonstrating that D188 charge and structure are relevant to its functional role.

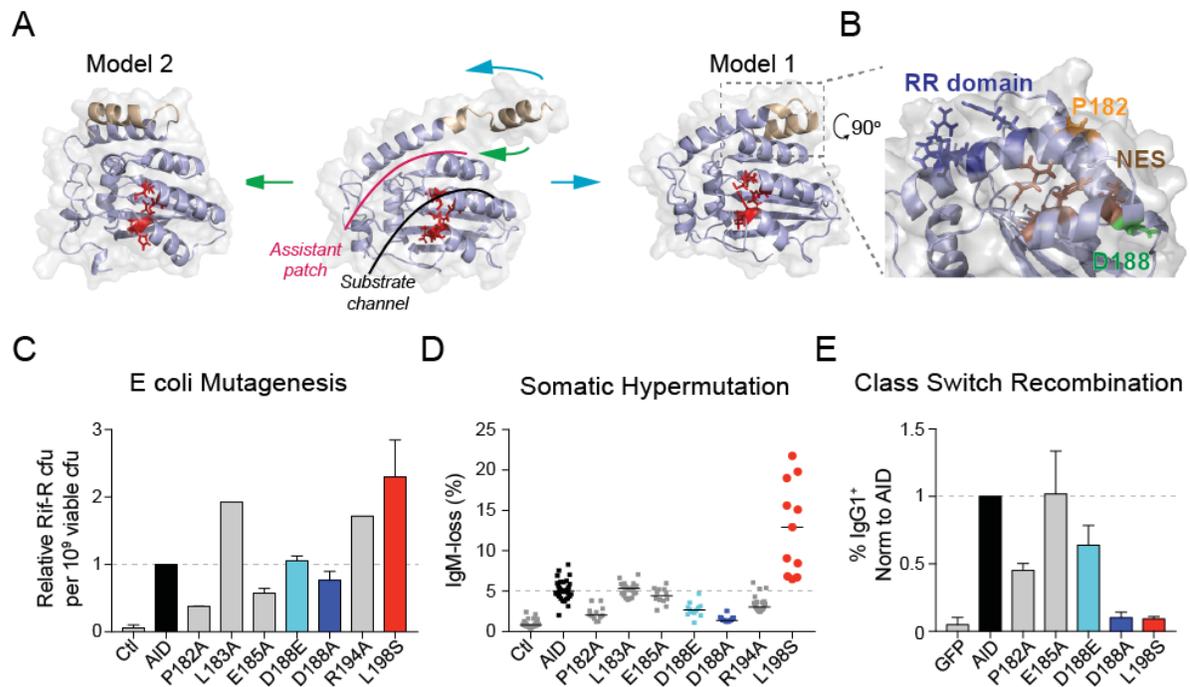


Figure 4.2. The E5 domain regulates AID catalytic and functional activities

(A) AID structural models from (Methot et al., 2015), highlighting AID E5 (wheat) and the catalytic pocket (red). E5 docking generates distinct structures, with the E5 remaining distant from the catalytic pocket. Assistant patch and substrate channel are from (Qiao et al., 2017). (B) Structural model of AID (as in A) highlighting structural elements involved in AID catalytic and functional activity. (C) Mutagenic activity in *E. coli*, measured by the relative frequency of rifampicin (Rif) resistant colonies arising from cultures expressing AID variants or empty vector (Ctrl). Means +SDs (bars) of median values obtained from 5 cultures/experiment, normalized to AID. Representative of 1-4 independent experiments/construct. (D) Somatic hypermutation activity, assayed by the relative IgM-loss accumulation in cultures of DT40 *Aicda*^{-/-} $\Delta\Psi\text{V}\lambda$ B cells complemented with AID variants-ires-GFP or empty vector (Ctrl). Medians (bars) from 12 cultures/construct are plotted. Representative of 1-3 independent experiments/construct. (E) Class switch recombination activity, monitored in naïve *Aicda*^{-/-} mouse primary B cells cultured with LPS and IL-4 and transduced with AID variants-ires-GFP or empty vector (Ctrl). The proportion of IgG1⁺ cells in the GFP⁺ population was determined 72 h after transduction. Means + SDs from 5-6 independent mice are shown, normalized to AID.

When all known AID E5 mutations are compiled (Table 4.1) it is clear that the 4 NES residues demonstrate the same phenotype of increased SHM activity but reduced CSR. Interestingly, D188 is the only E5 residue found to completely disconnect *E. coli*

mutagenesis from functional activity suggesting it may be the only E5 residue contributing to such a functional motif.

We conclude that residue D188 somehow mediates AID functional activity in B cells.

4.3.4. Fusions to AID C-terminus affect enzymatic activity

AID Δ E5 affects different aspects of AID: subcellular localization, catalytic activity, toxicity, SHM and CSR. This makes it difficult to distinguish which effects are directly related to its role in CSR. We sought to distinguish these effects by generating modular fusions of AID with the catalytic site mutant AID^{E58A}, a system we have previously used (Chapter 3).

Fusion of AID^{E58A} to the C-terminus of wt AID, or AID Δ E5, significantly reduced *E. coli* mutagenesis of both, though deamination activity was still detectable above AID^{E58A} background (Fig 4.3A). A similar reduction of *E. coli* mutagenesis occurred if AID Δ E5^{E58A} was used for the fusion instead, demonstrating that the E5 region itself did not modulate the catalytic activity of the fusion proteins. This suggests that the large C-terminal fusion exerts conformational or allosteric effects on the catalytic site. Interestingly, *E. coli* mutagenesis of AID7.3, an AID variant carrying three single amino acid replacements (K10E, T82I, E156G) that makes it catalytically hyperactive (Wang, NSMB, 2009), was not affected by the AID^{E58A} fusions in the exact same configuration (Fig. 4.3A). This result implies that the changes made in AID7.3 render it structurally resistant to allosteric effects from the C-terminus.

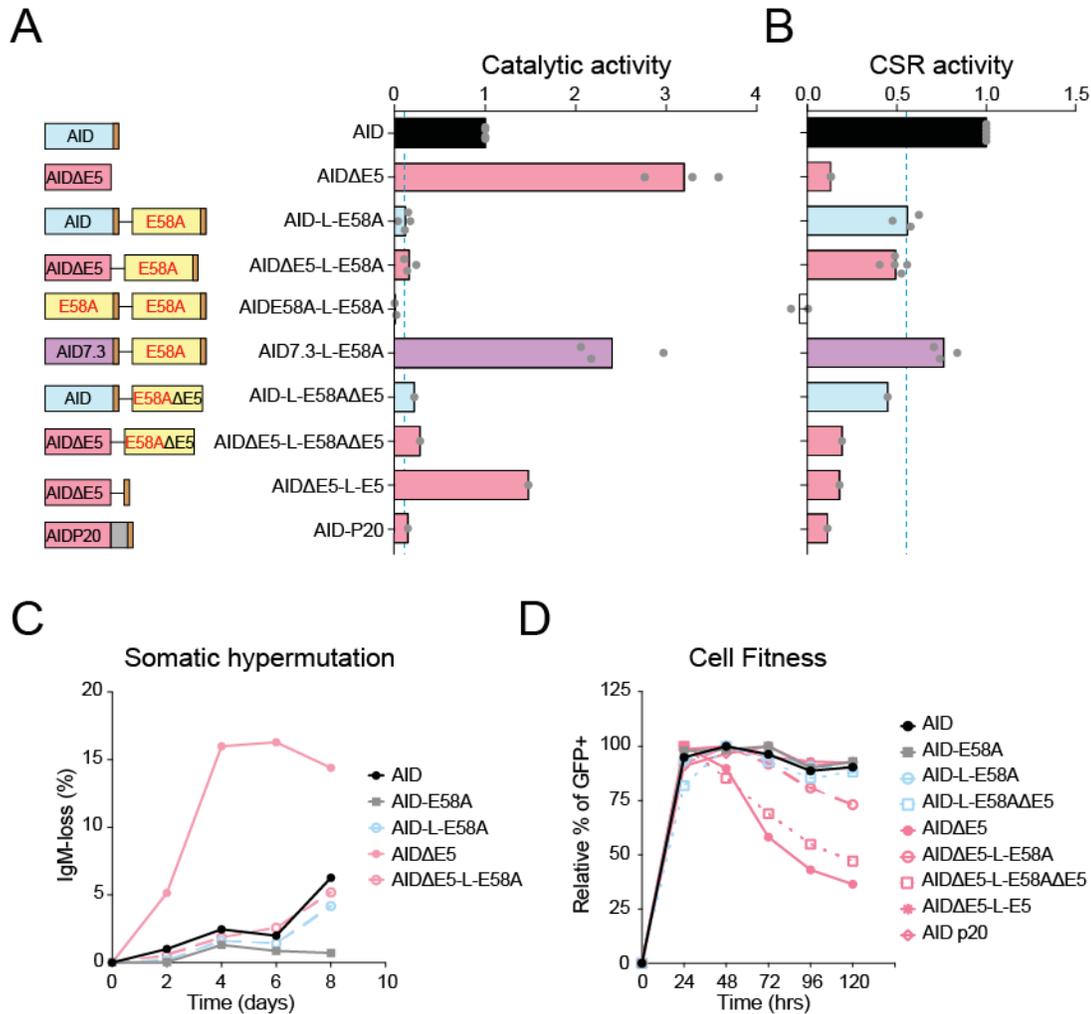


Figure 4.3. AID E5 structural integrity is necessary for CSR

(A) (Left) Schematic of AID fusion constructs. (Right) Relative *rpoB* mutation frequency in *E. coli* for various AID-variants. Means (bars) of medians (dots) from 1-3 independent experiments (5 clones/experiment) are shown. (B) Relative CSR capacity of AID variants expressed in AID-deficient CH12 mouse B cells. Means (bars) from 1-4 independent experiments (dots) are shown, normalized to AID. (C) Somatic hypermutation activity, assayed by IgM-loss over time in DT40 *Aicda*^{-/-} ΔΨVλ B cells complemented with AID variants-ires-GFP. (D) Effect of AID variants-ires-GFP on the growth of transduced AID-deficient CH12 B cells. The proportion of GFP+ cells monitored over time, normalized to max value.

Finally, we separated the E5 from the core deaminase fold of AID by inserting a flexible linker between them. This was sufficient to increase *E. coli* mutagenesis

compared with AID. Interestingly, AID-P20, an analogous AID variant found in a HIGM2 patient, which encodes a 34 amino acid insertion at the same site but of a different sequence, shows the opposite effect (Fig. 4.3A).

Together with the results using single point mutants, these results demonstrate that the structural integrity and conformation of the E5 can have important consequences on the catalytic activity of AID.

4.3.5. AID C-terminal integrity is critical for CSR

Despite reduced *E. coli* mutagenesis, we previously observed that the fusion of AID to AID^{E58A} was still capable of CSR activity in *Aicda*^{-/-} mouse B cells (chapter 3) and here find that it has a similar CSR activity in AID-deficient CH12 B cells (Fig 4.3B). Interestingly, the AIDΔE5 fusion to AID^{E58A} displayed CSR activity to the same level as AID-AID^{E58A} in complemented CH12 cells (Fig 4.3B). This ability of AIDΔE5-AID^{E58A} to induce CSR was associated with a normalization of SHM activity and abrogation of its negative effect on B cell fitness (Fig 4.3C,D). This result demonstrates that an inactive but full length AID can compensate for the defect of the AIDΔE5 for CSR, suggesting that it restores interactions lost in the truncated AID. However, AIDΔE5 and AIDΔE5-AID^{E58A} also differed in that the former was hyperactive and the latter was not.

We then sought to discriminate whether eliminating hyperactivity was sufficient to rescue CSR, using the fusion of AIDΔE5 to AIDΔE5^{E58A}. Despite normalized *E. coli* mutagenesis levels, AIDΔE5-AIDΔE5^{E58A} was unable to rescue CSR (Fig 4.3B). This fusion was still toxic, reducing B cell fitness, suggesting that it maintained high levels of off-target activity (Fig 4.3D) Nonetheless, this demonstrated that hyperactivity in and of itself is not responsible for the CSR defect, suggesting that the presence of an intact E5

is critical for CSR. However, the presence of E5 is not sufficient, since AID-P20 and the analogous variant with a flexible linker in front of E5 were inactive for CSR (Fig 4B). These two variants did not increase AID toxicity, suggesting that high levels of off-target activity are also not the cause of the CSR defect (Fig 4D).

All together, we manage to dissociate the CSR defect of AID Δ E5 from its increased *E. coli* mutagenesis and toxicity (Table 4.2), suggesting that the CSR defect is downstream from deamination. We conclude that AID E5 structural integrity and a specific conformation are necessary to mediate CSR, beyond its effect on modulating enzymatic activity.

Table 4.2. Comparison of AID fusion constructs

	E coli mut	SHM	toxicity	CSR
AID	+	+	+	+++
AID Δ E5	+++	+++	+++	+/-
AID-L-AID(E58A)	+	+	+	++
AID Δ E5-L-AID(E58A)	+	+	+	++
AID-L- Δ E5(E58A)	+	ND	+	++
AID Δ E5-L-AID Δ E5(E58A)	+	ND	+++	+/-
AID Δ E5-L-E5	++	ND	+	+/-
AID-P20	+	ND	+	+/-

4.4. Discussion

We define different roles for the E5 in regulating AID catalytic activity and functional activity.

First, we demonstrate that single point mutations within the E5, or C-terminal fusions, can alter its catalytic activity. This strongly argues for an allosteric role of E5 in modulating the catalytic domain of AID, we speculate that this is related to a certain flexibility that allows E5 to adopt different conformations. This is an important limitation of recent crystal structures of AID, solved without the E5 (Pham et al., 2016; Qiao et al.). Our molecular modeling suggests some possibilities to explain how E5 might modulate substrate accessibility. Notably, in model 1 the E5 is on the opposite side of the catalytic site, whereas in model 2 it comes closer to the active site and to helix α_6 (Fig 4.2A,B). Furthermore, in model 2, the E5 occupies the assistant patch, which seems to be necessary for activity on structured DNA and for CSR activity (Qiao et al., 2017). In future, using the AID crystal structures may provide a more precise base for testing the effects of E5 docking on the actual catalytic site structure, and may further explain the enzymatic activity effects we observe.

We also identify a residue, D188 within the E5, which is critical for AID functional activity, despite intact mutagenic activity in *E coli*. The exact role for D188 in B cells remains to be elucidated; however, we have previously shown that D188 is necessary for AID cytoplasmic retention, and that the D188A variant is more nuclear at steady state (Patenaude et al., 2009), excluding a defect in nuclear access to explain the functional defect in B cells. Molecular modeling suggests that D188 is exposed on the protein surface (Fig. 4.2B); thus, D188 may mediate a novel interaction or it may be

related to (or part of) the newly identified RR domain in the $\alpha 6$ helix of AID, which is necessary to license AID mutagenic activity at the chromatin (Chapter 3).

The NES of AID is required for CSR activity, and was previously linked to AID stability (Doi et al., 2009; Geisberger et al., 2009). Importantly, the role for the NES in CSR can be dissociated from nuclear export or CRM1 binding affinity (Ellyard et al., 2011), which strongly suggests that the NES promotes association of AID to additional factors during CSR.

Our results confirm a role for E5 in mediating functional interactions, but also imply that the E5 has a role downstream from deamination activity. This is demonstrated by our results showing that AID Δ E5 mutagenic activity still depends on Ser38 phosphorylation as well as the fact that AID Δ E5 produces mutations with a similar distribution and hotspot preference to wt AID (Barreto et al., 2003; Zahn et al., 2014). Our modular fusion constructs also demonstrate that the CSR defect of AID Δ E5 is distinct from its hyperactivity, increased SHM activity, or cytotoxicity. It still remains to be determined how these fusion constructs affect AID localization, and dominant negative activity, to see whether these effects can also be dissociated from CSR activity.

The lesions generated by AID Δ E5 undergo excessive end-resection and seem to be largely repaired by HR (Zahn et al., 2014). This can explain the CSR defects observed, as CSR requires NHEJ and cannot proceed via HR (Stavnezer et al., 2008). Furthermore, B cells that express AID Δ E5 have reduced UNG association at the S_{μ} compared to cells expressing wt AID (Ranjit et al., 2011; Sabouri et al., 2014; Zahn et al., 2014). This observation correlates with an increase in transitions versus

transversions at C:G pairs both at the IgV and at the S μ in B cells expressing AID Δ E5 (Barreto et al., 2003; Zahn et al., 2014). Together, these results suggest that the E5 can directly modulate downstream repair during both CSR and SHM. The more apparent effect on CSR may be due to the increased dependence of CSR on UNG activity (Rada et al., 2002b). Further work is needed to define whether UNG is directly recruited by the E5, or if other factors may directly associate with E5. In any case, our results provide good evidence that the E5 mediates functionally relevant interactions during antibody diversification.

4.5. Materials and Methods

4.5.1. Animals

Aicda^{-/-} mice (a gift from Dr. T. Honjo, Kyoto University, Japan) in C57BL6/J background were bred at the specific pathogens-free facility of IRCM. All animal work was approved by the IRCM animal protection committee in accordance to the guidelines of the Canadian Council for Animal Care.

4.5.2. DNA constructs

Retroviral vector pMXs human AID, AIDΔE5-ires-GFP, have been described (Zahn et al., 2014). Human AID fusions were assembled as BamHI-AID variant-EcoRI-Linker-HindIII-AID-E58A-XhoI cassettes into pTrcHisA (ThermoFisher) or pMX-ires-GFP (Cell Biolabs). The linker encoded a flexible (SGGGG)₃ peptide. AID variants were generated by PCR amplification with *ad hoc* oligonucleotides or by quick-change site-directed mutagenesis using KOD1 DNA polymerase (Toyobo Inc.). Oligonucleotide sequences are available on request.

4.5.3. Cell culture and transduction

CH12 cells, Plat-E and primary B lymphocytes were cultured in RPMI 1640 media (Wisent) at 37 °C with 5% (vol/vol) CO₂. Media were supplemented with 10% FBS (Wisent), 1% penicillin/streptomycin (Wisent), and 0.1 mM 2-mercaptoethanol (Bioshop). DT40 cells were culture in RPMI 1640 supplemented as above plus 1% chicken serum (Wisent). CH12 cells stably expressing shRNA against AID have been described (Cortizas et al., 2013). Naïve splenic B cells from *Aicda*^{-/-} or *Aicda*^{-/-} *Ung*^{-/-} mice were isolated from total splenocytes depleted with anti-CD43 microbeads in an

autoMACS cell separator (Miltenyi). For retroviral complementation of DT40 or CH12 cells, VSV-G, MLV gag-pol, and pMXs vectors (1:1:4 ratio, 2.5 µg DNA total) were transfected into HEK293 cells using Trans-IT LT-1 (Mirus Bio). Retrovirus for primary cell transduction was produced using Plat-E ecotropic packaging cells (Morita et al., 2000) transfected with pMXs vectors. For infections, 1 ml of HEK293 or 1.5 mL Plat-E supernatant at 48 h post-transfection was used to infect 10^6 B cells in 24-well plates, in the presence of 8 µg/mL polybrene, by spinning at 600 x g for 90 min at 32°C. Medium was replaced 4 h later.

4.5.4. Monitoring SHM and CSR

SHM was measured by fluctuation analysis of IgM-loss in DT40 *Aicda*^{-/-} ΔΨVλ cells (Arakawa et al., 2004) complemented by retroviral transduction with AID or mutants thereof, as described (Zahn et al., 2014). CSR to IgG1 in complemented mouse *Aicda*^{-/-} B cells was induced by adding 5 µg/mL LPS before and after the infection and 20 ng/mL mIL-4 (PeproTech) 4 h post-infection. CSR efficiency in the infected (GFP⁺) subpopulations was measured by flow cytometry using biotinylated anti-IgG1 (BD) followed by anti-biotin-allophycocyanin (Miltenyi Biotech) and propidium iodide to exclude dead cells. CSR in CH12 B cells was induced using CIT [1 µg/mL rat-anti-CD40 (clone 1C10, eBioscience), 10 ng/mL IL-4 and 1 ng/mL TGF-β1 (R&D Systems)]. The proportion of IgA⁺ cells was measured 3 days later by flow cytometry using anti-IgA conjugated with R-phycoerythrin (SouthernBiotech).

4.5.5. Deaminase activity

E. coli mutation assays were performed using the *Δung* BW310 strain, as described (Zahn et al., 2014). AID variants were subcloned as BamHI-XhoI fragments

into pTrcHisA (Invitrogen) to express His-AID fusions after Isopropyl β -D-1-thiogalactopyranoside induction. Mutation frequencies were calculated as the median number of cfu that survived rifampicin selection per 10^9 ampicillin-resistant cells from at least 3 experiments with 5 independent cultures per construct.

4.6. Acknowledgements

We thank L.C. Litzler for critically reading. We thank E. Massicotte and J. Lord with flow cytometry and E. Thivierge with animal care. This work was funded by a grant from the Canadian institutes of health research MOP130535 (to JMDN). SPM was funded by a doctoral training award from the Fonds de Recherche du Québec-Santé. JMDN holds a Canadian Research Chair tier 2. The authors have no conflicting financial interests.

CHAPTER 5: DISCUSSION

During my doctoral work we have used structure-function studies of AID in order to better understand the mechanisms regulating its subcellular transit, and to determine how these can modulate its activity. In this way, we have characterized the molecular mechanism for AID cytoplasmic retention, *via* eEF1A1 binding, defining how this association restricts AID activity in the nucleus (chapter 2). We have also uncovered and describe a licensing step for AID at the chromatin, which promotes efficient activity of AID by tethering it to the transcription elongation machinery (chapter 3). Finally, we describe how the structure of the E5 region, which mediates cytoplasmic retention of AID, is also important for regulating its catalytic activity and promoting CSR, downstream from DNA deamination (chapter 4). Interestingly, all these mechanisms are related by structural elements in and around the E5 region of AID.

5.1. Stepwise regulation of AID function

From my studies and the work of many others, we can posit a model for AID production, regulation and activity, occurring in a stepwise manner (Figure 5.1). Biogenesis of AID, its stabilization and structural maturation, require the functions of co-chaperones DnaJa1 and HSP90 (Orthwein et al., 2010; Orthwein et al., 2012). AID levels, and consequently its activity, are reduced after pharmacological inhibition of HSP90 or DnaJa1, as well as by genetic perturbation of DnaJa1. Conformational maturation of AID during its association with these molecular chaperones is likely important to promote functional catalytic core folding, but may also be necessary for

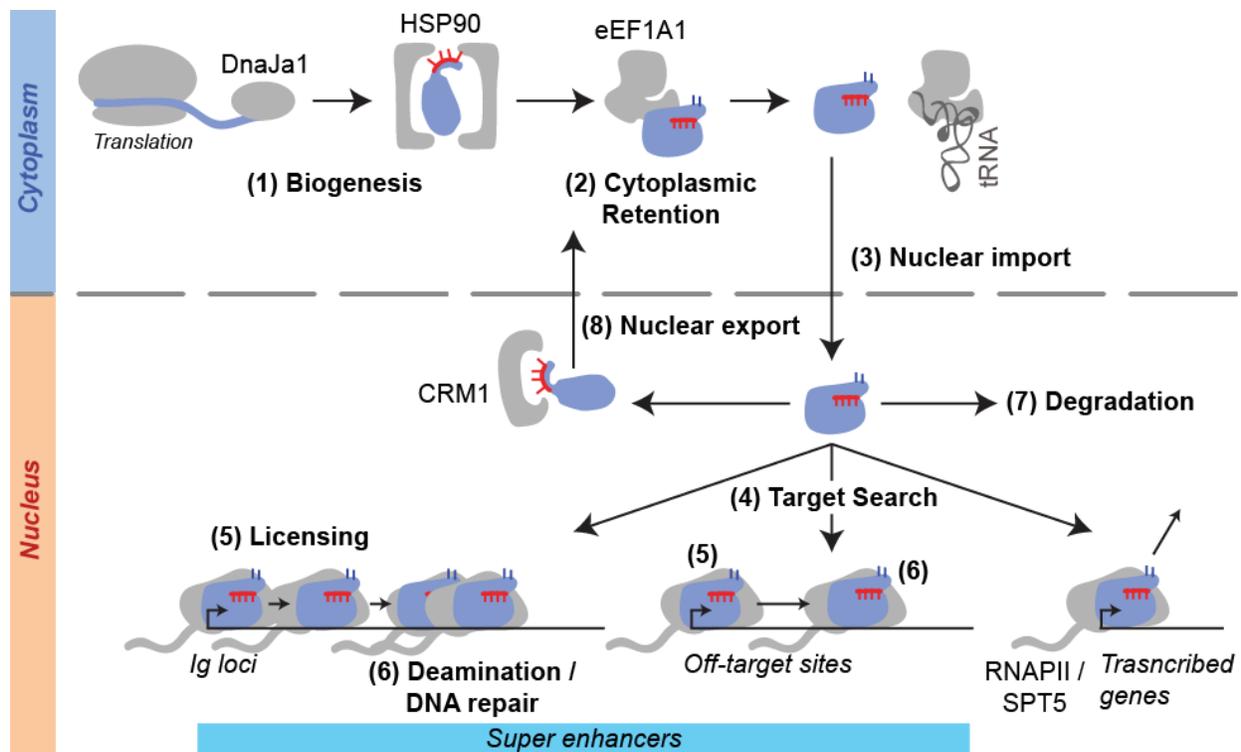


Figure 5.1. Stepwise model of AID activity.

Schematic for the stepwise model of AID activity (see text).

AID to associate with eEF1A1 (Methot et al., 2015). AID mutants that are unable to associate with eEF1A1 have increased association to Hsp90, and vice versa, demonstrating that these associations are mutually exclusive. Furthermore, cytoplasmic retention mutants cannot bind to eEF1A1, and eEF1A1 inhibition leads to AID accumulation in the nucleus and higher CSR activity, indicating that eEF1A1 retains fully folded AID in the cytoplasm (Hasler et al., 2011; Methot et al., 2015). Binding of tRNA to eEF1A1 seems to block AID association, providing a potential mechanism for AID release from eEF1A1 in order to permit active import into the nucleus (Ito et al., 2004; Methot et al., 2015; Patenaude et al., 2009). In the nucleus, AID can associate broadly to the chromatin throughout the genome (chapter 3) and it

has been shown to accumulate at the promoter thousands of genes (Yamane et al., 2011). Our results would suggest that broad association to chromatin is part of a search mechanism, with AID sampling the transcriptional environment at TSSs throughout the genome in order to locate its target genes, within the Ig loci (chapter 3). The genome-wide correlation of AID with Spt5, and the importance of SPT5 for AID association to RNAPII (Pavri et al., 2010), together with our demonstration that Spt5 is not only necessary but sufficient to recruit AID to the chromatin, would suggest that Spt5 causes the accumulation of AID around TSSs. However, this is not sufficient to determine deamination of the downstream genes. AID may only progress beyond the TSS at genes with the appropriate transcriptional environment. Though the exact mechanism for this transition remains to be elucidated, it is likely favoured at genes within super enhancers undergoing convergent and/or divergent transcription, as AID target sites are enriched for such elements (Meng et al., 2014; Pefanis et al., 2015; Qian et al., 2014). AID progression with the elongating transcription machinery requires a licensing step, dependent on an arginine-rich motif we have identified (chapter 3). Transcription can carry licensed AID into the gene body where it can deaminate ssDNA that is likely available during downstream transcriptional stalling and/or premature transcription termination (Kenter, 2012; Kodgire et al., 2013; Rajagopal et al., 2009; Sun et al., 2013a; Wang et al., 2014). Finally, as AID has been shown to associate with high affinity and low K_{off} to DNA (Larijani et al., 2007), we believe it remains bound after its deamination activity. This retention might favour clustered deaminations (Pham et al., 2003) and perhaps influence the downstream repair mechanisms, notably influencing CSR activity (Zahn

et al., 2014) and (chapter 4). This downstream function of AID likely requires interactions of unknown factors with the E5 region and is dependent on a specific conformation of the E5. The ability of AID to remain within the nucleus, and to access the chromatin is further limited by its intrinsic instability in the nucleus (Aoufouchi et al., 2008; Uchimura et al., 2011), and by its nuclear export *via* CRM1 (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004). AID activity may also be limited to the G1-phase of the cell cycle (Petersen et al., 2001; Schrader et al., 2007; Sharbeen et al., 2012). Accordingly, additional mechanisms regulating nuclear accumulation and/or stability seem to either promote G1 or restrict G2/S activity (He et al., 2015; Le and Maizels, 2015; Wang et al., 2017). The exact molecular mechanisms for this predominantly G1 activity are still not fully understood.

5.2. Nuclear export and cytoplasmic retention exclude different pools of AID from the nucleus

As mentioned (section 1.5.2), prior to this work the relationship between nuclear export and cytoplasmic retention was unclear. Studying endogenous, untagged AID, we demonstrated quite clearly that nuclear exclusion is mediated by cooperative functions of both CRM1 and eEF1A1 (Methot et al., 2015). Nonetheless, despite similar increases in nuclear AID levels, CRM1 inhibition with LMB does not increase AID CSR activity, whereas eEF1A1 inhibition with DidB does. This raises the possibility that only retention actually excludes an active pool of AID from the nucleus. The idea that nuclear export and cytoplasmic retention act on functionally

distinct pools of AID is in line with the fact that structural elements required for either mechanism are likely mutually exclusive. According to structures of CRM1 bound to NES regions (Dong et al., 2009), exposing the hydrophobic NES in order to fit into the substrate groove of CRM1 would require an “open” conformation of the E5. On the other hand, molecular modeling and docking suggests that the cytoplasmic retention would require a “closed” conformation of AID, with the NES buried in a hydrophobic groove on the catalytic core in order to expose the Aspartic acids 187 and 188 (Figure 5.2) (Methot et al., 2015).

The lack of increased CSR activity after LMB treatment could also be an artefact of CRM1 inhibition, as CRM1 is implicated in numerous cellular processes (Nguyen et al., 2012). An additional, intriguing possibility, could be that the structural elements provided by the E5 region during CSR (chapter 4), may require the “closed” conformation of AID. Therefore, the pool of AID that is evicted by nuclear export would be CSR deficient, but would perhaps retain SHM and off-target activity, as the NES mutants do (Zahn et al., 2014) and (chapter 4). In any case, additional work is needed to understand the functional relationship between nuclear export and cytoplasmic retention with regards to regulating AID activity.

Another outstanding question is: how is the association of AID and eEF1A regulated? As AID localization may be regulated by cyclin-dependent kinases (He et al., 2015), phosphorylation of eEF1A could potentially regulate cytoplasmic retention. As such, eEF1A contains a number of phosphorylation sites (Negrutskii et al., 2012), with phosphorylation at Serine 300 being particularly interesting, as it may regulate tRNA binding (Lin et al., 2010). It would therefore be interesting to dissect whether this

association is somehow regulated during distinct phases of the cell cycle and/or by post-translational modifications, and if this contributes to restricting AID activity to G1 phase.

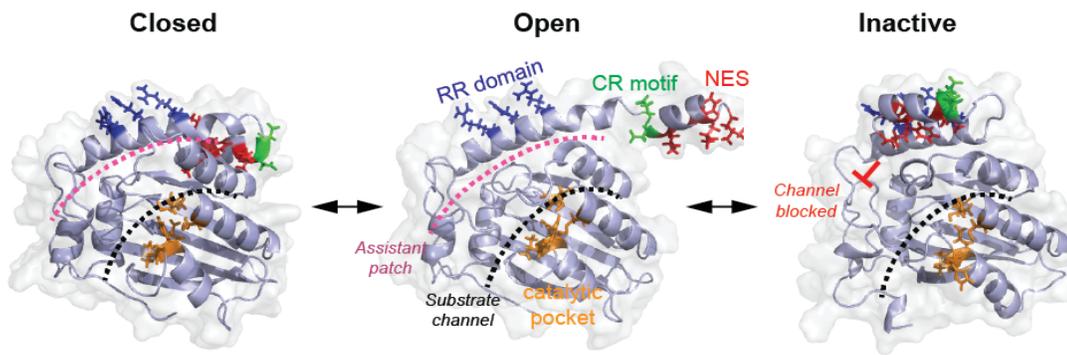


Figure 5.2. Structural modeling of AID regulation by E5.

Structure models for AID, generated by E5 docking on the catalytic core (Methot et al., 2015). Regulatory features studied in this thesis, are highlighted, along with the recently identified substrate channel and assistant patch (Qiao et al., 2017). The closed structure seems to favour cytoplasmic retention, and leaves both channels open, while E5 in the inactive structure would occupy the assistant patch.

5.3. Cytoplasmic retention enforces targeting by limiting AID nuclear accumulation

One of our goals in understanding the mechanism of cytoplasmic retention was to determine its biological relevance. To that end, we determined that cytoplasmic retention, mediated by eEF1A1 binding, not only seems to limit AID functional activity at the Ig locus, but also limits AID off-target activity. This was demonstrated by the increased proportion of IgH-cMyc translocations (Methot et al., 2015) and increased cytotoxicity (chapter 3) in cells expressing AID and treated with DidB. Notably, the rate of IgH-cMyc translocations was ~3 fold higher after DidB treatment, compared to

an increase of only ~1.5 fold for CSR activity (Methot et al., 2015). This discrepancy may be related to the sensitivity of our cellular model, or could indicate that retention limits off-target activity more than Ig targeted activity. In any case, cytoplasmic retention limits AID nuclear access and protects the cells from AID induced translocations and cytotoxicity.

Retention alone is not sufficient to restrict AID activity, as Reg- $\gamma^{-/-}$ B cells have reduced nuclear AID degradation, which leads to increased CSR but also increased IgH-cMyc translocations (Uchimura et al., 2011). Concurrently, promoting nuclear localization of AID using an AID-ER fusion activated with tamoxifen, leads to increased mutation frequency at the S μ region, as well as a number of off-target sites (Wang et al., 2017).

Based on our results with the R-mutants we propose that nuclear exclusion may actually enforce AID targeting. This is founded on the observation that the R-mutants lack on- or off- target activity in the context of full-length AID, yet do not affect the activity of an AID variant without E5, which is fully nuclear. This result suggests that the need for licensing can be bypassed when there is excessive nuclear AID (chapter 3). Based on these results, and our characterization of chromatin bound AID, we propose a model for AID targeting, where limited nuclear AID levels are necessary to enforce recruitment of AID to promoters by SPT5, thereby ensuring that AID is subjected to the licensing step at the transition from paused to elongating RNAPII at target genes (chapter 3).

This model is speculative, and based on the idea that SPT5 levels in the nucleus are sufficiently in excess of AID that there is limited unbound AID, and that

SPT5 only recruits AID to paused RNAPII at promoters. In a situation with an excess of nuclear AID, such as with AID Δ E5, it would then be possible for AID to associate with RNAPII, independent of direct SPT5 recruitment. This recruitment would no longer be restricted to paused, promoter associated RNAPII and AID could be directly recruited to stalled RNAPII within the gene body, bypassing the licensing step at the promoter. If this model were true, then the R-mutants, which seem to lose association to elongating RNAPII in live cells and S_{μ} association by ChIP (Chapter 3), should recover these defects in the context of AID Δ E5.

5.4. The E5 conformation regulates AID function

The structural elements studied during my thesis work have related to the E5 domain, which is unique to AID among the AID/APOBECs (Methot et al., 2015). Indirect approaches are required to define the functional contributions of AID E5 structure and conformation, as it is absent from the APOBECs and crystallization has so far required its removal (Pham et al., 2016; Qiao et al., 2017). Our work would suggest that the E5 region is dynamic, as it must assume at least two independent conformations during nuclear export and cytoplasmic retention. Mobility of the E5 is also in line with the contrasting effects that single point mutations within the E5 can have on catalytic activity (chapter 4). Together these results strongly argue that E5 is flexible and dynamic, features that can impinge on crystallization (Oldfield et al., 2013), making it unlikely that AID containing the E5 region will be crystallized on its own. As with intrinsically unstructured proteins (Dyson and Wright, 2005), it may be

possible to obtain E5 structural information using NMR techniques, or by purifying AID in complex with co-factors that stabilize its structure, such as eEF1A1, SPT5 or HSP90. Additionally, variants of AID such as AID7.3, may stabilize the structural confirmation of the protein, and could potentially permit crystallization of the full-length protein (chapter 4). In the meantime, E5-docking models using the crystal structures are informative, and should be refined using the high-resolution AID crystal structure (Qiao et al., 2017).

As mentioned, we believe that the E5 can adopt at least two distinct structural conformations, “open” or “closed”, and perhaps a third “inactive” conformation (Figure 5.2). The “inactive” conformation relates to the second structural model identified during our E5 docking, which does not seem to permit cytoplasmic retention (Methot et al., 2015). Interestingly, this structure places the E5 within the assistant patch, which seems to be necessary for structured DNA deamination, and would likely block such activity (Qiao et al., 2017). Furthermore, this assistant patch is believed to require the arginine motif, which we found is necessary for AID licensing (Chapter 3), raising two important questions: First, is the targeting defect of the R-mutants due to a defect in structured DNA deamination? Second, does the structured DNA deamination requirement extend beyond the S-regions? With the later being of particular interest to the field (Pucella and Chaudhuri, 2017).

Our data, especially with the R-mutants in the context of AID Δ E5 would argue that R174 and R178 are not essential to the assistant patch since when mutated there is no effect on AID Δ E5 activity. On the other hand, mutation of R171 does affect AID Δ E5 activity significantly, suggesting that it may be essential to assistant patch

function. In accordance, *in vitro* structured DNA deamination was almost completely abrogated by mutation of R171, while R174 and R178 mutations had intermediate effects (Qiao et al., 2017). Thus, structured deamination, involving R171 may be important for SHM and off-target activity, but R174 and R178 are not essential for such function, and licensing likely requires additional interactions.

5.5. Mechanistic link between CR and licensing

Another interesting observation is that the mutation D188A in AID, which almost completely disrupts cytoplasmic retention (Patenaude et al., 2009), mimics the functional defects we observed with the RR domain mutants (chapter 3 and 4). Mutating R174 also seems to reduce cytoplasmic retention to some extent, further supporting the possibility that retention and licensing are functionally linked (chapter 3). Nonetheless, confirming a role for D188 in licensing will require testing if D188A retains chromatin association and TSS recruitment, but loses gene body binding, similar to the R-mutants. Furthermore, as R171 and R178 do not seem to affect AID shuttling, retention and licensing are not completely overlapping.

Nonetheless, we can envision at least three possibilities for how retention and licensing could be linked (Figure 5.3 – see figure legend for details). The first possibility: AID bound to eEF1A1 in the cytoplasm becomes associated with a positive regulator of its activity (factor X). The second possibility: AID structural determinants for cytoplasmic retention and licensing are similar, but mechanistically unrelated. The third possibility: AID bound to eEF1A1 is blocked from associating to a

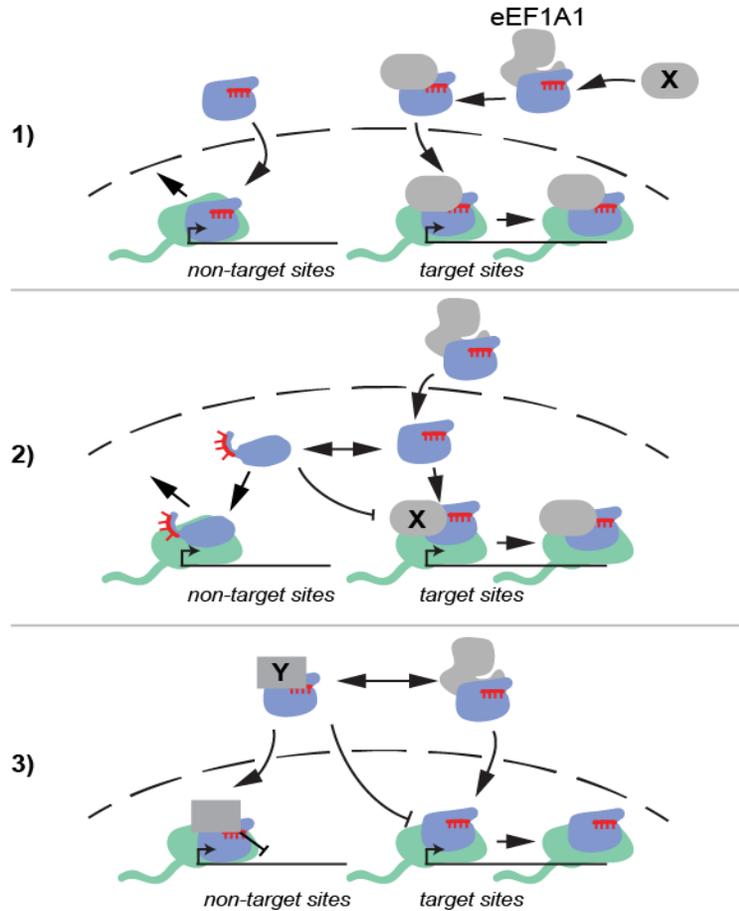


Figure 5.3. Mechanisms linking cytoplasmic retention and licensing.

Schemes for 3 potential mechanisms linking AID cytoplasmic retention and chromatin licensing are presented. **1)** Cytoplasmic retention promotes AID association to a positive regulator of activity. AID and factor X associate via eEF1A1, and shuttle together into the nucleus. Factor X has specificity for AID target genes, perhaps via association to a transcription factor, and therefore recruits AID to promoters that license association with elongating RNAPII. At off target sites, AID would dissociate due to the absence of factor X. **2)** Retention and licensing require a similar structural conformation but separate mechanisms, with distinct factors for each. Thus, eEF1A1 binding and binding to a licensing factor would require motifs but otherwise be unrelated. In this case, factor X would only be found at AID target sites, and likely facilitates licensing of the “closed” AID. The “closed” structure, and/or binding to factor X would also depend on the arginine motif. **3)** Retention prevents AID association with a negative regulator of licensing. This would ensure that only the pool of properly folded AID, released from eEF1A1 is able to deaminate the genome. Recruitment of AID with factor Y would inhibit licensing either directly or by limiting transcription elongation. Factor Y recruitment to target sites could be regulated by the transcription environment.

negative regulator of AID activity (factor Y), with only the small pool of AID released from eEF1A1 being competent to undergo licensing.

From our BioID results, the R-mutants lose association to a number of cytoplasmic proteins, with many involved in signalling pathways or translation initiation (Chapter 3), some of these could be candidates for factor X. For example, dishevelled 3 (DVL3), a Wnt/ β -catenin signalling scaffold re-localizes to the nucleus upon activation and seems to exert its function in the nucleus (Itoh et al., 2005). Another, Dicer, which processes dsRNA during formation of RNAi, has also been found to process dsRNA generated by convergent transcription (White et al., 2014), which is associated with AID activity (Meng et al., 2014). Our BioID also revealed a handful of proteins that associate more to the R-mutants, with a clear enrichment for certain chaperones and co-chaperones, including HSP90 (data not shown). Intriguingly, HSP90 has been shown to associate at promoter regions and to stabilize paused RNAPII, with Hsp90 inhibition increasing transcription at HSP90 target genes (Sawarkar et al., 2012). Therefore, HSP90 could be a potential factor Y candidate.

Discriminating between these possibilities will require further studies, starting with a mechanistic understanding of how AID licensing is established. The BioID candidates could be quickly tested using knockdown approaches to screen for effects on AID activity. If D188 is implicated in mediating licensing, then it should also lose associations, and repeating a BioID screen comparing D188A and the R-mutants compared to wt AID could help to refine the list of candidates.

5.6. The transcription environment favouring licensing

The main unresolved question regarding AID targeting is the molecular mechanism for how the R-mutants become disconnected from transcription elongation. Our evidence for defective association to transcription elongation, is that by BioID the R-mutants have reduced association with SPT6, and NAP1L4, two histone chaperones that promote transcription elongation (Del Rosario and Pemberton, 2008; Kaplan et al., 2003), along with the fact that they are enriched at the promoter, but absent in the gene body at the $S\mu$. As both SPT6 and Nap1L4 are histone chaperones, the licensing defect could relate to defective histone associations; however, we find equal enrichment for histones with the R-mutants by BioID.

One important caveat to our BioID data is that we did not detect a number of known AID interactions related to targeting and transcription elongation, including SPT5 (Pavri et al., 2010), the FACT complex, comprising SPT16 and SSRP1 (Aida et al., 2013), or the multi-subunit PAF complex (Willmann et al., 2012). This is unlikely to be a direct effect of the BirA* fusion on AID interactions, since the fusion protein was considerably active for CSR. Thus a lack of biotinylation may suggest that some of these interactions are indirect, perhaps as part of larger complexes, or that the BirA* is unable to access these factors due to biophysical constraints. Using a longer linker can extend the BirA* range (Kim et al., 2016), and could potentially help to better probe the transcriptional environment in the future.

Despite the correlation of SPT5 recruitment with AID association and activity at the Ig locus (Maul et al., 2014), we believe the R-mutants can still associate with SPT5, as evidenced by traditional colIPs and chromatin tethering experiments. Other known

factors still require further study to determine if they are involved in licensing. The PAF complex is particularly interesting, as it seems to have a role in recruiting AID to both the IgV and S-regions (Willmann et al., 2012). The PAF complex is a scaffold that coordinates histone modifications during transcription elongation (Jaehning, 2010), and is recruited by SPT5 phosphorylation in yeast (Liu et al., 2009). In mammals, this occurs during the transition from paused to elongating transcription (Yamada et al., 2006), which is the step at which we believe licensing occurs.

To further our understanding of licensing, and its relevance genome-wide, it will be important to determine whether the R-mutants are depleted from gene-bodies at the IgV region and at target sites outside the Ig locus. If this is the case, then comparing wt AID and R-mutants genome-wide could help to identify additional AID target sites, based on depletion of the R-mutants. By documenting the target sites of AID, we may be able to identify additional transcriptional or epigenetic features that define the sites, and that relate to AID activity. This information could help to glean the mechanism of licensing. Additionally, as yeast do not undergo transcription pausing (Adelman and Lis, 2012), and AID can mutate yeast genes (Gomez-Gonzalez and Aguilera, 2007; Taylor et al., 2013), it would also be interesting to test the R-mutants activity in yeast. If the R-mutants can mutate yeast genes, then they would likely have a fundamental defect in association to transcription elongation, whereas mutagenic activity would suggest that the licensing is actually a checkpoint, somehow established during transcription pausing.

An additional, unintended use for the R-mutants, could be in the design of CRISPR/Cas9 based deaminases (see section 1.4.5). As the R-mutants are defective

in targeting on their own, but retain mutagenic activity if tethered to genes, fusions to enzymatically deficient Cas9 could generate targeted deaminases that lack off-target activity, a current limitation of these systems (Yang et al., 2016).

5.7. Pharmacological manipulation of AID

Due to the nefarious secondary effects of AID activity, especially in cancer cells, there is interest in identifying drugs that can inhibit its activity. Importantly, aside from a compromised humoral immune response, and increased propensity for autoimmunity, AID deficient patients lack additional serious complications (Quartier et al., 2004), suggesting that blocking AID activity, especially short term, should be relatively safe. No direct inhibitors of AID have been identified to date; however, the recent crystal structures and dynamic modelling of the catalytic site (King et al., 2015; Pham et al., 2016; Qiao et al.), could help drive the identification of specific inhibitors using structure-guided docking (Lionta et al., 2014).

In the absence of direct inhibitors, HSP90 inhibitors can block AID biogenesis and reduce AID levels in B cells. Reduced AID levels result in reduced antibody diversification, as well as off-target activity measured at a BCR-ABL1 transgene (Orthwein et al., 2010). HSP90 inhibition can also limit AID activity *in vivo* (Montamat-Sicotte et al., 2015), demonstrating a therapeutic potential, and demonstrating the potential for translating *in vitro* discoveries to *in vivo* models. Co-administration HSP90 inhibitors with chemotherapeutics could be an interesting strategy to limit resistance in AID-expressing tumours.

Another interesting consideration for AID manipulation is that its activity is limiting for antibody diversification and can be boosted by increased expression or enzymatic activity (Orthwein et al., 2012; Robbiani et al., 2009; Uchimura et al., 2011; Wang et al., 2009b). Since we observe increased AID activity after inhibiting cytoplasmic retention, using DidB or similar drugs could provide a pharmacological method to boost AID activity *in vivo* (Methot et al., 2015) and (chapter 3). This could have therapeutic use to boost the immune response, or promote cytotoxicity in cancer cells.

One potential use for boosting the immune response could be in elderly patients, due to their reduced AID expression and humoral immune responses (Frasca et al., 2010; Khurana et al., 2012). Therefore increasing AID activity during vaccination could theoretically improve the humoral immune response and increase vaccination efficiency. Since AID expression in such patients is low, the potential side effects, and cytotoxicity associated with off-target AID activity should be minimized. As *Aicda*^{+/-} mice have a reduced antibody responses (Sernández et al., 2008; Takizawa et al., 2008), they could provide a model to test whether boosting AID activity can be beneficial when its expression is reduced.

With regards to cytotoxicity, it has been demonstrated that using an inhibitor of HR, which limits the repair of AID induced damage, can cause cell death in cancer cells that express AID (Lamont et al., 2013). AID is also able to deaminate DNA at telomeres, but UNG protects the telomeres from deleterious damage thereby protecting B cells (Cortizas, JEM, 2016). This could lend itself to synergistic effects if we could boost AID activity with DidB, while simultaneously inhibiting UNG, or HR, providing a highly specific mechanism to induce cell death in AID expressing cancer cells.

A number of clinical trials using DidB have been done against different types of cancer, but in most cases it has been highly toxic and ineffective (Hochster et al., 1998; Meng et al., 1998; Mittelman et al., 1999; Taylor et al., 1998). Interestingly, in lymphoma patients DidB does show some potential benefits (Kucuk et al., 2000), suggesting lymphoma cells are particularly sensitive, perhaps due to an unaccounted for effect of AID toxicity. The dose of DidB that can increase AID activity *in vitro*, 0.5-1 nM, is low compared to the doses that efficiently block translation in mouse leukemia cells, 10-100 nM (Robert et al., 2009). Nonetheless, translation inhibition using DidB has been used as a tolerable therapeutic strategy in mice (Robert et al., 2009). This suggests that tolerable doses of DidB could potentially inhibit cytoplasmic retention and boost antibody diversification or AID cytotoxicity. It remains to be tested if the observed effects of DidB from cell culture treatments (Methot et al., 2015) and (Chapter 3) are reproducible *in vivo*.

CHAPTER 6: CONCLUDING REMARKS

Here, we present structure-function studies on AID regulation, notably focussed on the C-terminal E5 region.

We identify the molecular mechanism of AID cytoplasmic retention, which is mediated by the translation factor eEF1A1 (Chapter 2) and dependent on a structural conformation related to the E5 region. Pharmacological inhibition of eEF1A1, promotes AID accumulation in the nucleus and increased AID activity, on- and off-target, demonstrating formally that nuclear exclusion limits AID activity. We also present work pertaining to the targeting of AID in the nucleus, by identifying an arginine motif in the α -helix 6 of AID, adjacent to the E5, which is necessary for AID activity in B cells. The arginine motif seems to nonetheless be dispensable for AID association to promoters along with the transcription machinery, highlighting additional requirements for AID activity. Though we have not yet elucidated the molecular mechanism, we propose that this motif is necessary for a licensing step that promotes productive recruitment of AID at target genes. Finally, by studying the functional requirements of the E5 region of AID, we further define its relevance to AID catalytic and functional activity, in particular highlighting structural requirements. This work has become more relevant in light of recently solved crystal structures for AID (Pham et al., 2016; Qiao et al., 2017), as they were solved without the E5 region.

All together this work highlights the critical functional role of the E5 region at multiple levels of AID regulation.

CHAPTER 7: BIBLIOGRAPHY

- Adelman, K., and J.T. Lis. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics* 13:720-731.
- Aida, M., N. Hamad, A. Stanlie, N.A. Begum, and T. Honjo. 2013. Accumulation of the FACT complex, as well as histone H3.3, serves as a target marker for somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America* 110:7784-7789.
- Alber, T., J.A. Bell, D.P. Sun, H. Nicholson, J.A. Wozniak, S. Cook, and B.W. Matthews. 1988. Replacements of Pro86 in phage T4 lysozyme extend an alpha-helix but do not alter protein stability. *Science* 239:631-635.
- Alt, F.W., T.K. Blackwell, and G.D. Yancopoulos. 1987. Development of the primary antibody repertoire. *Science* 238:1079-1087.
- Andersen, G.R., P. Nissen, and J. Nyborg. 2003. Elongation factors in protein biosynthesis. *Trends in biochemical sciences* 28:434-441.
- Andrulis, E.D., E. Guzman, P. Doring, J. Werner, and J.T. Lis. 2000. High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes Dev* 14:2635-2649.
- Aoufouchi, S., A. Faili, C. Zober, O. D'Orlando, S. Weller, J.-C. Weill, and C.-A. Reynaud. 2008. Proteasomal degradation restricts the nuclear lifespan of AID. *The Journal of experimental medicine* 205:1357-1368.
- Apostaei, A.I., and J.R. Trabalka. 2010. Review, Synthesis, and Application of Information on the Human Lymphatic System to Radiation Dosimetry for Chronic Lymphocytic Leukemia. *SENES Oak Ridge, Inc; Oak Ridge, TN*
- Arakawa, H., and J.M. Buerstedde. 2006. Dt40 gene disruptions: a how-to for the design and the construction of targeting vectors. *Subcell Biochem* 40:1-9.
- Arakawa, H., J. Hauschild, and J.M. Buerstedde. 2002. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* 295:1301-1306.
- Arakawa, H., H. Saribasak, and J.M. Buerstedde. 2004. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biol* 2:E179.
- Barbie, V., and M.P. Lefranc. 1998. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Exp Clin Immunogenet* 15:171-183.
- Barreto, V., B.R. Reina-San-Martin, A.R. Ramiro, K.M. McBride, and M.C. Nussenzweig. 2003. C-Terminal Deletion of AID Uncouples Class Switch Recombination from Somatic Hypermutation and Gene Conversion. *Molecular cell* 12:501-508.
- Barreto, V.M., Q. Pan-Hammarström, Y. Zhao, L. Hammarström, Z. Misulovin, and M.C. Nussenzweig. 2005. AID from bony fish catalyzes class switch recombination. *The Journal of experimental medicine* 202:733-738.

- Basso, K., and R. Dalla-Favera. 2010. BCL6: master regulator of the germinal center reaction and key oncogene in B cell lymphomagenesis. *Adv Immunol* 105:193-210.
- Basso, K., C. Schneider, Q. Shen, A.B. Holmes, M. Setty, C. Leslie, and R. Dalla-Favera. 2012. BCL6 positively regulates AID and germinal center gene expression via repression of miR-155. *J Exp Med* 209:2455-2465.
- Basu, U., J. Chaudhuri, C. Alpert, S. Dutt, S. Ranganath, G. Li, J.P. Schrum, J.P. Manis, and F.W. Alt. 2005. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* 438:508-511.
- Basu, U., F.-L. Meng, C. Keim, V. Grinstein, E. Pefanis, J. Eccleston, T. Zhang, D. Myers, C.R. Wasserman, D.R. Wesemann, K. Januszyk, R.I. Gregory, H. Deng, C.D. Lima, and F.W. Alt. 2011. The RNA Exosome Targets the AID Cytidine Deaminase to Both Strands of Transcribed Duplex DNA Substrates. *Cell* 144:353-363.
- Basu, U., Y. Wang, and F.W. Alt. 2008. Evolution of Phosphorylation-Dependent Regulation of Activation-Induced Cytidine Deaminase. *Molecular cell* 32:285-291.
- Beale, R.C.L., S.K. Petersen-Mahrt, I.N. Watt, R.S. Harris, C. Rada, and M.S. Neuberger. 2004. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *Journal of molecular biology* 337:585-596.
- Begum, N.A., A. Stanlie, M. Nakata, H. Akiyama, and T. Honjo. 2012. The histone chaperone Spt6 is required for activation-induced cytidine deaminase target determination through H3K4me3 regulation. *The Journal of Biological Chemistry* 287:32415-32429.
- Bennett, R.P., E. Diner, M.P. Sowden, J.A. Lees, J.E. Wedekind, and H.C. Smith. 2006. APOBEC-1 and AID are nucleo-cytoplasmic trafficking proteins but APOBEC3G cannot traffic. *Biochemical and biophysical research communications* 350:214-219.
- Bensaude, O. 2011. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* 2:103-108.
- Benson, M.J., L.D. Erickson, M.W. Gleeson, and R.J. Noelle. 2007. Affinity of antigen encounter and other early B-cell signals determine B-cell fate. *Curr Opin Immunol* 19:275-280.
- Betz, A.G., C. Rada, R. Pannell, C. Milstein, and M.S. Neuberger. 1993. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. *Proc Natl Acad Sci U S A* 90:2385-2388.
- Bhutani, N., J.J. Brady, M. Damian, A. Sacco, S.Y. Corbel, and H.M. Blau. 2010. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463:1042-1047.
- Blagodatski, A., V. Batrak, S. Schmidl, U. Schoetz, R.B. Caldwell, H. Arakawa, and J.-M. Buerstedde. 2009. A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation. *PLoS Genetics* 5:e1000332.

- Boehm, T., N. McCurley, Y. Sutoh, M. Schorpp, M. Kasahara, and M.D. Cooper. 2012. VLR-based adaptive immunity. *Annu Rev Immunol* 30:203-220.
- Boehm, T., and J.B. Swann. 2014. Origin and evolution of adaptive immunity. *Annu Rev Anim Biosci* 2:259-283.
- Bohn, M.F., S.M. Shandilya, J.S. Albin, T. Kouno, B.D. Anderson, R.M. McDougale, M.A. Carpenter, A. Rathore, L. Evans, A.N. Davis, J. Zhang, Y. Lu, M. Somasundaran, H. Matsuo, R.S. Harris, and C.A. Schiffer. 2013. Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure* 21:1042-1050.
- Bombardieri, M., F. Barone, F. Humby, S. Kelly, M. McGurk, P. Morgan, S. Challacombe, S. De Vita, G. Valesini, J. Spencer, and C. Pitzalis. 2007. Activation-induced cytidine deaminase expression in follicular dendritic cell networks and interfollicular large B cells supports functionality of ectopic lymphoid neogenesis in autoimmune sialoadenitis and MALT lymphoma in Sjogren's syndrome. *Journal of immunology* 179:4929-4938.
- Bothmer, A., D.F. Robbiani, N. Feldhahn, A. Gazumyan, A. Nussenzweig, and M.C. Nussenzweig. 2010. 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J Exp Med* 207:855-865.
- Bransteitter, R., P. Pham, M.D. Scharff, and M.F. Goodman. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences of the United States of America* 100:4102-4107.
- Brar, S.S., M. Watson, and M. Diaz. 2004. Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *The Journal of Biological Chemistry* 279:26395-26401.
- Buerstedde, J.-M., J. Alinikula, H. Arakawa, J.J. McDonald, and D.G. Schatz. 2014. Targeting Of Somatic Hypermutation By immunoglobulin Enhancer And Enhancer-Like Sequences. *PLoS Biology* 12:e1001831.
- Cantaert, T., J.-N. Schickel, J.M. Bannock, Y.-S. Ng, C. Massad, T. Oe, R. Wu, A. Lavoie, J.E. Walter, L.D. Notarangelo, W. Al-Herz, S.S. Kilic, H.D. Ochs, S. Nonoyama, A.H. Durandy, and E. Meffre. 2015. Activation-Induced Cytidine Deaminase Expression in Human B Cell Precursors Is Essential for Central B Cell Tolerance. *Immunity* 43:884-895.
- Casellas, R., U. Basu, W.T. Yewdell, J. Chaudhuri, D.F. Robbiani, and J.M. Di Noia. 2016. Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. *Nat Rev Immunol* 16:164-176.
- Casellas, R., A. Nussenzweig, R. Wuerffel, R. Pelanda, A. Reichlin, H. Suh, X.F. Qin, E. Besmer, A. Kenter, K. Rajewsky, and M.C. Nussenzweig. 1998. Ku80 is required for immunoglobulin isotype switching. *EMBO J* 17:2404-2411.

- Cattoretti, G., M. Büttner, R. Shaknovich, E. Kremmer, B. Alobeid, and G. Niedobitek. 2006. Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood* 107:3967-3975.
- Chandra, A., F. van Maldegem, S. Andrews, M.S. Neuberger, and C. Rada. 2013. Deficiency in spliceosome-associated factor CTNNB1 does not affect ongoing cell cycling but delays exit from quiescence and results in embryonic lethality in mice. *Cell cycle (Georgetown, Tex.)* 12:732-742.
- Chatterji, M., S. Unniraman, K.M. McBride, and D.G. Schatz. 2007. Role of Activation-Induced Deaminase Protein Kinase A Phosphorylation Sites in Ig Gene Conversion and Somatic Hypermutation. *Journal of immunology (Baltimore, Md. : 1950)* 179:5274-5280.
- Chaudhuri, J., C. Khuong, and F.W. Alt. 2004. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* 430:992-998.
- Chaudhuri, J., M. Tian, C. Khuong, K. Chua, E. Pinaud, and F.W. Alt. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 422:726-730.
- Chen, J., Z. Zhang, L. Li, B.C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, and Z. Liu. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* 156:1274-1285.
- Chen, K.M., E. Harjes, P.J. Gross, A. Fahmy, Y. Lu, K. Shindo, R.S. Harris, and H. Matsuo. 2008. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* 452:116-119.
- Cheng, H.-L., B.Q. Vuong, U. Basu, A. Franklin, B. Schwer, J. Astarita, R.T. Phan, A. Datta, J. Manis, F.W. Alt, and J. Chaudhuri. 2009. Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. *Proceedings of the National Academy of Sciences of the United States of America* 106:2717-2722.
- Chester, A., A. Somasekaram, M. Tzimina, A. Jarmuz, J. Gisbourne, R. O'Keefe, J. Scott, and N. Navaratnam. 2003. The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *EMBO J* 22:3971-3982.
- Conticello, S.G., K. Ganesh, K. Xue, M. Lu, C. Rada, and M.S. Neuberger. 2008. Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNB1. *Molecular cell* 31:474-484.
- Conticello, S.G., C.J. Thomas, S.K. Petersen-Mahrt, and M.S. Neuberger. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 22:367-377.
- Cortizas, E.M., A. Zahn, M.E. Hajjar, A.M. Patenaude, J.M. Di Noia, and R.E. Verdun. 2013. Alternative End-Joining and Classical Nonhomologous End-Joining

- Pathways Repair Different Types of Double-Strand Breaks during Class-Switch Recombination. *Journal of immunology* 191:5751-5763.
- Cortizas, E.M., A. Zahn, S. Safavi, J.A. Reed, F. Vega, J.M. Di Noia, and R.E. Verdun. 2016. UNG protects B cells from AID-induced telomere loss. *The Journal of experimental medicine* 213:2459-2472.
- Couzens, A.L., J.D. Knight, M.J. Kean, G. Teo, A. Weiss, W.H. Dunham, Z.Y. Lin, R.D. Bagshaw, F. Sicheri, T. Pawson, J.L. Wrana, H. Choi, and A.C. Gingras. 2013. Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Science signaling* 6:rs15.
- Crews, C.M., J.L. Collins, W.S. Lane, M.L. Snapper, and S.L. Schreiber. 1994. GTP-dependent binding of the antiproliferative agent didemnin to elongation factor 1 alpha. *The Journal of Biological Chemistry* 269:15411-15414.
- Crews, C.M., W.S. Lane, and S.L. Schreiber. 1996. Didemnin binds to the protein palmitoyl thioesterase responsible for infantile neuronal ceroid lipofuscinosis. *Proceedings of the National Academy of Sciences of the United States of America* 93:4316-4319.
- Crouch, E.E., Z. Li, M. Takizawa, S. Fichtner-Feigl, P. Gourzi, C. Montano, L. Feigenbaum, P. Wilson, S. Janz, F.N. Papavasiliou, and R. Casellas. 2007. Regulation of AID expression in the immune response. *The Journal of experimental medicine* 204:1145-1156.
- Das, R., and D. Baker. 2008. Macromolecular modeling with rosetta. *Annu Rev Biochem* 77:363-382.
- Davies, D.R., and H. Metzger. 1983. Structural basis of antibody function. *Annu Rev Immunol* 1:87-117.
- de Yébenes, V.G., L. Belver, D.G. Pisano, S. González, A. Villasante, C. Croce, L. He, and A.R. Ramiro. 2008. miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *The Journal of experimental medicine* 205:2199-2206.
- de Yébenes, V.G., and A.R. Ramiro. 2006. Activation-induced deaminase: light and dark sides. *Trends Mol Med* 12:432-439.
- Del Rosario, B.C., and L.F. Pemberton. 2008. Nap1 links transcription elongation, chromatin assembly, and messenger RNP complex biogenesis. *Mol Cell Biol* 28:2113-2124.
- Delbos, F., S. Aoufouchi, A. Faili, J.C. Weill, and C.A. Reynaud. 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., A. De Smet, A. Faili, S. Aoufouchi, J.C. Weill, and C.A. Reynaud. 2005. Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. *J Exp Med* 201:1191-1196.
- Di Niro, R., S.J. Lee, J.A. Vander Heiden, R.A. Elsner, N. Trivedi, J.M. Bannock, N.T. Gupta, S.H. Kleinstein, F. Vigneault, T.J. Gilbert, E. Meffre, S.J. McSorley, and

- M.J. Shlomchik. 2015. Salmonella Infection Drives Promiscuous B Cell Activation Followed by Extrafollicular Affinity Maturation. *Immunity* 43:120-131.
- Di Noia, J., and M.S. Neuberger. 2002. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* 419:43-48.
- Di Noia, J.M., and M.S. Neuberger. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* 76:1-22.
- Dickerson, S.K., E. Market, E. Besmer, and F.N. Papavasiliou. 2003. AID mediates hypermutation by deaminating single stranded DNA. *The Journal of experimental medicine* 197:1291-1296.
- Doi, T., L. Kato, S. Ito, R. Shinkura, M. Wei, H. Nagaoka, J. Wang, and T. Honjo. 2009. The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* 106:2758-2763.
- Dominguez, P.M., M. Teater, N. Chambwe, M. Kormaksson, D. Redmond, J. Ishii, B.Q. Vuong, J. Chaudhuri, A.M. Melnick, A. Vasanthakumar, L.A. Godley, F.N. Papavasiliou, O. Elemento, and R. Shakhovich. 2015. DNA Methylation Dynamics of Germinal Center B Cells Are Mediated by AID. *Cell reports* 12:2086-2098.
- Dong, J., R.A. Panchakshari, T. Zhang, Y. Zhang, J. Hu, S.A. Volpi, R.M. Meyers, Y.-J. Ho, Z. Du, D.F. Robbani, F. Meng, M. Gostissa, M.C. Nussenzweig, J.P. Manis, and F.W. Alt. 2015. Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. *Nature* 525:134-139.
- Dong, X., A. Biswas, K.E. Süel, L.K. Jackson, R. Martinez, H. Gu, and Y.M. Chook. 2009. Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* 458:1136-1141.
- Dorsett, Y., K.M. McBride, M. Jankovic, A. Gazumyan, T.-H. Thai, D.F. Robbani, M. Di Virgilio, B.R. Reina-San-Martin, G. Heidkamp, T.A. Schwickert, T.R. Eisenreich, K. Rajewsky, and M.C. Nussenzweig. 2008. MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity* 28:630-638.
- Dunnick, W.A., J.T. Collins, J. Shi, G. Westfield, C. Fontaine, P. Hakimpour, and F.N. Papavasiliou. 2009. Switch recombination and somatic hypermutation are controlled by the heavy chain 3' enhancer region. *J Exp Med* 206:2613-2623.
- Durandy, A., C. Hivroz, F. Mazerolles, C. Schiff, F. Bernard, E. Jouanguy, P. Revy, J.P. DiSanto, J.F. Gauchat, J.Y. Bonnefoy, J.L. Casanova, and A. Fischer. 1997. Abnormal CD40-mediated activation pathway in B lymphocytes from patients with hyper-IgM syndrome and normal CD40 ligand expression. *Journal of immunology* 158:2576-2584.

- Dyson, H.J., and P.E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197-208.
- Early, P., H. Huang, M. Davis, K. Calame, and L. Hood. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* 19:981-992.
- Eccleston, J., C. Yan, K. Yuan, F.W. Alt, and E. Selsing. 2011. Mismatch repair proteins MSH2, MLH1, and EXO1 are important for class-switch recombination events occurring in B cells that lack nonhomologous end joining. *Journal of immunology* 186:2336-2343.
- Ehrenstein, M.R., and M.S. Neuberger. 1999. Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *EMBO J* 18:3484-3490.
- Ellestad, K.K., and B.G. Magor. 2005. Evolution of transcriptional enhancers in the immunoglobulin heavy-chain gene: functional characteristics of the zebrafish Emu3' enhancer. *Immunogenetics* 57:129-139.
- Ellyard, J.I., A.S. Benk, B. Taylor, C. Rada, and M.S. Neuberger. 2011. The dependence of Ig class-switching on the nuclear export sequence of AID likely reflects interaction with factors additional to Crm1 exportin. *European journal of immunology* 41:485-490.
- Endo, Y., H. Marusawa, K. Kinoshita, T. Morisawa, T. Sakurai, I.m. Okazaki, K. Watashi, K. Shimotohno, T. Honjo, and T. Chiba. 2007. Expression of activation-induced cytidine deaminase in human hepatocytes via NF- κ B signaling. *Oncogene* 26:5587-5595.
- Endo, Y., H. Marusawa, T. Kou, H. Nakase, S. Fujii, T. Fujimori, K. Kinoshita, T. Honjo, and T. Chiba. 2008. Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. *Gastroenterology* 135:889-898- 898.e881-883.
- Epeldegui, M., Y.P. Hung, A. McQuay, R.F. Ambinder, and O. Martínez-Maza. 2007. Infection of human B cells with Epstein-Barr virus results in the expression of somatic hypermutation-inducing molecules and in the accrual of oncogene mutations. *Molecular immunology* 44:934-942.
- Feldhahn, N., N. Henke, K. Melchior, C. Duy, B.N. Soh, F. Klein, G. von Levetzow, B. Giebel, A. Li, W.-K. Hofmann, H. Jumaa, and M. MUschen. 2007. Activation-induced cytidine deaminase acts as a mutator in BCR-ABL1-transformed acute lymphoblastic leukemia cells. *The Journal of experimental medicine* 204:1157-1166.
- Feldman, S., R. Wuerffel, I. Achour, L. Wang, P.B. Carpenter, and A.L. Kenter. 2017. 53BP1 Contributes to Igh Locus Chromatin Topology during Class Switch Recombination. *Journal of immunology* 198:2434-2444.

- Frasca, D., A. Diaz, M. Romero, A.M. Landin, M. Phillips, S.C. Lechner, J.G. Ryan, and B.B. Blomberg. 2010. Intrinsic defects in B cell response to seasonal influenza vaccination in elderly humans. *Vaccine* 28:8077-8084.
- Frasca, D., E. Van der Put, R.L. Riley, and B.B. Blomberg. 2004. Reduced Ig class switch in aged mice correlates with decreased E47 and activation-induced cytidine deaminase. *Journal of immunology* 172:2155-2162.
- Fritz, E.L., B.R. Rosenberg, K. Lay, A. Mihailović, T. Tuschl, and F.N. Papavasiliou. 2013. A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. *Nature Immunology*
- Ganesh, K., S. Adam, B. Taylor, P. Simpson, C. Rada, and M.S. Neuberger. 2011. CTNNB1 is a novel nuclear localization sequence-binding protein that recognizes RNA-splicing factors CDC5L and Prp31. *The Journal of Biological Chemistry* 286:17091-17102.
- Gauwerky, C.E., K. Huebner, M. Isobe, P.C. Nowell, and C.M. Croce. 1989. Activation of MYC in a masked t(8;17) translocation results in an aggressive B-cell leukemia. *Proc Natl Acad Sci U S A* 86:8867-8871.
- Gazumyan, A., K. Timachova, G. Yuen, E. Siden, M. Di Virgilio, E.M. Woo, B.T. Chait, B.R. Reina-San-Martin, M.C. Nussenzweig, and K.M. McBride. 2011. Amino-terminal phosphorylation of activation-induced cytidine deaminase suppresses c-myc/IgH translocation. *Molecular and cellular biology* 31:442-449.
- Geisberger, R., C. Rada, and M.S. Neuberger. 2009. The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence. *Proceedings of the National Academy of Sciences of the United States of America* 106:6736-6741.
- Gilfillan, S., A. Dierich, M. Lemeur, C. Benoist, and D. Mathis. 1993. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* 261:1175-1178.
- Gold, M.R. 2002. To make antibodies or not: signaling by the B-cell antigen receptor. *Trends Pharmacol Sci* 23:316-324.
- Gomez-Gonzalez, B., and A. Aguilera. 2007. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc Natl Acad Sci U S A* 104:8409-8414.
- Greeve, J., A. Philipsen, K. Krause, W. Klapper, K. Heidorn, B.E. Castle, J. Janda, K.B. Marcu, and R. Parwaresch. 2003. Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas. *Blood* 101:3574-3580.
- Gruber, T.A., M.S. Chang, R. Sposto, and M. Muschen. 2010. Activation-Induced Cytidine Deaminase Accelerates Clonal Evolution in BCR-ABL1-Driven B-Cell Lineage Acute Lymphoblastic Leukemia. *Cancer Research* 70:7411-7420.
- Guo, P., M. Hirano, B.R. Herrin, J. Li, C. Yu, A. Sadlonova, and M.D. Cooper. 2009. Dual nature of the adaptive immune system in lampreys. *Nature* 459:796-801.

- Hadjiagapiou, C., F. Giannoni, T. Funahashi, S.F. Skarosi, and N.O. Davidson. 1994. Molecular cloning of a human small intestinal apolipoprotein B mRNA editing protein. *Nucleic Acids Res* 22:1874-1879.
- Hakim, O., W. Resch, A. Yamane, I. Klein, K.-R. Kieffer-Kwon, M. Jankovic, T. Oliveira, A. Bothmer, T.C. Voss, C. Ansarah-Sobrinho, E. Mathe, G. Liang, J. Cobell, H. Nakahashi, D.F. Robbiani, A. Nussenzweig, G.L. Hager, M.C. Nussenzweig, and R. Casellas. 2012. DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature* 484:69-74.
- Han, L., W. Mao, and K. Yu. 2012. X-ray repair cross-complementing protein 1 (XRCC1) deficiency enhances class switch recombination and is permissive for alternative end joining. *Proc Natl Acad Sci U S A* 109:4604-4608.
- Han, L., S. Masani, and K. Yu. 2010. Cutting edge: CTNNB1 is dispensable for Ig class switch recombination. *Journal of immunology (Baltimore, Md. : 1950)* 185:1379-1381.
- Han, L., S. Masani, and K. Yu. 2011. Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc Natl Acad Sci U S A* 108:11584-11589.
- Hardy, R.R., and K. Hayakawa. 2001. B cell development pathways. *Annu Rev Immunol* 19:595-621.
- Harris, L.J., S.B. Larson, K.W. Hasel, J. Day, A. Greenwood, and A. McPherson. 1992. The three-dimensional structure of an intact monoclonal antibody for canine lymphoma. *Nature* 360:369-372.
- Harris, L.J., S.B. Larson, K.W. Hasel, and A. McPherson. 1997. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry* 36:1581-1597.
- Harris, R.S., and J.P. Dudley. 2015. APOBECs and virus restriction. *Virology* 479-480:131-145.
- Hasham, M.G., N.M. Donghia, E. Coffey, J. Maynard, K.J. Snow, J. Ames, R.Y. Wilpan, Y. He, B.L. King, and K.D. Mills. 2010. Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. *Nature Immunology* 11:820-826.
- Hasler, J., C. Rada, and M.S. Neuberger. 2011. Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1 α (eEF1A). *Proceedings of the National Academy of Sciences of the United States of America* 108:18366-18371.
- Hasler, J., C. Rada, and M.S. Neuberger. 2012. The cytoplasmic AID complex. *Seminars in immunology* 24:273-280.
- Hay, F.C., M.G. Hull, and G. Torrigiani. 1971. The transfer of human IgG subclasses from mother to foetus. *Clin Exp Immunol* 9:355-358.

- He, B., N. Raab-Traub, P. Casali, and A. Cerutti. 2003. EBV-encoded latent membrane protein 1 cooperates with BAFF/BLyS and APRIL to induce T cell-independent Ig heavy chain class switching. *Journal of immunology* 171:5215-5224.
- He, M., E.M. Cortizas, R.E. Verdun, and E. Severinson. 2015. Cyclin-Dependent Kinases Regulate Ig Class Switching by Controlling Access of AID to the Switch Region. *Journal of immunology (Baltimore, Md. : 1950)* 194:4231-4239.
- Hein, K., M.G. Lorenz, G. Siebenkotten, K. Petry, R. Christine, and A. Radbruch. 1998. Processing of switch transcripts is required for targeting of antibody class switch recombination. *J Exp Med* 188:2369-2374.
- Henikoff, S., J.G. Henikoff, A. Sakai, G.B. Loeb, and K. Ahmad. 2009. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome research* 19:460-469.
- Hochster, H., R. Oratz, D.S. Ettinger, and E. Borden. 1998. A phase II study of Didemnin B (NSC 325319) in advanced malignant melanoma: an Eastern Cooperative Oncology Group study (PB687). *Invest New Drugs* 16:259-263.
- Hodgkin, P.D., J.H. Lee, and A.B. Lyons. 1996. B cell differentiation and isotype switching is related to division cycle number. *J Exp Med* 184:277-281.
- Holden, L.G., C. Prochnow, Y.P. Chang, R. Bransteitter, L. Chelico, U. Sen, R.C. Stevens, M.F. Goodman, and X.S. Chen. 2008. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* 456:121-124.
- Hozumi, N., and S. Tonegawa. 1976. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc Natl Acad Sci U S A* 73:3628-3632.
- Hsu, H.-C., Y. Wu, P. Yang, Q. Wu, G. Job, J. Chen, J. Wang, M.A.V. Accavitti-Loper, W.E. Grizzle, R.H. Carter, and J.D. Mountz. 2007. Overexpression of activation-induced cytidine deaminase in B cells is associated with production of highly pathogenic autoantibodies. *Journal of immunology (Baltimore, Md. : 1950)* 178:5357-5365.
- Hsu, H.-C., P. Yang, Q. Wu, J.H. Wang, G. Job, T. Guentert, J. Li, C.R. Stockard, T.-v.L. Le, D.D. Chaplin, W.E. Grizzle, and J.D. Mountz. 2011. Inhibition of the catalytic function of activation-induced cytidine deaminase promotes apoptosis of germinal center B cells in BXD2 mice. *Arthritis and rheumatism* 63:2038-2048.
- Hu, W., N.A. Begum, S. Mondal, A. Stanlie, and T. Honjo. 2015. Identification of DNA cleavage- and recombination-specific hnRNP cofactors for activation-induced cytidine deaminase. *Proc Natl Acad Sci USA* 112:5791-5796.
- Hu, Y., I. Ericsson, K. Torseth, S.P. Methot, O. Sundheim, N.B. Liabakk, G. Slupphaug, J.M. Di Noia, H.E. Krokan, and B. Kavli. 2013. A combined nuclear and nucleolar localization motif in activation-induced cytidine deaminase (AID) controls immunoglobulin class switching. *Journal of molecular biology* 425:424-443.

- Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J Mol Graph* 14:33-38, 27-38.
- Iliakis, G., T. Murmann, and A. Soni. 2015. Alternative end-joining repair pathways are the ultimate backup for abrogated classical non-homologous end-joining and homologous recombination repair: Implications for the formation of chromosome translocations. *Mutat Res Genet Toxicol Environ Mutagen* 793:166-175.
- Imai, K., Y. Zhu, P. Revy, T. Morio, S. Mizutani, A. Fischer, S. Nonoyama, and A. Durandy. 2005. Analysis of class switch recombination and somatic hypermutation in patients affected with autosomal dominant hyper-IgM syndrome type 2. *Clin Immunol* 115:277-285.
- Ito, S., H. Nagaoka, R. Shinkura, N.A. Begum, M. Muramatsu, M. Nakata, and T. Honjo. 2004. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proceedings of the National Academy of Sciences of the United States of America* 101:1975-1980.
- Itoh, K., B.K. Brott, G.U. Bae, M.J. Ratcliffe, and S.Y. Sokol. 2005. Nuclear localization is required for Dishevelled function in Wnt/beta-catenin signaling. *J Biol* 4:3.
- Jaehning, J.A. 2010. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta* 1799:379-388.
- Jansen, J.G., P. Langerak, A. Tsaalbi-Shtylik, P. van den Berk, H. Jacobs, and N. de Wind. 2006. Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J Exp Med* 203:319-323.
- Jeevan-Raj, B.P., I. Robert, V. Heyer, A. Page, J.H. Wang, F. Cammas, F.W. Alt, R. Losson, and B.R. Reina-San-Martin. 2011. Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. *The Journal of experimental medicine* 208:1649-1660.
- Jiang, C., J. Foley, N. Clayton, G. Kissling, M. Jokinen, R. Herbert, and M. Diaz. 2007. Abrogation of lupus nephritis in activation-induced deaminase-deficient MRL/lpr mice. *Journal of immunology (Baltimore, Md. : 1950)* 178:7422-7431.
- Jiricny, J. 2013. Postreplicative mismatch repair. *Cold Spring Harb Perspect Biol* 5:a012633.
- Jonkers, I., and J.T. Lis. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* 16:167-177.
- Kanehiro, Y., K. Todo, M. Negishi, J. Fukuoka, W. Gan, T. Hikasa, Y. Kaga, M. Takemoto, M. Magari, X. Li, J.L. Manley, H. Ohmori, and N. Kanayama. 2012. Activation-induced cytidine deaminase (AID)-dependent somatic hypermutation requires a splice isoform of the serine/arginine-rich (SR) protein SRSF1. *Proceedings of the National Academy of Sciences of the United States of America* 109:1216-1221.
- Kaplan, C.D., L. Laprade, and F. Winston. 2003. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301:1096-1099.

- Kasahara, Y., H. Kaneko, T. Fukao, T. Terada, T. Asano, K. Kasahara, and N. Kondo. 2003. Hyper-IgM syndrome with putative dominant negative mutation in activation-induced cytidine deaminase. *J Allergy Clin Immunol* 112:755-760.
- Kawaguchi, Y., H. Nariki, N. Kawamoto, Y. Kanehiro, S. Miyazaki, M. Suzuki, M. Magari, H. Tokumitsu, and N. Kanayama. 2017. SRSF1-3 contributes to diversification of the immunoglobulin variable region gene by promoting accumulation of AID in the nucleus. *Biochem Biophys Res Commun* 485:261-266.
- Keim, C., D. Kazadi, G. Rothschild, and U. Basu. 2013. Regulation of AID, the B-cell genome mutator. *Genes & development* 27:1-17.
- Kenter, A.L. 2012. AID targeting is dependent on RNA polymerase II pausing. *Seminars in immunology* 24:281-286.
- Kenter, A.L., S. Kumar, R. Wuerffel, and F. Grigera. 2016. AID hits the jackpot when missing the target. *Curr Opin Immunol* 39:96-102.
- Kessner, D., M. Chambers, R. Burke, D. Agus, and P. Mallick. 2008. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 24:2534-2536.
- Khacho, M., K. Mekhail, K. Pilon-Larose, A. Pause, J. Côté, and S. Lee. 2008. eEF1A is a novel component of the mammalian nuclear protein export machinery. *Molecular biology of the cell* 19:5296-5308.
- Khurana, S., D. Frasca, B. Blomberg, and H. Golding. 2012. AID activity in B cells strongly correlates with polyclonal antibody affinity maturation in-vivo following pandemic 2009-H1N1 vaccination in humans. *PLoS Pathog* 8:e1002920.
- Kim, D.I., S.C. Jensen, K.A. Noble, B. Kc, K.H. Roux, K. Motamedchaboki, and K.J. Roux. 2016. An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell* 27:1188-1196.
- King, J.J., C.A. Manuel, C.V. Barrett, S. Raber, H. Lucas, P. Sutter, and M. Larijani. 2015. Catalytic Pocket Inaccessibility of Activation-Induced Cytidine Deaminase Is a Safeguard against Excessive Mutagenic Activity. *Structure (London, England : 1993)* 23:615-627.
- Kitamura, S., H. Ode, M. Nakashima, M. Imahashi, Y. Naganawa, T. Kurosawa, Y. Yokomaku, T. Yamane, N. Watanabe, A. Suzuki, W. Sugiura, and Y. Iwatani. 2012. The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat Struct Mol Biol* 19:1005-1010.
- Klaus, G.G., M.B. Pepys, K. Kitajima, and B.A. Askonas. 1979. Activation of mouse complement by different classes of mouse antibody. *Immunology* 38:687-695.
- Klein, I.A., W. Resch, M. Jankovic, T. Oliveira, A. Yamane, H. Nakahashi, M. Di Virgilio, A. Bothmer, A. Nussenzweig, D.F. Robbiani, R. Casellas, and M.C. Nussenzweig. 2011. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147:95-106.

- Klemm, L., C. Duy, I. Iacobucci, S. Kuchen, G. von Levetzow, N. Feldhahn, N. Henke, Z. Li, T.K. Hoffmann, Y.-m. Kim, W.-K. Hofmann, H. Jumaa, J. Groffen, N. Heisterkamp, G. Martinelli, M.R. Lieber, R. Casellas, and M. MÜschen. 2009. The B Cell Mutator AID Promotes B Lymphoid Blast Crisis and Drug Resistance in Chronic Myeloid Leukemia. *Cancer Cell* 16:232-245.
- Knight, J.D., G. Liu, J.P. Zhang, A. Pasculescu, H. Choi, and A.C. Gingras. 2015. A web-tool for visualizing quantitative protein-protein interaction data. *Proteomics* 15:1432-1436.
- Kodgire, P., P. Mukkawar, S. Ratnam, T.E. Martin, and U. Storb. 2013. Changes in RNA polymerase II progression influence somatic hypermutation of Ig-related genes by AID. *The Journal of experimental medicine* 210:1481-1492.
- Kohler, K.M., J.J. McDonald, J.L. Duke, H. Arakawa, S. Tan, S.H. Kleinstein, J.M. Buerstedde, and D.G. Schatz. 2012. Identification of core DNA elements that target somatic hypermutation. *Journal of immunology* 189:5314-5326.
- Kohli, R.M., S.R. Abrams, K.S. Gajula, R.W. Maul, P.J. Gearhart, and J.T. Stivers. 2009. A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *The Journal of Biological Chemistry* 284:22898-22904.
- Kohli, R.M., and Y. Zhang. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502:472-479.
- Komor, A.C., A.H. Badran, and D.R. Liu. 2017. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* 168:20-36.
- Komor, A.C., Y.B. Kim, M.S. Packer, J.A. Zuris, and D.R. Liu. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533:420-424.
- Komori, T., A. Okada, V. Stewart, and F.W. Alt. 1993. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* 261:1171-1175.
- Koning, F.A., E.N. Newman, E.Y. Kim, K.J. Kunstman, S.M. Wolinsky, and M.H. Malim. 2009. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* 83:9474-9485.
- Kornblihtt, A.R., I.E. Schor, M. Alló, G. Dujardin, E. Petrillo, and M.J. Muñoz. 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature reviews. Molecular cell biology* 14:153-165.
- Kotani, A., N. Kakazu, T. Tsuruyama, I.M. Okazaki, M. Muramatsu, K. Kinoshita, H. Nagaoka, D. Yabe, and T. Honjo. 2007. Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. *Proc Natl Acad Sci U S A* 104:1616-1620.

- Kothapalli, N., D.D. Norton, and S.D. Fugmann. 2008. Cutting edge: a cis-acting DNA element targets AID-mediated sequence diversification to the chicken Ig light chain gene locus. *Journal of immunology* 180:2019-2023.
- Kracker, S., M. Di Virgilio, J. Schwartzentruber, C. Cuenin, M. Forveille, M.C. Deau, K.M. McBride, J. Majewski, A. Gazumyan, S. Seneviratne, B. Grimbacher, N. Kutukculer, Z. Herceg, M. Cavazzana, N. Jabado, M.C. Nussenzweig, A. Fischer, and A. Durandy. 2015. An inherited immunoglobulin class-switch recombination deficiency associated with a defect in the INO80 chromatin remodeling complex. *J Allergy Clin Immunol* 135:998-1007 e1006.
- Krokan, H.E., and M. Bjoras. 2013. Base excision repair. *Cold Spring Harb Perspect Biol* 5:a012583.
- Kucuk, O., M.L. Young, T.M. Habermann, B.C. Wolf, J. Jimeno, and P.A. Cassileth. 2000. Phase II trial of didemnin B in previously treated non-Hodgkin's lymphoma: an Eastern Cooperative Oncology Group (ECOG) Study. *Am J Clin Oncol* 23:273-277.
- Kuraoka, M., T.M. Holl, D. Liao, M. Womble, D.W. Cain, A.E. Reynolds, and G. Kelsoe. 2011. Activation-induced cytidine deaminase mediates central tolerance in B cells. *Proceedings of the National Academy of Sciences of the United States of America* 108:11560-11565.
- Kuwahara, K., S. Fujimura, Y. Takahashi, N. Nakagata, T. Takemori, S. Aizawa, and N. Sakaguchi. 2004. Germinal center-associated nuclear protein contributes to affinity maturation of B cell antigen receptor in T cell-dependent responses. *Proceedings of the National Academy of Sciences of the United States of America* 101:1010-1015.
- Lackey, L., Z.L. Demorest, A.M. Land, J.F. Hultquist, W.L. Brown, and R.S. Harris. 2012. APOBEC3B and AID have similar nuclear import mechanisms. *Journal of molecular biology* 419:301-314.
- Lambert, L.J., S. Walker, J. Feltham, H.J. Lee, W. Reik, and J. Houseley. 2013. Etoposide induces nuclear re-localisation of AID. *PloS one* 8:e82110.
- Lamont, K.R., M.G. Hasham, N.M. Donghia, J. Branca, M. Chavaree, B. Chase, A. Breggia, J. Hedlund, I. Emery, F. Cavallo, M. Jasin, J. Rüter, and K.D. Mills. 2013. Attenuating homologous recombination stimulates an AID-induced antileukemic effect. *The Journal of experimental medicine* 210:1021-1033.
- Land, A.M., E.K. Law, M.A. Carpenter, L. Lackey, W.L. Brown, and R.S. Harris. 2013. Endogenous APOBEC3A DNA cytosine deaminase is cytoplasmic and nongenotoxic. *J Biol Chem* 288:17253-17260.
- Langlois, M.A., R.C. Beale, S.G. Conticello, and M.S. Neuberger. 2005. Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res* 33:1913-1923.

- Larijani, M., and A. Martin. 2007. Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. *Molecular and cellular biology* 27:8038-8048.
- Larijani, M., A.P. Petrov, O. Kolenchenko, M. Berru, S.N. Krylov, and A. Martin. 2007. AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. *Molecular and cellular biology* 27:20-30.
- Larson, E.D., D.W. Bednarski, and N. Maizels. 2008. High-fidelity correction of genomic uracil by human mismatch repair activities. *BMC Mol Biol* 9:94.
- Le, Q., and N. Maizels. 2015. Cell Cycle Regulates Nuclear Stability of AID and Determines the Cellular Response to AID. *PLoS Genetics* 11:e1005411.
- Lee, C.G., K. Kinoshita, A. Arudchandran, S.M. Cerritelli, R.J. Crouch, and T. Honjo. 2001. Quantitative regulation of class switch recombination by switch region transcription. *J Exp Med* 194:365-374.
- Lee-Theilen, M., A.J. Matthews, D. Kelly, S. Zheng, and J. Chaudhuri. 2011. CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. *Nat Struct Mol Biol* 18:75-79.
- Leuenberger, M., S. Frigerio, P.J. Wild, F. Noetzli, D. Korol, D.R. Zimmermann, C. Gengler, N.M. Probst-Hensch, H. Moch, and M. Tinguely. 2010. AID protein expression in chronic lymphocytic leukemia/small lymphocytic lymphoma is associated with poor prognosis and complex genetic alterations. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 23:177-186.
- Li, G., C.A. White, T. Lam, E.J. Pone, D.C. Tran, K.L. Hayama, H. Zan, Z. Xu, and P. Casali. 2013. Combinatorial H3K9acS10ph Histone Modification in IgH Locus S Regions Targets 14-3-3 Adaptors and AID to Specify Antibody Class-Switch DNA Recombination. *Cell reports* 5:702-714.
- Lin, K.W., I. Yakymovych, M. Jia, M. Yakymovych, and S. Souchelnytskyi. 2010. Phosphorylation of eEF1A1 at Ser300 by TbetR-I results in inhibition of mRNA translation. *Curr Biol* 20:1615-1625.
- Lindqvist, L., F. Robert, W. Merrick, H. Kakeya, C. Fraser, H. Osada, and J. Pelletier. 2010. Inhibition of translation by cytotrienin A--a member of the ansamycin family. *RNA (New York, N.Y.)* 16:2404-2413.
- Lionta, E., G. Spyrou, D.K. Vassilatis, and Z. Cournia. 2014. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem* 14:1923-1938.
- Liso, A., D. Capello, T. Marafioti, E. Tiacci, M. Cerri, V. Distler, M. Paulli, A. Carbone, G. Delsol, E. Campo, S. Pileri, L. Pasqualucci, G. Gaidano, and B. Falini. 2006. Aberrant somatic hypermutation in tumor cells of nodular-lymphocyte-predominant and classic Hodgkin lymphoma. *Blood* 108:1013-1020.

- Liu, G., J. Zhang, B. Larsen, C. Stark, A. Breitkreutz, Z.Y. Lin, B.J. Breitkreutz, Y. Ding, K. Colwill, A. Pasculescu, T. Pawson, J.L. Wrana, A.I. Nesvizhskii, B. Raught, M. Tyers, and A.C. Gingras. 2010. ProHits: integrated software for mass spectrometry-based interaction proteomics. *Nat Biotechnol* 28:1015-1017.
- Liu, M., J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, and D.G. Schatz. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451:841-845.
- Liu, Y., L. Warfield, C. Zhang, J. Luo, J. Allen, W.H. Lang, J. Ranish, K.M. Shokat, and S. Hahn. 2009. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol Cell Biol* 29:4852-4863.
- Lorenz, M., S. Jung, and A. Radbruch. 1995. Switch transcripts in immunoglobulin class switching. *Science* 267:1825-1828.
- Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
- Luby, T.M., C.E. Schrader, J. Stavnezer, and E. Selsing. 2001. The mu switch region tandem repeats are important, but not required, for antibody class switch recombination. *J Exp Med* 193:159-168.
- MacDuff, D.A., Z.L. Demorest, and R.S. Harris. 2009. AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic Acids Research* 37:1854-1867.
- Maeda, K., S.K. Singh, K. Eda, M. Kitabatake, P. Pham, M.F. Goodman, and N. Sakaguchi. 2010. GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. *The Journal of Biological Chemistry* 285:23945-23953.
- Mahdaviani, S.A., A. Hirbod-Mobarakeh, N. Wang, A. Aghamohammadi, L. Hammarstrom, M.R. Masjedi, Q. Pan-Hammarstrom, and N. Rezaei. 2012. Novel mutation of the activation-induced cytidine deaminase gene in a Tajik family: special review on hyper-immunoglobulin M syndrome. *Expert Rev Clin Immunol* 8:539-546.
- Manis, J.P., Y. Gu, R. Lansford, E. Sonoda, R. Ferrini, L. Davidson, K. Rajewsky, and F.W. Alt. 1998. Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *J Exp Med* 187:2081-2089.
- Manis, J.P., J.C. Morales, Z. Xia, J.L. Kutok, F.W. Alt, and P.B. Carpenter. 2004. 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. *Nature Immunology* 5:481-487.
- Marco, E., S. Martín-Santamaría, C. Cuevas, and F. Gago. 2004. Structural basis for the binding of didemnins to human elongation factor eEF1A and rationale for the potent antitumor activity of these marine natural products. *Journal of medicinal chemistry* 47:4439-4452.

- Marino, D., M. Perkovic, A. Hain, A.A. Jaguva Vasudevan, H. Hofmann, K.M. Hanschmann, M.D. Muhlebach, G.G. Schumann, R. Konig, K. Cichutek, D. Haussinger, and C. Munk. 2016. APOBEC4 Enhances the Replication of HIV-1. *PLoS One* 11:e0155422.
- Martin, R.M., J.L. Brady, and A.M. Lew. 1998. The need for IgG2c specific antiserum when isotyping antibodies from C57BL/6 and NOD mice. *J Immunol Methods* 212:187-192.
- Mateyak, M.K., and T.G. Kinzy. 2010. eEF1A: thinking outside the ribosome. *The Journal of Biological Chemistry* 285:21209-21213.
- Matsuda, F., and T. Honjo. 1996. Organization of the human immunoglobulin heavy-chain locus. *Adv Immunol* 62:1-29.
- Matsuda, T., K. Bebenek, C. Masutani, F. Hanaoka, and T.A. Kunkel. 2000. Low fidelity DNA synthesis by human DNA polymerase-eta. *Nature* 404:1011-1013.
- Matsumoto, Y., H. Marusawa, K. Kinoshita, Y. Endo, T. Kou, T. Morisawa, T. Azuma, I.-m. Okazaki, T. Honjo, and T. Chiba. 2007. Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nature Medicine* 13:470-476.
- Matthews, A.J., S. Husain, and J. Chaudhuri. 2014a. Binding of AID to DNA does not correlate with mutator activity. *Journal of immunology (Baltimore, Md. : 1950)* 193:252-257.
- Matthews, A.J., S. Zheng, L.J. DiMenna, and J. Chaudhuri. 2014b. Regulation of immunoglobulin class-switch recombination: choreography of noncoding transcription, targeted DNA deamination, and long-range DNA repair. *Adv Immunol* 122:1-57.
- Maul, R.W., Z. Cao, L. Venkataraman, C.A. Giorgetti, J.L. Press, Y. Denizot, H. Du, R. Sen, and P.J. Gearhart. 2014. Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from ex vivo-activated cells. *The Journal of experimental medicine* 211:2297-2306.
- Maul, R.W., T. MacCarthy, E.G. Frank, K.A. Donigan, M.P. McLenigan, W. Yang, H. Saribasak, D.E. Huston, S.S. Lange, R. Woodgate, and P.J. Gearhart. 2016. DNA polymerase iota functions in the generation of tandem mutations during somatic hypermutation of antibody genes. *J Exp Med* 213:1675-1683.
- Maul, R.W., H. Saribasak, S.A. Martomo, R.L. McClure, W. Yang, A. Vaisman, H.S. Gramlich, D.G. Schatz, R. Woodgate, D.M. Wilson, and P.J. Gearhart. 2011. Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nature Immunology* 12:70-76.
- McBlane, J.F., D.C. van Gent, D.A. Ramsden, C. Romeo, C.A. Cuomo, M. Gellert, and M.A. Oettinger. 1995. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* 83:387-395.

- McBride, K.M., V. Barreto, A.R. Ramiro, P. Stavropoulos, and M.C. Nussenzweig. 2004. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *The Journal of experimental medicine* 199:1235-1244.
- McBride, K.M., A. Gazumyan, E.M. Woo, V.M. Barreto, D.F. Robbiani, B.T. Chait, and M.C. Nussenzweig. 2006. Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* 103:8798-8803.
- McBride, K.M., A. Gazumyan, E.M. Woo, T.A. Schwickert, B.T. Chait, and M.C. Nussenzweig. 2008. Regulation of class switch recombination and somatic mutation by AID phosphorylation. *The Journal of experimental medicine* 205:2585-2594.
- McCarthy, H., W.G. Wierda, L.L. Barron, C.C. Cromwell, J. Wang, K.R. Coombes, R. Rangel, K.S.J. Elenitoba-Johnson, M.J. Keating, and L.V. Abruzzo. 2003. High expression of activation-induced cytidine deaminase (AID) and splice variants is a distinctive feature of poor-prognosis chronic lymphocytic leukemia. *Blood* 101:4903-4908.
- Meng, F.-L., Z. Du, A. Federation, J. Hu, Q. Wang, K.-R. Kieffer-Kwon, R.M. Meyers, C. Amor, C.R. Wasserman, D. Neuberg, R. Casellas, M.C. Nussenzweig, J.E. Bradner, X.S. Liu, and F.W. Alt. 2014. Convergent Transcription at Intragenic Super-Enhancers Targets AID-Initiated Genomic Instability. *Cell* 159:1538-1548.
- Meng, L., N. Sin, and C.M. Crews. 1998. The antiproliferative agent didemnin B uncompetitively inhibits palmitoyl protein thioesterase. *Biochemistry* 37:10488-10492.
- Methot, S.P., and J.M. Di Noia. 2017. Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination. *Adv Immunol* 133:37-87.
- Methot, S.P., L.C. Litzler, F. Trajtenberg, A. Zahn, F. Robert, J. Pelletier, A. Buschiazzo, B.G. Magor, and J.M. Di Noia. 2015. Consecutive interactions with HSP90 and eEF1A underlie a functional maturation and storage pathway of AID in the cytoplasm. *The Journal of experimental medicine* 212:581-596.
- Meyers, G., Y.-S. Ng, J.M. Bannock, A. Lavoie, J.E. Walter, L.D. Notarangelo, S.S. Kilic, G. Aksu, M. Debré, F. Rieux-Laucat, M.E. Conley, C. Cunningham-Rundles, A.H. Durandy, and E. Meffre. 2011. Activation-induced cytidine deaminase (AID) is required for B-cell tolerance in humans. *Proceedings of the National Academy of Sciences of the United States of America* 108:11554-11559.
- Miller, N.W., M.A. Ryczyn, M.R. Wilson, G.W. Warr, J.P. Naftel, and L.W. Clem. 1994. Development and characterization of channel catfish long term B cell lines. *Journal of immunology* 152:2180-2189.
- Mittelman, A., H.G. Chun, C. Puccio, N. Coombe, T. Lansen, and T. Ahmed. 1999. Phase II clinical trial of didemnin B in patients with recurrent or refractory anaplastic

- astrocytoma or glioblastoma multiforme (NSC 325319). *Invest New Drugs* 17:179-182.
- Mond, J.J., Q. Vos, A. Lees, and C.M. Snapper. 1995. T cell independent antigens. *Curr Opin Immunol* 7:349-354.
- Mondal, S., N.A. Begum, W. Hu, and T. Honjo. 2016. Functional requirements of AID's higher order structures and their interaction with RNA-binding proteins. *Proc Natl Acad Sci USA* 113:E1545-E1554.
- Montamat-Sicotte, D., L.C. Litzler, C. Abreu, S. Safavi, A. Zahn, A. Orthwein, M. Muschen, P. Oppezso, D.P. Munoz, and J.M. Di Noia. 2015. HSP90 inhibitors decrease AID levels and activity in mice and in human cells. *Eur J Immunol* 45:2365-2376.
- Montgomery, D.W., A. Celniker, and C.F. Zukoski. 1987. Didemnin B: an immunosuppressive cyclic peptide that stimulates murine antibody responses in vivo and in vitro. *Transplantation proceedings* 19:1295-1296.
- Morgan, H.D., W. Dean, H.A. Coker, W. Reik, and S.K. Petersen-Mahrt. 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *The Journal of Biological Chemistry* 279:52353-52360.
- Morita, S., T. Kojima, and T. Kitamura. 2000. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther* 7:1063-1066.
- Morrison, J.R., C. Paszty, M.E. Stevens, S.D. Hughes, T. Forte, J. Scott, and E.M. Rubin. 1996. Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. *Proc Natl Acad Sci U S A* 93:7154-7159.
- Mu, Y., C. Prochnow, P. Pham, X.S. Chen, and M.F. Goodman. 2012. A structural basis for the biochemical behavior of activation-induced deoxycytidine deaminase class-switch recombination-defective hyper-IgM-2 mutants. *The Journal of Biological Chemistry* 287:28007-28016.
- Muñoz, D.P., E.L. Lee, S. Takayama, J.-P. Coppé, S.-J. Heo, D. Boffelli, J.M. Di Noia, and D.I.K. Martin. 2013. Activation-induced cytidine deaminase (AID) is necessary for the epithelial-mesenchymal transition in mammary epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America* 110:E2977-2986.
- Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102:553-563.
- Muramatsu, M., V.S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N.O. Davidson, and T. Honjo. 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *The Journal of Biological Chemistry* 274:18470-18476.

- Murphy, K., P. Travers, M. Walport, and C. Janeway. 2012. *Janeway's immunobiology*. Garland Science, New York. xix, 868 p. pp.
- Nabel, C.S., H. Jia, Y. Ye, L. Shen, H.L. Goldschmidt, J.T. Stivers, Y. Zhang, and R.M. Kohli. 2012. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature chemical biology* 8:751-758.
- Nadassy, K., S.J. Wodak, and J. Janin. 1999. Structural features of protein-nucleic acid recognition sites. *Biochemistry* 38:1999-2017.
- Nambu, Y., M. Sugai, H. Gonda, C.-G. Lee, T. Katakai, Y. Agata, Y. Yokota, and A. Shimizu. 2003. Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science (New York, N.Y.)* 302:2137-2140.
- Navaratnam, N., J.R. Morrison, S. Bhattacharya, D. Patel, T. Funahashi, F. Giannoni, B.B. Teng, N.O. Davidson, and J. Scott. 1993. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J Biol Chem* 268:20709-20712.
- Negrutskii, B., D. Vlasenko, and A. El'skaya. 2012. From global phosphoproteomics to individual proteins: the case of translation elongation factor eEF1A. *Expert Rev Proteomics* 9:71-83.
- Nemazee, D. 2017. Mechanisms of central tolerance for B cells. *Nat Rev Immunol* 17:281-294.
- Nguyen, K.T., M.P. Holloway, and R.A. Altura. 2012. The CRM1 nuclear export protein in normal development and disease. *Int J Biochem Mol Biol* 3:137-151.
- Nieuwenhuis, P., and D. Opstelten. 1984. Functional anatomy of germinal centers. *Am J Anat* 170:421-435.
- Nishida, K., T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K.Y. Hara, Z. Shimatani, and A. Kondo. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353:
- Notarangelo, L.D., M. Duse, and A.G. Ugazio. 1992. Immunodeficiency with hyper-IgM (HIM). *Immunodeficiency Rev* 3:101-121.
- Nowak, U., A.J. Matthews, S. Zheng, and J. Chaudhuri. 2011. The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. *Nature Immunology* 12:160-166.
- Nozawa, R.S., L. Boteva, D.C. Soares, C. Naughton, A.R. Dun, A. Buckle, B. Ramsahoye, P.C. Bruton, R.S. Saleeb, M. Arnedo, B. Hill, R.R. Duncan, S.K. Maciver, and N. Gilbert. 2017. SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. *Cell* 169:1214-1227 e1218.
- Nutt, S.L., P.D. Hodgkin, D.M. Tarlinton, and L.M. Corcoran. 2015. The generation of antibody-secreting plasma cells. *Nat Rev Immunol* 15:160-171.

- Okazaki, I.-m., H. Hiai, N. Kakazu, S. Yamada, M. Muramatsu, K. Kinoshita, and T. Honjo. 2003. Constitutive expression of AID leads to tumorigenesis. *The Journal of experimental medicine* 197:1173-1181.
- Okazaki, I.-m., K. Okawa, M. Kobayashi, K. Yoshikawa, S. Kawamoto, H. Nagaoka, R. Shinkura, Y. Kitawaki, H. Taniguchi, T. Natsume, S.-I. Iemura, and T. Honjo. 2011. Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America* 108:7920-7925.
- Oldfield, C.J., B. Xue, Y.Y. Van, E.L. Ulrich, J.L. Markley, A.K. Dunker, and V.N. Uversky. 2013. Utilization of protein intrinsic disorder knowledge in structural proteomics. *Biochim Biophys Acta* 1834:487-498.
- Orthwein, A., and J.M. Di Noia. 2012. Activation induced deaminase: how much and where? *Seminars in immunology* 24:246-254.
- Orthwein, A., S.M. Noordermeer, M.D. Wilson, S. Landry, R.I. Enchev, A. Sherker, M. Munro, J. Pinder, J. Salsman, G. Dellaire, B. Xia, M. Peter, and D. Durocher. 2015. A mechanism for the suppression of homologous recombination in G1 cells. *Nature* 528:422-426.
- Orthwein, A., A.-M. Patenaude, E.B. Affar, A. Lamarre, J.C. Young, and J.M. Di Noia. 2010. Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. *The Journal of experimental medicine* 207:2751-2765.
- Orthwein, A., A. Zahn, S.P. Methot, D. Godin, S.G. Conticello, K. Terada, and J.M. Di Noia. 2012. Optimal functional levels of activation-induced deaminase specifically require the Hsp40 DnaJa1. *The EMBO journal* 31:679-691.
- Pallares, N., J.P. Fripiat, V. Giudicelli, and M.P. Lefranc. 1998. The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. *Exp Clin Immunogenet* 15:8-18.
- Pan-Hammarstrom, Q., A.M. Jones, A. Lahdesmaki, W. Zhou, R.A. Gatti, L. Hammarstrom, A.R. Gennery, and M.R. Ehrenstein. 2005. Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells. *J Exp Med* 201:189-194.
- Pancer, Z., C.T. Amemiya, G.R. Ehrhardt, J. Ceitlin, G.L. Gartland, and M.D. Cooper. 2004. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 430:174-180.
- Parker, D.C. 1993. T cell-dependent B cell activation. *Annu Rev Immunol* 11:331-360.
- Parkin, J., and B. Cohen. 2001. An overview of the immune system. *Lancet* 357:1777-1789.
- Pasqualucci, L., G. Bhagat, M. Jankovic, M. Compagno, P. Smith, M. Muramatsu, T. Honjo, H.C. Morse, M.C. Nussenzweig, and R. Dalla-Favera. 2008. AID is required for germinal center-derived lymphomagenesis. *Nature Genetics* 40:108-112.

- Pasqualucci, L., R. Guglielmino, J. Houldsworth, J. Mohr, S. Aoufouchi, R. Polakiewicz, R.S.K. Chaganti, and R. Dalla-Favera. 2004. Expression of the AID protein in normal and neoplastic B cells. *Blood* 104:3318-3325.
- Pasqualucci, L., Y. Kitaura, H. Gu, and R. Dalla-Favera. 2006. PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proceedings of the National Academy of Sciences of the United States of America* 103:395-400.
- Pasqualucci, L., P. Neumeister, T. Goossens, G. Nanjangud, R.S. Chaganti, R. Küppers, and R. Dalla-Favera. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412:341-346.
- Patenaude, A.-M., A. Orthwein, Y. Hu, V.A. Campo, B. Kavli, A. Buschiazzi, and J.M. Di Noia. 2009. Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nature structural & molecular biology* 16:517-527.
- Pauklin, S., and S.K. Petersen-Mahrt. 2009. Progesterone inhibits activation-induced deaminase by binding to the promoter. *Journal of immunology (Baltimore, Md. : 1950)* 183:1238-1244.
- Pauklin, S., I.V. Sernandez, G. Bachmann, A.R. Ramiro, and S.K. Petersen-Mahrt. 2009. Estrogen directly activates AID transcription and function. *The Journal of experimental medicine* 206:99-111.
- Pavri, R., A. Gazumyan, M. Jankovic, M. Di Virgilio, I. Klein, C. Ansarah-Sobrinho, W. Resch, A. Yamane, B.R. Reina-San-Martin, V. Barreto, T.J. Nieland, D.E. Root, R. Casellas, and M.C. Nussenzweig. 2010. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 143:122-133.
- Pefanis, E., J. Wang, G. Rothschild, J. Lim, J. Chao, R. Rabadan, A.N. Economides, and U. Basu. 2014. Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* 514:389-393.
- Pefanis, E., J. Wang, G. Rothschild, J. Lim, D. Kazadi, J. Sun, A. Federation, J. Chao, O. Elliott, Z.-P. Liu, A.N. Economides, J.E. Bradner, R. Rabadan, and U. Basu. 2015. RNA Exosome-Regulated Long Non-Coding RNA Transcription Controls Super-Enhancer Activity. *Cell* 161:774-789.
- Peled, J.U., F.L. Kuang, M.D. Iglesias-Ussel, S. Roa, S.L. Kalis, M.F. Goodman, and M.D. Scharff. 2008. The biochemistry of somatic hypermutation. *Annual Review of Immunology* 26:481-511.
- Perelson, A.S., and G.F. Oster. 1979. Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. *J Theor Biol* 81:645-670.
- Perry, R.P., and D.E. Kelley. 1970. Inhibition of RNA synthesis by actinomycin D: characteristic dose-response of different RNA species. *J Cell Physiol* 76:127-139.

- Peters, A., and U. Storb. 1996. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* 4:57-65.
- Petersen, S., R. Casellas, B.R. Reina-San-Martin, H.T. Chen, M.J. Difilippantonio, P.C. Wilson, L. Hanitsch, A. Celeste, M. Muramatsu, D.R. Pilch, C. Redon, T. Ried, W.M. Bonner, T. Honjo, M.C. Nussenzweig, and A. Nussenzweig. 2001. AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature* 414:660-665.
- Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature* 418:99-103.
- Phalipon, A., A. Cardona, J.P. Kraehenbuhl, L. Edelman, P.J. Sansonetti, and B. Corthesy. 2002. Secretory component: a new role in secretory IgA-mediated immune exclusion in vivo. *Immunity* 17:107-115.
- Pham, P., S.A. Afif, M. Shimoda, K. Maeda, N. Sakaguchi, L.C. Pedersen, and M.F. Goodman. 2016. Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification. *DNA Repair* 43:48-56.
- Pham, P., R. Bransteitter, J. Petruska, and M.F. Goodman. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 424:103-107.
- Phan, T.G., and S.G. Tangye. 2017. Memory B cells: total recall. *Curr Opin Immunol* 45:132-140.
- Popp, C., W. Dean, S. Feng, S.J. Cokus, S. Andrews, M. Pellegrini, S.E. Jacobsen, and W. Reik. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463:1101-1105.
- Prochnow, C., R. Bransteitter, M.G. Klein, M.F. Goodman, and X.S. Chen. 2007. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* 445:447-451.
- Pucella, J.N., and J. Chaudhuri. 2017. AID Invited to the G4 Summit. *Mol Cell* 67:355-357.
- Qian, J., Q. Wang, M. Dose, N. Pruett, K.-R. Kieffer-Kwon, W. Resch, G. Liang, Z. Tang, E. Mathe, C. Benner, W. Dubois, S. Nelson, L. Vian, T.Y. Oliveira, M. Jankovic, O. Hakim, A. Gazumyan, R. Pavri, P. Awasthi, B. Song, G. Liu, L. Chen, S. Zhu, L. Feigenbaum, L. Staudt, C. Murre, Y. Ruan, D.F. Robbani, Q. Pan-Hammarström, M.C. Nussenzweig, and R. Casellas. 2014. B Cell Super-Enhancers and Regulatory Clusters Recruit AID Tumorigenic Activity. *Cell* 159:1524-1537.
- Qiao, Q., L. Wang, F.-L. Meng, J.K. Hwang, F.W. Alt, and H. Wu. AID Recognizes Structured DNA for Class Switch Recombination. *Molecular Cell*
- Qiao, Q., L. Wang, F.L. Meng, J.K. Hwang, F.W. Alt, and H. Wu. 2017. AID Recognizes Structured DNA for Class Switch Recombination. *Mol Cell* 67:361-373 e364.

- Qiao, Y., X. Han, G. Guan, N. Wu, J. Sun, V. Pak, and G. Liang. 2016. TGF-beta triggers HBV cccDNA degradation through AID-dependent deamination. *FEBS Lett* 590:419-427.
- Quackenbush, J. 2002. Microarray data normalization and transformation. *Nat Genet* 32 Suppl:496-501.
- Quartier, P., J. Bustamante, O. Sanal, A. Plebani, M. Debré, A. Deville, J. Litzman, J. Levy, J.-P. Fermand, P. Lane, G. Horneff, G. Aksu, I. Yalçın, G. Davies, I. Tezcan, F. Ersoy, N. Catalan, K. Imai, A. Fischer, and A.H. Durandy. 2004. Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to Activation-Induced Cytidine Deaminase deficiency. *Clinical immunology (Orlando, Fla.)* 110:22-29.
- Quinlan, E.M., J.J. King, C.T. Amemiya, E. Hsu, and M. Larijani. 2017. Biochemical regulatory features of AID remain conserved from lamprey to humans. *Mol Cell Biol*
- Rada, C., J.M. Di Noia, and M.S. Neuberger. 2004. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* 16:163-171.
- Rada, C., M.R. Ehrenstein, M.S. Neuberger, and C. Milstein. 1998. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* 9:135-141.
- Rada, C., J.M. Jarvis, and C. Milstein. 2002a. AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proceedings of the National Academy of Sciences of the United States of America* 99:7003-7008.
- Rada, C., G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, and M.S. Neuberger. 2002b. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* 12:1748-1755.
- Rai, K., I.J. Huggins, S.R. James, A.R. Karpf, D.A. Jones, and B.R. Cairns. 2008. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135:1201-1212.
- Rajagopal, D., R.W. Maul, A. Ghosh, T. Chakraborty, A.A. Khamlichi, R. Sen, and P.J. Gearhart. 2009. Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *The Journal of experimental medicine* 206:1237-1244.
- Ramiro, A.R., M. Jankovic, E. Callen, S. Difilippantonio, H.T. Chen, K.M. McBride, T.R. Eisenreich, J. Chen, R.A. Dickins, S.W. Lowe, A. Nussenzweig, and M.C. Nussenzweig. 2006. Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. *Nature* 440:105-109.
- Ramiro, A.R., M. Jankovic, T.R. Eisenreich, S. Difilippantonio, S. Chen-Kiang, M. Muramatsu, T. Honjo, A. Nussenzweig, and M.C. Nussenzweig. 2004. AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* 118:431-438.

- Ramiro, A.R., P. Stavropoulos, M. Jankovic, and M.C. Nussenzweig. 2003. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nature Immunology* 4:452-456.
- Ranjit, S., L. Khair, E.K. Linehan, A.J. Ucher, M. Chakrabarti, C.E. Schrader, and J. Stavnezer. 2011. AID binds cooperatively with UNG and Msh2-Msh6 to Ig switch regions dependent upon the AID C terminus. *Journal of immunology (Baltimore, Md. : 1950)* 187:2464-2475.
- Ratcliffe, M.J. 2006. Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development. *Dev Comp Immunol* 30:101-118.
- Ravetch, J.V., and J.P. Kinet. 1991. Fc receptors. *Annu Rev Immunol* 9:457-492.
- Refsland, E.W., M.D. Stenglein, K. Shindo, J.S. Albin, W.L. Brown, and R.S. Harris. 2010. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res* 38:4274-4284.
- Reina-San-Martin, B., J. Chen, A. Nussenzweig, and M.C. Nussenzweig. 2007. Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP1-/- B cells. *Eur J Immunol* 37:235-239.
- Revy, P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A.G. Ugazio, N. Brousse, M. Muramatsu, L.D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer, and A. Durandy. 2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 102:565-575.
- Rhodes, D., and R. Giraldo. 1995. Telomere structure and function. *Curr Opin Struct Biol* 5:311-322.
- Richard, D.J., E. Bolderson, and K.K. Khanna. 2009. Multiple human single-stranded DNA binding proteins function in genome maintenance: structural, biochemical and functional analysis. *Crit Rev Biochem Mol Biol* 44:98-116.
- Robbiani, D.F., A. Bothmer, E. Callen, B.R. Reina-San-Martin, Y. Dorsett, S. Difilippantonio, D.J. Bolland, H.T. Chen, A.E. Corcoran, A. Nussenzweig, and M.C. Nussenzweig. 2008. AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* 135:1028-1038.
- Robbiani, D.F., S. Bunting, N. Feldhahn, A. Bothmer, J. Camps, S. Deroubaix, K.M. McBride, I.A. Klein, G. Stone, T.R. Eisenreich, T. Ried, A. Nussenzweig, and M.C. Nussenzweig. 2009. AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Molecular cell* 36:631-641.
- Robbiani, D.F., S. Deroubaix, N. Feldhahn, T.Y. Oliveira, E. Callen, Q. Wang, M. Jankovic, I.T. Silva, P.C. Rommel, D. Bosque, T. Eisenreich, A. Nussenzweig, and M.C. Nussenzweig. 2015. Plasmodium Infection Promotes Genomic Instability and AID-Dependent B Cell Lymphoma. *Cell* 162:727-737.

- Robbiani, D.F., and M.C. Nussenzweig. 2013. Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. *Annual review of pathology* 8:79-103.
- Robert, F., M. Carrier, S. Rawe, S. Chen, S. Lowe, and J. Pelletier. 2009. Altering chemosensitivity by modulating translation elongation. *PLoS one* 4:e5428.
- Robertson, A.B., A. Klungland, T. Rognes, and I. Leiros. 2009. DNA repair in mammalian cells: Base excision repair: the long and short of it. *Cell Mol Life Sci* 66:981-993.
- Rocha, P.P., R. Raviram, Y. Fu, J. Kim, V.M. Luo, A. Aljoufi, E. Swanzey, A. Pasquarella, A. Balestrini, E.R. Miraldi, R. Bonneau, J. Petrini, G. Schotta, and J.A. Skok. 2016. A Damage-Independent Role for 53BP1 that Impacts Break Order and Igh Architecture during Class Switch Recombination. *Cell Rep* 16:48-55.
- Rogozin, I.B., L.M. Iyer, L. Liang, G.V. Glazko, V.G. Liston, Y.I. Pavlov, L. Aravind, and Z. Pancer. 2007. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nature Immunology* 8:647-656.
- Rolink, A.G., C. Schaniel, J. Andersson, and F. Melchers. 2001. Selection events operating at various stages in B cell development. *Curr Opin Immunol* 13:202-207.
- Romanello, M., D. Schiavone, A. Frey, and J.E. Sale. 2016. Histone H3.3 promotes IgV gene diversification by enhancing formation of AID-accessible single-stranded DNA. *The EMBO journal* 35:1452-1464.
- Rosenberg, B.R., C.E. Hamilton, M.M. Mwangi, S. Dewell, and F.N. Papavasiliou. 2011. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. *Nat Struct Mol Biol* 18:230-236.
- Roth, D.B. 2014. V(D)J Recombination: Mechanism, Errors, and Fidelity. *Microbiol Spectr* 2:
- Rothkamm, K., I. Kruger, L.H. Thompson, and M. Lobrich. 2003. Pathways of DNA double-strand break repair during the mammalian cell cycle. *Mol Cell Biol* 23:5706-5715.
- Rouaud, P., C. Vincent-Fabert, A. Saintamand, R. Fiancette, M. Marquet, I. Robert, B. Reina-San-Martin, E. Pinaud, M. Cogne, and Y. Denizot. 2013. The IgH 3' regulatory region controls somatic hypermutation in germinal center B cells. *J Exp Med* 210:1501-1507.
- Roux, K.J., D.I. Kim, M. Raida, and B. Burke. 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* 196:801-810.
- Rush, J.S., M. Liu, V.H. Odegard, S. Unniraman, and D.G. Schatz. 2005. Expression of activation-induced cytidine deaminase is regulated by cell division, providing a mechanistic basis for division-linked class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* 102:13242-13247.

- Sabouri, S., M. Kobayashi, N.A. Begum, J. Xu, K. Hirota, and T. Honjo. 2014. C-terminal region of activation-induced cytidine deaminase (AID) is required for efficient class switch recombination and gene conversion. *Proc Natl Acad Sci USA* 111:2253-2258.
- Sakaguchi, N., T. Kimura, S. Matsushita, S. Fujimura, J. Shibata, M. Araki, T. Sakamoto, C. Minoda, and K. Kuwahara. 2005. Generation of High-Affinity Antibody against T Cell-Dependent Antigen in the Ganp Gene-Transgenic Mouse. *Journal of immunology (Baltimore, Md. : 1950)* 174:4485-4494.
- Sale, J.E., D.M. Calandrini, M. Takata, S. Takeda, and M.S. Neuberger. 2001. Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. *Nature* 412:921-926.
- Saribasak, H., R.W. Maul, Z. Cao, R.L. McClure, W. Yang, D.R. McNeill, D.M. Wilson, 3rd, and P.J. Gearhart. 2011. XRCC1 suppresses somatic hypermutation and promotes alternative nonhomologous end joining in Igh genes. *J Exp Med* 208:2209-2216.
- Saribasak, H., R.W. Maul, Z. Cao, W.W. Yang, D. Schenten, S. Kracker, and P.J. Gearhart. 2012. DNA polymerase zeta generates tandem mutations in immunoglobulin variable regions. *J Exp Med* 209:1075-1081.
- Saribasak, H., N.N. Saribasak, F.M. Ipek, J.W. Ellwart, H. Arakawa, and J.M. Buerstedde. 2006. Uracil DNA glycosylase disruption blocks Ig gene conversion and induces transition mutations. *Journal of immunology* 176:365-371.
- Sato, Y., H.C. Probst, R. Tatsumi, Y. Ikeuchi, M.S. Neuberger, and C. Rada. 2010. Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *J Biol Chem* 285:7111-7118.
- Sawarkar, R., C. Sievers, and R. Paro. 2012. Hsp90 globally targets paused RNA polymerase to regulate gene expression in response to environmental stimuli. *Cell* 149:807-818.
- Sawasdichai, A., H.T. Chen, N. Abdul Hamid, P.S. Jayaraman, and K. Gaston. 2010. In situ subcellular fractionation of adherent and non-adherent mammalian cells. *J Vis Exp*
- Schrader, C.E., J.E.J. Guikema, E.K. Linehan, E. Selsing, and J. Stavnezer. 2007. Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *Journal of immunology (Baltimore, Md. : 1950)* 179:6064-6071.
- Schroeder, H.W., Jr. 2006. Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Dev Comp Immunol* 30:119-135.
- Scornik, J.C., H. Cosenza, W. Lee, H. Kohler, and D.A. Rowley. 1974. Antibody-dependent cell-mediated cytotoxicity. I. Differentiation from antibody-independent cytotoxicity by "normal" IgG. *Journal of immunology* 113:1510-1518.

- Sernández, I.V., V.G. de Yébenes, Y. Dorsett, and A.R. Ramiro. 2008. Haploinsufficiency of activation-induced deaminase for antibody diversification and chromosome translocations both in vitro and in vivo. *PloS one* 3:e3927.
- Severi, F., A. Chicca, and S.G. Conticello. 2011. Analysis of reptilian APOBEC1 suggests that RNA editing may not be its ancestral function. *Mol Biol Evol* 28:1125-1129.
- Shanbhag, N.M., I.U. Rafalska-Metcalf, C. Balane-Bolivar, S.M. Janicki, and R.A. Greenberg. 2010. ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell* 141:970-981.
- Sharbeen, G., C.W.Y. Yee, A.L. Smith, and C.J. Jolly. 2012. Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase. *The Journal of experimental medicine* 209:965-974.
- Shen, H.M., A. Peters, B. Baron, X. Zhu, and U. Storb. 1998. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science (New York, N.Y.)* 280:1750-1752.
- Shen, H.M., A. Tanaka, G. Bozek, D. Nicolae, and U. Storb. 2006. Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice. *Journal of immunology* 177:5386-5392.
- Shimatani, Z., S. Kashojiya, M. Takayama, R. Terada, T. Arazoe, H. Ishii, H. Teramura, T. Yamamoto, H. Komatsu, K. Miura, H. Ezura, K. Nishida, T. Ariizumi, and A. Kondo. 2017. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat Biotechnol* 35:441-443.
- Shimizu, T., H. Marusawa, Y. Endo, and T. Chiba. 2012. Inflammation-mediated genomic instability: roles of activation-induced cytidine deaminase in carcinogenesis. *Cancer Sci* 103:1201-1206.
- Shinkura, R., S. Ito, N.A. Begum, H. Nagaoka, M. Muramatsu, K. Kinoshita, Y. Sakakibara, H. Hijikata, and T. Honjo. 2004. Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nature Immunology* 5:707-712.
- Shteynberg, D., E.W. Deutsch, H. Lam, J.K. Eng, Z. Sun, N. Tasman, L. Mendoza, R.L. Moritz, R. Aebersold, and A.I. Nesvizhskii. 2011. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol Cell Proteomics* 10:M111 007690.
- Simpson, L.J., and J.E. Sale. 2003. Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line. *EMBO J* 22:1654-1664.
- Singh, S.K., K. Maeda, M.M.A. Eid, S.A. Almofty, M. Ono, P. Pham, M.F. Goodman, and N. Sakaguchi. 2013. GANP regulates recruitment of AID to immunoglobulin variable regions by modulating transcription and nucleosome occupancy. *Nature communications* 4:1830.

- Skourti-Stathaki, K., K. Kamieniarz-Gdula, and N.J. Proudfoot. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 516:436-439.
- Skuse, G.R., A.J. Cappione, M. Sowden, L.J. Metheny, and H.C. Smith. 1996. The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing. *Nucleic Acids Res* 24:478-485.
- Stanlie, A., M. Aida, M. Muramatsu, T. Honjo, and N.A. Begum. 2010. Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* 107:22190-22195.
- Stanlie, A., N.A. Begum, H. Akiyama, and T. Honjo. 2012. The DSIF subunits Spt4 and Spt5 have distinct roles at various phases of immunoglobulin class switch recombination. *PLoS Genetics* 8:e1002675.
- Staszewski, O., R.E. Baker, A.J. Ucher, R. Martier, J. Stavnezer, and J.E. Guikema. 2011a. Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Mol Cell* 41:232-242.
- Staszewski, O., R.E. Baker, A.J. Ucher, R. Martier, J. Stavnezer, and J.E.J. Guikema. 2011b. Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Molecular cell* 41:232-242.
- Stavnezer, J. 1996. Immunoglobulin class switching. *Curr Opin Immunol* 8:199-205.
- Stavnezer, J., and C.T. Amemiya. 2004. Evolution of isotype switching. *Semin Immunol* 16:257-275.
- Stavnezer, J., J.E. Guikema, and C.E. Schrader. 2008. Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 26:261-292.
- Stavnezer, J., and C.E. Schrader. 2014. IgH chain class switch recombination: mechanism and regulation. *Journal of immunology* 193:5370-5378.
- Stenglein, M.D., H. Matsuo, and R.S. Harris. 2008. Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *J Virol* 82:9591-9599.
- Storb, U. 2014. Why does somatic hypermutation by AID require transcription of its target genes? *Advances in immunology* 122:253-277.
- Sun, J., C.D. Keim, J. Wang, D. Kazadi, P.M. Oliver, R. Rabadan, and U. Basu. 2013a. E3-ubiquitin ligase Nedd4 determines the fate of AID-associated RNA polymerase II in B cells. *Genes & development* 27:1821-1833.
- Sun, Q., Y.P. Carrasco, Y. Hu, X. Guo, H. Mirzaei, J. Macmillan, and Y.M. Chook. 2013b. Nuclear export inhibition through covalent conjugation and hydrolysis of Leptomycin B by CRM1. *Proceedings of the National Academy of Sciences of the United States of America* 110:1303-1308.
- Supek, F., and B. Lehner. 2017. Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. *Cell* 170:534-547 e523.

- Suurmond, J., and B. Diamond. 2015. Autoantibodies in systemic autoimmune diseases: specificity and pathogenicity. *J Clin Invest* 125:2194-2202.
- Swaminathan, S., L. Klemm, E. Park, E. Papaemmanuil, A. Ford, S.-M. Kweon, D. Trageser, B. Hasselfeld, N. Henke, J. Mooster, H. Geng, K. Schwarz, S.C. Kogan, R. Casellas, D.G. Schatz, M.R. Lieber, M.F. Greaves, and M. MUschen. 2015. Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. *Nature Immunology* 16:766-774.
- Ta, V.-T., H. Nagaoka, N. Catalan, A.H. Durandy, A. Fischer, K. Imai, S. Nonoyama, J. Tashiro, M. Ikegawa, S. Ito, K. Kinoshita, M. Muramatsu, and T. Honjo. 2003. AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nature Immunology* 4:843-848.
- Takizawa, M., H. Tolarová, Z. Li, W. Dubois, S. Lim, E. Callen, S. Franco, M. Mosaico, L. Feigenbaum, F.W. Alt, A. Nussenzweig, M. Potter, and R. Casellas. 2008. AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *The Journal of experimental medicine* 205:1949-1957.
- Tang, E.S., and A. Martin. 2007. Immunoglobulin gene conversion: synthesizing antibody diversification and DNA repair. *DNA Repair (Amst)* 6:1557-1571.
- Taylor, B.J., S. Nik-Zainal, Y.L. Wu, L.A. Stebbings, K. Raine, P.J. Campbell, C. Rada, M.R. Stratton, and M.S. Neuberger. 2013. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* 2:e00534.
- Taylor, S.A., D.J. Giroux, K.A. Jaeckle, T.J. Panella, S.R. Dakhil, and S.C. Schold. 1998. Phase II study of Didemnin B in central nervous system tumors: a Southwest Oncology Group study. *Invest New Drugs* 16:331-332.
- Teng, B., C.F. Burant, and N.O. Davidson. 1993. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* 260:1816-1819.
- Teng, G., P. Hakimpour, P. Landgraf, A. Rice, T. Tuschl, R. Casellas, and F.N. Papavasiliou. 2008. MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* 28:621-629.
- Tepper, S., J. Jeschke, K. Böttcher, A. Schmidt, K. Davari, P. Müller, E. Kremmer, P. Hemmerich, I. Pfeil, and B. Jungnickel. 2016. PARP activation promotes nuclear AID accumulation in lymphoma cells. *Oncotarget* 7:13197-13208.
- Thomas-Claudepierre, A.-S., E. Schiavo, V. Heyer, M. Fournier, A. Page, I. Robert, and B.R. Reina-San-Martin. 2013. The cohesin complex regulates immunoglobulin class switch recombination. *The Journal of experimental medicine* 210:2495-2502.
- Thomas-Claudepierre, A.S., I. Robert, P.P. Rocha, R. Raviram, E. Schiavo, V. Heyer, R. Bonneau, V.M. Luo, J.K. Reddy, T. Borggrefe, J.A. Skok, and B. Reina-San-Martin. 2016. Mediator facilitates transcriptional activation and dynamic long-range

- contacts at the IgH locus during class switch recombination. *J Exp Med* 213:303-312.
- Tiller, T., M. Tsuiji, S. Yurasov, K. Velinzon, M.C. Nussenzweig, and H. Wardemann. 2007. Autoreactivity in human IgG⁺ memory B cells. *Immunity* 26:205-213.
- Trachsel, H. 1996. Isolation and characterization of eukaryotic translation initiation, elongation and termination factors. In O. Resnekov, and A. Gabain, editors. Springer Berlin Heidelberg, 165-176.
- Treanor, B. 2012. B-cell receptor: from resting state to activate. *Immunology* 136:21-27.
- Ucher, A.J., S. Ranjit, T. Kadungure, E.K. Linehan, L. Khair, E. Xie, J. Limauro, K.S. Rauch, C.E. Schrader, and J. Stavnezer. 2014. Mismatch repair proteins and AID activity are required for the dominant negative function of C-terminally deleted AID in class switching. *Journal of immunology (Baltimore, Md. : 1950)* 193:1440-1450.
- Uchimura, Y., L.F. Barton, C. Rada, and M.S. Neuberger. 2011. REG- γ associates with and modulates the abundance of nuclear activation-induced deaminase. *The Journal of experimental medicine* 208:2385-2391.
- Vaidyanathan, B., W.-F. Yen, J.N. Pucella, and J. Chaudhuri. 2014. AIDing Chromatin and Transcription-Coupled Orchestration of Immunoglobulin Class-Switch Recombination. *Frontiers in immunology* 5:120.
- Victoria, G.D., and M.C. Nussenzweig. 2012. Germinal centers. *Annu Rev Immunol* 30:429-457.
- Victoria, G.D., T.A. Schwickert, D.R. Fooksman, A.O. Kamphorst, M. Meyer-Hermann, M.L. Dustin, and M.C. Nussenzweig. 2010. Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell* 143:592-605.
- Vuong, B.Q., and J. Chaudhuri. 2012. Combinatorial mechanisms regulating AID-dependent DNA deamination: interacting proteins and post-translational modifications. *Seminars in immunology* 24:264-272.
- Vuong, B.Q., K. Herrick-Reynolds, B. Vaidyanathan, J.N. Pucella, A.J. Ucher, N.M. Donghia, X. Gu, L. Nicolas, U. Nowak, N. Rahman, M.P. Strout, K.D. Mills, J. Stavnezer, and J. Chaudhuri. 2013. A DNA break- and phosphorylation-dependent positive feedback loop promotes immunoglobulin class-switch recombination. *Nature Immunology* 14:1183-1189.
- Wada, T., T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, S. Hirose, S. Sugimoto, K. Yano, G.A. Hartzog, F. Winston, S. Buratowski, and H. Handa. 1998. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & development* 12:343-356.

- Wagner, S.D., C. Milstein, and M.S. Neuberger. 1995. Codon bias targets mutation. *Nature* 376:732.
- Wang, L., R. Wuerffel, S. Feldman, A.A. Khamlichi, and A.L. Kenter. 2009a. S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *The Journal of experimental medicine* 206:1817-1830.
- Wang, M., C. Rada, and M.S. Neuberger. 2010. Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *The Journal of experimental medicine* 207:141-153.
- Wang, M., Z. Yang, C. Rada, and M.S. Neuberger. 2009b. AID upmutants isolated using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity. *Nature structural & molecular biology* 16:769-776.
- Wang, Q., K.-R. Kieffer-Kwon, T.Y. Oliveira, C.T. Mayer, K. Yao, J. Pai, Z. Cao, M. Dose, R. Casellas, M. Jankovic, M.C. Nussenzweig, and D.F. Robbiani. 2017. The cell cycle restricts activation-induced cytidine deaminase activity to early G1. *The Journal of experimental medicine* 214:49-58.
- Wang, X., M. Fan, S. Kalis, L. Wei, and M.D. Scharff. 2014. A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation. *Nature communications* 5:4137.
- Ward, I.M., B. Reina-San-Martin, A. Oлару, K. Minn, K. Tamada, J.S. Lau, M. Cascalho, L. Chen, A. Nussenzweig, F. Livak, M.C. Nussenzweig, and J. Chen. 2004. 53BP1 is required for class switch recombination. *J Cell Biol* 165:459-464.
- Watashi, K., G. Liang, M. Iwamoto, H. Marusawa, N. Uchida, T. Daito, K. Kitamura, M. Muramatsu, H. Ohashi, T. Kiyohara, R. Suzuki, J. Li, S. Tong, Y. Tanaka, K. Murata, H. Aizaki, and T. Wakita. 2013. Interleukin-1 and tumor necrosis factor- α trigger restriction of hepatitis B virus infection via a cytidine deaminase activation-induced cytidine deaminase (AID). *J Biol Chem* 288:31715-31727.
- Weinstein, P.D., A.O. Anderson, and R.G. Mage. 1994. Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms. *Immunity* 1:647-659.
- Westermann, J., and R. Pabst. 1992. Distribution of lymphocyte subsets and natural killer cells in the human body. *Clin Invest* 70:539-544.
- White, E., M. Schlackow, K. Kamieniarz-Gdula, N.J. Proudfoot, and M. Gullerova. 2014. Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. *Nat Struct Mol Biol* 21:552-559.
- Wickramasinghe, V.O., P.I. McMurtrie, A.D. Mills, Y. Takei, S. Penrhyn-Lowe, Y. Amagase, S. Main, J. Marr, M. Stewart, and R.A. Laskey. 2010. mRNA export from mammalian cell nuclei is dependent on GANP. *Curr Biol* 20:25-31.
- Wiesendanger, M., B. Kneitz, W. Edelmann, and M.D. Scharff. 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* 191:579-584.
- Willmann, K.L., S. Milosevic, S. Pauklin, K.-M. Schmitz, G. Rangam, M.T. Simon, S. Maslen, M. Skehel, I. Robert, V. Heyer, E. Schiavo, B.R. Reina-San-Martin, and S.K. Petersen-Mahrt. 2012. A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *The Journal of experimental medicine* 209:2099-2111.
- Wilson, T.M., A. Vaisman, S.A. Martomo, P. Sullivan, L. Lan, F. Hanaoka, A. Yasui, R. Woodgate, and P.J. Gearhart. 2005. MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. *J Exp Med* 201:637-645.
- Woyach, J.A., A.J. Johnson, and J.C. Byrd. 2012. The B-cell receptor signaling pathway as a therapeutic target in CLL. *Blood* 120:1175-1184.
- Wu, X., and J. Stavnezer. 2007. DNA polymerase beta is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination. *J Exp Med* 204:1677-1689.
- Wu, X., and Y. Zhang. 2017. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet*
- Wuerffel, R., L. Wang, F. Grigera, J. Manis, E. Selsing, T. Perlot, F.W. Alt, M. Cogne, E. Pinaud, and A.L. Kenter. 2007. S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* 27:711-722.
- Xu, X., H.C. Hsu, J. Chen, W.E. Grizzle, W.W. Chatham, C.R. Stockard, Q. Wu, P.A. Yang, V.M. Holers, and J.D. Mountz. 2009. Increased expression of activation-induced cytidine deaminase is associated with anti-CCP and rheumatoid factor in rheumatoid arthritis. *Scandinavian journal of immunology* 70:309-316.
- Xu, Z., Z. Fulop, G. Wu, E.J. Pone, J. Zhang, T. Mai, L.M. Thomas, A. Al-Qahtani, C.A. White, S.R. Park, P. Steinacker, Z. Li, J. Yates, 3rd, B. Herron, M. Otto, H. Zan, H. Fu, and P. Casali. 2010. 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. *Nat Struct Mol Biol* 17:1124-1135.
- Xue, K., C. Rada, and M.S. Neuberger. 2006. The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in *msh2*^{-/-} *ung*^{-/-} mice. *The Journal of experimental medicine* 203:2085-2094.
- Yamada, T., Y. Yamaguchi, N. Inukai, S. Okamoto, T. Mura, and H. Handa. 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell* 21:227-237.
- Yamanaka, S., K.S. Poksay, K.S. Arnold, and T.L. Innerarity. 1997. A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme. *Genes Dev* 11:321-333.
- Yamane, A., W. Resch, N. Kuo, S. Kuchen, Z. Li, H.-w. Sun, D.F. Robbiani, K. McBride, M.C. Nussenzweig, and R. Casellas. 2011. Deep-sequencing identification of the

- genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nature Immunology* 12:62-69.
- Yan, C.T., C. Boboila, E.K. Souza, S. Franco, T.R. Hickernell, M. Murphy, S. Gumaste, M. Geyer, A.A. Zarrin, J.P. Manis, K. Rajewsky, and F.W. Alt. 2007. IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* 449:478-482.
- Yang, L., A.W. Briggs, W.L. Chew, P. Mali, M. Guell, J. Aach, D.B. Goodman, D. Cox, Y. Kan, E. Lesha, V. Soundararajan, F. Zhang, and G. Church. 2016. Engineering and optimising deaminase fusions for genome editing. *Nat Commun* 7:13330.
- Yeap, L.-S., J.K. Hwang, Z. Du, R.M. Meyers, F.-L. Meng, A. Jakubauskaitė, M. Liu, V. Mani, D. Neuberg, T.B. Kepler, J.H. Wang, and F.W. Alt. 2015. Sequence-Intrinsic Mechanisms that Target AID Mutational Outcomes on Antibody Genes. *Cell* 163:1124-1137.
- Young, J.C., V.R. Agashe, K. Siegers, and F.U. Hartl. 2004. Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol* 5:781-791.
- Yu, K., F. Chedin, C.-L. Hsieh, T.E. Wilson, and M.R. Lieber. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature Immunology* 4:442-451.
- Yuseff, M.I., P. Pierobon, A. Reversat, and A.M. Lennon-Dumenil. 2013. How B cells capture, process and present antigens: a crucial role for cell polarity. *Nat Rev Immunol* 13:475-486.
- Zahn, A., A.K. Eranki, A.-M. Patenaude, S.P. Methot, H. Fifield, E.M. Cortizas, P. Foster, K. Imai, A.H. Durandy, M. Larijani, R.E. Verdun, and J.M. Di Noia. 2014. Activation induced deaminase C-terminal domain links DNA breaks to end protection and repair during class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* 111:E988-997.
- Zan, H., and P. Casali. 2013. Regulation of Aicda expression and AID activity. *Autoimmunity* 46:83-101.
- Zan, H., J. Zhang, S. Ardeshtna, Z. Xu, S.-R. Park, and P. Casali. 2009. Lupus-prone MRL/faslpr/lpr mice display increased AID expression and extensive DNA lesions, comprising deletions and insertions, in the immunoglobulin locus: concurrent upregulation of somatic hypermutation and class switch DNA recombination. *Autoimmunity* 42:89-103.
- Zarrin, A.A., F.W. Alt, J. Chaudhuri, N. Stokes, D. Kaushal, L. Du Pasquier, and M. Tian. 2004. An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nat Immunol* 5:1275-1281.
- Zarrin, A.A., C. Del Vecchio, E. Tseng, M. Gleason, P. Zarin, M. Tian, and F.W. Alt. 2007. Antibody class switching mediated by yeast endonuclease-generated DNA breaks. *Science* 315:377-381.

- Zeng, X., D.B. Winter, C. Kasmer, K.H. Kraemer, A.R. Lehmann, and P.J. Gearhart. 2001. DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol* 2:537-541.
- Zheng, S., B.Q. Vuong, B. Vaidyanathan, J.-Y. Lin, F.-T. Huang, and J. Chaudhuri. 2015. Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA. *Cell* 161:762-773.