Université Paris Descartes

B3MI : Biochimie, Biothérapies, Biologie Moléculaire et l'Infectiologie

Institut Necker Enfants Malades/INSERM U783 « Développement du système immunitaire »

L'hypermutation somatique des gènes des immunoglobulines:

Corrélation avec le cycle cellulaire et contribution des voies de réparation mutagènes

Par Marija ZIVOJNOVIC

Thèse de doctorat de Immunologie

Dirigée par Claude Agnès REYNAUD et Sébastien STORCK

Présentée et soutenue publiquement le 26/11/2013

Devant un jury composé de :
RADMAN Miroslav, PU-PH et Professeur émérite, PRESIDENT DU JURY
KANNOUCHE Patricia, Directeur de l'équipe, RAPPORTEUR
KHAMLICHI Ahmed Amine, Directeur de l'équipe, RAPPORTEUR
COGNE Michel, PU-PH et Directeur de l'équipe, EXAMINATEUR
REVY Patrick, CR1, EXAMINATEUR

Résumé (en français) :

Pour augmenter l'affinité des anticorps sécrétés en réponse à un antigène, les gènes d'immunoglobulines subissent l'hypermutation somatique, une mutagénèse adaptative initiée par l'action de l'*a*ctivation-*i*nduced cytidine *d*eaminase (AID). L'uracile provenant de la désamination des cytosines par cette enzyme est réparé de façon erronée par la suite : si il est pris en charge par l'*u*racile *N*-*g*lycosylase (UNG), enzyme à l'origine d'une réparation poursuivie habituellement par des composantes de la voie du "base-excision repair", il reste à sa place un site abasique franchissable par les ADN polymérases translésionnelles avec un taux d'erreur très élevé. Si le mésappariement U:G est reconnu par la voie du « mismatch repair », le brin d'ADN entourant le U est dégradé puis néo-synthétisé par une autre ADN polymérase translésionnelle particulièrement mutagène en face des bases T et A, la polymérase eta.

Nous avons proposé que le choix entre ces deux voies de réparation mutagènes puisse être régulé en fonction du cycle cellulaire: les mutations des paires A:T seraient introduites dans les gènes d'immunoglobulines par la voie du mismatch repair en phase G1 alors que la voie erronée d'UN introduirait les autres mutations lors de la phase S. Nous sommes parvenus à restreindre l'activité de l'AID à deux parties distinctes du cycle, la phase G1 ou les phases S/G2/M, et nous avons établi le fonctionnement de ce système dans le modèle murin. De façon surprenante, nous avons détecté un taux de mutation proche du bruit de fond chez toutes les souris dont l'AID opérait uniquement dans les phases S/G2/M. Par contre, les souris dont l'AID a été restreinte en G1 présentaient un spectre de mutation diversifié sur les quatre bases et similaire au normal. A la lumière de ces résultats, nous proposons que les lésions introduites tout au long du cycle par l'AID soient diversifiées par les acteurs de l'hypermutation somatique pendant la phase G1, alors que les lésions seraient soit réparées de façon fidèle en dehors de cette phase-là, soit de faible impact.

Afin d'expliquer le biais de brin dans l'hypermutation somatique observé pour les mutations sur les bases A :T, nous proposons pour l'ADN polymérase eta un rôle inhabituel de réparation du brin portant la « lésion », et non de synthèse translésionnelle classique en face de cette lésion. Nous avons analysé le profil, le taux et la distribution des mutations introduites par Pol eta sur un oligonucléotide cible pour l'hypermutation, qui a été inséré au locus des immunoglobulines et utilisé pour l'établissement des souris knock-in avec un fond génétique déficient ou non en UNG. Nos résultats, selon lesquels Pol eta continue de cibler le brin codant indépendamment de la localisation des « points d'entrée » en forme d'uraciles, contredisent les rapports déjà publiés sur ce sujet. De façon inattendue, nos résultats mettent en évidence une coopération entre les voies UNG et et les activités endonucléasique du mismatch repair, fournissant la cassure simple brin qui va permettre d'initier la resynthèse à fort taux d'erreur à l'origine de la mutagénèse A/T. Ces résultats résolvent aussi le paradoxe de la non-participation apparente du complexe effecteur du mismatch repair (Mlh1/Pms2) dans le processus d'hypermutation, en proposant qu'il fonctionne en redondance avec UNG, dans une distribution des tâches qui dépend du contexte de la séquence ciblée et de la densité du processus de deamination.

Mots clés (français) : lymphocyte B, hypermutation somatique, gènes des immunoglobulines, maturation d'affinité, AID, polymérases translesionnelles, cycle cellulaire, dégrons

Abstract:

Somatic hypermutation is a localized mutagenesis, essentially targeted to the immunoglobulin V region, and occurring during the immune response. This process is triggered by AID (activation-induced cytidine deaminase) that deaminates cytosines into uracils at the Ig locus. This lesion is further processed by Ung or the Msh2-Msh6 complex, with an abnormal outcome for both pathways that results in an increased mutation load. The Msh2-Msh6 complex recruits Pol eta to generate a short patch DNA synthesis with mostly mutations at A and T bases.

To get further insight into this error-prone repair process, we have generated hypermutation substrates consisting in an A/T oligonucleotide of 100 bases with or without 3 cytidines in its core region, inserted by knock-in at the heavy chain Ig locus. Our aim was to compare the mutation frequency, distribution and mutation profile of substrates with C on either the coding or the non-coding strand on WT or Ung-deficient background, taking into account that Pol eta is a preferred A to G mutator. Our results suggest that Pol eta resynthesis may proceed on the coding strand, whatever the strand localization of the uracil, thus contradicting previous reports. Unexpectedly, our results revealed a cooperation between the Ung pathway and the endonuclease activity of the mismatch repair, with both of them providing the single-strand nick that allows initiation of the error-prone process that generates mutations at A and T bases. These results resolve the apparent paradox of the non-involvement of the mismatch repair effector complex (Mlh1-Pms2) in hypermutation, by proposing that it works redundantly with UNG, in a distribution of tasks that will depend upon the sequence context and the intensity of deamination activity.

We have also constructed cell cycle restricted mutants of AID, to study in which phase of the cell cycle this atypical, mismatch repair driven, error-prone synthesis is taking place. Using the Fucci restriction system (degrons based on Cdt1 or Geminin peptides), we have generated AID constructs with proper restriction in either G1 or S/G2/M phases. These retroviral constructs have been used to transduce mouse hematopoietic stem cells from either AID -deficient mice and to restore immunodeficient animals, in order to analyze their immune response. We report that restriction of AID expression in S/G2/M part of the cycle yielded only background mutation frequency, while AID operating in the G1 phase is able to generate an equal proportion of A/T and G/C mutations at the Ig loci, thus demonstrating that uracils generated in G1 are substrates for both Ung- and mismatch repair pathways.

Keywords : B lymphocytes, somatic hypermutation, immunoglobulin genes, affinity maturation, activation-induced deaminase, translesional polymerases, cell cycle, degrons

Aux professeurs de biologie qui m'ont marquée au long de mon parcours scolaire,

Slobodanka Dumović

Gordana Dronjak

Biljana Nikolić

Nada Šerban

Remerciements

Commençons par expliquer pourquoi je rédige mes remerciements en français alors que toute la thèse est en anglais : le français restera la langue de ma période de doctorat et en tout de mes 7 dernières années passées en France : années de doute, de joie, des nouveaux débuts et des profonds changements personnels.

Je tiens à remercier ma directrice de thèse, Claude-Agnès Reynaud et le co-directeur scientifique de l'unité, Jean-Claude Weill de m'avoir permis d'éffectuer la thèse au sein de leur laboratoire, de m'avoir aidé tout au long de mon parcours scientifique aussi bien que personnel par des discussions fructueuses et un soutien sans réserve.

Je garde un sentiment de remerciement profond pour mon encadrant direct pendant mon stage de Master et tout au long de mon doctorat, Sébastien Storck. C'était loin d'être facile, et beaucoup de fois je me posais des questions si c'était mon karma (ou le tien – insert smiley here !) – mais quelques soupirs et quelques « bon couraggio » prononcés, et tout le malheur s'envole ! Merci de m'avoir appris d'être persévérante, plus tolérante et plus humble (même s'il reste encore du travail à faire….)

Un grand merci aussi à Olivier Albagli-Curiel pour son aide apporté lors de la mise en route des expériences touchant aux retrovirus, les transductions et l'enceinte P2. Je trouve notre collaboration sur le project du cycle cellulaire extrêmement enrichissante et je te remercie chaleureusement pour ton soutien scientifique et moral.

Je veux remercier ici aussi tous les membres du laboratoire U783 pour leur soutien et bonne humeur pendant ces quatre années :, Damiana et à son bout de choux, mes voisines du bureau pour nos échanges artistiques, culinaires et artisanales ; Marc, mon compatriote dans la souffrance apportée par la thèse ; Rémi, pour toute aide informatique et pratique dans les moment cruciaux ; Sandra, pour son bonne humeur et son esprit imbattable… dans tous les sens du mot ! à Matthieu, pour sa musique enchanteresse et pour son pouvoir d'être partout à tout moment, labo, hôpital, EFS, famille… à Barbara, pour son esprit philosophique à l'italienne… à Yi et à Davide, pour nos discussions à travers nos repas ou les cartons de déménagement, à Giulia pour la danse et pour son esprit rigoleur, à Delphine pour nos bons souvenirs du cours Pasteur. Un merci à tous les membres anciens du labo que j'ai eu le plaisir de croiser pendant ma thèse : Said, Ismael, Laïla, Vanessa, Amélie, Rosa et Lan.

Ma chère Annie, notre Lady Di Smet, ma « maman à la française » qui ne me laissait pas passer Noël toute seule (même si ce Noël-là, je ne le fête pas...) ; François pour son soutien moral et pratique, pour l'aventure de son cours Pasteur et pour d'innombrables fois

qu'il m'a épargné le stress d'une manipulation... très particulière, on sait laquelle ; à Lucie, pour nos chit-chats sans cesse et des échanges pour tout ce qui est le sujet des filles ; à Simon, mon voisin de bureau de gauche, pour tous les livres qui m'a faits découvrir, toutes les séries dont il m'a fait accro, toutes les recettes qui ont fait saliver le labo entier et toutes les conneries autour d'un certain chat nommé Neko.

Je ne vais pas oublier ici Fred, pour nos rigolades sur les paillasses, pour sa rigueur scientifique et aussi pour ses super-pouvoirs de faire la science, faire du sport, jouer dans un groupe et être papa à la fois ; et enfin à Valérie, ma très chère amie, merci d'être là aux moments difficiles – nous en avons eu beaucoup les deux pendant ses quatre ans et demi, ma belle étrangère !

Je tiens à remercier tous mes amis, français, serbes et de toutes les autres nationalités qui me demanderont un annexe entier pour les dénombrer, mais une pensée spéciale pour ceux qui m'ont épaulé aux moments cruciaux : Lucie, Emma, Ludivine, Pascale, Victoria, Ni Ra, Emili, Kristel, Emilija i Ksenija, Rosa, Donovan, Louise, Gilbert et Germán. Bande de thésards et non-thésards, vous m'aviez supporté et vous m'avez fait rire, je me trouve très chanceuse d'avoir eu votre connaissance. Je m'excuse à tous ceux à qui j'ai manqué d'exprimer ma gratitude ici.

Bien sûr, un grand merci à mes parents et à mon frère Marko qui, même à distance, arrivaient à m'accabler d'amour et de compréhension, à me persuader ou dissuader si nécessaire, à m'écouter et essayer de comprendre tous les hauts et les bas d'un doctorat dans un pays étranger. Vous savez bien à quel point je vous aime et à quel point je suis paniquée de présenter devant vous, pour la première fois de ma vie.

Je ne pourrais pas oublier de remercier mon compagnon, Nikola, pour sa présence et pour son caractère tellement complémentaire du mien, justement aux moments quand il fallait : rédiger nos thèses en même temps promettait d'être un enfer, mais nous étions là l'un pour l'autre et nous voilà réussis. Devant la période encore plus difficile qui nous attend, nous en sortirons ensemble. Comme nous l'avons fait depuis une éternité déjà.

Table of Contents

Table des illustrations

Avant-propos

"It takes all the running you can do to keep in the same place"

These are the words the Red Queen used to explain to Alice how Wonderland runs; indeed, one cannot help the feeling of being in Wonderland once they enter into the intricacies of vertebrate adaptive immune system. Immunity "Wonderland" is built on the constant arms race between the innovative mechanisms of **pathogen** infection and the **host** recognition/reaction system. Both sides multiply their arms by using the one key parameter: **diversity**. If one side succeeds in evading the other by means of diversified attack or dissimulation mechanisms, this means only one thing for the winner: keep in the same place, or in more biological terms, survive.

Keeping in the same place means as well that a pathogen cannot support diversity by entirely changing all parts of its molecular machinery that already proved to work well in the course of evolution. Some parts of the pathogen will inevitably remain the same and hence even resemble among pathogen species. The immune system knows how to take advantage of these invariable components, termed **PAMPs (pathogen-associated molecular pattern**). **Innate immune response**, one arm of the vertebrate immune system that proved to be extremely efficient in recognizing these generalized patterns, engages a set of effector molecules and cells to quickly eradicate the invading pathogen. Therefore, the innate component of the immune system is the first defense line when pathogen gains access through the organism entry points, like skin, respiratory and digestive ways, or blood.

On the other hand, extremely short generation time allows the pathogen to quickly overwhelm the innate response and, more importantly, change rapidly its tactics for the attack and escape the immune defense. In some cases pathogens can dissimulate their PAMPs or resist the destruction mechanism deployed on them by the innate immune system. Therefore, a more targeted action is required to recognize the pathogen separately from the standard surface characters and elicit highly specific combating tactics. This is the moment where **adaptive immune system** comes in and helps the organism not only clear the pathogen, but also perpetuate a memory of it and mount a defensive response up to several times faster in case of a repeated infection.

Here, the demand of the Red Queen seems unfeasible for a vertebrate adaptive immune system. How to stay in the run when running means to cope with the infinite diversity of all pathogen moieties (**antigens**), by recognizing each of them with extreme specificity? Immunoglobulin and T-cell receptor genes, encoding the molecular effectors of

the adaptive immune system had only one thing left to do: diversify as fast as the ever changing pathogen in hope that they are innovative enough to win the arms race.

Characteristics of the adaptive immune response

An adaptive immune response takes place through two phases: recognition phase allows specialized antigen-presenting cells (APC) to take up, process and present the antigen to the adequate mediator cells, T and B cells; an effector phase will allow the mediator cells that had specifically recognized the antigen to differentiate into effector cells and fight for antigen elimination. In the way we have just described it, an immune response implicate both arms of an individual's immune system: the innate system comprising macrophages, granulocytes, NK cells and dendritic cells which play the role of APCs, and the adaptive immune system, including the humoral and the cellular type of adaptive responses that will launch the effector phase.

Humoral immunity provides for combative molecular complexes, **antibodies or immunoglobulins** which specifically bind and counteract the harmful effect of non-self molecules or microbes. Antibodies can exist as cell surface proteins in the membrane of the **B lymphocyte** that produces them, and recognize specific antigens as B-cell antigen-receptors. Once a B-cell has bound a specific antigen and receives help from other mediator cells, it becomes activated and differentiates into an effector plasma cell that will secrete the antibody in body fluids and blood. Once the antibody binds to the antigen, this immune complex can be either eliminated by complement factors, or phagocyted by macrophages and granulocytes more efficiently than bare pathogens thanks to the recognition of the antibody portion of the immune complex through special **FcR receptors** on the surface of the macrophage's membrane.

The cellular component of the adaptive immune system comprises **T-cells**, which can adopt a regulatory, suppressor or effector role. T-cells can detect and bind antigen through their membrane bound **T-cell receptor (TCR)**. However, TCR can recognize antigen only if it is presented to the T-cell by an APC, like a motile dendritic cells that has encountered the antigen while sampling for them in peripheral tissues. Once they recognize the specific antigen, T-cells will undergo the differentiation from mediator to effector cells and become either helper T-cells or cytotoxic T-cells. Cytotoxic T-cells specifically target somatic cells that got already infected or transformed by the antigen-bearing pathogen and eliminate them by enzyme-mediated lysis. Helper T-cell play a part in activating B-cells, helping them to become antibody producers and directing the best suited antibodies according to the previously estimated nature of the invading antigen (parasites, viruses or bacteria).

Unlike innate immune responses, adaptive immune reactions display the following properties:

- they take time to establish (up to 5 days to detect antibodies in the blood after the exposure to the infectious agent);
- they are by far more specific than innate reactions;
- they have memory: they recognize better the infectious agent from repeated infections and maintain faster responses against it through time.

Since the B-cell antigen receptor (or antibody) and T-cell receptor are the molecular entities supposed to specifically bind a portion of the antigen (termed **epitope** or **antigenic determinant**), genes that code for these proteins are the units of molecular evolution at the somatic level that takes place in B and T cells. A variety of diversification mechanisms act on these genes. Both categories undergo **V(D)J recombination** that generates a repertoire of pre-formed BCR and TCR specificities before encountering the antigen. This **pre-formed primary repertoire** of BCRs can specifically bind an antigen, but weakly and with low affinity. Only antibody genes are subjects for further diversification processes: **class switch recombination** that enhances antibody effector function, **gene conversion** and **somatic hypermutation** that further diversify the antigen-binding portion of the BCR/antibody. An antigen-governed selection process, termed **affinity maturation**, is carried out to keep alive and amplify the BCR-bearing cells that show enhanced antigen recognition and binding upon the diversification of the antibody genes. Therefore, with each pathogen invasion, we assist to a true Darwinian evolution at the micro-scale of the B cell: BCR/antibody-coding genes are induced to accumulate mutations, whose effect is further subjected to selective evolutionary pressure.

A bit of history on origins of antibody diversification

Origins of molecular immunology emerged with the proposition of Ehrlich that antibodies, already studied at the chemical level, exist as side chains combined in a preformed receptor on the cell surface (Ganesh and Neuberger, 2011). If a fortuitous encounter of a preformed antibody and an antigen that fits extremely well its binding sites takes place, the individual is protected by counteracting the noxious effects of the antigen-bearing entity. The protective effect of this encounter was questioned once Landsteiner showed that antibodies can be raised against any antigen, existing in nature as well as against chemically synthesized compound without any toxic effect. However, Landsteiner's discovery was a very first glimpse into the vertiginous diversity of the antibody repertoire, perfectly adapted to cope not only with nature versatility but everything different from a self, be it natural or artificial. For a while, the scientific community concentrated on the enigma of how antibodies as proteins are capable of yielding this variety. Pauling (Pauling, 1940) and Haurowitz (Haurowitz F. and Breinl F., 1930) proposed an antigen induced-fit, suggesting that antibody changes its binding site at the contact with the antigen to fit it better, as an inducible enzyme will do with the cognate substrate. On the other hand, Jerne had a different theory in mind: paralleling the

survival of an individual with a selective choice of an antibody fitting the best the antigen, he hypothesized that antibodies with preformed specificities are subject to a natural selection governed by the antigen; the best fitting antibody would be reproduced within a cell and secreted in large quantities (Jerne N. K., 1955). The part on protein replication seemed to undermine the more or less accepted Jerne's idea of natural selection operating in antibody production and expansion. Actually, It was already known at Jerne's time that antibodies were produced by plasma cells, as experimentally proven by Fagraeus (Fagreus A., 1948). In 1950s Burnet modified Jerne's theory by shifting the natural selection focus from antibodies to antibody-producing cells, turning it into clonal selection: therefore, one preformed antibody specificity corresponded to a single and unique antibody-producing cell; if the antigen-antibody fit was good, the corresponding antibody-producing cell was selected and further expanded, as Burnet suggested, in perinatal period (Burnet F. M., 1957). Nossal and Lederberg brought experimental confirmation to Burnet's theory by showing that there is not a single plasma B cell bearing two antibody specificities; furthermore, Lederberg adjusted Burnet's theory in 1959 by proposing this clonal selection to be maintained through life-long encounters with the given antigen, by having DNA constantly mutating genome-wide in the selected antibody producer cell and bringing about for the first time the term of somatic mutations in B-cells (Lederberg, 1959).

The origin of pre-formed antibody diversity became the major debate in the scientific community; ones claimed that each antibody's specificity is encoded by a gene in total B-cell DNA, defending the so-called germinal theory, while the others argued against this enormous number of genes required to cover the antigen diversity in the naive B cell and supported the idea of modifications in somatic B cell DNA, generating antibody diversity. Talmage showed that the number of antibody coding genes is not necessarily of the same order of magnitude as antigen diversity (Talmage D. W., 1959); if one remembers Ehrllich's theory of combined chains making a for an antibody, different chain combinations account for different specificities. Tonegawa and Hozumi further contributed to unveiling the diversity-generating mechanism by showing that V and C genes are located on the same chromosome but are physically separated in mouse embryo cells; in hybridoma cells V and C genes are joined together in a primary diversification process termed V(D)J recombination (Hozumi and Tonegawa, 2004). Therefore, defenders of both germinal and somatic theory were right. Further genetic screening characterized differential gene expression of recombination-activating genes (RAG), which led to the discovery of the recombinases specific for this diversification process (Schatz et al., 1989).

Body of evidence soon showed that V(D)J recombination cannot account for the increasing antibody diversity during subsequent encounters with antigen. Lederberg had already offered his somatic mutation theory, and Brenner and Milstein developed it further in 1966 by limiting it to the V- coding region of immunoglobulin genes (Brenner and Milstein, 1966). Restriction enzymes were emerging at the time, so they provided a mechanism for a

local entry in immunoglobulin DNA by a DNA-cleaving enzyme that would leave a break, allowing for the further entry of an error-prone polymerase that is capable of introducing errors at all four bases while copying the undamaged strand. In 1970 Weigert and Cohn backed up this somatic theory descendant with experimental evidence of extremely localized amino acid variations within the variable region of antibodies coming from the hybridomas established after different immunizations with the same antigen (Weigert et al., 1970). These amino acid variations translated into point mutations at the DNA level of Ig genes. This last finding marked the turning point where the focus of studies on somatic hypermutation definitively shifted from protein to DNA sequence; hypermutating Ig gene sequence became thoroughly dissected, either in the endogenous locus by Gearhart and colleagues or by using the immunoglobulin transgenes as in the Storb group. With the appearance of immortalized and transformed B-cell lines, research on somatic hypermutation became increasingly accessible, eventually leading to the landmark discovery of AID, the initiating enzyme of somatic hypermutation predicted by Brenner and Milstein 30 years before.

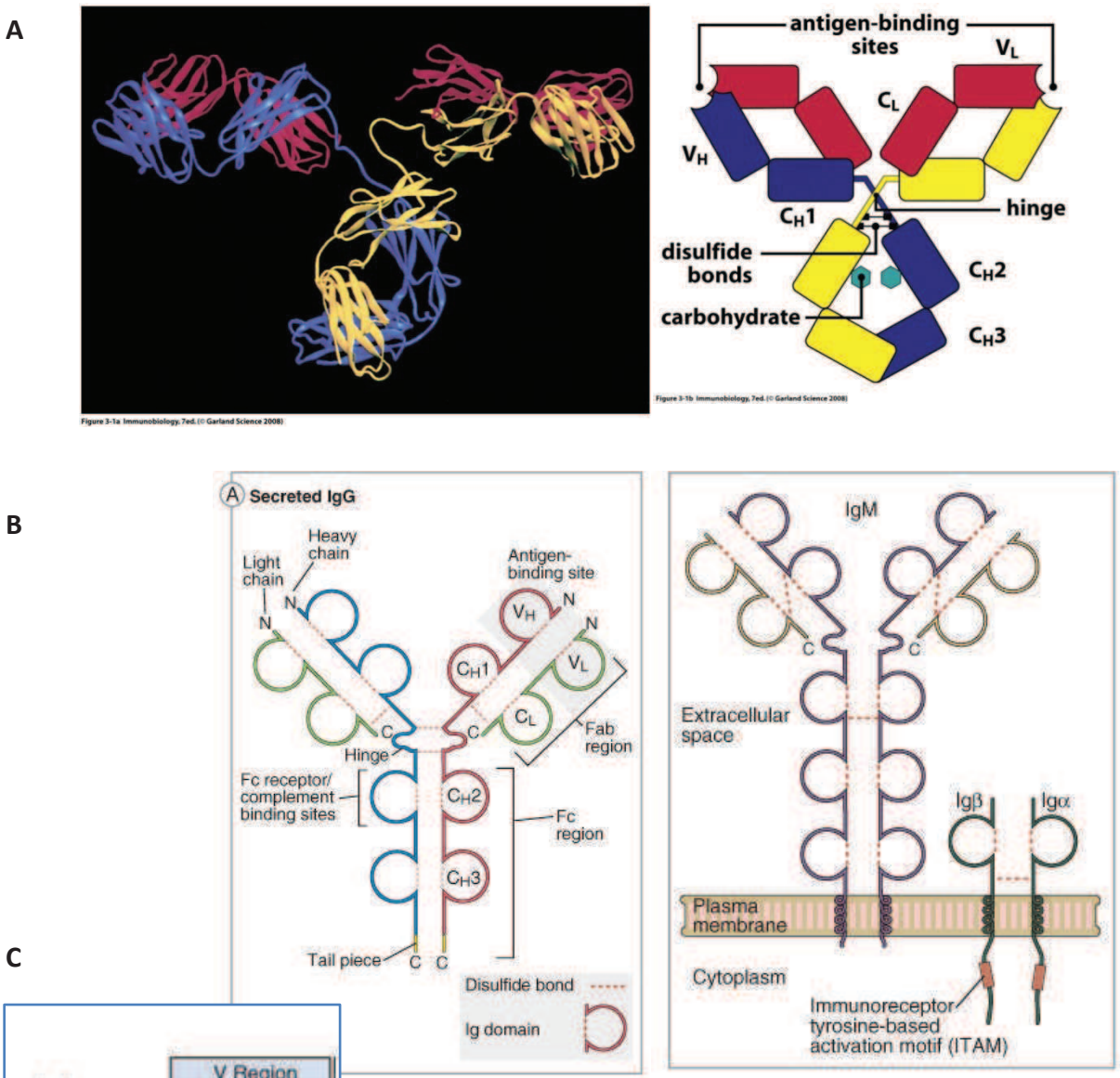


Figure 1. Multiple views of antibody structure.
A) Tridimensional structure of an antibody with the color code corresponding to the schematic view. B) Two-dimensional structure of an secreted antibody (left panel) and B-cell receptor (BCR) tethered in the B-cell membrane (right panel), accompanied with signaling invariant chains Igα and Igβ. C) Close view on variable region of IgH and IgL: boxes outlined with dotted line indicate positions of CDR1, CDR2 and CDR3 (the last one located at the VDJ and VJ junction of IgH and IgL, respectively). CH – heavy chain constant region, CL – light chain constant region, VH – heavy chain variable region, VL – light chain variable region.
Adapted from following sources: (Abbas, Abul K et al., 2010; Murphy et al., 2012)

Part One : General Introduction

Chapter I : Introduction to B cell biology

I .1 BCR receptors/Antibodies

B-cell fulfills the role of humoral immunity mediator through the molecular structure unique for this cell type – antibody or immunoglobulin (Murphy et al., 2012). These specific molecules can be anchored in the B cell membrane with their transmembrane domain, followed by a cytoplasmic tail usually involved in further signaling steps. The contact with antigen determinants takes place at the extreme end of their free extracellular half. This form is expressed through the major part of B-cell ontogenesis and it is termed B-cell antigen receptor (BCR). However, in the final differentiation of B-cell towards their effector function, the expression of BCR is modified in terms of losing their transmembrane domain but keeping the same specificity and affinity for the eliciting antigen. Terminally differentiated B-cells, termed plasma cells, can finally secrete antibodies in the bloodstream following the loss of transmembrane domain. This way, the former "extracellular" half keeps browsing for the cognate antigen and binds it tightly once found, while the former membrane-orientated domain instructs other participants in the immune response towards the most appropriate way of antigen elimination.

The antibody protein is composed of four polypeptide chains in total, two of them called **heavy chains (IgH)** and two **light chains (IgL)**. Four polypeptide chains are linked with disulfide bonds in a way presented on the Figure 1B.: two heavy chains are linked between them, and each heavy chain is linked to only one light chain. Disulfide bonds exist even within a single polypeptide chain, allowing the maintenance of a **specific immunoglobulin protein domain**, the specific structure that can be recognized in many other proteins: TCR receptors, co-BCR invariant chains, cytokine receptors, Fc receptors and adhesion molecules. When assembled, antibody molecule resembles the Figure 1A. C-termini of both chain types, IgH and IgL, are facing each other, as well as N-termini. The tips of the antibody's arms, composed of aligned N-termini of a heavy and a light chains, come in direct contact with the antigenic determinant. This means one antibody molecule can bind two identical epitopes simultaneously. The "stalk" region of an antibody is anchored in the membrane if it contains an additional transmembrane domain. The presence of two more immunoglobulin-like invariant chain, **Igα and Igβ** that assemble in the proximity of this surface immunoglobulin is necessary for a full B-cell receptor function because these

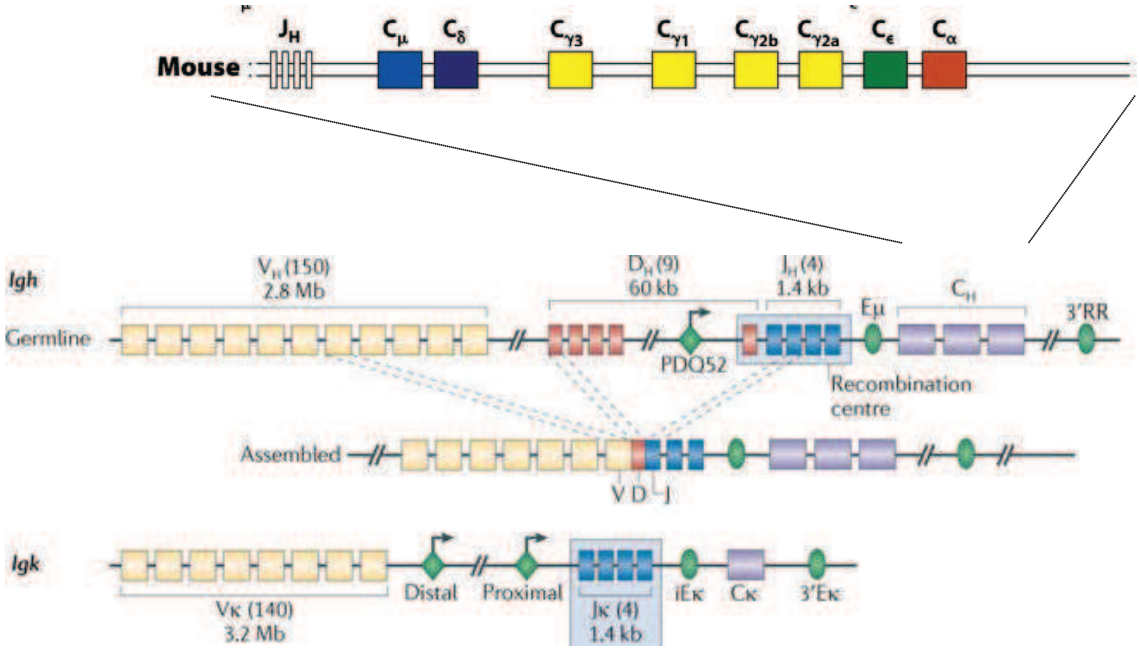
accessory molecules act as signal transmitters to the tyrosine kinases of the BCR signaling cascade through their cytoplasmic tails (Figure 1B). When the immunoglobulin is in the soluble form, the stalk region is the one that communicates with the precise set of receptors on the surface of other immune cells and mediates the clearance of this antibody-antigen complex, termed immune complex.

When comparing the amino-acid sequence of two IgM antibodies found in the same organism, as well as comparing IgM antibodies between individuals, striking differences at the level of intra-species amino acid variability are found only at the level of the N-terminal region, therefore termed **variable (V-) region**. This is to be expected, since N-termini of all antibodies must accommodate a huge versatility of antigens, existing in nature as well as synthetic. For a given antibody class, a term that will be discussed soon, the rest of the antibody sequence is identical, and named **constant (C-) region** (Figure 1B).

The C-region of the heavy chain is composed of three to four immunoglobulin domains (depending on the antibody class), and the C-region of the light chain contains only one immunoglobulin domain. As mentioned, in all organisms of the same species, there are different classes of antibodies that one can think of as different "flavors" of an antibody's constant region. Each antibody class triggers a very specific response for antigen evacuation or destruction, so that a precise antibody class is produced at a site where this type of clearance is the most appropriate.

If the amino-acid sequence of variable regions of different B-cell clones are compared, one can establish histograms of amino-acid residue variability for the given position within the V-region of both polypeptide chains (variability plots, Chapter II, Figure 18.A). Three regions of extreme variability within a polypeptide V-region can be discerned; these are termed hyper-variable regions 1-3 (HV1, HV2, HV3). Surrounding amino-acid sequences with more conserved content are considered as framework regions (from the beginning of the N-terminus: FR1, FR2, FR3 and FR4). When the three-dimensional structure of the antibody became available, hyper-variable region could finally be mapped: they corresponded to six flexible loops (three from IgH, three from IgL) protruding away from the compact tertiary structure. When antigen-antibody V-region complex was examined by X-ray crystallography, one could see that these were the major antibody contact points with the specific epitope, whose shape perfectly accommodated the one of the antigenic determinant (accounting for the other names as complementarity-determining regions, or CDRs). The secondary structure of the framework regions serves to shape and rigidify the V-region spatial organization, and does not withstand any radical amino-acid replacement that may result in improper folding of the V-region immunoglobulin domain (Figure 1C).

Figure 2. The structure of mouse IgH and Igκ genes. For more details, see text. Eμ : intronic enhancer, 3'RR: 3' regulatory region. Adapted from: (Murphy et al., 2012; Schatz and Ji, 2011)



I.1.1. Mouse immunoglobulin gene organization

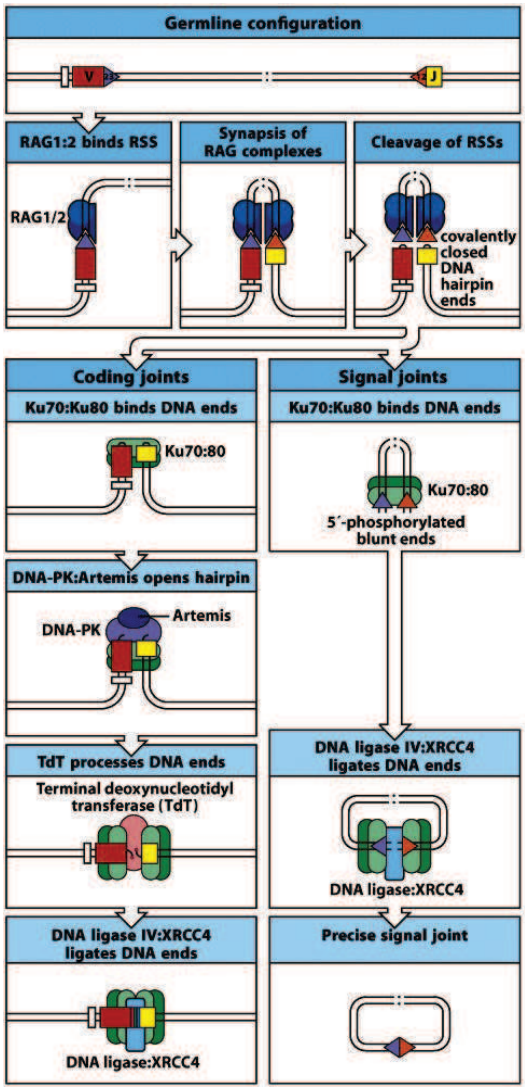
The heavy chain polypeptide is coded by the heavy chain locus (IgH) on the chromosome 12 (Figure 2). This locus contains coding sequences (exons) interspersed with pseudogenes and intronic regions. Both variable and constant regions of the heavy chains are coded by the exons in this locus. There are three groups of variable region exons: V(ariable), D(iversity) and J(oining) that make up for the whole heavy chain variable locus. In the mouse, there are around 150 different V_H exons, 9 D_H exons and 4 J_H exons. When all variable exons are present in the locus, its configuration is considered to be germline. Each V_H exon is preceded by a corresponding promoter (P_V1, P_V2, etc.) and a leader (L-) sequence. This leader peptide directs the immunoglobulin to the endoplasmic reticulum for membrane expression or secretion.

Genes coding for the heavy chain constant region can be found downstream of the outmost 3' J_H exon. In the mouse, there are 8 of them, and they are named μ, δ, γ, α, ε. They code for different antibody class or **isotype**: IgM, IgD, IgG1, IgG3, IgG2, IgG2b, IgA and IgE, responsible for different effector functions. Each gene coding for the constant region contains several exons, each exon coding for a single immunoglobulin domain or the flexible hinge within the constant region. Some constant genes have 3 exons, others have 4. Additional coding sequences can be found downstream of regular exons that code for the transmembrane and cytoplasmic domain, with a second polyadenylation signal. Alternative splicing directs the formation of the secreted or the membrane form of the heavy chain.

Each C_H gene segment, except for the Cδ gene, is preceded by a specific 5' I_H-exon and S_H-region. S_H regions are the sites of initiation from class switch recombination (CSR) process (detailed in Chapter II.6.1). All S_H region are composed of repetitive sequences, mostly pentamers like GAGCT and GGGGT; however, each S_H region is different and can contain the two mentioned sequences organized in tandem repeats or interspersed with other sequence motifs, creating complex patterns. The $I\text{-}_H$ exon was detected from germline transcripts initiated from the previously unknown C_H-germline promoter sequences prior to class switch recombination. This transcription is shown to be necessary for the efficient opening of the locus and to facilitate access to the CSR-initiating factors. These sterile germline transcripts contain I_H exons and C_H exons in the mature form, but since the I_H exon contains many stop codons in all reading frames, no protein products of these transcripts are detected.

There are two different light chain loci: one, named κ, is found at the chromosome 6 and comprises 140 Vκ genes and 4 Jκ genes. Only one Cκ gene is found. The variable region of light chains does not contain any D exons. The other light chain locus, λ, is found on the chromosome 16 and is the smallest Ig system known; it comprises three functional Vλ exons and three Jλ. Each J_λ exons is immediately followed by the corresponding C_λ gene, disrupting the classical clustering of J exons as in heavy or κ chain locus. Each light chain C gene codes

Figure 3 . Mechanism of V(D)J recombination.
For more details, see text. RSS: recognition signal sequences; DNA-PK: DNA protein kinase. Source: (Murphy et al., 2012)



for only one immunoglobulin region of the light chain constant domain, for which there is no specific effector function. This limited V_λ and J_λ exon diversity accounts for a disequilibrium in the usage of κ and λ light chains in mouse antibody assembly (κ:λ = 95:5 % of mouse serum Ig).

If looked at closely, V_H genes can be grouped into families under the condition that each family member shares 80% of the sequence identity. Altogether, mouse V_H can be regrouped in 3 clans and a total of 14 families, together with 3 different families of D_H exons.

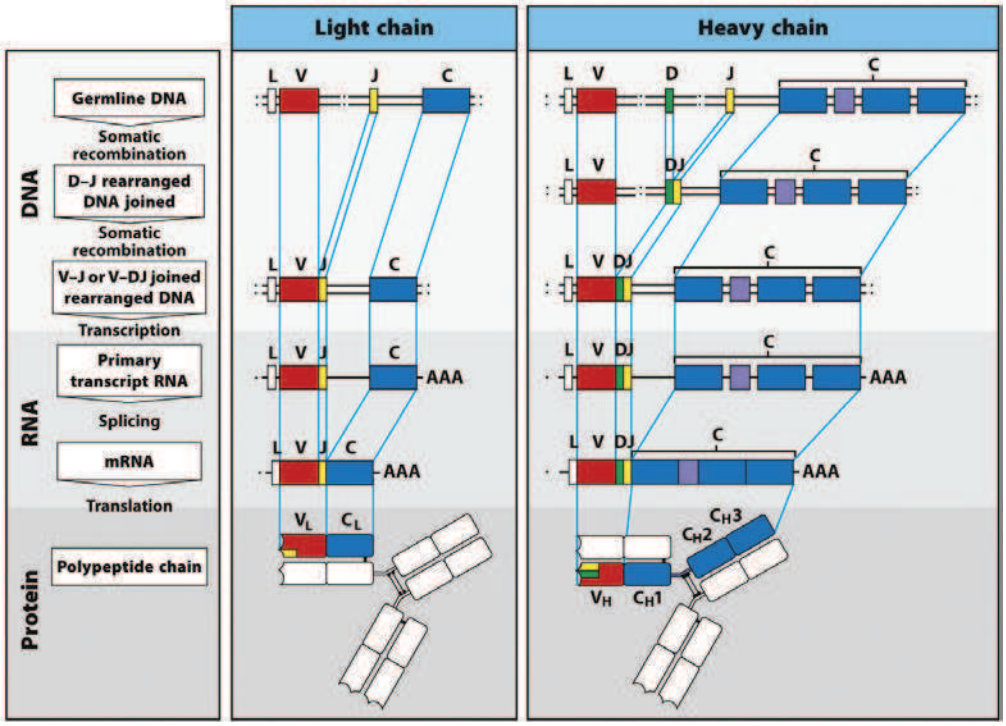
As already mentioned, HV1, HV2 and HV3 hypervariable regions are comprised within variable regions of IgH and IgL chains. Their positions relative to V, J and D_H coding sequences are known (Figure 1C): while HV1 and HV2 are entirely embedded within the V exon, HV3 in the heavy chain encompasses the whole D exon and comprises as well V-D and D-J junction. In the light chain, HV3 is exactly at the V-J junction. When all diversification mechanisms at the junction sites are taken into account, no wonder HV3 region is considered as the most variable one.

I.1.2. Primary diversification of immunoglobulin genes

One of the hallmarks of B-cell development is V(D)J recombination, a somatic recombination event that rearranges V, D and J exons at the heavy chain locus and V and J exons in light chain(s). Bringing one V, one D and one J tightly together is the prerequisite for synthesis, assembly, and expression at the cell surface of a complete BCR. V(D)J recombination proceeds in a stepwise manner, each joining event characterizing a B-cell development stage (Figure 3). First, D and J exons at the heavy chain locus are joined, followed by V_H-D_H joining. The heavy chain locus of all B-cells starting from their late pro-B development stage is therefore irreversibly rearranged. After the first checkpoint in B-cell development where this μ heavy chain is briefly expressed at the cell surface to test its functionality, one of light chain's V and J exons are joined and a full BCR is expressed.

V(D)J recombination proceeds by double-strand cleavage within specific signaling sequences that are positioned at 3' of the chosen D exon and 5' of the chosen J exon, separated by dozen kilobases. These signaling sequences ensure the directionality of exon joining throughout the whole process. The cleavage is carried out by the recombination-activating gene products, RAG1 and RAG2. Covalently closed hairpin DNA ends, products of the DNA cleavage by RAG proteins, are resolved further, resulting in formation of P nucleotides (P for palindromic, due to the Artemis-mediated asymmetrically opening of DNA hairpins) and allowing the addition of N nucleotides (N for nucleotide, randomly added by terminal deoxynucleotidyl transferase TdT, template-independent DNA polymerase). Finally, unaligned protruding ends of the coding joints are re-filled and ligated (Figure 3). In conclusion, this primary diversification of immunoglobulin genes is a compulsory step in B-cell development that takes place independently of antigen presence. In absence of RAG1 or

Figure 4. Steps in transcription, splicing and protein synthesis of heavy and light antibody chain. Source: (Murphy et al., 2012)



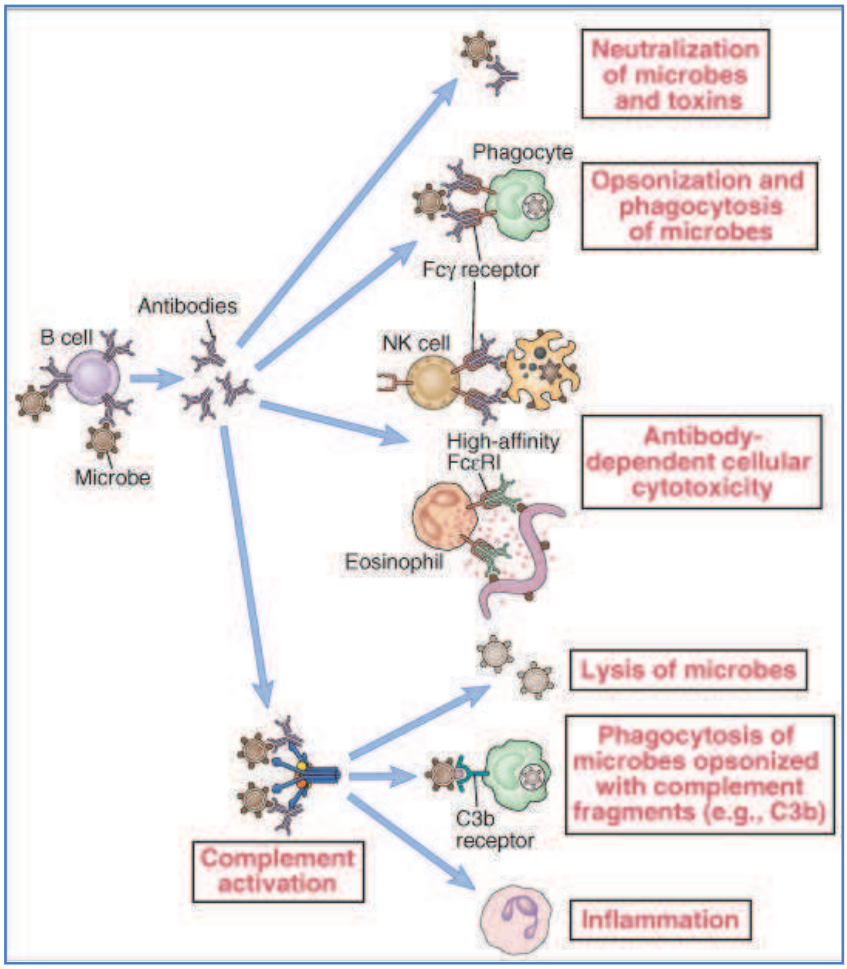
RAG2 protein, B cells as well as T-cells are blocked in their development at the earliest stage since they cannot express rearranged B-cell or T-cell receptor, predisposing the concerned individual to a high susceptibility to infections in a form of primary immunodeficiency syndrome (severe combined immunodeficiency, SCID)(Brandt and Roth, 2008; Shinkai et al., 1992). V(D)J recombination is as well a source of three types of diversity in immunoglobulin gene sequences, 1) a combinatorial diversity which is provided by random joining of a V, J and D_H sequences within a locus, 2) another layer of combinatorial diversity by pairing rearranged heavy and light chains. After the heavy chain is rearranged and checked for activity, large pre-B cell divides actively and gives rise to daughter cells which will all have the same rearranged heavy chain but each one of them can proceed to their own, unique rearrangement of one of the light chains in order to finally assemble them in the BCR; 3) a junctional diversity due to the addition of uncoded nucleotides (P and N) at the sites of D_H:J_H junction and less extensively at the sites of V_H:D_H junction (due to the short time frame of TdT expression in late pro-B cells). If all three diversity components are taken into account, the final result of theoretical total diversity is astonishing ($5x10^{13}$ combinations!).

I.1.3. Transcritpional regulation of Ig genes

V(D)J recombination provided for the immediate vicinity of the variable region exons in Tonegawa's experiments, thus allowing the synthesis of a final RNA transcript of the Ig genes. Another consequence of the V(D)J recombination is bringing the rearranged V promoter into proximity of Ig locus enhancers, the first one localized in the intron between the last J and C gene (IgH and Igκ intronic enhancer, not found in λ locus) and the second one found downstream of the last C gene (IgH and Igκ 3' regulator element or 3'RR and Igλ 2-4 enhancer, Figure 2). Therefore, this promoter of the rearranged V exon will be the preferred V promoter from which high-level transcription of the rearranged V-region genes will start. Ig promoters will thus drive Ig gene transcription only in B-lineage cells.

In the primary RNA transcript of immature or naive mature B-cell, the whole intronic region downstream of the rearranged JH exon is transcribed altogether with the unrearranged J_H exons left (if any), and a C_H gene (Figure 4) . In case of immature B-cells, transcription will stop immediately after Cμ. In case of a mature B-cell, both Cμ-containing and Cμ+C_δ-containing transcripts can be detected. C_δ is directly joined to the rearranged JH at the RNA level by alternative splicing and yields a membrane IgD that is the hallmark of mature B-cells. Transmembrane domain-coding sequence, coded by one of the C_H exons, will be included in the primary RNA transcript depending on the differential usage of the 5' or 3' polyA signals that flank this exon within C_H coding gene segment. If 3' polyA signal is used, synthesized Ig molecule will be directed to the cell membrane and tethered. The factors that govern the choice of polyA signals, either in C_δ inclusion within the primary transcript or transcription of transmembrane domain-coding exon still remain to be discovered.

Figure 5. Many effector functions of antibody constant region. See text for more details. Fcγ, FcεRI receptors – receptor for antibody IgG and IgE constant region; NK cell – natural killer cell; C3b receptor– complement receptor 1.
Source: (Abbas, Abul K et al., 2010)



The same splicing process takes place in maturation of κ or λ primary transcripts. As already mentioned, B-cell can switch antibody classes from IgM/IgD by replacing them with another constant region-coding gene. The mechanism bringing that alternative CH gene at the DNA level in the proximity of rearranged V(D)J region is a part of secondary diversification processes of immune repertoire and is dictated by antigen nature.

I.1.4. Secondary diversifications of immune repertoire

When a B-cell in mice or humans comes across the cognate antigen and receives help from the cognate T-helper cell in the secondary lymphoid organs (Chapter I.2.3c), an immediate choice is made between two ways to act against the antigen. The first one is to develop immediately into an effector cell and secrete low affinity, mostly neutralizing IgM antibodies (which buys more time for the more adapted immune response to take place). The second choice is to populate a germinal center where clones of activated B-cells are triggered to mutate their variable region-coding genes (by somatic hypermutation) and exchange their IgM class BCR against some other heavy chain class (by class-switch recombination) that, once secreted, acts concertedly with other innate immune factors in order to completely eradicate the pathogen. Both molecular mechanisms are in the heart of diversification processes of the secondary immune repertoire, since they take place in mature B-cells and only in the presence of a cognate antigen. In other species like birds and rabbits, another antigen-independent diversification process takes place: a gene conversion, which creates additional diversity in the V-region by introducing chunks of different V-pseudogenes within the rearranged V exon. All three processes depend on a single initiating enzyme and share some similar steps, so they will be discussed in more molecular details in the next chapter.

I.1.5. Roles of different Ig classes

The most logical consequence of an antibody binding an antigen determinant is to block the pathological effect of the antigen if the antigenic determinant happens to be a toxin or a viral coat protein. This antibody function is called neutralization (Figure 5). In the mouse, the most efficient neutralizing isotype is IgG1 and to a lesser degree, IgG2a. Other possible outcomes for an antigen within the immune complex is to be more easily phagocytosed by macrophages *via* their Fcγ receptors; this effect of antibody binding is called opsonisation. The best in this class is mouse IgG1. The third outcome is antibody-dependent cellular toxicity (ADCC), a process mediated by natural killer cells (NK) or eosinophils. Finally, antigen-antibody complexes may be recognized by the factors of the complement system, whose activation cascade leads to further lysis of the microbe or phagocytosis *via* receptors for complement fragments that opsonized the microbe. In mice, IgM is the most efficient antibody for this function; beside IgM, complement activation is showed for IgG2a and IgG2b. IgA remains the mediator of mucosal immunity, by being transported through the cells

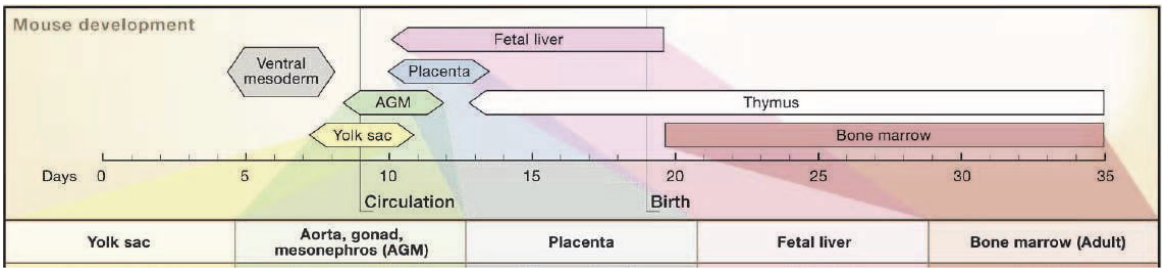


Figure 6. Different locations of hematopoietic stem cells throughout mouse development. AGM – aorta gonad mesonephros. Source: (Orkin and Zon, 2008)

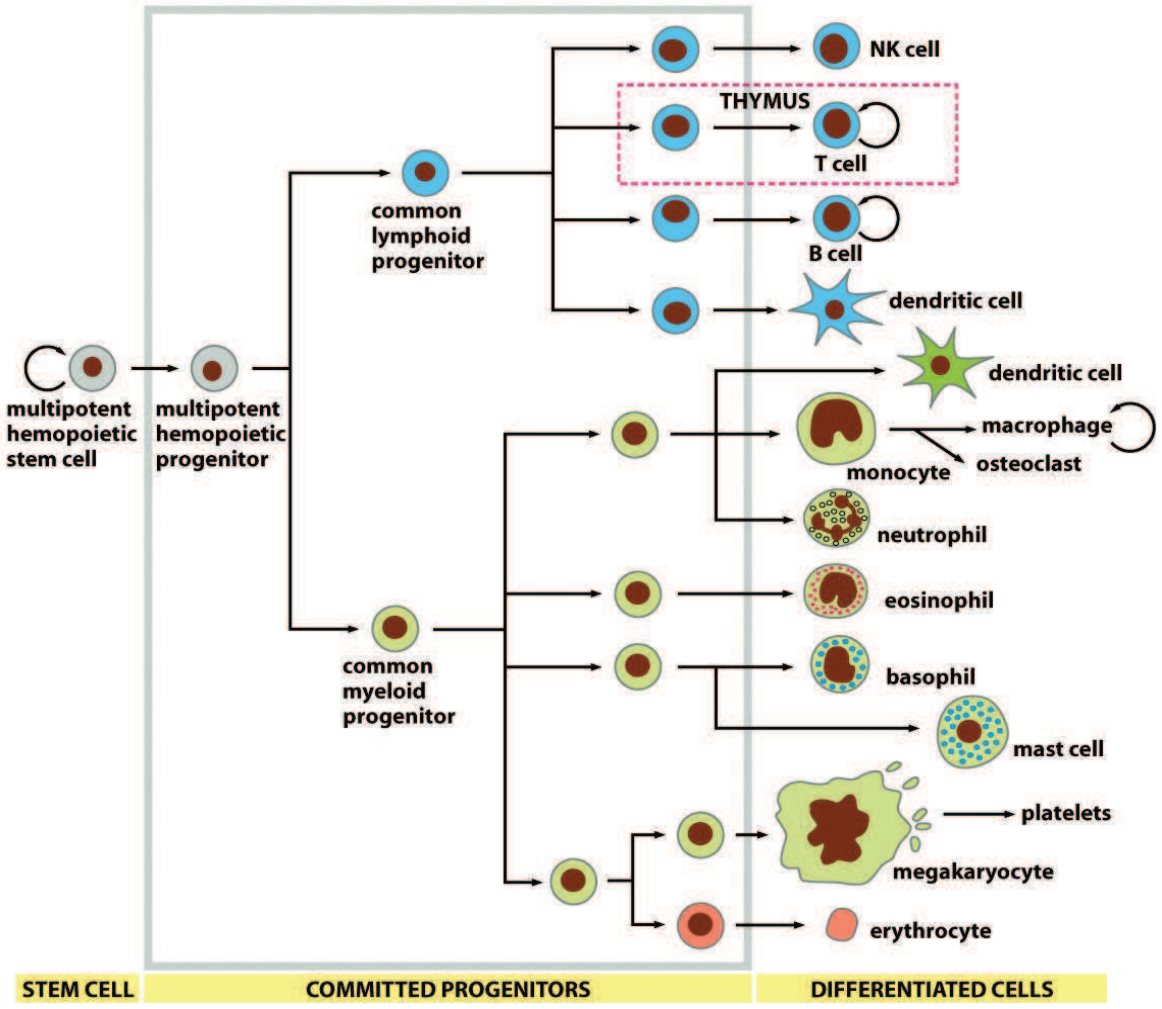


Figure 7. Schematic representation of hematopoietic stem cell differentiation. Grey line: putative unilateral differentiation pathways with progressive lost of multipotency, still under investigation. Source: (Alberts et al., 2002)

in epithelial barrier into body fluids like milk or mucus. If secreted alone, it forms the first immune barrier by inhibiting antigen adherence; when coupled to antigen in an immune complex, it enhances the complex evacuation through excretion in the lumen.

I.2 B cell ontogenesis : from HSC to mature naïve B cell.

I.2.1. Hematopoietic stem cells: preserve the source and provide for the diversity.

All immune system cells stem from one precursor, hematopoietic stem cell (HSC, Figure 5) that allows the renewal of the whole blood cell pool throughout a lifetime. These stem cells keep the potential to differentiate into a few different lineages but their differentiation potential is not to be compared to the pluripotency of an embryonic stem cell that can give rise to cells from all three germ layers, ectoderm, mesoderm or endoderm. The location of these specific multipotent precursors in organs changes throughout the development, along with the intrinsic properties of HSC specific for the concerned site in terms of self-renewal and potential to differentiate. In mice, hematopoietic stem cells derive from ventral mesoderm, with the first wave arising in the form of a hemangioblast in yolk sac and aorta-gonad mesonephros (AGM) region starting from 7 days (E7.5) post-fertilization. The primary role of these cells is to produce red blood cells and endothelial cells in the next 3-4 days of mouse embryo development. Controversy lingers about the relationship between yolk sac primitive HSC and definite adult HSC, concerning whether the latter descends from the former. By E12, it is the fetal liver that takes over the role of a major blood cell production site in a scheme that resembles more adult differentiation towards myeloid and lymphoid lineages. Hematopoietic stem cells will colonize thymus, spleen and it is only until a week after birth that hematopoiesis shifts definitively to bone marrow, which will remain lymphocyte production site until death (Orkin and Zon, 2008).

In order to characterize which fraction of isolated bone marrow cells contains true, long-term HSC, the most exploited way was to carry out adoptive bone marrow transfer into irradiated recipients. High-dose whole body irradiation leads to death of all proliferating cells: epithelial cells, germinal cells and especially hematopoietic precursors, which is the main raison of a fatal bone marrow failure and subsequent death. A single HSC from a immunologically compatible donor mice will suffice to repopulate the bone marrow of irradiated host (Kondo et al., 2003). Both HSC from fetal liver and adult bone marrow of a donor mouse, even though they differ in surface markers and in the mature blood cell types they are able to produce, are capable of colonizing the irradiated host. Fetal liver HSC were shown to reconstitute donor blood cells more successfully than adult bone marrow HSC, indicating a larger pool of long-term HSC (up to seven times more) with improved expansion, homing ability and wider differentiation panel. Moreover, it was shown that 25% of fetal liver

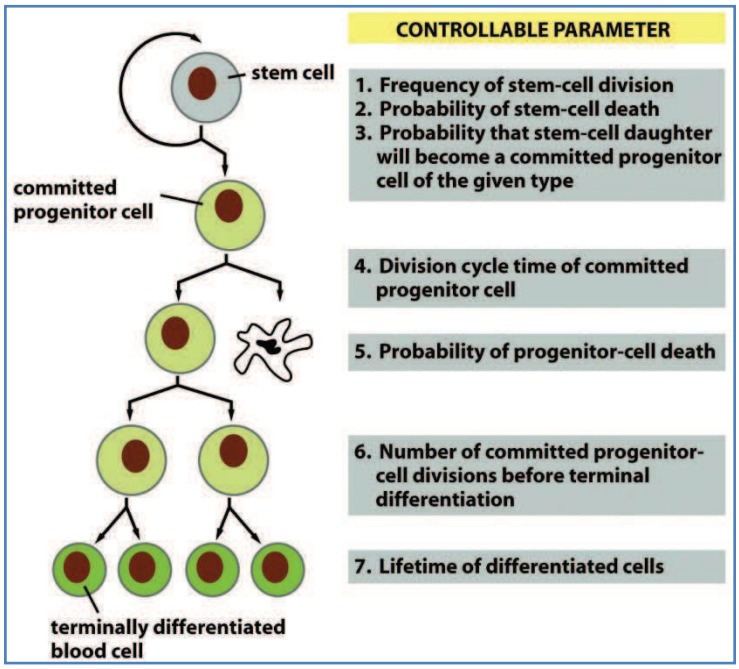
HSC are actively cycling cells while this category comprises no more than 4% of total adult bone marrow HSC (Morrison 1994, Morrison and Weissman, 1995).

The research of HSC biology is still ongoing in order to unveil the major factors that allow HSC to maintain their multipotency at the same time as to self-renew and maintain their number roughly stable. The decision-making events that promote differentiation over self-maintenance are yet to be discovered. However, several findings have already allowed *ex vivo* long-term culture and expansion of HSC, which is an inevitable step towards gene therapy interventions and transplantation. A modest expansion is achieved by growing HSC in defined cytokine milieu that stimulates division without loss of self-renewal (IL-3, IL-6, IL-11, SCF, Flt-3 ligand, thrombopoïetin TPO, Notch ligands, insulin-growth factor 2, etc.). Several other factors are coming into light (pleîotropin , angiopoietin-like protein 5, co-culture with mesenchymal stromal cells (Robinson et al., 2006)The choice of these factors was partially deduced from the analysis of the hematopoietic stem cells "niche" in bone marrow: in a apparently disorganized tissue, a "niche" is defined by the immediate surroundings of the stem cell entity that provide the signals to maintain its multipotency through cell-to-cell contact and local cytokine secretion. In case of hematopoietic stem cells, these are osteoblasts and bone marrow stromal cells. A stem cell tends to lose its stem cell character as soon as it leaves the particular stem cell niche. One example of necessity for the cell-to-cell contact in order to maintain the stem cell identity is illustrated by the Kit-Kit ligand signaling pathway, Kit being a tyrosine kinase at the surface of hematopoietic stem cells and kit-ligand (stem cell factor SCF or steel factor) being a membrane-bound molecule of the surrounding stromal cells. Kit is one of the markers of long-term HSC and very early progenitors within a so-called LSK subset (lineage-negative Lin^-, stem cell antigen positive $Sca1^+$ and Kit^{hi}) (Welner et al., 2008)

However, cell culture in presence of soluble factors is only transient and short-termed, since HSC eventually differentiate and die. The second means is to transduce HSC with a retroviral vector encoding transcription factors that induce a burst of in vitro HSC proliferation, like HOXB4 (up to 80-fold, (Antonchuk et al., 2002) or HOXA10 homeodomain fused to N-terminal domain of nucleoporin 98 (NUP98-HOXA10hd fusions (Ohta et al., 2007). The expansion post-transduction is impressive (1000-fold) over a short period (6-10 days) and promising for future therapy challenges (Sorrentino, 2004).

Long-term HSC after a division may give rise to a precursor that retains the multipotency but loses the ability to self-renew. This multi-potent progenitor cell (MPP) requires signaling through another cell-surface tyrosine kinase receptor, FMS-related tyrosine kinase 3 (Flt-3), triggered by the membrane-bound Flt-3 ligand on the surrounding stromal cells. This signaling pathway favors further differentiation towards two novel oligopotent lineage precursors, common lymphoid precursor (CLP) and common myeloid precursor (CMP, Figure 7). Further differentiation is depicted in a linear, stepwise manner that may not be perfectly respected in reality in terms of lineage "fidelity", because cases have been

Figure 8. Parameters of the HSC differentiation which can be potentially modulated by colony-stimulating factors *in vitro* as *in vivo*. Source : (Alberts et al., 2002)



observed in which myeloiod precursors retain the capacity to differentiate towards B cells (Hardy et al., 2007). For a cell precursor to be considered as lineage-restricted, two conditions are required: the loss of capacity to develop into other blood cell types at that specific stage and the irreversibility of this loss. For the moment, only CLP and CMP show a high degree of lineage commitment both *in vitro* and *in vivo*. Unlike true HSC, committed progenitors have to undergo many divisions (many, but finite), with the progeny of each one being driven towards a more differentiated stage with a smaller capacity to proliferate and a relatively short lifespan. Even with these limitations, a huge number of fresh blood cells is produced from a single division of a hematopoietic stem cell, which explains the HSC scarcity in total bone marrow cell count (1 in 10 000) and reduces the necessity for them to divide frequently. Each division brings in the risk of introducing mutations and telomere shortening (Kondo et al., 2003), indicating a limited self-renewal potential even for the stem cells. Indeed, the majority of HSC appears quiescent, and several dye-dilution assays for proliferation indicate that HSC may divide only five times in a whole mouse lifespan (Zhang et al., 2012). Moreover, the efficiency of bone marrow reconstitution of irradiated mice worked less well with the hematopoietic stem cells that were driven into cycling prior to transplantation.

Red blood cells and megakaryocytes, precursors for platelets, descend from MEP (megacaryocyte-erythroid progenitor), the further step of commitment after CMP while mast cells, eosinophils, neutrophils, monocytes/macrophages and conventional dendritic cells stem from GMP (granulocyte-macrophage progenitor). CLP is supposed to yield B and T lymphocytes, as well as plasmacytoid dendritic cells and natural killer cells. There is an ongoing debate about even earlier progenitors of CLP, such as ELP (early lymphoid progenitor) or LMPP (lymphoid-primed multipotent progenitors), the latter designated upon the expression of early transcripts that are hallmarks of mature blood cell gene expression, a phenomenon known as "priming" (Hardy et al., 2007; Zhang et al., 2012). As already mentioned, intermediates between the common progenitor and finally differentiated cells retain a residual proliferation capacity without requiring the absolute contact with the surrounding cells. This allowed the creation of an assay of colony formation in a semi-solid medium: in this surrounding, the progeny cells cannot migrate but tend to accumulate and form a colony. The aspect of these colonies differs accordingly to the subsequent differentiation of the progeny and has contributed to identify and name the intermediate precursors as colony-forming units (GM-CFU, E-CFU, etc).

This colony-forming capacity also provided the functional assay for examinating the impact of different factors that orientate blood cells precursors towards a particular lineage (colony-stimulating factors or CSF). Some of the above-mentioned factors that modulate hematopoietic stem cell survival, division and uncommitted fate were discovered this way and subsequently tested on the true stem cells. CSFs are capable of regulating several parameters of the hematopoietic stem cell culture as indicated in the Figure 8. However, they are not the

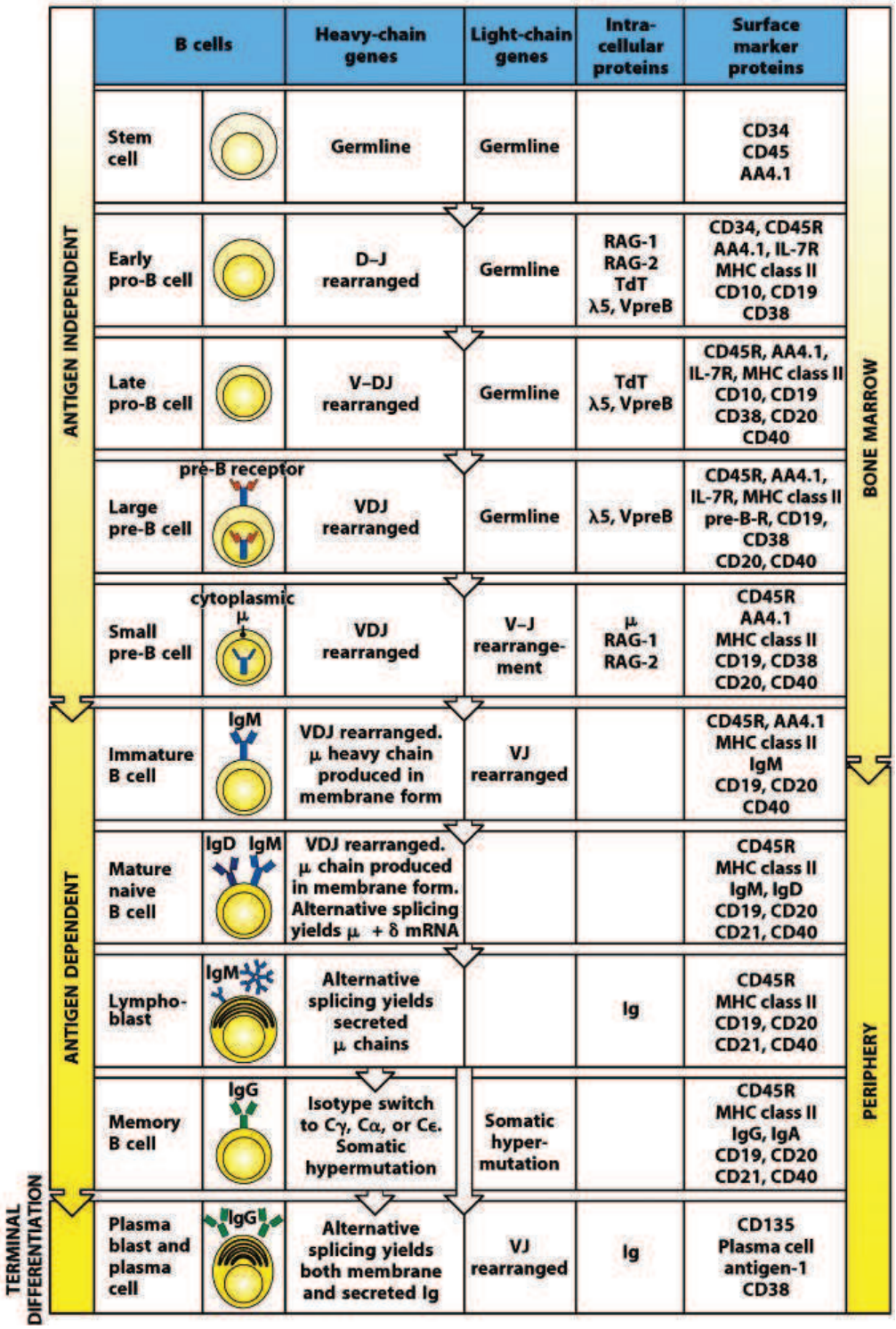


Figure 9. Schema of B-cell development from hematopoietic stem cell to differentiated plasma cell. More details in text. Source: (Murphy et al., 2012)

sole governors of the hematopoietic stem cell fate: it seems that a single HSC maintains a stochastic factor in deciding to orientate towards one lineage or the other. Even in the most carefully purified HSC culture, the difference in sizes as well as in the characters of the developing colonies can sometimes be striking. It seems that, as a whole, the systems perpetuates a reproducible development pattern, but at the individual cell level the process of commitment and programming to cell division can sometimes be considered as random (Alberts et al., 2002).

I.2.2. From common lymphoid progenitors to mature naive B cells

B-cell lymphopoiesis continues in the bone marrow of bones like femur, tibia and to some extent in arm/fore limb bones throughout the lifetime. This is an antigen-independent process and includes progressive, stepwise changes in cytoplasmic determinants, surface markers and Ig gene configuration, as well as the proliferation and survival capacity of its intermediates (Figure 9). The survival is especially at stake since unsuccessful rearrangements and some of the negative selection processes take their toll in the bone marrow until immature B lymphocytes are unleashed in the bloodstream.

Further steps of B-cell differentiation from common lymphoid progenitor involves **pro-B cell**, characterized by the initiation of Ig gene rearrangement, the V(D)J recombination (Chapter I.1.2). The very first B-cell governing transcription factors, PU.1, E2A and EBF emerge at the stage of a pro-B cell, initiating the expression of another transcription factor, Pax5 that acts as lineage-stabilizing factor, irreversibly committing this hematopoietic intermediate to the B lineage. Other targets of these early B-cell transcription factors are the two recombination-activating genes, RAG1 and RAG2 necessary for V, D and J exon rearrangement. In the first cascade of recombination events, early pro-B cell rearranges one D with one J segment of heavy chain immunoglobulin loci on both chromosomes. At this time, the expression of TdT polymerase is detectable, allowing for the N-nucleotides insertion. After DJ recombination, RAG1 and RAG2 proceed to rearrange heavy chain V exon with rearranged DJ segment, marking the transition towards **late pro-B cell**. This time, the recombination proceeds at only one chromosome. Small number of N-nucleotides is inserted at this stage as the expression of Tdt is turned off early in mice. If V exon is joined in frame with DJ segments, an entire μ heavy chain is synthesized, coupled with a surrogate invariant light chains (termed λ5 and VpreB) since the light chain loci have not yet recombined, and sent to the cell surface for a test of its integrity and functionality. The expression of this so-called pre-B receptor is the first checkpoint in B-cell development, marking the transition from late pro-B cell to a large pre-B cell. If the VDJ rearrangement on this chromosome is unsuccessful, it switches to the second heavy chain allele. If this second rearrangement fails, the late pro-B cell is destined to die by apoptosis. Around 45% pro-B cells die at this stage.

If VDJ rearrangement at one of the chromosome is successful, the newly differentiated **large pre-B cells** start to actively proliferate for two reasons. The first one is to reduce but not to totally bring to an end the expression of RAG proteins. This way further rearrangement at the other heavy chain allele is prevented if the first one is successful. RAG-2 protein expression is conditioned by its G1/S-dependent phosphorylation that targets it to degradation outside of the G0/G1 phase (§ Chapter III.2). This is as well the probable trigger for allelic exclusion, an epigenetic event that turns off the heavy chain allele that is not to be used for the assembly of the future BCR. The second reason is to amplify the effect of the successful first recombination event in order to increase the chances of immature B cell formation, since the second fortuitous joining event is to be expected in the light chain loci in **small pre-B cells**, the progeny of large pre-B cells. RAG proteins continue to operate in small pre-B cells by joining VL and JL in one of the light chain loci, κ or λ. If the recombination goes on successfully from the start, a complete BCR is allowed to be assembled and is expressed at the cell surface, with the definitive shutdown of RAG protein expression and allelic exclusion of the other light chain alleles. However, if the first joining event fails, the RAG recombinases go on rearranging the remaining V exons of the same allele; if in vain, the other allele is tested for the rearrangement. If recombination completely fails at the kappa light chain locus, the RAG proteins will attack the lambda light chain locus, allele by allele.

At last, an **immature cell** derived from the small pre-B cell is formed, expressing a surface IgM (sIgM) and two pan-B cell markers, B220 (CD45R) and CD19. CD19 appears as a consequence of Pax5 transcription factor expression and therefore persists from the earliest B cell precursors onwards. The B220 marker is being expressed prior to heavy chain rearrangement, thus earlier than CD19. The expression of the sIgM is considered to be the second checkpoint in B-cell ontogenesis, since fully formed immature cells are exposed to a negative selection if they bind to a self-antigen. If some binding and downstream signaling do occur, the immature cell undergoes one of the following scenarios: receptor editing, apoptosis or forced unresponsiveness (anergy).

This is roughly the scheme that hematopoietic stem cells follow once they get engaged to the B-lymphoid lineage, be it in fetal liver or in the bone marrow. However, beside larger replenishing potential of fetal liver HSC when compared to adult HSC, the sole fetal HSC retain the capacity of producing another type of differentiated B-cells, the so called **B-1 subset** (Hardy et al., 2007; Kondo et al., 2003). These B cells express high levels of the CD5 surface marker, high IgM and low IgD. They can be found mainly in the peritoneal and pleural cavities, and make up to 5% of the total B cell pool in the mouse. B-1 cells are capable of self-renewal. They are considered to be "natural antibody" producers – a majority of circulating IgM being secreted by these cells even in the absence of infection. The BCR of these cells is pre-selected to bind some self-antigens that largely resemble environmental antigens as those found in the microbial gut flora. It seems that exposure to these types of antigen has positively selected B-1 cells with an apparent auto-reactivity that is not

pathogenic, but might even be involved in cleaning up the senescent cells (Hardy et al., 2007; Shapiro-Shelef and Calame, 2005). They react extremely quickly upon antigen encounter, migrating towards spleen and differentiating into antibody-secreting cells without requirement for T-cell help (for more on this subject, §Chapter I.2.3b).

Hematopoietic stem cells can be found in direct contact with osteoblasts in the endosteum, the inner lining of the bone tissue. While generating oligopotent progenitors and other intermediary species, cells are dividing and pushing their progeny towards the central blood sinus in the middle of the bone marrow, where the non-autoreactive immature cells are released. Once in the bloodstream that carries them to the spleen, B cells shift from immature towards the mature stage by undergoing two consecutive transitional stages, T1 and T2. T1 cells are defined as IgM^{hi} IgD^{-} subset that receive a strong signal to home to spleen, but did not gain the ability to recirculate (Pillai and Cariappa, 2009). These transitional cells are discharged in the white pulp of the spleen after being carried by the blood flow through the central arteriole and marginal sinus. Until now, the only known requirement for T1 cell survival is the so-called tonic BCR signaling: even though the cognate antigen is not yet encountered, this non-specific and continual activation of the signaling pathway downstream of the BCR is indispensable. The second requirement is the ability to recirculate through the follicles of the spleen white pulp in which the other key molecule for terminal differentiation is encountered, the BAFF survival factor. If an immature lymphocyte fails to enter a follicle due to the fierce competition between other immature cells and preferential entry for mature B-cells, it dies shortly after release from the bone marrow. However, if the immature cell succeeds in entering the follicle, it needs both BCR tonic signaling and BAFF-BAFF receptor interaction in order to become T2 transitional cells that express $IgM^{hi}IgD^{+}$ markers. However, some other immaturity markers are kept at the surface, like AA4.1, known to be expressed by some early multipotent progenitors. Every 3 days, 50% of the short-lived immature B cell population dies due to the impossibility to fulfill these requirements (Charles A Janeway et al., 2001). T1 and T2 cells are considered to proceed through additional rounds of negative selection in the spleen. Finally, with a different combination of extracellular ligands and the evolving strength of BCR signaling, T2 B cells transit towards mature B cells, expressing both $IgD^{hi}IgM^{low}$, termed B-2 cells to distinguish them from the B-1 subset. Among B-2 cells, two distinct subsets are further identified depending on their localization once they recirculate through the spleen: **B-2 cells or follicular B cells (Fo-B)** that recirculate through primary lymphoid follicles in lymph nodes and spleen, and marginal zone B cells that home exclusively around the marginal sinus at the spleen white pulp-red pulp border (Chapter I.2.3a).

Marginal zone cells (MZ B-cells) remain resting and long-lived, with a capacity to self-renew but also with a potential of being replenished from marginal zone B cell precursors in case of rapid MZ B cell depletion.

Their BCR seems to bind some self-antigens and common environmental antigens, with low affinity. This low level of BCR signaling is considered to be favored during some positive selection process in spleen while presenting specific self-antigens to T2 transitional cells newly homed to spleen. Their main role is considered to be getting in a direct contact with blood-borne pathogens, recognizing mostly T-cell independent antigens. Their most remarkable property is to respond rapidly to antigens, making them the very first B-cells that differentiate into plasma cells upon the antigen entrance, constituting an innate-like first line of defense (Hardy et al., 2007; Shapiro-Shelef and Calame, 2005).

Mature naive follicular B cells continue to circulate *via* the bloodstream throughout the body, spending some time in draining lymph nodes or spleen white pulp. As detailed in the Chapter I.2.3a, these organs posses an architecture that enhances the probability of encounter between three major factors necessary for naive B cell activation:

- priming of B lymphocytes by a specific antigen *via* their BCR receptor in a direct contact with antigen,
- naive T-helper cell differentiation through a contact with professional antigen-presenting cell (dendritic cells, mostly) *via* their epitope-bound MHC II – TCR contact,
- and the cognate interaction between the T-helper cell TCR and epitope-bound MHC II molecule at the B cell surface, which ultimately activates a mature B cell that is no longer considered naive.

These interactions allow for the further differentiation of activated B cell towards effector cells that sustain the function of humoral immunity: short-lived antibody-secreting plasma cells, long-lived plasma cells and memory B cells.

I.2.3. Germinal center reaction and affinity maturation

I.2.3a. Secondary lymphoid organs: a place to be seen for follicular naive B cells

Among all lymphoid organs, one can distinguish those that serve for the development and education of the major immune system cell types (the primary lymphoid organs) and others in which those cells can complete their development, acquire their definitive phenotype and get engaged in immune responses (secondary lymphoid organs). Bone marrow and thymus (the development organ for T cells, even though their early precursors arise in bone marrow and migrate towards thymus) belong to the category of primary lymphoid organs. We shall briefly detail the structure of the secondary lymphoid organs, which contribute to the formation of plasma cells and memory B cells through germinal center reactions: lymph nodes, mucosal-associated lymphoid tissue (MALT) and spleen.

- Lymph node (Figure 10).

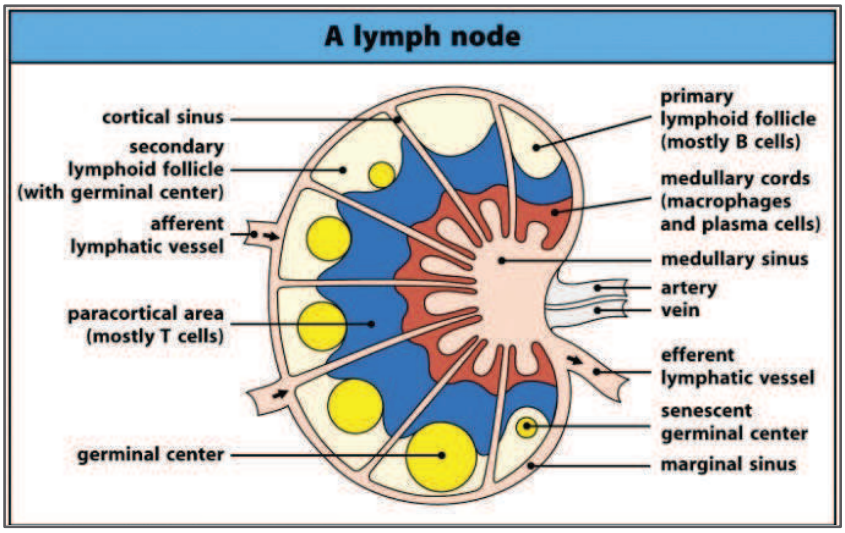
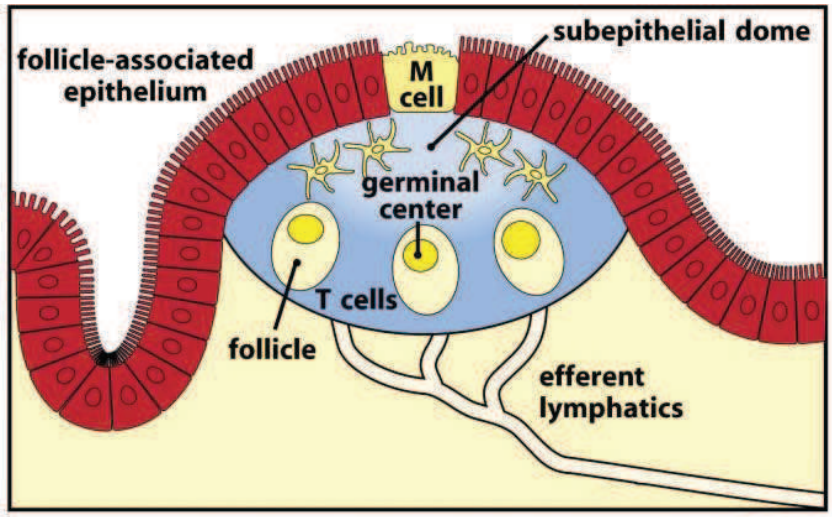


Figure 10. Schema of lymph node structure.
Source : (Murphy et al., 2012)

Figure 11. Structure of a Peyer's patch in ileum. M cell – microfold cells.
More details in text.
Source: (Murphy et al., 2012)



Lymph nodes are strategically positioned alongside lymphatic vessels, allowing the lymphatic fluid to be drained. One can imagine the lymph vessel network within the lymph node as a capillary network, where the afferent vessel branches in many side channels, vascularizing the cortex horizontally and vertically, which converge towards the medullary sinus and are drained through the efferent lymphatic vessels. Two major anatomical regions can be distinguished within the lymph node: the cortex and the medulla. The cortex is delimited from the medulla by a paracortical area containing diffusely arranged T-cells, (called **T-cell zone**). Lymphatic vessels bring in the free soluble antigen and **antigen-presenting cells** that have captured the antigen in the extracellular space within the tissues, allowing them to be "trapped" for a while in the node, searching for the T cell that can recognize the same antigen. Blood vessels vascularize the lymph node as well. A specialized blood vessel type, high-endothelial venules (HEV), allow lymphocytes that have been carried with the bloodstream to mingle between the endothelial cells and to get "squeezed" out on the outer side of the vessel wall. This is the way that mature naive follicular B-cells use to recirculate through the lymph node. Once outside of the HEV, B-cells directly come across T-cells within the paracortical area. Antigen-primed B-cells migrate quite quickly towards the follicle zone of the lymph node if there were no retention mechanisms: peptides coming from the internalized and processed antigen are being exhibited on the B-cell MHC-II, at the same time allowing for activation of adhesion molecule and chemokine receptors that literally anchor them in the T-cell zone. This way, the chances of encountering the cognate T-cells are maximized; if otherwise, the odds would be between 1 in 10 000 and 1 in 1 000 000.

At the outer border of T-cell zone, a so-called B-cell zone can be found, being organized in primary follicles. If a B-cell undergoes a cognate reaction, it returns to the primary follicle from the T-cell zone and transforms it into a secondary lymphoid follicle with a germinal center in its core. After several weeks of proliferation leading to production of memory B-cells and plasmocytes, the germinal center reaction wanes, degenerate and become senescent. However, persisting germinal center-like structures were observed even 8 months after immunization with a specific antigen type, so-called particulate antigen (sheep red blood cells, vesicular stomatitis virus, protein-coated beads) (Gonzalez et al., 2011).

- Illeal Peyer's patches (Figure 11).

Peyer's patches within the small intestine wall make up for an important the part of mucosa-associated lymphoid tissue, along with adenoid glands and tonsils. While the major role of the lymph node is to bring in contact cells that specifically recognize lymph-borne antigens (that entered through epiderm), Peyer's patches screen for the antigens that are brought in with food or changing microbial flora. They are localized immediately beneath the intestinal epithelium, forming a cupola-like structure tightly filled with primary and secondary follicles and surrounded by a T-cell zone. Specialized M-cells (for microfold, referring to their apical protrusions) are found in the follicle-associated intestinal epithelium that have the role to sample the antigens caught on the inner mucosal lining of the intestine, phagocyte

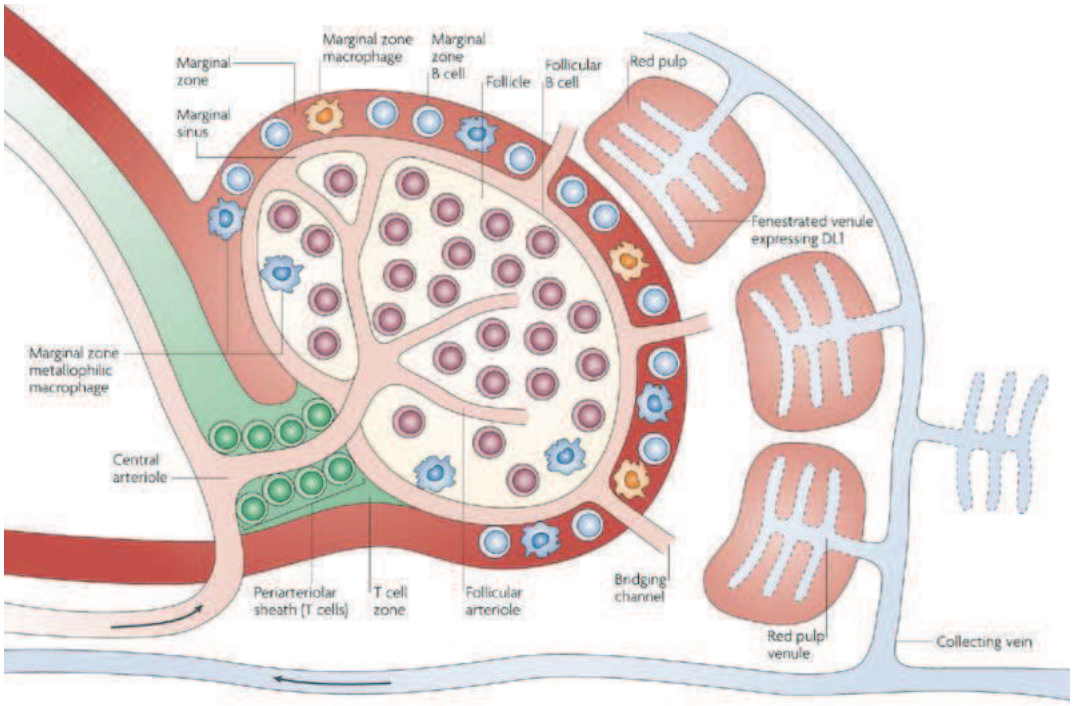


Figure 12. A schematic view of the anatomy of the spleen.
Source: (Pillai and Cariappa, 2009)

them and transfer them across the whole cell body to be released on the cell's basal side. Immediately beneath the epithelium is a network of dendritic cells that take up the antigen and pursue the antigen presentation in the T-cell zone. In the well-irrigated Peyer's patches, B and T lymphocytes enter by bloodstream and leave after germinal center reaction through the efferent lymphatic channels.

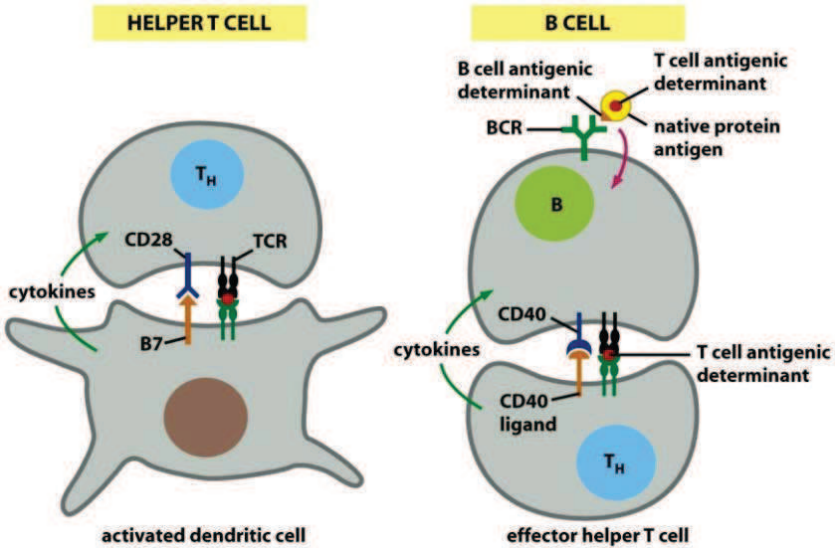
- Spleen (Figure 12).

Spleen is not directly connected with the lymphatic vessel network, so all the antigens that circulate through it are blood-borne. When one takes into account diverse roles of spleen, it is makes it easier to imagine its organisation: spleen serves as a disposal for used red and abnormal blood cells, with the important network of iron retrieval and stocking; spleen stocks as well platelets, polynuclear neutrophils and specialized spleen macrophages that eliminate directly the antigen brought by the bloodstream and are able to tackle huge pathogens like protozoa. The majority of spleen tissue is made up of **red pulp** in which red blood cells are deposited for destruction. In-between red pulp, dispersed spots of white pulp can be found, grouped around the trabecular arteries bringing in the fresh blood. These radially disposed vessels branch into transversally orientated central arterioles enveloped in the first layer of T-cells making up the **periarterioral lymphoid sheath (PALS)**. Central arterioles branch beyond T-cell zone into sinuses, specialized fenestrated blood vessels that lack the classical rigid organization of the capillary network. Surrounded by these sinuses, a B-cell zone can be found, composed of the follicles surrounded by the so-called **marginal zone** that separates them from the surrounding marginal blood sinuses. PALS, follicles and marginal zone constitute together the **spleen white pulp**. Once again, spleen organization allows bringing in contact mature B-cells with T-cells from PALS primed for the cognate antigen, and on the other side, the traffic of the immature B-cells that complete their differentiation within the follicles and marginal sinuses of the white pulp in order to become sedentary marginal zone B-cells or follicular mature B-2 cells.

I.2.3b. The first requirement: antigen

The versatility of antigens that an individual's B-cell repertoire is able to recognize is amazing: protein/sugar/lipid antigens; natural or synthetic low-molecular weight chemical compounds, complex antigens containing multiple chemically diverse epitopes (virus particles, other species' blood cells...), antigens with unique or highly repetitive moieties... As we have seen so far, blood/lymph circulation system and secondary lymphoid organs have evolved to match the incoming antigen from a defined entry point (epiderm, extracellular space, inhalation, ingestion, blood) with the most appropriate adaptive immune response. The recognition of a specific antigen can be mediated with or without T-cell help, and the outcomes of further naive B-cell differentiation are different. Two classes of antigens can be distinguished whether the necessity for the T-cell help exists or not. The ones that need the

Figure 13. Most important cell-to-cell interactions for B-cell activation : activation of the cognate T-cell by interaction with dendritic cell (left) and activation of B-cell through triple interaction with T-cell. Source: (Alberts et al., 2002)



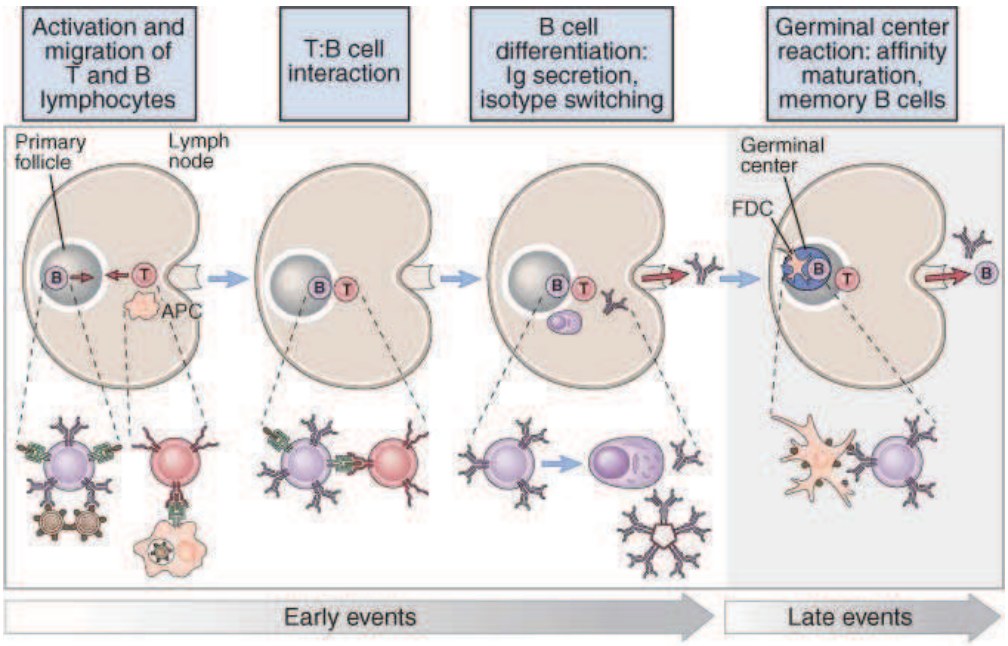


Figure 14. Cognate B-T cell interaction. FDC – follicular dendritic cell. More details in text. Source : Abbas, Abul K et al., 2010)

intervention of T-helper lymphocytes to activate B-cells are called **T- or thymus-dependent antigens**, and comprise mostly protein antigens or antigens carrying a protein component. **T-independent antigens** are divided into two subgroups, TI-1 and TI-2. TI-1 class comprises antigens that can induce B cell proliferation but manage to bypass the BCR signaling pathway. These antigens are called B-cell mitogens; they will activate any B-cell, mature or immature, independently of its BCR specificity, like bacterial lypopolysaccharide (LPS) that signals through innate-immunity receptors like TLR-4 (Toll-like receptors). The TI-2 class of antigens, like pneumococcal polysaccharides, possesses repetitive moieties that are able to cross-link the BCR of a mature cell and send a downstream signal strong enough to trigger antibody production. This type of antigen is recognized by special subsets of B- cells trained to react quickly and without outside help for development into effector cells (B-1 cells, marginal zone B cells). However, only one type of antigen can initiate a germinal-center reaction, somatic hypermutations and most importantly, B-cell memory: the ones whose recognition is mediated by T-cell help, and those are the ones that we will take into account when tackling the key issue of this study.

I.2.3c. Cognate B cell-T cell encounter

At this point, we shall concentrate more closely on the prerequisites for the germinal center reaction. For it to happen and for a B-cell to survive, the B lymphocytes have to be present at the right place and at the right time. Once more, three major actors must juxtapose: the cognate antigen, the mature naive follicular B-cell specific for the antigen and the T-cell primed with the cognate antigen thanks to the previous interaction between the MHC-II of a dendritic cell and a compatible TCR (Figure 13). After having acquired the mature phenotype, B-cell may encounter its cognate antigen while recalculating or interacting with dendritic cells and macrophages. Antigen-bound BCR is internalized; antigen is processed and presented through MHC-II. As we have seen previously, the structure of the secondary lymphoid organs facilitates direct contact with recalculating primed B-cells and T-cells within different T-cell zones. The first direct contact between the two is established through epitope-presenting MHC-II of the B-cell and TCR of the primed cognate T-cell. The second contact, delivering another necessary signal for the B-cell activation, happens through CD40:CD40 ligand, the latter being expressed at the T-cell surface upon the TCR- B cell MHC-II contact. Finally, the last activation step is engaged, with T-cell producing specific cytokines (IL-4, TNF, IFN-γ) in a polarized manner, exclusively at the site of TCR-MHC-II interaction (Figure 14). These cytokines bind to the appropriate receptors at the surface of the B cell and trigger the signaling cascade that creates a certain number of decisions within B-cell transcription networks: first one, expression of AID protein, and the second one, sterile transcription of the IgH constant region genes, μ and the one that B-cell is supposed to switch to and which is mostly appropriate to respond to the given antigen. After this triple embrace between B and

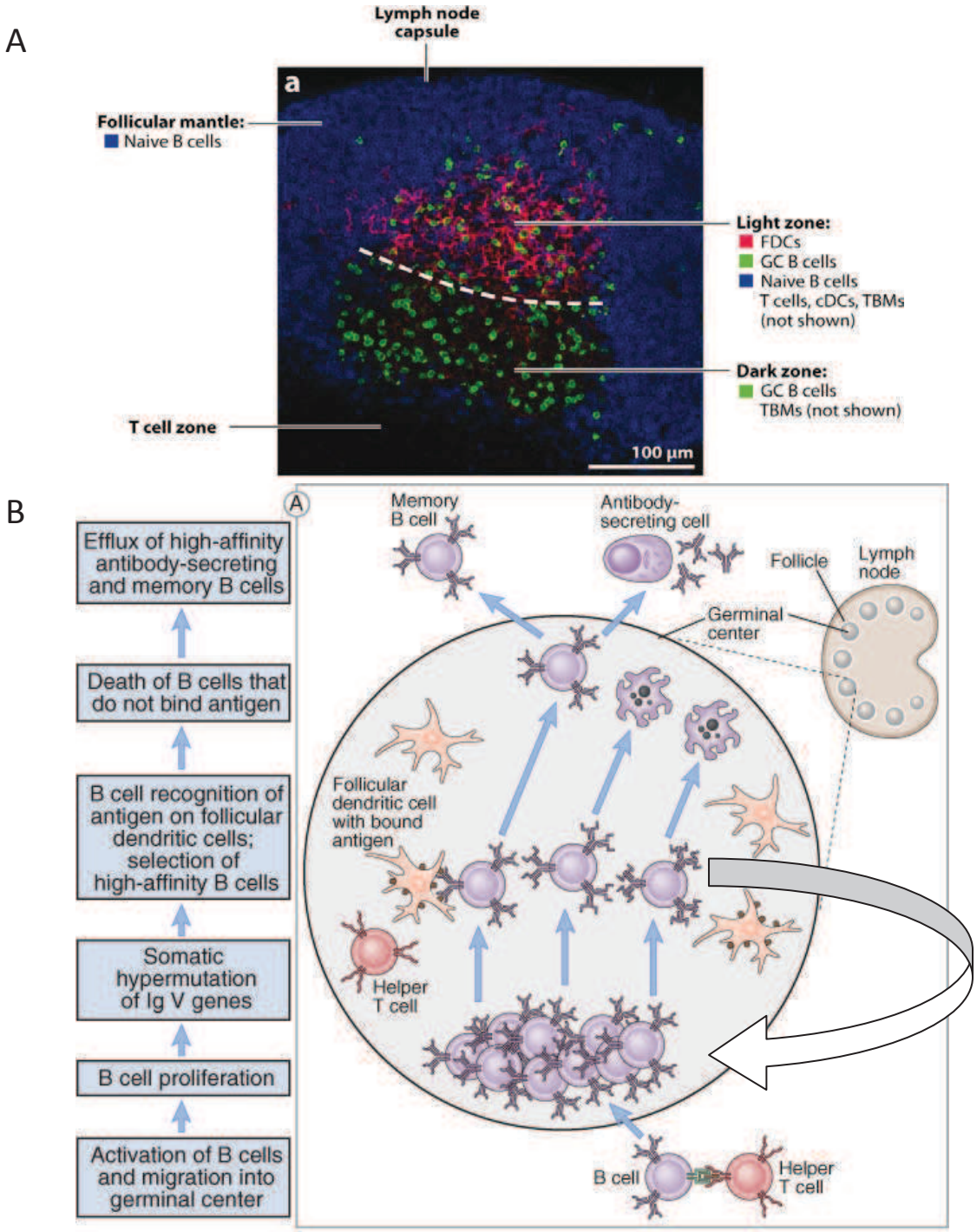


Figure 15. Germinal center reaction *in vivo* and in schema. A) Micrograph of a secondary follicle within a lymph node with distinct germinal center histological zones and labeling of key cell populations. FDC: follicular dendritic cells, GC: germinal center, TBM: tingible body macrophages. Source: (Victora and Nussenzweig, 2012)

B) Schematic overview of B-cell passage through germinal center. The "flow" of B-cells through germinal centeris never straightforward; there are several rounds of proliferation and hypermutation, followed by FDC and follicular T cell-mediated selection before a B-cell which received all necessary survival signals is released. Here, no distinction between histological zones since proliferation seems to happen in both light and dark zone. Source: (Abbas, Abul K et al., 2010)

T-cell, B cells can form extrafollicular foci if they undergo immediate proliferation and IgM production. These extracellular foci can sometimes be observed alongside PALS in the spleen. Some evidence exists that these cell can undergo an extrafollicular isotype switch (Shapiro-Shelef and Calame, 2005). Otherwise, an activated B-cell migrates into the follicle zone of the secondary lymphoid organ and engages itself into a germinal center reaction.

I.2.3d. The germinal center architecture and reaction

A germinal center is traditionally considered to be a *de novo* formed structure that welcomes activated B-cells. Instead, both B and T cells which recognize different epitopes physically linked within a present antigen colonize a primary follicle and transform it into a secondary follicle. Primary follicles already contain follicular dendritic cells, a special type of dendritic cells organizing future germinal centers, making up a network of their protrusions where naive B-cell of diverse BCR specificity are constantly moving.

Activated B-cells that enter the primary follicle engage immediately into successive rounds of massive proliferation, pushing other naive B cells towards the follicle periphery, forming the follicular mantle. On the micrographs showing active germinal centers, one can separate two zones surrounded by the mantle: a dark zone and a light zone. Dark zone, closer to the T-cell zone, is formed of massively proliferating B-cells, while the impression of "light" texture of the light zone, situated towards the lymph node capsule or spleen marginal zone, is given by the mesh formed from the protrusions of follicular dendritic cells. For a long time, it was considered that a passage through the dark zone towards the light zone was unidirectional: however, this old model was revised after discovering that proliferation is ongoing in both zones (Gonzalez et al., 2011; Victora and Nussenzweig, 2012). Moreover, this anatomical separation fades as time goes by. GC B-cells acquire a phenotype distinct from naive B-cell: they express no more IgD but they express high levels of n-glycolylneuraminic acid (which serves as a ligand for GL-7 antibody, the marker of activated T-cells as well), Fas receptor (CD95) and bind to peanut agglutinin (PNA).

In germinal centers, the Burnet's clonal selection theory comes clearly into light: during the heavy proliferation, the rearranged variable Ig genes acquire random mutations that will affect the affinity of GC B-cell BCR: modification can vary from slight to extreme, changing the BCR's affinity for the better or for the worse. **This process, termed somatic hypermutation, is the central subject of this work.**

After the first cycling round, B cells enter the competition for the intact antigens presented on the surface of follicular dendritic cells (Figure 15). Follicular dendritic cells are antigen-presenting cells that capture the antigen but not in the same manner as conventional, migratory dendritic cells: this GC-specific antigen, recognized by some pre-existing antibody, is enclosed within immune complexes-coated bodies or iccosomes, that persist on the surface of the FDC by being presented *via* complement and Fc receptors. These intact antigen

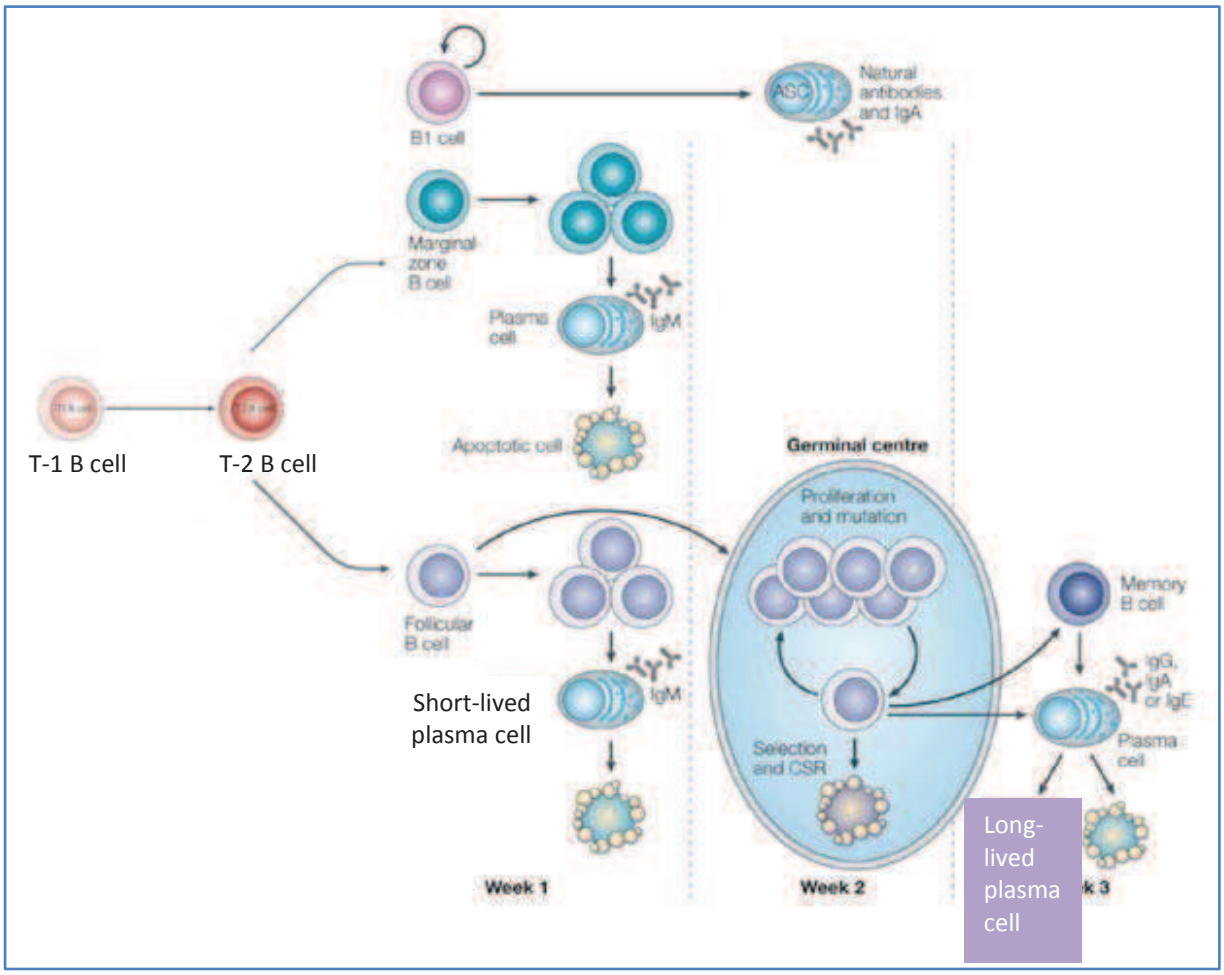


Figure 16. The destiny of the immature cell from transitional to effector B-cells. ASC: antibody secreting cells; CSR – class switch recombination. Adapted from: (Pillai and Cariappa, 2009; Shapiro-Shelef and Calame, 2005)

complexes can keep activating cognate B cells for much more than a week after administration, justifying their nomination for long-term antigen reservoirs.

The quantity of this presented antigen is limited, so the competition is once again ferocious: only the cells with marked increase in affinity against the cognate antigen will receive survival signals, potentially from follicular helper T-cells, which allows them to re-enter cell cycle and try to modify the BCR of their progeny. **This phenomenon in which the somatic hypermutation alternates with affinity-based selection in order to yield B-cells expressing a BCR with high affinity for the cognate antigen is named affinity maturation.** Affinity maturation does not only allow the mutagenic shift (amino acid change in BCR peptide sequence, submitted to a further selection) but also a B-cell clone drift (change in the present germinal center B-cell repertoire by the outgrowth of previously poorly represented clones, which got selected and undergo further somatic hypermutation) (Berek and Milstein, 1987).

The B-cells bearing mutations that result in an incorrectly folded BCR, unchanged/weakened BCR affinity or rendered auto-reactive, will undergo immediate apoptosis, and their bodies will be cleared by tingible-body macrophages. In fact, it is considered that the default fate of the GC B-cell is to die by apoptosis, knowing that Fas is highly expressed on the surface and that the expression of anti-apoptotic genes like bcl-2 is turned off (Victora and Nussenzweig, 2012). If GC-B cells are placed in culture without any stimulation factors, they die within few hours. However, if a minimum of stimulation signals is provided, GC-B cells are capable to keep proliferating in culture for few more days.

After rounds of affinity-based selection, germinal center B cells can commit to several fates. One is becoming a plasma cell precursor, named **plasmablast**, that will have its transcription programs modified once again: proliferating cells cease to cycle, increase their endoplasmatic reticulum and Golgi aparatus to accommodate the incredible antibody production and become **secreting plasma cells** (Figure 16**)**. These effector cells either stay within lymph nodes, small fraction of them migrates to the spleen and majority lodges in the bone marrow, constantly releasing highly specific antibodies in the bloodstream. Exhaustion of these cells' endoplasmatic reticulum and emerging pro-apoptotic signals will make them undergo apoptosis in few days after the final differentiation step (~2 weeks from the infection onset). However, some of these cells will manage to uncover a survival niches within bone marrow that will allow them to persist in secreting low levels of soluble antibodies (referred to as persistent antibody) into the bloodstream up to one year after infection (Radbruch et al., 2006).

As we have already discussed, another modification of the GC B-cell transcription program turns them into **memory B** cells that keep exhibiting modified high-affinity BCR at the surface and keep patrolling down the bloodstream until the next encounter with the familiar antigen. A memory B cell is a true Methusalem among long-lived cell populations that is capable of spanning the individual's lifetime. These cells remain quiescent and

continue to circulate or stay in the lymph node. Two layers of this memory response in mice are described in case of immunization with a particulate antigen: switched memory B-cell that differentiate immediately into plasmocytes and contribute to the immediate protection, and IgM+ memory B-cells that return to GC and undergo proliferation and class switch recombination in order to replenish the switched memory pool (Dogan et al., 2009).

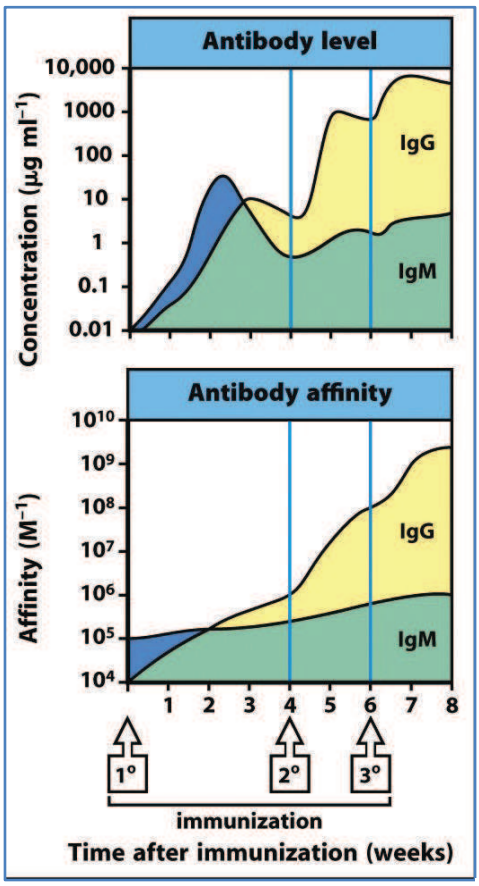


Figure 17. Dynamics of the humoral immune response during the successive encounters with an antigen. Source: (Murphy et al., 2012)

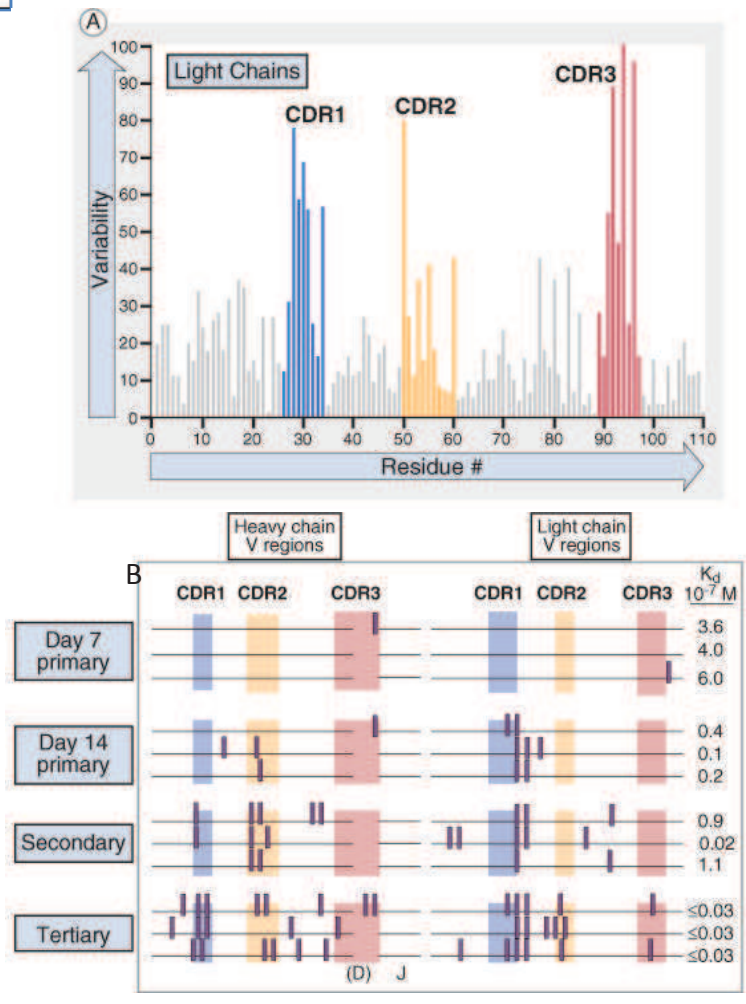


Figure 18. Changes in amino acid variability in association with the affinity for the antigen. A) amino acid variability plots; B) Kd (dissociation constant of an antibody), the inverse measure of the antibody affinity, drops as B-cell accumulates mutations in both framework and CDR regions. These mutations are the subject of cell-based selection in germinal centers. CDR – complementarity determining region. Source: (Abbas, Abul K et al., 2010)

Chapter II: Somatic hypermutation, the real-time adaptive mutagenesis

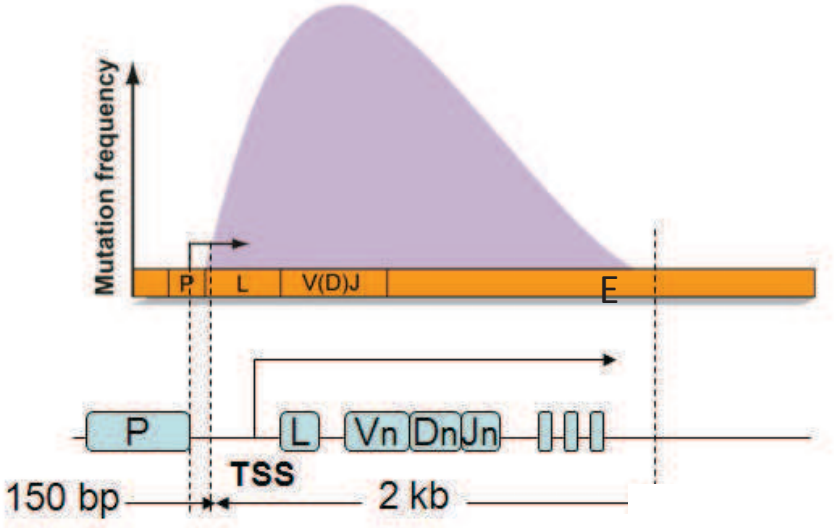
The clearest picture one can get about the advantages conferred by the secondary diversification of the immune repertoire is when studying successive immunizations. When an individual gets in contact with a never-seen-before pathogen (e.g. when a mouse is immunized for the first time), the primary immune response is triggered and activated B-cells secrete IgM starting from one week post-immunization. Only two weeks post-immunization do IgG antibodies barely attain the peak of maximal IgM concentration (Figure 17). However, secreted IgM antibodies are polyvalent and of low affinity for the given antigen, while IgG affinity for the eliciting antigen improves with time. The picture is radically different during the second antigen encounter: predominant antibodies in serum are switched IgG, IgA or IgE, depending on the antigen nature. Switched antibody titers peak colossally high and in much shorter time following the secondary immunization (less than a week). Switched antibodies improve their affinity in quicker pace than during the first immunization, and even more after a third immunization. This dynamics of humoral immunity beautifully illustrates all the differentiation stages of a mature B-cell, from activation to terminal differentiation: IgM secretion in extrafollicular foci, affinity maturation and class switching in a germinal center, differentiation towards plasma cell and memory cell, the "wake-up" call for memory B-cell with switched BCR upon secondary immunization.

For modifications affecting antibody affinity, the most concerned regions of the antibody molecule are the IgH and IgL hypervariable regions HV1, HV2 and HV3 (respectively CDR1, CDR2 and CDR3, Figure 1C), with the last one having the largest amino-acid variability. If one looks at the genomic regions in hybridoma cells (fusions of a myeloma and GC B-cell clones) encompassing hypervariable regions and simultaneously measures antibody affinity at different time points (Berek and Milstein, 1987), one may observe a peculiar phenomenon presented in the Figure 18. Nucleotide substitutions (along with infrequent, small insertions and deletions) accumulate with time after primary immunization, and even more after secondary and tertiary immune responses. Most mutations cluster in hypervariable regions (especially HV1), but a few of them occur in the V framework as well.

As a result of concomitant affinity-based selection and further mutagenesis of a selected BCR, a sharp decrease in the dissociation constant of the antibody is detected, reflecting the net increase of the antibody affinity. These mutagenic events underlying antibody affinity improvement following the clonal selection are named somatic hypermutation (SHM).

The phenomenon is somatic since these mutations do not occur in the germ lines; mutations will die with the individual, but the somatic hypermutation retains characteristics of

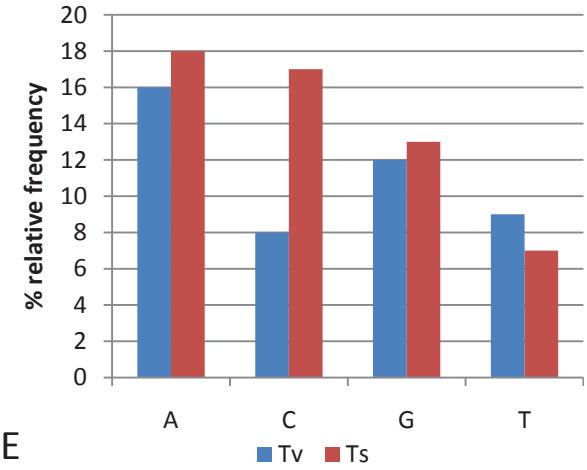
Figure 19. Properties of somatic hypermutation : A) mutated regions in IgH locus (no S_H or C_H genes shown) ; B) mutation profile, with red letters indicating transitions ; C) graphical representation of pooled transitions or transversion per base ; D) relative frequency of individual mutations at A:T base pairs in coding strand; E) individual mutations at C:G base pairs as read off the coding strand. Ts-transition, Tv: transversion, Pyr-pyrimidine, Pur-purine. TSS: transcription start site. Adapted from: (Di Noia and Neuberger, 2007) (Peled et al., 2008)



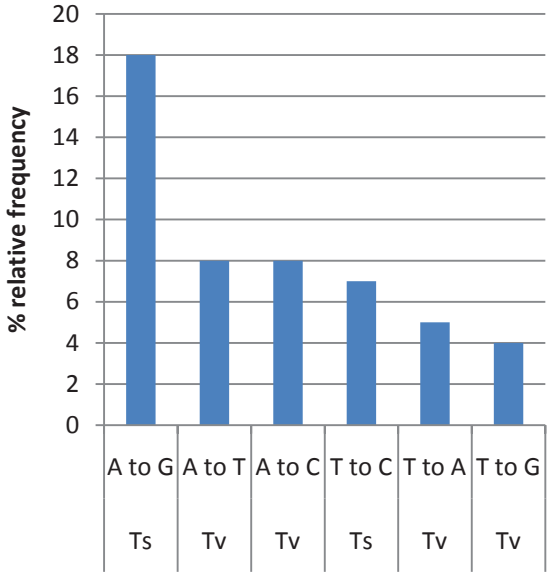
B

From \ To	Pyr T	Pyr C	Pur G	Pur A	
Pyr T		7	4	2	13
Pyr C	16		2	4	22
Pur G	7	6		15	28
Pur A	3	12	22		37

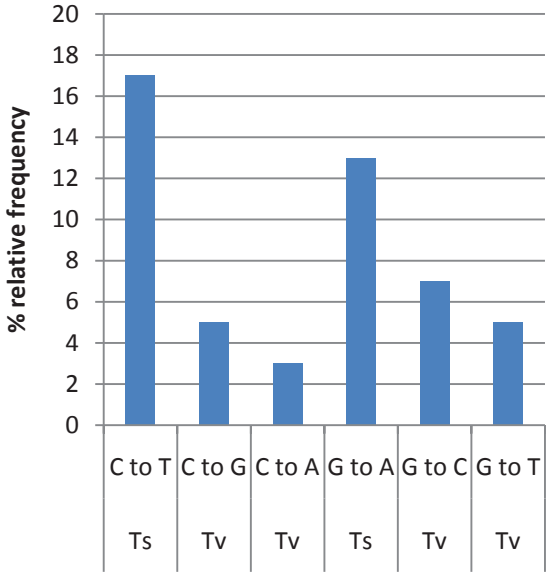
C



D



E



Darwinian evolution on a micro scale because the B-cell progeny inherits these mutations and undergoes further rounds of selection.

The prefix "hyper" comes from the fact that mutations occur every 1000 basepairs/cell/division, which is 10^6 times higher than the frequency of spontaneous mutagenesis in the genome. In the classical turnover of hypermutating B-cells in germinal center, this means that every B–cell survivor leaves the germinal center with up to 10-15 mutations in V-regions and about 2-3 amino-acid substitutions (per each replacement mutation, three or four silent mutations are acquired) (Honjo and Alt, 1995), mostly targeted to the hypervariable regions. Deleterious mutations in form of stop codons or mutations that impair antibody folding do occur, but requirement for constant BCR expression and signaling in the competitive environment of the germinal center allows them to be purged from the V sequence .

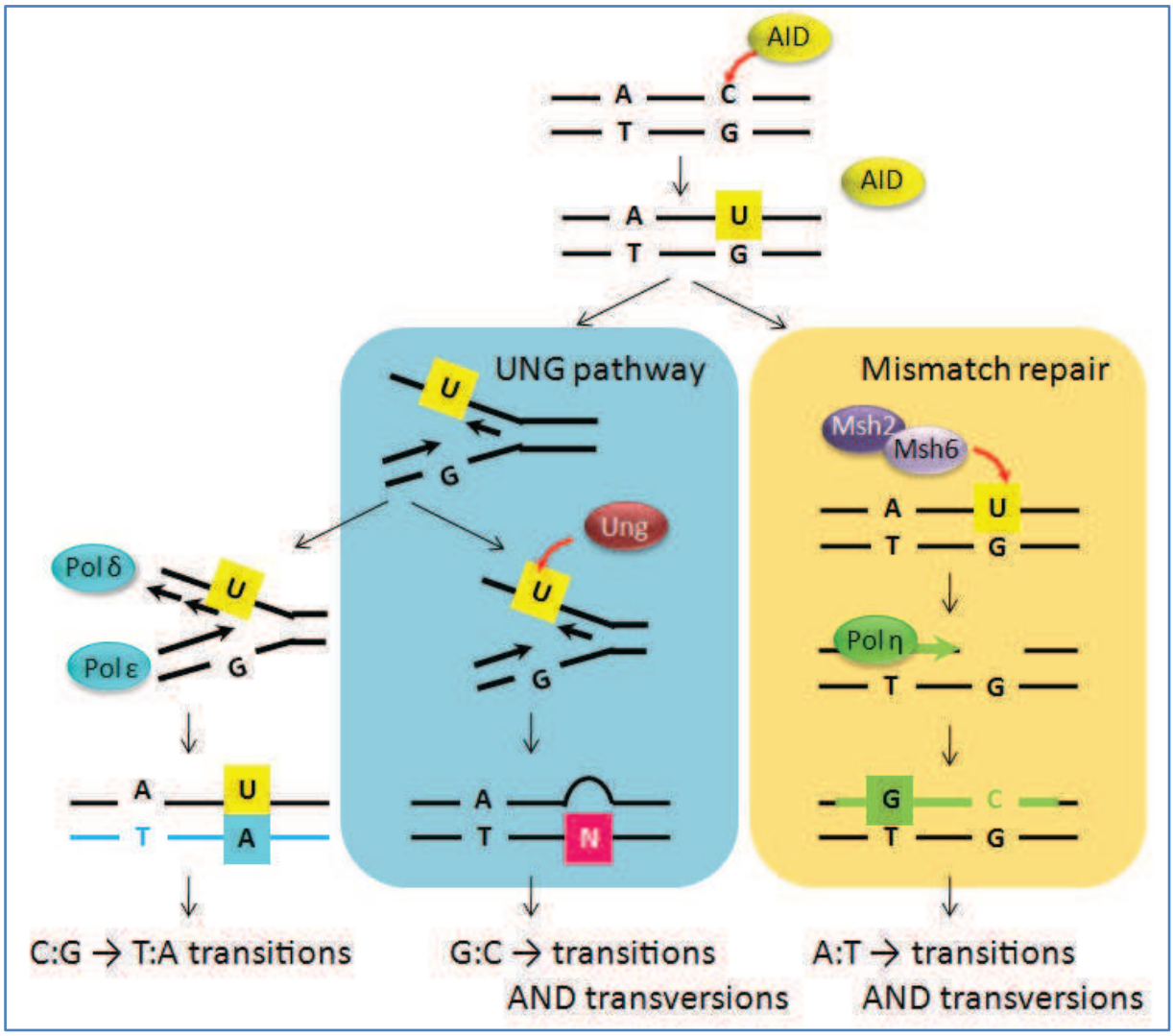
II. 1 Somatic hypermutation span and major properties

Once regions beyond V-exon and V(D)J junctions in hybridomas became accessible by DNA sequencing, the entire span of somatic hypermutation could be appreciated (Figure 19) It appeared that mutations start to accumulate at a sharp boundary 150 bp downstream of the promoter of the rearranged V exon, around the transcription start site (TSS). They peak at the level of the V(D)J region and gradually drop down in frequency some 1-2 kb downstream of the TSS, becoming undetectable in the 3' enhancer region. A second wave of mutation is observed downstream of the I_H promoters, extending within the switch region. However, mutations are never detected in Cμ genes and rarely in Cλ(Honjo and Alt, 1995).

Mutations in parts coding for variable regions are highly biased after rounds of affinity-based selection, but mutations in regions that are not subjected to selection (J intron, S regions, passenger transgenes) offer the less skewed mutation pattern of somatic hypermutation but are not totally exempt of other influences (proximity of the control regions, base composition). The major properties of somatic hypermutation are the following:

- In activated GC B-cells, all four bases are mutated. A is the most mutated base, followed by G, C and the least mutating T;
- A and T basepairs grouped together (A:T) mutate at the same rate as G and C basepairs together (G:C);
- as recorded from the non-transcribed strand (referred to also as top strand or coding strand) G mutates equally as cytidines, but between A:T basepairs a particular bias is observed since mutations seem to occur twice more often at A than at T positions ;
- concerning different types of mutation, the overall SHM pattern is biased for transition mutations (purine to purine or pyrimidine to pyrimidine) at all base positions.

Figure 20. Widely adopted model of somatic hypermutation in Ig genes, consisting of initiation phase in which AID deaminates cytosines, and resolving phase where uracil is handled by mutagenic pathways, giving rise to the mitations at all four bases. AID: activation-induced deaminase, UNG: uracil-N-glycosylase, Pol ε, Pol δ: replicative polymerases; Msh6-Msh2: mismatch repair sensor complex; Pol η: main A:T mutator and translesional polymerase.



- Transitions represent a little more than half of total mutations, while randomly they should occur in only one third of the cases (4 transition types versus 8 transversion mutation types;
- mutations tend to concentrate within mutational hotspots; many of G:C mutations *in vivo* occur in WR**C**Y and R**G**YW motifs (R for purine, Y for pyrimidine, W for "weakly" paired bases A or T, underlined base being mutated) (Rogozin and Kolchanov, 1992) while A:T mutations arise in W**A** and to a lower extent at **T**W motif (Rogozin et al., 2001). However, one thing to keep in mind is that mutations are never strictly confined to these hotspots, nor do all hotspot consensus sequences in IgV region accumulate mutations;
- transcription is necessary for somatic hypermutation of IgV genes, and the rate of transcription correlates to the rate of mutation (Peters and Storb, 1996).

Since the experimental confirmation of Lederberg's proposition on mutating mechanisms enhancing the best antibody-antigen fit (McKean et al., 1984), somatic hypermutation has for a long time remained an intriguing phenomenon since the factors responsible for the introduction of mutations had given much trouble to researchers in quest to identify them. The ultimate mutator protein that initiates the whole process was long sought for until the landmark 1999 paper of Tasuku Honjo's group (Muramatsu et al., 1999) introduced activation-induced cytidine deaminase (AID) on the scene and forever changed the landscape of research on somatic hypermutation.

II.2. Current model of molecular mechanisms underlying somatic hypermutation

Even back at the time when the Ig gene mutator or recombinase was still unknown, the following steps were agreed upon as indispensable for the somatic hypermutation to proceed (Figure 20):

- **the initial entry in one of DNA strands**: on the basis of widely adopted models in the scientific community, a cytidine deaminating enzyme AID introduces uracils along the region that is subject to hypermutation in Ig genes. Uracil in DNA is a frequent product of spontaneous deaminations and as such is quickly treated by classical repair pathways, but these pathways are somewhat subverted from their role in high-fidelity repair when recruited onto Ig genes during somatic hypermutation;
- **a subsequent spreading of mutations on the surrounding bases**: studies on knock-out mice suggest that uracil recognition happens the same way as in canonical error-free pathways. The base excision repair component uracil-N-glycosylase (UNG) recognizes and cuts off the uracil base, leaving **abasic site** (AP site) behind. Mismatch

- repair pathway recognizes the uracil within the U:G mismatch and proceeds in degrading the DNA surrounding the lesion. Both pathways rely on the special class of highly error-prone translesional polymerases to either restart the replication fork that stalled opposite the abasic site, or to fill in the gap after strand degradation. These polymerases have the intrinsic ability to misincorporate when facing certain template nucleotides. The joint action of these mutagenic repair pathways along with translesional polymerase recruitment accounts for mutations on all four template bases, even though the entry was initially made through cytosine only.

II.3. Initiating phase: AID as a mutator

The milestone paper of (Brenner and Milstein, 1966) predicted a "restriction enzyme" that made its way in DNA and allowed for a single-strand patch degradation and error-prone gap filling by some polymerase lacking proofreading qualities. Usual suspects, the proteins known to incise DNA like RAG1 (Zheng et al., 1998) were checked for their participation, but none of them seemed to contribute to the process.

In 1999 in Honjo's lab, Muramatsu identified a novel B-cell protein based on the subtractive hybridization of transcriptional programs between a resting B-lymphoma cell line and the same line induced to undergo isotype switching (Muramatsu et al., 1999). This new 24-kDa protein belonged to a APOBEC family (ApoB mRNA editing component) of RNA deaminases, whose member APOBEC1 in vertebrates was known to deaminate cytidines within the RNA transcript of apolipoprotein B, a component of low-density lipoprotein (Davidson and Shelness, 2000). By changing C to U in this RNA, a premature stop codon was created and allowed for translation of a shorter apolipoprotein B isoform. This newbee APOBEC member, dubbed *a*ctivation-*i*nduced cytidine *d*eaminase (AID) was shown shortly after to be indispensable for somatic hypermutation and class-switch recombination due to the complete impairment of these processes in AID-deficient mice (Muramatsu et al., 2000), as well as by the discovery of human AID deficiency in sufferers of hyper-IgM syndrome variant (Revy et al., 2000). The SHM mutator protein was finally unveiled.

II.3.1 Biochemical properties of AID activity

AID possesses a unique enzymatic activity in vertebrates since so far known cytidine deaminations in DNA happen either spontaneously or induced by chemical reagents (nitrosic acid, bisulfite). AID shares cytidine deaminase active domain with members of APOBEC family and with other bacterial cytidine deaminases acting on free cytosines (Neuberger et al., 2003). Different AID deletion mutants allowed for dissection of a C-terminal domain which is indispensable for class switch recombination and an amino-terminal domain responsible for

SHM and gene conversion, another Ig gene diversification mechanism (Barreto et al., 2003; Shinkura et al., 2004) (Chapter II.6). Therefore, two functions are physically unliked in AID protein and their respective deletion mutants do not influence the other diversification process (Shinkura et al., 2004). CSR-inducing domain overlaps with the specific AID nuclear export sequence and intimately links AID's capacity to perform class switch with AID shuttling capacity (Geisberger et al., 2009).

According to the latest consensus, AID deaminates cytidines within DNA. Several pieces of evidence account for this specific enzymatic activity within a family of RNA deaminases, since:

- AID was found to mutate endogenous genes and episomal substrates when expressed in bacteria (Petersen-Mahrt et al., 2002) and endogenous V-regions in hybridoma transfected with AID-encoding vectors (Martin et al., 2002);
- AID was found localized in chromatin of S_H regions when assaying for class switch recombination in cells, and these regions do not exist in mature RNA (Ranjit et al., 2011);
- *in vitro*, AID binds and mutates single stranded DNA but never dsDNA or RNA-DNA hybrids (Bransteitter et al., 2003; Dickerson et al., 2003; Pham et al., 2003);
- *in vitro* AID mutation pattern was soon confirmed by mutation profiling of *Ung*$^{-/-}$*Msh2*$^{-/-}$ mice lacking all mutagenic pathways that further diversify AID-generated uracil (Rada et al., 2002a). The Ig genes of these mice allow the analysis of *in vivo* AID footprint, which matches perfectly RGYW/WRCY focusing in somatic hypermutation hotspots *in vivo*.

One should also mention that, if Apobec-1 was the only known member of the family at the time of AID discovery, the eventual characterization of Apobec-3 genes, which are markedly expanded in primates and are all DNA-based cytidine deaminases, and the later appearance of Apobec-1 in evolution compared to other members of the family, has turned the activity of Apobec-1 into the exception and not the rule (Conticello et al., 2007).

II.3.2. AID expression is induced upon B-cell activation and controlled at post-transcriptional level

The limiting factor for SHM and CSR to occur is the AID expression level. Consequently, both processes can be regulated by fine-tuning of AID expression or AID activity at multiple levels. It is always the story about being 1) **at the right place on the genomic level** (Ig locus with all issues on SHM/CSR targeting, Chapter II.5) as well as **at the cellular level** (nucleus where AID catalyzes versus cytoplasm where it is more stable but inoperative); 2) **at the right time** (AID expression is under transcriptional repression outside

of activated B cell status; 3**) at the right amount**, not too little (appropriate transcriptional activation, post-translational modification that enhances SHM and CSR targeting) but not too much either, for the consequences can be pathological (post-transcriptional regulation).

Mouse *Aicda* gene is localized on chromosome 6 and encodes for 5 exons, all of them included in the mature RNA transcript. In humans, four regions were identified to govern *hAicda* locus expression (Tran et al., 2010): region 1 contains immediate upstream core promoter, region 2, which is identified within first *hAicda* intron, contains binding sites for Myb and E21 transcriptional inhibitors but also a B-cell transcriptional activator Pax5-binding site. In non-activated B-cells, these two regions recruit factors that act in convergent manner to maintain AID expression at extremely low basal level. Only upon B-cell activation, T-cell secreted cytokines (IL-4, TNFβ, interferon-γ) induce a signaling cascade that allows for expression and binding of transcriptional activators Stat6, NFκB and C/EBP to the sites in the enhancer region 4, localized 7 kb upstream of *Aicda* transcription start site. No known transcription factor binding sites were identified within region 3, localized in the 3' untranslated region.

Therefore, AID is expressed at the adequate levels for somatic hypermutation and isotype switching only in activated mature B-cells. It is detected at 10-fold lower levels in normal mouse embryonic stem cells and primordial and mature germinal cells where it has been proposed to participate in modifying DNA methylation patterns (Fritz and Papavasiliou, 2010; Wu and Zhang, 2010).

Aicda transcripts possess miRNA-binding sites within 3' untranslated region that allows binding of miRNA-151 and miRNA181b. Binding of the first miRNA is known to be upregulated in class switch recombination (Dorsett et al., 2008; Teng et al., 2008) and probably contributes to the AID level decrease in post-germinal center reactions, preventing the unneeded AID from further accumulating in the cell and inflicting genomic damage. Expression of the 181b miRNA was downregulated in CSR *in vitro*, potentially allowing for AID amounts to reach a peak when it is necessary for CSR to occur (de Yébenes et al., 2008).

II.3.3. AID activity is modulated by protein localization, stability and post-translational modifications

AID is considered to be small enough to passively diffuse through nuclear pore complex, but this could never be observed (McBride et al., 2004). It was generally thought that AID actively shuttles between nucleus and cytoplasm, since it is predominantly cytoplasmic in lymphocyte B-cell lines (Aoufouchi et al., 2008a; Patenaude et al., 2009; Rada et al., 2002b). However, this view is being questioned after a correlation was proposed between AID localization in the DT40 cell line and the cell cycle status of DT40 (Ordinario et al., 2009).

Several mechanisms contribute to AID localization in the cytoplasm:

- active CRM-1 mediated export using nuclear export sequence in AID C-terminus (McBride et al., 2004);
- active nuclear import *via* yet unidentified import factors (karyopherin-α importin, CTNNBL1, GANP) using a putative conformational nuclear localization sequence within AID (Patenaude et al., 2009);
- cytoplasmic retention by sequestering AID in cytoplasmic complex with EF1α (Häsler et al., 2011; Patenaude et al., 2009).

The equilibrium of these three DNA localization mechanisms regulates the level of operative AID. This level can be maintained constant, or it can be further modulated to increase the AID level for the purposes of diversification downstream of some signaling event or depending on the cell cycle.

Furthermore, the AID protein is stabilized upon leaving ER by being funneled from DnaJa1-Hsp40 complex (Orthwein et al., 2012) into a complex with Hsp90. The latter AID partner permanently checks for any misfolding, correcting it by ATP-hydrolysis and protecting AID from proteasomal degradation in cytoplasm (Orthwein et al., 2010). However, AID is found to have shorter half-life in nucleus than in cytoplasm due to nuclear proteasomal degradation, mediated by both ubiquitin-independent pathway (Uchimura et al., 2011) and ubiquitination-dependent mechanism (Aoufouchi et al., 2008a). Therefore, modulating AID stability is another way of controlling AID levels in hypermutating cells.

Several phosphorylation sites were identified, but currently only S38, the phosphorylation target of protein kinase A (Basu et al., 2005) was found to associate with RPA protein (Chaudhuri et al., 2004; McBride et al., 2008). This phosphorylated AID-RPA association upregulates CSR and SHM and accounts for AID targeting to Ig variable genes (see further). The function *in vivo* of of other AID phosphoforms awaits further characterization.

II.4. Resolving phase: mutagenic repair pathways are not just another oxymoron

After spontaneous or chemically induced deamination of a cytidine, the resulting uracil within DNA will not impede normal processes of replication or transcription. If uracil is left intact at the deamination site, replicative polymerases will elongate opposite to the uracil base without stalling and give rise to G:C transition mutations in newly synthesized DNA molecules (C → T and G → A). However, due to its premutagenic potential, uracil in DNA is efficiently removed by two DNA repair processes: base-excision repair (BER) and mismatch repair (MMR). Some 100-200 uracils that are generated per day in the human

Steps in repair	BER		MMR factors..... ....in complexes:	
Removal of....	spontaneous deaminations and dUTP incorporations		mismatches and replication slipping errors	
Damage recognition	UDGs: UNG 1 and 2, TDG, SMUG, MBD4		MSH2 -MSH6	MutSα
			MSH2 -MSH3	MutSβ
Entry into DNA	APE1 : 5' endonuclease		PMS2 -MLH1	MutLα
			PMS1 -MLH1 (?)	MutLβ
Handling the neighbouring DNA	long patch	short patch:1bp	EXO-1: 5'→3' activity	
	FEN1/DNaseIV	Pol β as dRPase		
Gap filling	PCNA + Pol β	Pol β	PCNA +Pol δ	
Ligation	lig 3/XRCC1	ligase 1	ligase 1	

Table T1. A summary of conventional base-excision repair and mismatch repair.

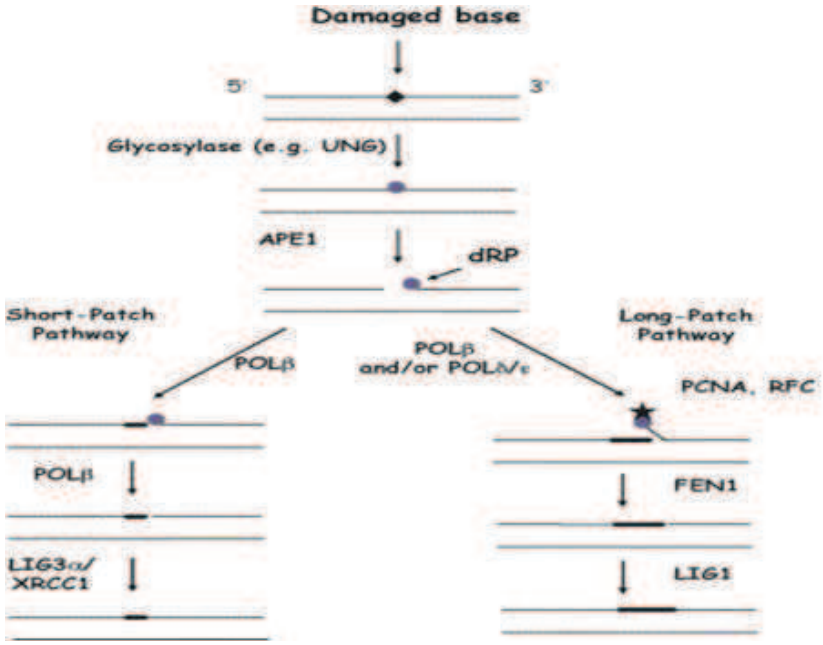


Figure 21. Schematic representation of base-excision repair (BER) pathway.
The lesion recognition step and entry into DNA are common for two branches of BER pathway : one of uracil glycosylases recognizes the uracil base and excises the N-glycosyl bond, leaving an abasic site. APE-1 endonuclease excises the phosphodiester bond at 5' of the abasic site, while in short-patch BER repair polymerase β removes the remaining deoxyribose-phosphate and fills the one nucleotide gap by inserting the correct one. LIG3/XRCC1 complex proceed to ligation. In case of an altered abasic sites difficult for deoxyribose-phosphate elimination by polymerase β, PCNA is recruited to the site and allows for 2-13 nucleotide-long patch synthesis by polymerase β. The existing strand is displaced and progressively degraded by exonuclease complex FEN1/DNaseIV. Patch is finally religated by ligase 1. *Source:* (Wilson and Bohr, 2007)
APE-1: apurinic/apyrimidinic endonculease 1; XRCC1: x-ray cross-complementing protein 1;
PCNA: proliferating cell nuclear antigen ; FEN1: flap endonuclease.

genome (Kavli et al., 2007; Lindahl, 1982) are thus repaired quickly and in an error-free manner. These repair processes appear partially modified in somatic hypermutation: uracils administered by AID are undoubtedly being recognized the way they are used to. Nevertheless, not only a fraction of them escapes and keeps on being fixed, but the outcomes of uracil handling get more diversified, with even more mutations introduced in the flanking sequence. How do C:G transversions and surrounding A:T mutations occur in the hypermutating region?

II.4.1. Canonical repair pathways

BER- and MMR-implicated enzymes remained the all-time favorites in deciphering the somatic hypermutation puzzle, so a brief overview of each repair mechanism is given in captions of Figures 21 and 22. Both pathways follow the canonical workflow when facing a DNA damage that they are specialized for (Table T1):

- damage recognition or modification that makes it more convenient to handle;
- entry in the DNA strand containing the damage after ensuring which one of the two is the damaged one;
- removal or displacement of neighboring DNA segment;
- gap filling by a high-fidelity DNA polymerase;
- ligation of the synthesized patch and the old DNA sequence.

Different UDG glycosylases. Most frequent lesions that can be easily removed by short-patch conventional BER are hydrolytic depurinations (leaving an apurinic site), spontaneous deaminations with a strong premutagenic potential (cytosine deamination generating uracil, 5-methyl cytosine deamination generating thymidine, etc.) and dUTP misincorporation by replicative polymerases when facing template A. Different uracil DNA glycosylases (UDGs), like UNG (uracil-N-glycosylase) 1 which is mitochondrial, UNG2, which is nuclear, and SMUG (single-strand selective monofunctional uracil-DNA glycosylase) traditionally attack uracil in DNA. Thymidine opposite to G (T:G) and less frequently uracil facing G (U:G) can be as well recognized by thymine DNA glycosylase (TDG) and methyl-CpG binding domain protein (MBD-4). All uracil DNA glycosylases leave an abasic site (AP site) as the common intermediate after deglycosylation. However, requirements for substrate configuration and limiting steps are not the same among glycosylases: UNG2 (called UNG for the sake of simplicity in further text) can attack uracil in both single and double stranded form, whereas TDG needs to make sure that T is mispaired with G in the other strand before proceeding to base excision.

Strand discrimination in canonical mismatch repair. In contrast to uracil DNA glycosylases that tightly fit uracil/thymidine within their active site, the mismatch recognition component MutSα is capable of identifying a non-Watson-Crick pairing geometry within a DNA heteroduplex. However, since mismatch repair identifies mispaired bases that are not

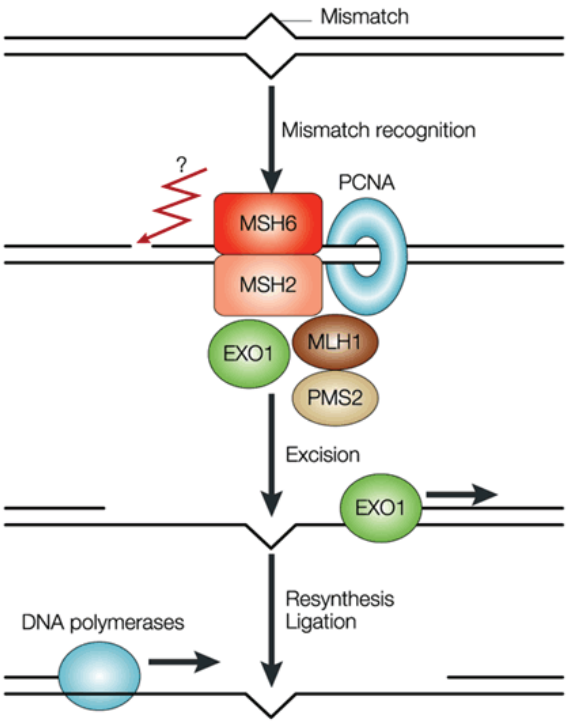


Figure 22. Model for eukaryote mismatch repair mechanism (MMR).
Mismatch repair recognizes typical post-replicative errors like small deletion/insertion loops due to polymerase slippage or single nucleotide misincorporations. MutSα heterodimer (MSH2 and MSH6 for **M**ut**S** **h**omologs) recognizes the mismatch englobed in double-stranded DNA and recruits MutLα heterodimer (composed of **M**ut**L** **h**omolog MLH1 and **p**ost-**m**eiotic **s**egregation increased PMS2). The latter complex scans for clues that indicate which strand is newly synthesized and therefore carrying the replication error. Once established, exonuclease Exo1 makes its entry and degrades the DNA strand comprising the error, while the single stranded DNA is protected by RPA proteins. PCNA clamp loaded onto the free 3' end of the degraded strand recruits Pol δ in order to fill the gap in error-free manner. Lig1 seals the remaining nick after gap filling. MutSβ is another mismatch recognition complex (composed of MSH2 and MSH3) which recognizes insertion/deletion loops of more than two nucleotides and whose resolution converges with MutSα towards MutLα intervention. However, another secondary mismatch complex MutLγ is recently identified, comprising MLH1 and MLH3, and is known to participate in meiotic recombination or as a MutLα backup. A third resolving mismatch complex MutLβ is composed of MLH1 and PMS1 whose function is yet to be identified. Source : (Martin and Scharff, 2002)

unusual in DNA, the problem of identifying which base of the mismatch is wrong requires some additional information. Mismatch repair is considered to be a post-replicative mechanism, therefore the wrong base in the mismatch is probably the one which got misinserted by replicative polymerases. Dilemma of newly synthesized strand in bacteria is solved by MutH, the sole mismatch repair factor with no homologues in yeast or humans. Indeed, due to a small time window between completion of DNA synthesis and methylation of adenines in newly synthesized DNA strand in bacteria, MutH has enough time to detect and nick the transiently unmethylated strand on either side of the mismatch (Kunkel and Erie, 2005; Längle-Rouault et al., 1987). However, in the absence of adenine methylation in eukaryotes, it is thought that early interaction of MutLα and PCNA localized at the nearby 3' end of the priming strand may help the strand identification (Lindahl et al., 1997; Shibahara and Stillman, 1999; Umar et al., 1996). Knowing that a canonical mismatch repair acts on post-replicative errors, the second proposition is that MutLα complex takes advantage of detecting the free 3' end of leading strand or 3' or 5' termini of freshly synthesized Okazaki fragments to "sense" on which strand the replication is still ongoing (Holmes et al., 1990; Lacks et al., 1982). However, the 3' end of a discontinued strand can be quite far away; furthermore, the 3' end can be located on the 3' side of the mismatch, meaning that EXO-1 5'-3' degradation will proceed away from the mismatch. Recently, a third source of single strand nicks has been identified: once a remote 3' free end is detected, nicks can be administered in the vicinity of the mismatch (and also at its 5' side) by the PMS2 protein, which possesses a cryptic endonuclease activity (Kadyrov et al., 2006, 2007).

A new controversial issue emerged with a recent study on EXO-1 separation-of-function mutant (Schaetzlein et al., 2013) showing that the exonuclease activity of this enzyme is dispensable for canonical *in vivo* mismatch repair, meiosis, somatic hypermutation and class switch recombination; the only process that was downregulated in presence of this EXO1 mutant was double strand break repair. Therefore, it seems that only the structural function of EXO-1 is indispensable in somatic hypermutation, while some other exonuclease can be recruited as backup to create the gap.

II.4.2. Translesional polymerases

DNA damage usually hampers vital processes carried on during the cell cycle like replication, induces signal transduction at cell cycle checkpoints and arrests further progress through phases. Depending on the environmental cues and the overall condition of the cell, DNA damage can be tackled in two ways. It can be efficiently repaired error-free *via* classical multiprotein repair pathways. However, if damage is neglected until S-phase or inflicted in the course of replication, it can hamper the progression of the replication fork. If the stalling goes on for a while, replicative machinery can disassemble and both double strand breaks and

Family	Pol	Gene	Template damage for error-free TLS	Associated activities	3'→5' exo	Error rate on undamaged DNA	SHM phenotype in single KO mice
Y	**Rev1**	REV1	N^2-G adducts; AP site	dCTP transferase; catalytic function dispensable		10^{-2}-10^{-3}	decreased G:C Tv (Jansen et al., 2006)
B	**Pol ζ**	REV3 (POL Z)	all damages will do	Most frequent extender		10^{-4}	embryonic lethal (Bemark et al., 2000; Esposito et al., 2000a)
Y	**Pol η**	RAD30 (POLH)	CPD 8-oxoG ; G-cysplatin-G	accommodates base dimers in the active site	No	10^{-2}-10^{-3}	decreased A:T (Delbos et al., 2005; Martomo et al., 2005)
Y	**Pol ι**	RAD30B (POLI)	γHOPdG; 8-oxoG ; AP sites	Hoogsteen base-pairing favored; inserts G facing T; dRP lyase	No	10^{-1}-10^{-5}	no change (Delbos et al., 2005)
Y	**Pol κ**	DINB (POLK)	N^2-G adducts; γHOPdG	Extender; introduces -1 frameshifts	No	10^{-3}-10^{-4}	no change (Schenten et al., 2002)
A	**Pol θ**	POLQ	AP sites	DNA-dependent ATPase; dRP lyase		10^{-2}-10^{-3}	controversial

Table T2. DNA polymerases participating/proposed to participate in somatic hypermutation.
For comparison, fidelity of replicative polymerase with proofreading activity is 10^{-6}-10^{-7}; when replicases lack proofreading activity: 10^{-4}-10^{-5}.
Abbreviations: G- guanin; C-cytosin; T-thymine; AP sites – apurinic/apyrimidinic sites; ex. CPD – cyclobutan photoproducts; T^T – thymine dimers; G-cysplatin-G : intrastrand cross-linkage by cysplatin; dRP lyase – deoxyribose-phosphate lyase; γHOPdG –γ-hydroxy-1, N^2-propano-2'deoxyguanosine; TLS – translesion synthesis; Tv-transversions; Ts –transitions. Data pooled from(Casali et al., 2006; Foti and Walker, 2010; Goodman, 2002; Kunkel, 2003; Prakash et al., 2005; Weill and Reynaud, 2008)

chromosome translocations can be generated. Two options remain in a cellular decision-making scheme: either triggering apoptosis or inducing damage tolerance. The latter one was shown to be preferred over apoptosis even in highly developed eukaryotes. Damage tolerance is represented by two mechanisms, DNA lesion bypass and template switching. Not many details are available on template switching; it is generally considered as an error-free process that borrows the other newly synthesized daughter strand from in order to allow the replication fork to skip the blocking lesion. Lesion bypass can be also mediated by translesional DNA polymerases that are shown to be more error-prone than replicative polymerases. Once the lesion site bypassed by TLS polymerases, replication fork can resume and initial DNA damage can be further ignored but at the cost of introducing new potentially harmful mutation. DNA damage tolerance remains a risky interplay of odds for survival by successfully completing the cell cycle and for permanently altering an essential cell function.

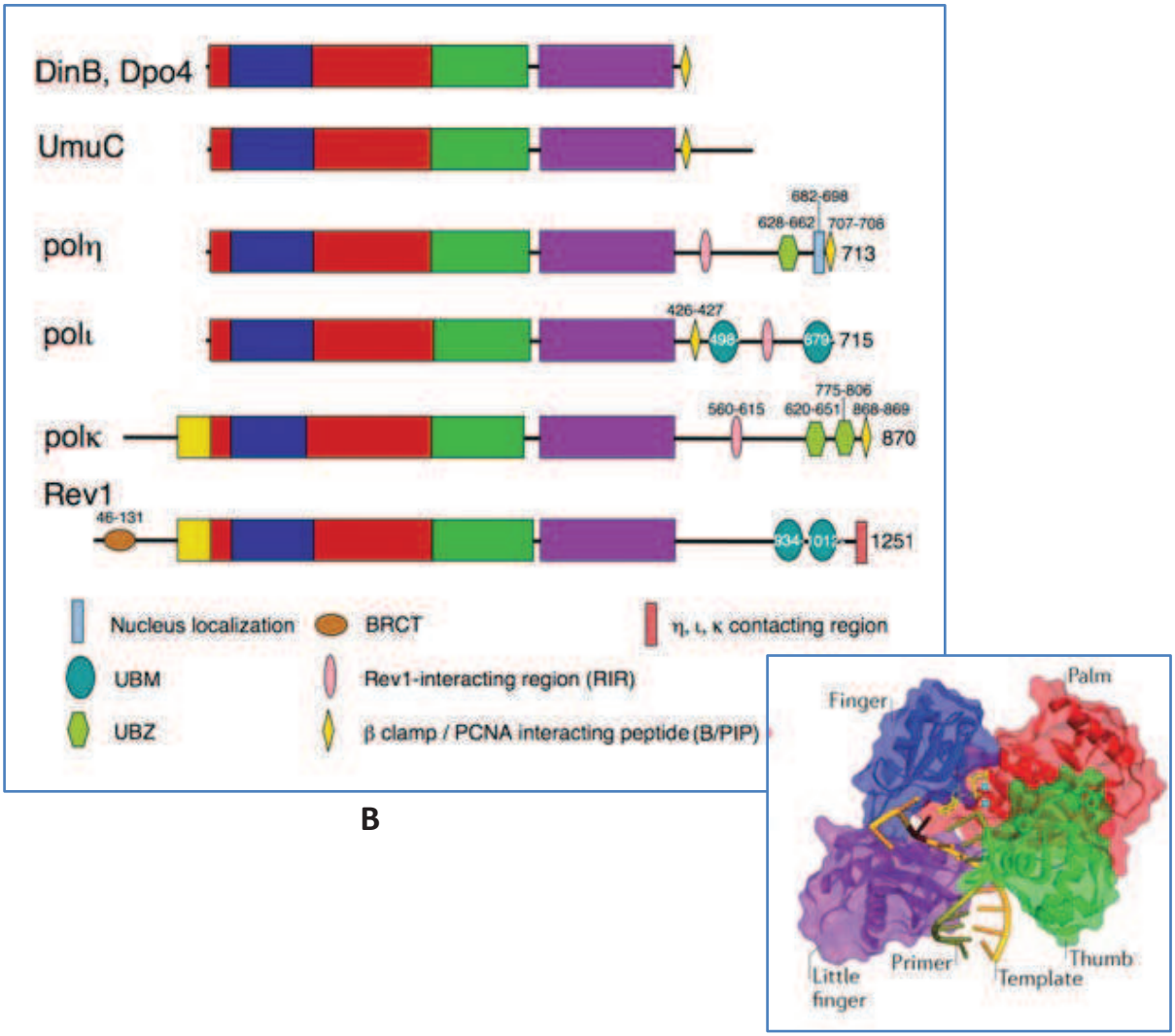
II.4.2a. Properties of translesional DNA polymerases

Translesional polymerases, grouped according to their cellular function, belong to distinct DNA polymerases families regarding their structure or fidelity. Rev1, Pol η, Pol κ and Pol ι belong to Y-family polymerases, Pol ζ belongs to B-family polymerases and Pol θ is among A-family polymerases. A summary of each polymerase classification, substrate specificity, associated activity and fidelity can be found in Table T2. We shall limit our introduction only to translesional polymerases that are confirmed or suspected to participate in somatic hypermutation, therefore we will left out Pol μ and λ in spite of the important place they occupy in other diversification processes in B-cells.

***Y-family polymerases*.** A boom in unraveling diverse members of Y-family polymerases occurred around 1999, quite some time after the two first member had been described, Umu complex (for UV-mutagenesis or Pol V) in bacteria and REV1 (reversionless locus) in *S. cerevisiae*. UmuD'C complex is responsible for SOS-bypass reaction in bacteria in response to UV mutagenesis or specific chemical treatments (Radman, 1974). However, eukaryotic polymerases have gone long way from Pol V and adapted to the specific but diverse roles in lesion bypass, with a common trait of being rather error-free when facing a cognate DNA damage subset. However, many of them remain error-prone when copying undamaged DNA, *in vitro* and *in vivo.*

PCNA (proliferating cell nuclear antigen) sliding clamp remains the principal docking platform for various TLS polymerases and provides for "polymerase switch" at the site of stalled replication fork by replacing a replicative with a translesional polymerase, and inversely when the fork is about to resume;

A



B

Figure 23. Structural properties of Y-family polymerases.
A) Schematic representation of different functional domains. The color code of the boxes corresponds to the domains shown in B. DinB, Dpo4, UmuC: bacterial Y-family polymerases. B) Crystal structure of the human Y-family DNA polymerase Pol η in a ternary complex with a cyclobutane pyrimidine dimer (CPD). The template strand is shown in rust colour, and the primer is olive green. The incoming dNTP is shown in yellow. The burgundy stick represents the position of the CPD. The small blue spheres represent the metal ions. The protein backbone is represented by ribbon surrounded by semi-transparent solvent accessible surface. Source: (Sale et al., 2012; Yang and Woodgate, 2007)

In spite of the reputation of being functionally divergent, Y-family polymerases share some general structural features:

- N-terminal 350-450 amino acids are well conserved since they shape the catalytic core. These residues are structured as to resemble a right hand: a palm, finger and a thumb (Figure 23). With finger and thumb smaller and stubbier than in replicative polymerases, the active site is much more solvent exposed and less constrained to Watson-Crick base pair geometry, hence more spacious. This allows for accommodation of bulky base adducts and crosslinks that would otherwise discourage replicative polymerases; however, a trade-off against this unusual catalytic site architecture is the smaller processivity and decreased fidelity of Y-family polymerases.
- C-terminal appendage is somewhat less conserved, and contains little finger domain (LF), polymerase-associated domain (PAD) that provides for contact among polymerases (discussed further), PIP boxes in some polymerases (PCNA-interacting peptide) allowing for the interaction with PCNA, UBM (ubiquitin-binding motifs) or UBZ (ubiquitin-binding Zn finger) that account for the specific interaction with K164 monoubiquitinylated PCNA. By contrast, the BRCT domain in REV1 N-terminus is the one responsible for PCNA binding.
- Little finger domain is the unique feature of Y- family polymerases; it contributes the most to the specificity of each Y polymerase, adopting different configurations when accommodating the most suited substrate for the catalytic activity of a given polymerase;
- none of the Y-family polymerases possesses the proofreading ~~3'~~5' exonuclease activity;
- among Y-polymerases, a task repartition when bypassing some of the lesions can be discerned: one polymerase is especially adept at incorporating an appropriate nucleotide opposite to the DNA damage (inserter), but another polymerase carries out the extension step from the unusual base pair (extender). When a DNA is not way too much distorted from Watson-Crick geometry after the insertion step, the same polymerase can carry out extension (Pol η, Pol ι, Rev1). Other Y-polymerases are almost or exclusively extenders (Pol κ) or require the extender activity of a polymerase from other family, like Pol ζ.

Other SHM candidates outside of Y-family. **Pol ζ (zeta)** is a B-family polymerase composed of two subunits, Rev3, the catalytic subunit and Rev7, the regulatory subunit. Many biochemical studies on this polymerase were done with the yeast enzyme since the cloning of the mammalian Rev3 was particularly tricky due to the its size. In yeast, Pol ζ is extremely inefficient in incorporating nucleotides opposite many lesions (except for thymidine dimers), but it turned out to be a perfect extender.

This was the unique case among TLS polymerases whose knockout in mice was not viable. Ever since, many models of REV3 downregulation were tested: Burkitt lymphoma cells BL2 knockout for REV3 had increased UV sensitivity and were delayed at S/G2 transition; *Rev3*$^{-/-}$ mouse embryonic fibroblasts barely divided; mice with Rev3 knockdown were healthy, with fewer B-cells in the spleen, weakened antibody affinity upon immunization and less mutation in IgV_H genes of memory B-cell, but the mutation pattern was unchanged from normal . Conditional knock-out in B-cells, in contrast to the last study, showed no impact on SHM, but a decreased CSR efficiency, increased chromosomal instability and impaired proliferation capacity (Schenten et al., 2009).

Pol θ (theta) conserves high domain homology with other members of the A-family polymerases; however, its fidelity is much lower. It is a preferential A inserter and is known for robust performance in bypass of an abasic site. Contrasting results on SHM participation were reported by three different teams. On the basis of the work performed by the team of P. Gearhart, it seems nevertheless safe to conclude that the contribution of Pol θ is minor, if any (Martomo et al., 2008; Masuda et al., 2006; Zan et al., 2005)

II.4.2b. PCNA: replication and repair tool belt

In eukaryotes, PCNA is a homotrimer loaded onto the DNA template-primer end junction by replication factor C (RFC) and in itself has no enzymatic activity, but it slides along the DNA axis and recruits partners by protein-protein interactions. However, its most important function is to help the loading of DNA replicative polymerases and enhance their processivity. Most PCNA partners, among them replicative polymerases, bind PCNA *via* a PCNA-binding peptide (PIP) domain. By so doing, PCNA maintains its role of "replication traffic cop" or "matchmaker protein" even in translesion bypass. When the replication machinery assembled around PCNA stalls at a DNA lesion, an event cascade is initiated *via* the Rad6/Rad18 complex in order to monoubiquitinylate PCNA at conserved Lys 164 residue (Hoege et al., 2002). This modification is considered to weaken PCNA-replicative polymerases interaction and favor high affinity binding of translesional polymerases to PCNA *via* their PIP boxes and ubiquitin-binding domain/Zn motifs. The event order in reverse polymerase switch remains speculative: PCNA can get deubiquitinated or polyubiquitinated at the same residue but *via* non-canonical K63 intra-Ub linking (considered to entail error-free template switching mechanism) or translesional polymerases in their turn can undergo ubiquitinylation and dissociate away from the clamp.

Another role of PCNA comes from its polarity when getting loaded onto the template-primer 3' end in replication: namely, RFC loads it always with PCNA's so-called C-side (side with protruding C-termini from homotrimers) orientated towards the 3' end. This C-side contains the hydrophobic pocket that accommodates different PIP boxes of PCNA partners, replicative polymerases included. This polarized orientation with PCNA C-side always facing

the 3' growing end of nascent DNA strand is thought to contribute to strand discrimination during mismatch repair. This explains early recruitment of PCNA during mismatch recognition outside of its "coordinator" function during gap filling with error-free polymerases.

Lastly, PCNA was shown to interfere when the mismatch repair detects a 3' free end situated 3' of the mismatch, therefore risking the degradation of the right strand but in the wrong direction due to EXO1 ~~5'~~3' orientated activity. Together with RFC and maybe dependent of the mismatch "nature", PCNA is shown to activate PMS2 cryptic endonuclease function that will allow for the novel incision at the 5' end of the mismatch and the entry of EXO1 (Kadyrov et al., 2006).

II.4.2c. Polymerase eta in the spotlight

Pol η was already known in *S. cerevisiae* as RAD30 gene product at the time when a study of human Pol η deficiency in xeroderma pigmentosum variant (XPV) patients appeared. These patients were extremely sensitive to sunlight (Johnson et al., 1999; Masutani et al., 1999). These two works provided the first clear demonstration of the eukaryotic cell preference for lesion bypass over apoptosis. Cyclo-butane pyrimidine dimer (CPDs) is a UV lesion shown to be difficult to resolve by nucleotide-excision repair. However, Pol η was shown to readily insert two adenines opposite thymidine dimers, qualifying as principal actor in suppressing this UV-induced damage. Pol η was shown to accurately bypass other lesions, like 8-oxo guanine and guanine-guanine intra-strand links due to the action of cysplatine. The synthesis fidelity opposite these lesions was the consequence of a unique active site among polymerases: while normally a replicative polymerase lodges one single template base in its active site and the previous one is pushed outside, Pol η accommodates both pyrimidine dimers when bypassing this UV lesion and takes care not to loop them out to avoid frameshifting, typical for Pol κ bypass. Pol η is not processive; it incorporates only a few bases past the lesion before dissociating from the strand, and dissociates even faster after a misincorporation (Washington et al., 2001).

Pol η carries out UV-lesion bypass in S-phase by being recruited to stalled replication forks, that are detectable in nucleus 8h maximum after irradiation by specific granulations, termed repair foci (Kannouche et al., 2001). This Pol η recruitment is dependent on ubiquitination of PCNA on conserved Lys164 (Kannouche et al., 2004a).

Whereas Pol η remains the rare enzyme capable of correctly bypassing CPD lesions, it retains high error rate (10^{-2}) opposite to this cognate lesion, as well as when copying undamaged template. Biochemical studies of Pol η reported a stunning ability to polymerize over two hundred base-long gaps in lacZ gene substrates (Matsuda et al., 2001) or V_{κ}ox gene substrates (Pavlov et al., 2002) in spite of its low processivity. Polη preferentially inserted dGTP when facing T, and to lesser extent dATP facing A. The efficiency of the former

misincorporation was however behind Pol ι, a champion in low fidelity synthesis opposite to thymidines.

With such pronounced inclination towards low-fidelity synthesis, it is clear that Pol η should be tightly controlled in eukaryote cells at multiple levels. Studies mostly done in *S. cerevisiae* showed transcriptional control of RAD30 gene (McDonald et al., 1997) and persistent proteasomal degradation of RAD30 gene product, somewhat loosened in case of UV irradiation when the protein was transiently stabilized (Skoneczna et al., 2007). Pol η can be phosphorylated in human cells, which enhanced translesion activity (Chen et al., 2008) or it can be in turn ubiquitinated (Bienko et al., 2010). It is thought that Pirh2 Ub ligase-mediated mono ubiquitination of Pol η weakens its interaction with monoUb-PCNA, therefore allowing to more "avid" polymerase, e.g. bypass extender, to access PCNA (Chen et al., 2008; Moldovan et al., 2007). Pol η can be polybiquitinated by Mdm2 Ub-ligase which targets it for proteasomal degradation (Jung et al., 2012). Another way of being recruited to repair foci is *via* Rev1, notorious for being necessary in UV lesion bypass independently of its polymerase function (Moldovan et al., 2007). Mouse Rev1 contains a unique domain in its C-terminal moiety allowing it to interact with other Y-family polymerases via their polymerase associated domain PAD domain (Guo et al., 2003) and at the same time with monoUb-PCNA, and might therefore serve as a scaffold for polymerase swapping while browsing for the right one to handle the lesion (Friedberg et al., 2005).

II.4.3. Properties of MMR-mediated mutagenesis in SHM

Knock-out studies in mice turned out to be the most straightforward and reliable way to figure out the implication of different repair factors in somatic hypermutation. Mutation rate and pattern in hypermutating regions were used as readouts to evaluate the impact of the absence of the manipulated factor (Table T3).

When the mismatch repair pathway in somatic hypermutation was dissected, some surprising discoveries allowed for the distinction of error-free "canonical MMR" and highly error-prone "mutagenic MMR" (non-canonical or SHM MMR). Both pathways share the essential sensor complex, MutSα, and both knock-out mice (*Msh2*$^{-/-}$ and *Msh6* $^{-/-}$) have the same SHM mutation profile with decreased relative frequency of A:T mutations (Table T3, discussed further in Chapter III.1). This mutation pattern is reproduced for EXO-1 and Polη as well, indicating that all four factors (MSH2, MSH6, EXO-1, Polη) contribute to the error-prone MMR generating A:T mutation load.

MutSβ (comprising Msh2-Msh3) does not seem to play a significant role. Surprisingly, the absence of the transducer complex MutLα (composed of Pms2 and Mlh1) also did not seem to result in a disequilibrium in the mutation pattern or a major change in the mutation rate.

Knockout genotype	Mutation rate	Mutation spectra	References
Ung $^{-/-}$	no change	slight G:C increase= decreased G:C Tv; increased G:C Ts; A:T no change	(Rada et al., 2002a)
SMUG $^{-/-}$	no change	no change	(Di Noia et al., 2006)
TDG $^{-/-}$	embryonic lethal		(Cortázar et al., 2011)
MBD4	no change	no change	(Bardwell et al., 2003)
APE-1 $^{-/-}$	embryonic lethal		(Xanthoudakis et al., 1996)
MSH2 $^{-/-}$	decrease	Total load decreased= decreased A:T; increased G:C Ts; no change G:C Tv	(Ehrenstein and Neuberger, 1999; Rada et al., 1998)
MSH3 $^{-/-}$	no change	no change	(Wiesendanger et al., 2000)
MSH6 $^{-/-}$	decrease	Total load decreased = decreased A:T; increased G:C Ts; no change G:C Tv	(Wiesendanger et al., 2000)
MLH1 $^{-/-}$	no change	no change	(Phung et al., 1999)
PSM2 $^{-/-}$	no change	no change	(Winter et al., 1998)
EXO-1 $^{-/-}$	decrease	Decreased A:T; increased G:C Ts; no change G:C Tv	(Bardwell et al., 2004)
PCNA^{K164R}		Abolished A:T	(Langerak et al., 2009)
Ung $^{-/-}$ *Msh2* $^{-/-}$	No change	No A:T; no G:C Ts	(Rada et al., 2004)
Ung $^{-/-}$ *Msh6* $^{-/-}$	No change	No A:T; no G:C Ts	(Shen et al., 2006)
Rev1 $^{-/-}$	No change	Decreased G:C	(Jansen et al., 2006)

Table T3. Summary of generic knock-out studies for different repair factors and accessory proteins and their impact on somatic hypermutation.

This opened the question on how the entry for EXO-1 is provided outside of the replication context of canonical mismatch repair; with following attempts to provide an answer:

- an interplay between error-prone mismatch repair and BER components was proposed, especially when two AID-induced uracils were found close to each other(Pham et al., 2003; Xue et al., 2006) . One may be already occupied by MutSα, but the second one can be readily processed by UNG and half-cut by APE-1 endonuclease introducing a single-stranded nick and providing for EXO-1 entry (Schanz et al., 2009). However, the absence of impact of Ung inactivation on the A/T mutation pattern speaks against this proposition (Chapter III.1).
- others suggest an intervention of NBS1-MRE11-Rad50, exonuclease complex in homologous recombination double-strand repair, shown to promote somatic hypermutation (Yabuki et al., 2005). Apparently, this complex can be recruited downstream of UNG action, so another scenario of cooperation between two pathways in SHM may be possible (Larson et al., 2005).

Further characterization of the nicking machinery may also shed more light on the strand discrimination of mutagenic MMR. In a context where the term of nascent strand is irrelevant (as in somatic hypermutation), mutagenic mismatch repair can introduce the incision in U-containing strand as well as in the unmutated strand. The lack of directionality was demonstrated for MutLα-mediated nicking of a heteroduplex DNA substrate containing a bubble in a system reconstituted with pure enzyme components (Pluciennik et al., 2010).

The gap filling in conventional mismatch repair is delegated to high fidelity, proofreading-capable replicative polymerase Polδ. However, error-free gap filling in somatic hypermutation is highly compromised by unusual partnership between MutSα component and error-prone Pol η translesional polymerase. Two studies revealed almost simultaneously the implication of Pol η in somatic hypermutation: by analyzing A:T mutagenesis hotspots in somatic hypermutation and comparing them with the *in vitro* Pol η pattern (Rogozin et al., 2001) and by assessing somatic hypermutation in XPV patients (Zeng et al., 2001).

XPV patients, although heterogeneous in different mutations that abolished Pol η expression, yielded overall pattern that closely resembled the one of Pol η knockout mice (Delbos et al., 2005; Martomo et al., 2005). Pol η deficiency did not impact general mutation and maintained the A/T bias. By contrast, A:T and G:C were importantly unbalanced, with residual 15% of mutations at A:T (against normal 50%), accounting for the involvement of another unknown polymerase in A:T mutagenesis. Pol ι was discarded as Pol η substitute since double knockout *Polh*$^{-/-}$ *Poli*$^{-/-}$ showed the same profile as *Polh*$^{-/-}$ only (Delbos et al., 2005). In *Polh* $^{-/-}$ residual A:T mutation pattern, A:T transversions prevailed over transitions with a Pol κ recognizable pattern. In a subsequent study of double knockout *Polh*$^{-/-}$ *Polk*$^{-/-}$ (Faili et al., 2009) A:T mutagenesis dropped to 7%, showing that Pol κ can be recruited in

SHM as back-up. To our knowledge, no new polymerase candidate accounted for the remaining 7% of A:T mutations.

A surprising discovery came out of Msh2/Polη double knockout mice analyses in which all A:T mutations were completely abolished (Delbos et al., 2007). Moreover, *Msh2*$^{-/-}$ and *Msh2*$^{-/-}$ *Polh*$^{-/-}$ mice shared similar properties, like 50% decrease in overall mutation load and clustering of remaining mutations within G/C hotspots (more in Chapter III.1). This report imposed several conclusions:

- Pol η is the sole contributor of A and T mutations in mice; A:T mutagenesis takes place through joint intervention of MutSα and Polη;
- If A:T mutations are reduced but persist up to 15% in Msh2 absence and are entirely abolished in Msh2/Pol η double knockout, this means that in the absence of mismatch signal complex Pol η is recruited in a A:T mutagenesis salvage pathway *via* UNG. It is not known whether this may occur in physiological conditions.
- As long as MutSα is operational in the cell, it will remain the principal inducer of A:T mutagenesis independently of UNG presence. If Pol η is missing, other translesional polymerases (like Pol κ) may get involved but will always depend on MutSα direction.

Pol η recruitment in mutagenic MMR. Biochemical approaches had shown that MSH2-MSH6 associate with Pol η in the chromatin of cultured cells and that this complex even improves the rate of Polη-catalyzed *in vitro* synthesis (Goodman and Scharff, 2005; Wilson et al., 2005). Another piece of the puzzle came from a PCNA knock-in mouse designed to replace the conserved Lys 164 residue in mouse PCNA. Unmodified PCNA clamp loader was long ago shown to be the partner in error-free MMR (Umar et al., 1996). PCNA knock-out mice are not viable. However, this PCNA ^{K164R} mice generated a drastic drop in A:T mutations load in somatic hypermutation (Langerak et al., 2007). PCNA ^{K164R} mutant is known for Lys164 importance in PCNA ubiquitinylation (Hoege et al., 2002) and for the mutant inability to recruit polymerase η in UV-irradiation assays (Kannouche et al., 2004a), suggesting a role for ubiquitinylated PCNA in polymerase η enrollment in SHM MMR in just the same manner as in polymerase switch. However, residual A:T mutagenesis in PCNA ^{K164R} knock-in mice is proven to be independent of PCNA-ubiquitinylation and further dropped to the level of A:T mutation is *Ung*$^{-/-}$*Msh2*$^{-/-}$ mice when PCNA ^{K164R} mice were crossed with *Polh*$^{-/-}$ mice (Krijger et al., 2011). Namely, Ub-PCNA is dispensable even in classical translesion across UV and cisplatin-induced lesions, although the translesion efficiency and rate are significantly reduced (Hendel et al., 2011).

This casting for error-prone MMR actors is not restricted to the isolated case of SHM. In response to oxidative damage, BER operates at the damage site along with Ub-PCNA and Pol η recruitment which was relying exclusively on MSH2-MSH6 (Zlatanou et al., 2011). When DNA damage is inflicted by alkylating reagents, creating base intermediates which normally bind MSH2-MSH6 *in vitro*, there is an increase in PCNA ubiquitinylation and Pol η

recruitment to the chromatin highly dependent on MMR proficiency of the cell (Peña-Diaz et al., 2012). Unlike in oxidative damage, this Ub-PCNA-Pol η engagement in alkylating damage handling was impossible without MLH1 and therefore MutLα complex. In the same study, a situation of SHM MMR facing U:G mismatch was simulated *in vitro*: substrates containing U/G mismatches were repaired in presence of MMR-proficient B-cell extracts whose BER pathway and replicative polymerases were inhibited. MutLα complex was required for this artificial substrate repair and only in its presence did Ub-PCNA accumulate.

The length of Pol η synthesized patch. Another mystery of Pol η recruitment concerns the length of the gap it is able to fill in during somatic hypermutation. Aforementioned studies tried to deduce this patch length *in vivo* according to the mutation spread measured from the fixed deamination site. It seems that mutations spread away from the deamination site in both directions, predicting a filled patch size of 57 bp (Unniraman and Schatz, 2007) or up to 200 bp (Frieder et al., 2009). Others estimated it to be no more than 20-30 bp (Chahwan et al., 2012). It remains unclear what restricts this patch length *in vivo* since EXO-1 5'→ 3' nibbling in cell extracts can go as far as 150 bp (Genschel et al., 2002).

II.4.4. Potential outcomes of UNG intervention in SHM

Genetic studies of knock-out mice for different BER-implicated glycosylases had provided clear evidence for UNG involvement in somatic hypermutation (Table T3). Another potential candidate is thymidine glycosylase whose knock-out phenotype is embryonic lethal, probably due to its implication in epigenetic processes (Cortázar et al., 2011; Cortellino et al., 2011). SMUG was shown not to be highly expressed in mutating B-cell, to process uracils in rather error-free manner with a very limited capacity to supplement UNG-induced mutagenesis in somatic hypermutation (Di Noia et al., 2006; Rada et al., 2004). MBD-4 knock-out had no major impact on somatic hypermutation pattern or mutation load.

Ung$^{-/-}$ mice showed no change in total mutation frequency compared to wild type mice, with an important decrease in G:C transversions and only slight decrease in A:T mutation load. This G:C transversion decrease was in line with the results obtained with Rev1$^{-/-}$ mice (Table T3), showing a dramatic depletion of G:C → C:G transversions which is consistent with Rev1 deoxycytidyl-transferase activity. The mystery on G:C transversion persists since no G:C → T:A inverter polymerase has been identified.

G:C transversions were always considered as hallmark of UNG-mediated active mutagenesis whereas G:C transition were thought to arise exclusively as replicative polymerases elongate the nascent strand across uracils that escaped other SHM pathways. However, GC transitions can be actively generated even during UNG-mediated mutagenesis, if the recruited translesional polymerase applies the "A-rule" (inserting A no matter the nature

of the facing base) when polymerizing across the abasic site (Reynaud et al., 2009; Weill and Reynaud, 2008).

Mutation pattern of UNG knockout mice provided other interesting conclusions:

- mutagenic MMR seems fully independent in generating A:T mutations;
- strand nicking by APE-1 downstream of UNG is dispensable for MMR processing of uracils, since MMR-generated A:T mutations remain unchanged in UNG absence. Either another glycosylase can take over the role of uracil removal and subsequent APE-1 recruitment at the abasic site in somatic hypermutation (a putative role for thymidine glycosylase) or other propositions on strand nicking may come into play (see paragraph on MMR strand discrimination).

Regarding other BER components downstream of glycosylase intervention in canonical pathway, all of them remain essential for individual's survival as indicated by the absence of BER-related cancers and embryonic lethal phenotype of all knockout mice (APE-1, Polβ, XRCC1). Irradiation chimera mice transplanted with Polβ knock-out cells was shown to affect more class-switch recombination than somatic hypermutation (Esposito et al., 2000); furthermore, some B-cell lines with enhanced hypermutation on G:C basepairs seem to have downregulated Polβ (Poltoratsky et al., 2007) and therefore promoted UNG-mutagenesis on G:C. XRCC1 heterozygous mice had 50% increase in total mutation load and unchanged mutation pattern (Saribasak et al., 2011), adding to the proof of error-free BER in ongoing somatic hypermutation in V(D)J region.

II.5 Targeting of somatic hypermutation

Due to the high risk of introducing deleterious mutations elsewhere in the genome, AID was thought to be strictly targeted to the Ig loci. Therefore, somatic hypermutation targeting used to be the synonym for AID targeting. Other non-Ig genes were reported to be mutated by AID (Shen et al., 1998) but it was not until the seminal work of Schatz group (Liu et al., 2008) that AID was acknowledged to mutate on the genome-wide scale. Further deep-sequencing studies reported even bigger number of AID genomic targets (Yamane et al., 2011) but the doubt persists whether this technique overestimates AID "anchor" sites in the genome.

Another objective of the study by Liu et al. and another group (Hasham et al., 2010) was to temperate the risk of AID-mediated whole genome mutagenesis by showing the importance of repair pathways in correcting AID-induced mutations outside of the scope of Ig variable region genes. AID was first shown to mutate genome-wide at a basal level which is 50-100 times lower than in the Ig loci. The control of the mutation load (as some mutagenesis persists in genes like BCL-6, CD83 or Pim-1) or total elimination of mutations were dependent on classical repair pathways (MMR, BER, homologous recombination). By

contrast, the effect of AID-mediated mutagenesis in the immunoglobulin genes is further exacerbated by the action of specific error-prone repair at this region. Outcomes of AID-mediated mutagenesis became the matter of region-specific equilibrium between error-prone and error-free repair: in non-Ig genes, the error-free repair largely prevails; in Ig gene diversification, error-free repair is not absent but is overthrown by the error-prone mechanisms (Liu and Schatz, 2009). Hence the need to reformulate the term of somatic hypermutation targeting, since the problem became two-headed: there is AID targeting and there is error-prone repair pathway favoritism at the Ig loci. For the moment, we shall concentrate on the AID targeting, based on three major levels:

1) **at the global level**, where Ig gene diversification contrasts the rest of the genome that is mutated at the basal level. The common denominator of Ig genes and AID-mutated genes elsewhere in the genome was the high transcription level. Somatic hypermutation was already shown to depend on transcription that provides to AID the substrate it catalyzes the best: single stranded DNA within transcription bubbles. However, there were highly expressed genes that were not shown to mutate; therefore, cis-acting sequences of Ig genes become the favorites for AID targeting. Known transcriptional regulator sequences, such as IgV promoter, intronic enhancer and 3' enhancer did not prove to be sufficient to target SHM (Betz et al., 1994; van der Stoep et al., 1998); however, new locus-specific sequences were identified to be necessary and sufficient for AID recruitment to the Ig locus in chicken DT40 cell line separately from transcription requirements (Blagodatski et al., 2009; Kothapalli et al., 2008), but these sequences await validation in mammal models.

   The second means of attracting AID specifically to the Ig locus were trans-acting factors, proteins that are able to bind it and retain it in the chromatin context of Ig genes. This quest for B-cell specific trans-acting factors was somewhat discouraged by the studies of eukaryotic systems other than B-cell lines (and of prokaryotic systems) in which ectopically expressed or endogenous AID was mutating transcription cassettes that contained Ig genes or other substrates like GFP (Bachl and Olsson, 1999; Petersen-Mahrt et al., 2002; Yoshikawa et al., 2002).

   Another way for AID to access the Ig locus was proposed by the means of interacting with E2A, a typical B-cell transcription factor. E2A binds to E-boxes, sequences found in Ig locus. Multiple E-boxes inserted within an Ig transgene increased somatic hypermutation of the transgene (Michael et al., 2003). Even though these sequences were able to modulate SHM frequency and therefore participate in targeting, the direct interaction of E2A and AID is yet to be demonstrated.

2) **at the regional level**, by delimiting the boundaries of SHM at Ig loci (no mutations in promoters and C-genes). There is a clear requirement for a promoter; IgH or Igκ promoters themselves are dispensable if replaced by other strong promoters. 5' boundary of somatic hypermutation would shift the same way as the upstream promoter is moved; if a promoter is inserted just 5' of the C-genes, they would start acquiring mutations as well. These observations allowed further study of AID interactions with RNA polymerase II complex and proved that AID associates with this complex only when transcription is stalling *via* the specific stalling factor Spt5 (Pavri et al., 2010). Another aspect of regional targeting is the apparent difference in targeting rearranged VDJ exons/J intron compared to S-region repeats. S-region, beside being enriched in hotspot sequences, are G-C rich as well, which accounts for a number of secondary structures forming in this region. Among them are DNA structures called R-loops that are formed when a transcribed DNA strand pairs with nascent RNA. R-loops appear to be the recognition motifs for AID, which can further attack looped-out non-transcribed DNA strand.

   These structures do not form in V genes. So far, two targeting requirements for V-region genes are identified: transcription bubbles which are the only means of providing ssDNA in this region of the Ig loci, and interaction of phosphorylated AID at Ser38 with RPA, which binds single-stranded regions in rearranged VDJ genes and recruits AID at the same time (Chaudhuri et al., 2004).

   Protein factors known to participate in AID targeting in other diversification processes are further detailed in section dealing with class switch recombination for a simple reason that most of them have validated interactions with AID when recruiting it to the S-regions. Further evidence is needed to confirm the direct implication of the same factors in targeting variable region genes *in vivo*.

3) **at the local level**, comprising targeting of AID to the hotspots. As we have already mentioned, not all hotspots sequences are targeted and cytidines outside of hotspots mutate as well. However, one group identified a motif in AID protein sequences that modulates recognition of C within WRCY classical motif and when replaced with the corresponding sequence from APOBEC3G, the preference for the hotspot changed along with the reduction in CSR and SHM (Kohli et al., 2010; Wang et al., 2010).

Unregulated AID activity is proven to yield pathological consequences: T and B-cell tumors due to AID overexpression in transgenic and bone marrow transplantation models (Komeno et al., 2010; Okazaki et al., 2003), double strand breaks initiated by AID which contribute to c-Myc-IgH translocations (Matsumoto et al., 2007), other off-target AID-dependent DSBs in non-Ig genes that contribute to germinal center-derived lymphomagenesis (Pasqualucci et al., 2008; Robbiani et al., 2009).

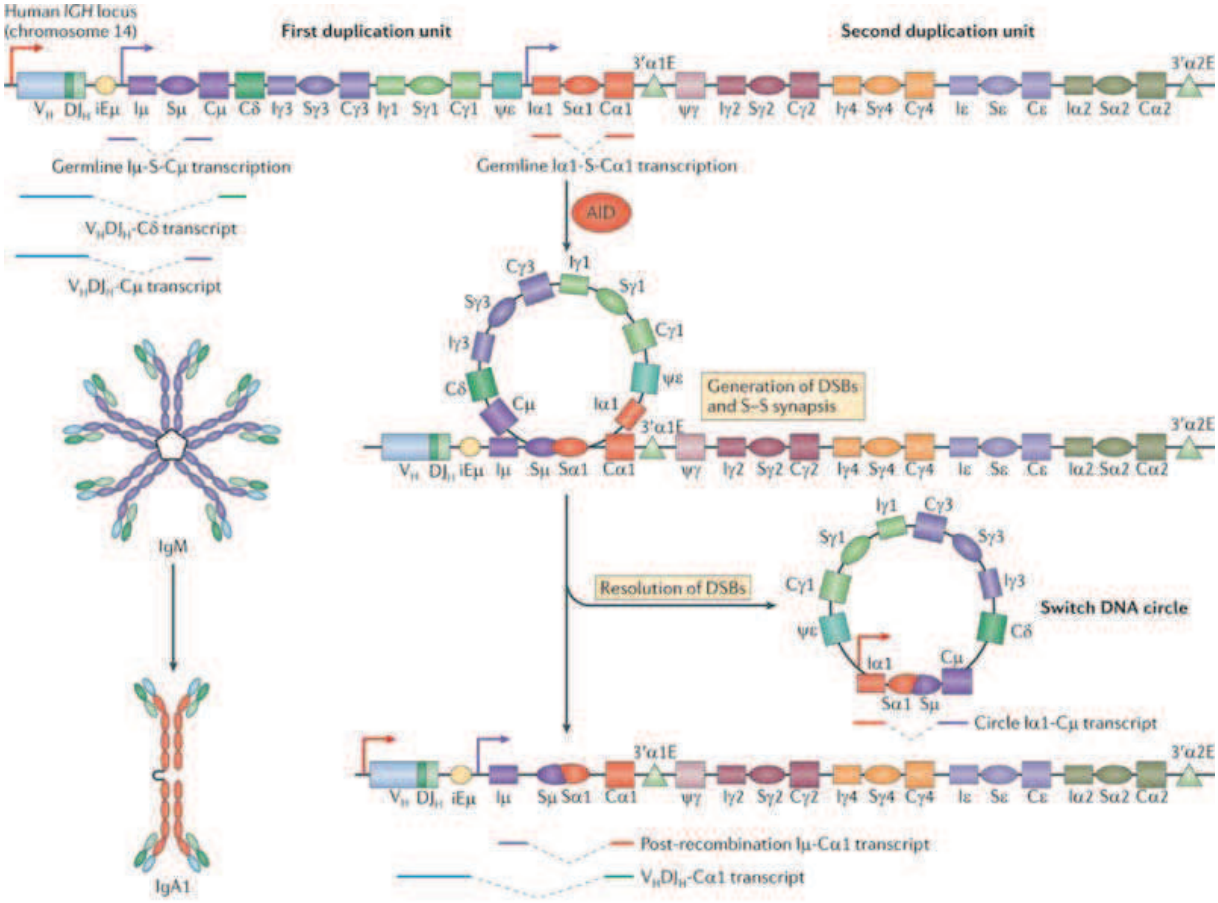


Figure 24. Class-switch DNA recombination (CSR) exchanges the gene encoding the immunoglobulin heavy chain constant region (CH) with one of a set of downstream CH genes. The figure depicts CSR between Sμ and Sα1 in the human immunoglobulin heavy chain (IGH) locus. ψ, pseudogene; 3′αE, 3′ α enhancer; iEμ, IGH intronic enhancer; IH, IGH intervening region. Source: (Xu et al., 2012)

II.6 Other secondary diversification processes with a same denominator: AID

II.6.1. Class-switch recombination

Class-switch recombination is the molecular mechanism for isotype switching, allowing for modification of antibody constant region and yielding new functions in antigen clearance. It can happen outside of germinal centers as well as during GC reaction. This modification happens by literally swapping the Cμ gene with another C gene located downstream, therefore bringing it immediately after the J_H intron and the Eμ enhancer and closest to rearranged V(D)J segment (Figure24).

Isotype switch is disrupted in *Aicda* knockout mice, indicating that AID-mediated cytosine deamination is required for the reaction onset. AID was shown to deaminate cytosines within or near the S region preceding the Cμ gene (donor region) and S-region upstream of specific C gene towards which isotype switch takes place (acceptor region).

Two signals elicit class switch recombination: primary signals that induce the expression of AID and other factors implicated in CSR, and secondary signals that contribute to the choice of the C-gene coding for the new isotype. Primary signals include CD40-CD40L interaction with helper T-cells in case of T-dependent antigens. Coupled signals from simultaneously activated TLR-4 and cross-linked BCRs in case of T-independent antigens (eg., bacterial lypopolysaccharide LPS) can also serve as primary signals for CSR. Secondary signals comprise cytokine requirements defining which acceptor S-region is to be selected and therefore direct class switch towards one of IgGs, IgE or IgA.

Two other requirements for isotype class switch are proliferation (B cells have to undergo a minimum number of divisions) and ongoing transcription over C_H genes directed from the corresponding promoter regions upstream of I_H exons (Chapter I.1.1). All acceptor I_H promoters contain cytokine-responsive elements, and the onset of the transcription from a single I_H promoter is dependent on the nature of cytokine signaling. Transcription is constitutive from Iμ promoters. Nowadays one can take advantage of this process and induce the switch of isolated mouse splenic B cells in culture by applying the recombinant cytokine cocktail known to stimulate cell proliferation and shift the switching towards one particular isotype (LPS only to induce transcription of Iγ3 promoter and switching to IgG3, LPS and IL-4 towards IgG1 and IgE, TGFβ for IgG2b and IgA synthesis, IFNγ for IgG2). This assay is particularly useful to delineate the capacity of mutant or modified CSR factors to restore the switch *in vitro* in cultured B cells deficient for the tested factor.

The synthesis of these sterile germline transcripts contributes to CSR at several levels:

- the chromatin structure at the donor and acceptor S-region is relaxed and allows for the access of initially recruited proteins, like 14-3-3 protein with its intrinsic affinity for binding 5'-AGCT-3' sequences (subset of WRCY motifs). This motif is highly enriched in the core of both acceptor and donor S-regions. It was demonstrated that AID interacts with this protein (Xu et al., 2010) through a C-terminal peptide sequence within its nuclear export signal, finally connecting dots on why AID NES and C-terminus are essential for class switch recombination to occur (Ranjit et al., 2011) .
- During the ongoing transcription mediated by RNA polymerase II transcription complex, the nascent RNA transcript hybridizes with the transcribed DNA strand, looping out the non-transcribed strand to form the R-loops. It is thought that these DNA secondary structures can impede transcription progression (Kenter, 2012), while recruiting factors to the stalled site. Other way of inducing transcription pausing may be RNA II polymerase collision with another polymerase molecule, elongating the transcript from an antisense promoter in S-region. In any case, one of these "stalling" factors, Spt5, is shown to interact and recruit AID onto the stalled RNA Pol II complex (Pavri et al., 2010). AID can be imagined to tether RNA Pol II complex *via* Spt5 and continue to "ride" it as the transcription goes on, scanning for the precise spot within the S-region where it can further interact with 14-3-3 protein.
- the whole I_H-S_H-C_H region is being transcribed, after which this sterile transcript is spliced to I_H-C_H mature transcript. Although this final transcript is non-coding, this splicing step is vital for CSR. This RNA processing *in situ* might account for the recruitment of splicing factors PTBP2 (Nowak et al., 2011) and others, known to further interact with AID. Enrollment of RNA exosome in S-region transcription has finally explained the mystery of how AID accesses the transcribed strand, occupied by the hybridized RNA: RNA exosome simply degrades it and opens the way for AID to act (Basu et al., 2011).

The major difference between SHM and CSR is found once AID is done with cytosine deamination: during CSR, uracils are processed to yield double strand breaks in Sμ and acceptor S region. This processing is however still not explained in detail, but it is considered that UNG contributes to uracil base removal and subsequent endonuclease recruitment that will introduce single-strand breaks. Two SSB created this way from the uracils not far from each other create a double strand break in one S-region. Remote SSB within a single S-region are postulated to be further processed with participation of MMR factors and some translesional polymerases in order to give rise to blunt or nearly blunt double strand breaks (Stavnezer et al., 2008). The MRE11-RAD50-NBS1 complex that is an early genome-wide detector of double strand breaks, triggers a phosphorylation cascade *via* ATM kinase to enrich phosphorylated DSB histone marks γH2AX in the C_H region and allow the entry of the

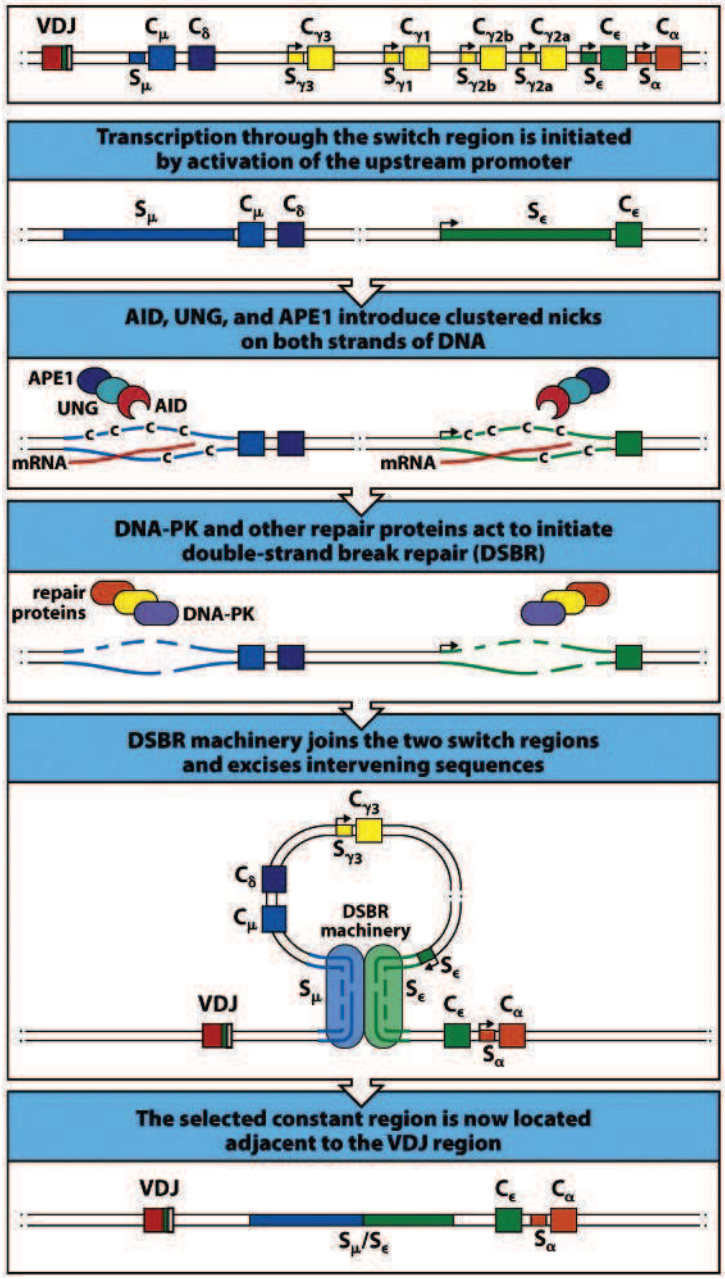


Figure 25. Scheme of class switch recombination. More details in the text. DSBR: double strand break repair. Source: (Murphy et al., 2012)

classical non-homologous end-joining pathway (NHEJ). NHEJ starts by the double-strand break sensor protein Ku70/Ku80 surrounding the free DNA ends and loading DNA protein kinase that will recruit other downstream effector proteins of NHEJ, Cernunnos and XRCC4/Ligase4. Finally, the intervening region between the double strand breaks in two S-regions will be removed and DNA ends will be joined into a hybrid Sμ/Sx region in a joint action of XRCC4 and ligase 4 (Figure 25).

Only few SHM factors are essential in isotype switching. UNG-deficient mice showed considerable reduction in switched isotypes, accounting for the molecular basis of hyper-IgM syndrome 4 in humans (Imai et al., 2003). MutSα and MutLα knockouts were shown to have reduced CSR, and it has been proposed that mismatch repair is involved in the gap-filling process processing two distant nicks into a double-strand break (Stavnezer and Schrader, 2006). The absence of different SHM translesional polymerases, among which only the mutagenic signature of pol η is recognizable in S-regions (Faili et al., 2004), do not affect CSR.

Another form of AID targeting to S-region is by modifying the existing epigenetic marks. Specific histone modification H3K9me3, reputed as repressive, was widespread in IgH locus in resting B cells until the initiation of CSR induction, when it was retained only in donor S region. This histone modification is shown to impair CSR if ablated in a B cell line (Stanlie et al., 2010) and can tether a protein complex which can further bind AID (Jeevan-Raj et al., 2011) . In CSR, only the 3' regulatory region (cis-acting region located far downstream of CH genes, 30kb in size and comprising four hypersensitive DNA sites) was identified as the one necessary and sufficient to trigger AID deamination independently of its role in transcription (Rouaud et al., 2013; Vincent-Fabert et al., 2010) .

II.6.2. Gene conversion

This form of diversification process of immunoglobulin genes takes place in chicken and rabbit B-cells to generate the pre-immune repertoire. Chicken variable region genes, contained within one single light chain and a heavy chain loci, are much less numerous compared to mouse and humans. In λ light chain locus, there is only one functional Vλ and Jλ, preceded by 25 non-functional V pseudogenes (Reynaud et al., 1987). These pseudogenes lack 5' or 3' end of the coding region in order to be operative. AID was discovered to be responsible for triggering gene conversion (Arakawa et al., 2002) by deaminating cytosine in coding V-exon . These deaminated cytosines are further processed to become a single or double strand break, allowing for unilateral transfer of genetic material between different parts of pseudogenes and corresponding parts in the single coding Vλ gene (Figure 26). These non-reciprocally exchanged regions can range from 10 to 100 base pairs. The pseudogene sequence remains intact after transfer and is available for novel rounds of gene conversion. Factors other than AID known to participate in this intra-chromosomal homologous

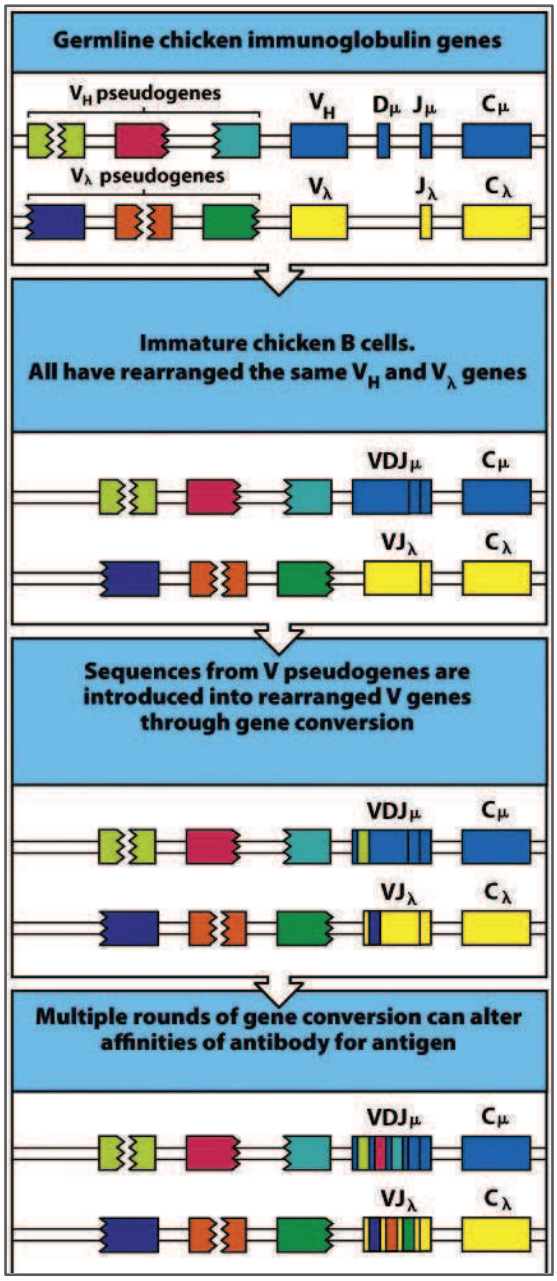


Figure 26. Overview of gene conversion in chicken B-cells. See text for detail.. Source: (Murphy et al., 2012)

recombination are Rad54, XRCC2/3 and Rad51B. In DT40 avian cells deficient for the last two factors, gene conversion is remodeled into classical somatic hypermutation on the V gene that is normally only minor in chicken B cells (Sale et al., 2001). Deletion of all pseudogenes similarly restricted AID-induced modifications to point mutations, thus providing a cellular model of somatic hypermutation that is easy to manipulate but whose results should be taken with caution, due to the prevalence of G:C mutation in the total pattern.

Chapter III : A place for cell cycle in somatic hypermutation

III.1. Hypothesis of cell cycle-mediated coordination of mutagenic repair pathways involved in SHM

The absolute mutation rate in WT mice is considered to be a balance between 1) active mutagenesis by mutagenic repair pathways, 2) error-free repair that still operates in hypermutating regions and 3) uracil "ignorance" in which none of the two uracil handling mechanisms takes place and where uracils are replicated giving rise to G:C transitions ("passive" mutagenesis). Change in the total load reflects a disequilibrium between active error-prone and error-free pathways, while further increase in absolute G:C transition frequency may be explained by more G:U mispairs escaping both mutagenesis and error-free repair, and lingering in DNA before they are fixed in replication. Altogether, it makes one error-free pathway, two mutagenic mechanisms, replication as a shortcut and five different outcomes (A:T transitions and transversions, G:C transitions and transversions, high fidelity U repair), all against a single substrate: a cytosine deaminated by AID into uracil. A question that one cannot help asking is how do all different repair pathways split the work on AID-generated uracil in somatic hypermutation?

What is the precise mechanism by which UNG generates mutations? Two alternatives are suggested, the most praised one includes UNG as the sole component of classical BER pathway engaged in somatic hypermutation. Other less accepted alternative involves some downstream players as in long-patch BER. Both alternatives of Ung processing may account for the way UNG uses to recruit pol η and generate residual A:T mutations in the non-physiological context of Msh2 absence.

First proposed model starts with UNG-mediated processing of the single-stranded uracil, towards which UNG shows slight catalytic preference (Kavli et al., 2005). Once the target is recognized, UNG glycosylates the N-bond and leaves an **abasic site**. This partially

processed deamination site can be found in multiple **single-stranded configurations**: 1) in the transcription bubble, just after AID deamination; 2) prior to gap filling in MMR repair action, after the strand containing a mismatch is degraded; 3) during replication, when a lag between replicative helicase and replicative machinery leaves template DNA in single-stranded form. This last proposition is accepted as the main pathway of G:C mutagenesis governed by UNG, where replication factory stalled at the abasic site seeks help of a mutagenic translesional polymerase (Rev1 or another), whose misincorporation gives rise to G:C mutations in general (Di Noia and Neuberger, 2007; Weill and Reynaud, 2008).

Second proposition is that uracil recognition by UNG happens while it is still in double strand DNA. Once recognized by UNG, a subsequent **long-patch repair** would be initiated with a remarkable difference – instead of PCNA and Pol β/δ, mono-ubiquitinylated PCNA and a translesional polymerase would be recruited (Krijger et al., 2009; Otterlei et al., 1999).

***A role for a partially processed deamination site in mutagenic MMR: cooperative pathways*.** Lasting debate about partially processed deamination site comes from the fact that we do not really know to which extent the error-free BER processing of uracil can be exploited in somatic hypermutation. Many studies hold their focus on two immediate outcomes of potential post-UNG processing: the abasic site and the subsequent incision of the strand, requiring an endonuclease or dRP lyase activity downstream of UNG and giving rise to single strand break. Other groups defend the possibility of joined efforts of UNG and MMR factors in generating Ig gene mutations by using one or both of these processing intermediates. Otherwise said, error-prone pathways would rather cooperate.

In this light, two reports placed the abasic site as a major event next to U:G detection in mismatch-induced mutagenesis. An abasic site in the middle of MMR processing in SHM would appear in the non-degraded strand prior to gap filling. One possibility is it would arise as a previous MutSα choice of sparing the U-containing strand when deliberating about U:G mismatch. UNG produces AP site from U left in the non-degraded strand (Frieder et al., 2009). The other possibility is that an abasic site can emerge prior to gap filling as a consequence of AID-mediated targeting of a single-stranded cytosine recognized in the strand spared from EXO-1-degradation. Authors of this proposition previously "inculpated" error-prone MMR pathway for generating a portion of G:C transversions, until then thought of as an exclusive right of UNG-recruited Rev1 polymerase (Krijger et al., 2009). In both cases, gap filling across the abasic site would be the trigger for PCNA ubiquitinylation and a translesional polymerase recruitment, e.g. Rev1 as first-choice inserter (yielding G:C transversions) and Pol η as a extender (generating A:T mutations).

Single strand break will remain the Gordian knot of the Ig gene diversification processes: if introduced in both strands not far away from each other, it can yield a double strand break so much sought for in class switch recombination and so much dreaded

elsewhere in the genome. If introduced in the only functional V gene in chicken, the gene conversion has half-way succeeded; and if introduced at a reasonable distance from the double-stranded U:G mismatch already recognized by the MutSα, this UNG-generated single strand break can spare MutLα from seeking the free 3' end (Schanz et al., 2009) or constantly "waking" the endonuclease activity in PMS2. While ones favor classical UNG partner APE-1 as single strand "nickase" in somatic hypermutation, others bet on MRE11/RAD50 complex and its susceptibility to assume AP-lyase function behind UNG and even stimulate mutagenic repair (Larson et al., 2005; Yabuki et al., 2005).

Mismatch repair impinges on UNG in SHM: competitive pathways. Our group had adopted a competitive model of interaction between mutagenic pathways based on data interpretation from *Ung*$^{-/-}$, *Msh2*$^{-/-}$ and *Msh2*$^{-/-}$ *Polh*$^{-/-}$ mice, (Delbos et al., 2007; Reynaud et al., 2009). Firstly, as long as the MMR sensor complex is present, A :T mutagenesis is independent of UNG according to the results from *Ung*$^{-/-}$ mice (Chapter II.4.4). Error-free MMR does not seem to be engaged during somatic hypermutation, otherwise mutation load in *Ung*$^{-/-}$ mice would be decreased.

On the contrary, MSH2 deficiency reduces overall mutation by 50%, a decrease that combines largely reduced A:T mutagenesis with relative increase of G:C transition, while G:C transversions remained the same. This counter-intuitive imbalance in relative frequency of G:C transitions is further explained by their clustering to AID hotspot sequences, as if AID-generated uracils remained protected from UNG-mediated error-free repair within hotspots until the replication fork arrival. Other non-protected uracils in background that would be normally handled by MutSα, are efficiently repaired by error-free base excision repair and therefore contribute to the general decrease in mutation frequency. Therefore, mutation pattern in *Msh2*$^{-/-}$ mice confirm there are both BER error-free mechanism and UNG-mediated region-specific mutagenesis operating in somatic hypermutation and they both employ UNG for the onset (Saribasak and Gearhart, 2012). The reason for different management of UNG-generated intermediates may be embedded in the sequence context of processed uracils (Pérez-Durán et al., 2012). As a result, it seems that the presence of MMR factors greatly downplay UNG activity in both error-prone and error-free handling of uracils.

Third point, residual A:T mutagenesis in *Msh2*$^{-/-}$ mice was totally absent in *Msh2*$^{-/-}$ *Polh*$^{-/-}$ mice, indicating that in the non-physiological context of Msh2 absence, Pol η kept on generating remaining A:T mutations, only this time being recruited by UNG. This "uracil overtaking" by one mutagenic pathway when the other is absent, together with MMR impinging on UNG activity, fostered the idea that these pathways act in competitive manner and that their outcomes are better balanced if one confines the other to temporally separated action frames.

MutSα requirement for a mismatched base embedded in double stranded DNA would make it an ideal candidate to act on uracils generated in G1 phase of dividing germinal center B-cells. UNG is known to act on both double-stranded and single-stranded uracil (Kavli et al., 2005); many cultured cell lines specifically upregulate UNG expression from the background level starting from G1/S phase transition (Chapter II.4.1); and translesional polymerases would perform their standard role of bypassing the non-instructional abasic site generated by UNG during replication. All these arguments converge towards the idea that UNG is better suited to induce mutagenesis in S-phase. With Pol η as a sole contributor to A:T mutagenesis downstream of MMR-mediated recruitment, all A:T mutations would arise in G1 phase, while G:C mutagenesis would be limited to the S phase.

If one could manage to limit the activity of a single SHM effector to the phase considered "suboptimal" for its performance according to our model, one might expect to detect an important disbalance in A:T vs. G:C mutagenesis as an outcome (Chapter III.3, Figure 28). This change in frequency on A:T or G:C basepairs depending on the restriction to G1 or S phase is a reliable readout that would allow us to confirm or infirm our hypothesis. Before going into technical rationale of how this phase-specific restriction of a SHM factor might be done, let us take a quick tour of the cell cycle in general in order to understand the layers of regulation in this cell cycle-regulated coordination of somatic hypermutation.

III.2. Cell cycle in brief.

From fertilized egg cell until death, cells undergo the same cyclic succession of processes that allows them to duplicate their mass, their genetic material, maintain genetic integrity of daughter cells and divide. Many cells are terminally differentiated in adulthood but others like epidermal cells, hair cells, cellular lining of the gut, blood cells and germ cells continue to perpetuate this cycle in highly controlled manner in order to renew the used up tissue. It is therefore logical for these cells to subject all house-keeping processes to the cycling rhythm of divisions in order to optimize the whole mechanism and carry it out with outmost precision. In the case of large pre-B-cells, proliferation can be exploited to rather downregulate some key development factors like RAG2, which is stable only in G1 and degraded in other phases, in order to control the V(D)J recombination (Chapter I.1.2 and I.2.2).

Cell cycle phases. Cell cycle is organized around three major events: genetic material duplication, equal share of genetic material between two future cell poles and splitting up all cellular material in two. First process is carried out during the period termed S phase for synthesis period; others two succeed one another during official cell division process, mitosis. In rapidly dividing embryo, cells literally jump from mitosis to DNA duplication, not even taking the time to double cell mass to maintain the size of daughter cells. However, two additional "gap" phases exist in adult dividing cells: gap phase preceding the DNA synthesis

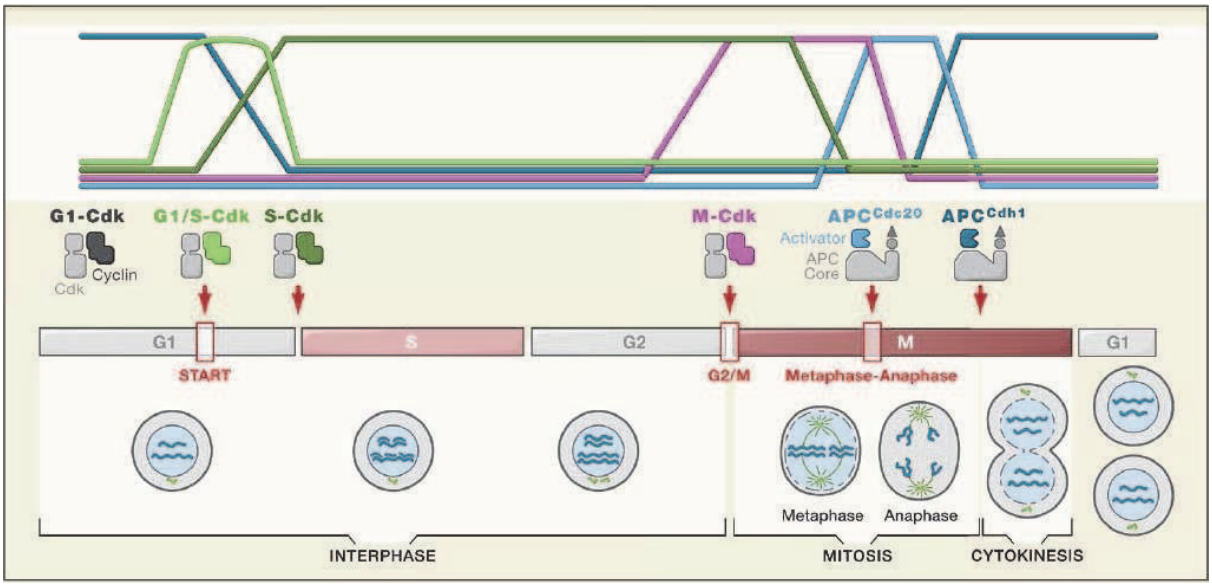


Figure 27. Overview of the eukaryotic cell cycle phases, major events and different cyclin-Cdk complex expression. Adapted from: (Morgan, 2008)

Cyclin-Cdk complex	Vertebrates		Role
	Cyclin	CDK	
G1-CDK	cyclin D (D1, D2, D3)	CDK4, CDK6	cell cycle entry if environment is favorable
G1/S CDK	cyclin E	CDK2	transition to S; DNA replication and centrosome duplication
S-CDK	cyclin A	CDK2, CDK1	DNA replication, centrosome duplication and mitosis
M-CDK	cyclin B	CDK1	mitosis

Table T4. Summary of cyclins and cyclin-dependent kinases (Cdk) that build different complexes and their role according to the timing of their activity. Source : (Alberts et al., 2002)

called G1 and the second gap phase between S phase and mitosis, named G2 phase. Those two additional phases are implemented in the "long" version of cell cycle allowing the cell to estimate the quality of the accomplished work and integrate environmental stimuli into further decision-making in the cell cycle.

Cell cycle checkpoints. Control points are identified between some phases of the cell cycle at which the cell will arrest further progress, try to solve problem and, if failing at that, trigger programmed cell death. The first checkpoint is at the G1/S transition, where the cell checks for free deoxynucleotide level and availability of enzymes, nutrients and growth factors before launching into DNA duplication. Second checkpoint is reached when entering G2/M transition, when the cell takes time to verify if the total DNA is duplicated only once and whether any other DNA damage persists, in which case post-replicative mechanisms will be engaged. The final control point is spindle assembly checkpoint, just before anaphasis (separation of sister chromatids and pulling each one toward opposite cell poles) when the cell makes sure that every sister chromatid is lined up and attached by a spindle fiber, which guarantees to each daughter cell the exact same quantity of genetic material as in the mother cell before duplication.

***Cell cycle effector proteins*.** The famous "cell cycle engine" that ensures the timing and succession of cell cycle phases is made up of a regulatory system of proteins: cyclin, cyclin-dependent kinases (Cdk) and Cdk-inhibitory proteins (CIP) specific for each cell cycle phase. Central actors in this system are cyclin-dependent kinases that can inactivate or activate their substrates by phosphorylation within a defined phase. However, Cdks are inactive and with no defined substrate specificity unless bound to cyclins, which can be considered as regulatory subunits of these kinase complexes. The quantity of CDK within a dividing cell is in most cases constant; it is the abundance of their cyclin partner that varies according to the cell cycle phase (Figure 27).

In absence of cyclins, Cdks are inhibited by Cdk-inhibitory protein. If a signaling cascade gives a green light for a transition into the next stage, the corresponding cyclin is upregulated just before the transition and persists throughout. CIPs are then removed and the cyclin can associate with the corresponding CDK within the active complex, promoting the movement through the cycle.

The three major classes of cyclin-Cdk complexes whose activity oscillates with phase succession are: G1/S cyclin-Cdk complex, S cyclin-Cdk and M cyclin-Cdk complex with their name indicating their position in the cycle. G1 complex controls the G1 phase length, leaving more time to the cell to integrate the information from the environment to ensure that all conditions are favorable to entail the risky business of DNA duplication and cellular material segregation. The list of cyclins and cyclin-dependent kinases that make up for these complexes in vertebrates is given in the Table T4.

M-CDKs are degraded if the signal is positive at metaphase-anaphase transition. With no Cdk-cyclin complexes left then to supervise the rest of mitosis, another indispensable component of cell cycle machinery comes into spotlight: anaphase-promoting complex or cyclosome (APC/C), an ubiquitin ligase that is activated from mid-mitosis (anaphase) until the end of G1. Once activated, it targets S and M-cyclins for ubiquitinylation and degradation, allowing for the completion of M phase and safely landing daughter cells in the new G1 phase.

Cell cycle regulators. The protein level of cyclins, Cdk kinase inhibitors and other regulatory factors in cycle are under permanent control of ubiquitin-dependent degradation. This layer of control is achieved through two classes of cell cycle ubiquitin ligases, the aforementioned APC/C and SCF complex (SKP1-CUL1-F box protein). Both complexes have similarly organized structures, containing common invariable components making up the core of the complex and a variable substrate-specific subunit (named activator for APC/C and one of F box proteins for SCF). Taking for example the Skp2/SCF ubiquitin ligase complex: SCF is permanently active throughout the whole cycle, in contrast to APC/C which needs activation to become functional. SCF can associate with one of its variable substrate-recognition component, S-phase kinase associated protein 2 (Skp2). Skp2/SCF has a detailed list of potential targets (among which G1, G1/S and S cyclins and negative cell cycle regulators like CKI), but these will remain untouched unless phosphorylated at a precise point of the cycle by one of Cdks. Therefore, while degradation of APC targets begins immediately after APC activation, ever-active SCF-F box complexes depend on timed phosphorylation of their substrates in order to target them to degradation and allow progression through S and G2 phases. The window of SCF targeted ubiquitination spans late G1 until early mitosis.

Dependent on phosphorylation or not, the portion of the protein sequence that is sufficient and necessary to bring about the recognition by SCF or APC/C complexes is named degron. An example would be Myc protein, which is stabilized in many human tumors due to the mutation of two residues in its Myc box 1 (MB1) degron. In normal conditions, phosphorylation of MB1 residues leads to recognition by FWB7/SCF complex and subsequent proteasomal degradation of Myc. The second example is cyclin B from M complex: its N-terminal domain harbors a destruction box (D-box) which is recognized by the APC/C complex, making it a degradation substrate starting from anaphase.

General scheme of eukaryotic cell-cycle control. Cyclin-CDK phosphorylating activity is almost absent in G1, while it is highly active in S, G2 and M phase. While in G1, appropriate mitogen signals and information on cell size allow further stabilization of G1 cyclins, which associate with G1 Cdks and help activate the gene expression of G1/S and S cyclins, as well as other factors participating in S-phase events. Being assembled the first, G1/S complex phosphorylates S complex-inhibitory CKIs, therefore providing the first substrates for SCF-mediated ubiquitination. In G1/S transition and throughout S phase, G1/S and S complexes collaborate on initiation of DNA replication and duplication of the spindle

poles. Whereas M-cyclins start to accumulate already in G2 phase, their cognate M-CDKs remain inactivated by another form of cell cycle "brakes" – inhibitory non-Cdk kinases, which phosphorylate inhibitory sites on M-CDKs and prevent them from associating with M-cyclins M-complexes. These inhibitory kinases are targeted by another F-box β-TRCP/SCF complex at the onset of M phase, while M-CDKs get rid of inhibitory phosphates, and, within the M-complex, promote the mitotic spindle assembly and other preparations for chromosome segregation. After one last check on whether sister chromatids are properly attached to the spindle, the M-complex triggers its own extermination by phosphorylating and therefore activating APC/cyclosome ubiquitin ligase. APC/C associates with different activators for targeting all M-cyclins and securin, a protein whose destruction separates sister chromatids and unleashes the spindle contraction. APC/C keeps ubiquitinating throughout G1 phase until its substrate-specific subunit is inactivated by G1 cyclin-Cdk complex at the beginning of the next cell cycle.

The activities of APC/C and SCF complexes seem to be perfectly delineated throughout the cycle. Even though there is no direct experimental evidence, it is thought that there is a tight interplay between the two, allowing them to work separate shifts during the cell cycle watch.

III.3. Timing of individual SHM players during cell cycle

Ig locus in general. E2A transcription factor is one of the first transcription factors to regulate pro-B cell genes (Chapter I.2.2). Its binding to E-box motifs, which are found in Ig gene loci, does not alter Ig gene transcription but is suspected to have a role in AID targeting (Chapter II.5). A cell model for visualizing position of rearranged Igλ locus in the nucleus of constitutively mutating DT40 cell line allowed the observation that transfected E2A does colocalize with Igλ genes. Furthermore, colocalization studies indicate that this association takes place only in G1 phase on the basis of correlating cell cycle phase with observed cell nuclear radius (Yabuki et al., 2009)..

AID. AID expression level during the cell cycle does not seem to change according to the cell cycle analysis of AID-EGFP knock-in BL2 cells (Aoufouchi et al., 2008b). Furthermore, this study showed that AID is longer-lived in cytoplasm compared to nuclear location, where it is ubiquitinylated and targeted for degradation. Endogenous AID in stimulated mouse splenic B cells is present in all phases, even though it is difficult to establish whether the expression level varies significantly (Schrader et al., 2007).

Concerning cell cycle phase-specific activity of AID, two studies examined the evolution of AID-deamination products, mutations and single-strand breaks. Immediately after upregulation of AID expression in the Burkitt's lymphoma cell line BL2 inducible for somatic hypermutation, mutations are introduced as a single strand event in rearranged V_H gene during G1 and G2/M phases. However, mutations become fixed in DNA only if

introduced in G1 phase. Frequency of mutations introduced in G2/M was reduced 24h after mutation triggering in G2/M sorted cells (Faili et al., 2002). Only G1-arrested activated splenic B-cells, fully proficient for AID, were able to generate double strand breaks in Sμ regions (Schrader et al., 2007).

When considering the possibility of cell cycle-dependent localization of AID in the cell, not much is reported on sub-cellular localization of endogenous AID in dividing germinal center B cells. AID localization was studied in Ramos B-cell line either by AID-GFP chimeric protein overexpression (Rada et al., 2002b) or by endogenous AID detection (Patenaude et al., 2009); in AID-YFP transfected DT40 chicken cell line (Ordinario et al., 2009) and in AID-EGFP-transfected BL2 cell line (Aoufouchi et al., 2008b). All studies indicated the predominant cytoplasmic localization of AID except for (Ordinario et al., 2009), whose authors tried to take a freeze frame shot of AID-YFP in DT40 nucleus by treating cells with nuclear-export drug leptomycin B and localizing YFP fluorescence after the treatment. While one can debate whether this is the most physiological way of catching a "snapshot" of AID localization, authors managed to show that during the short half life of AID in nucleus, it is at least more stable in G1 phase nuclei than compared to S phase nuclei.

***Pol η*.** Polη is usually recruited at the site of UV-induced lesion and replication stalling, therefore in S-phase replication foci. As commented in (Phillips and Sale, 2009), formation of foci does improve local concentration of molecules and enhances reaction kinetics, but the molecule localization in itself does not necessarily imply the ongoing activity in the foci. In the previously described model of rearranged Igλ locus visualization in DT40 nuclei, Pol η was found to colocalize with this locus preferentially in G1 phase (Ordinario et al., 2009). Furthermore, focal organization of Polη was shown to emerge upon UV irradiation of permanently G1 phase-arrested cell population, independently of PCNA or nucleotide excision factors recruitment. It was postulated that Pol η might be just setting up in G1 foci and getting ready in case of UV-bombing with clustered, dense lesions (Soria et al., 2009). In the same report, authors showed that pol η accessed thymidine dimers independently of foci formation in all phases of the cell cycle in the U2OS cell line.

Uracil DNA glycosylases. : Expression of different uracil-glycosylases during cell cycle was assessed in several studies. MBD4 (Hardeland et al., 2007) and SMUG (Nilsen et al., 2001) were shown to be constitutively expressed and active during the cell cycle. However, major differences in temporal pattern of expression were attributed to UNG2 and TDG (Hardeland et al., 2007): UNG levels increased above the background at G1/S boundary, reached the peak in early S and started to gradually decline from mid-S onwards, reaching the null level at the end of DNA replication. By contrast, TDG protein level has the exactly opposite behavior: it starts to accumulate in G2, reaches its peak in G1 and is kept under proteasomal degradation control from G1/S transition until G2. Interestingly, UNG was found to be phosphorylated at residues within a motif highly resembling the G1/S cyclin

phosphodegron (Hagen et al., 2008). This UNG phosphoform is further ubiquitinylated and subsequently degraded (Fischer et al., 2004; Hagen et al., 2008). With the additional proof that cyclin-dependent kinases can phosphorylate UNG *in vitro* (Hagen et al., 2008) and that UNG ubiquitinylation was inhibited by cyclin-dependent kinase inhibitors (Fischer et al., 2004), one cannot help recalling the interplay of SCF and APC/C regulated expression of oscillating cell cycle effectors. However, ubiquitin-ligase for UNG and TDG are still unknown.

After all is said on the subject of UNG cell cycle-dependent expression, one must notice that all data were collected in different cell lines using not so physiological means to bring the cells to cell cycle arrest. A key question for UNG-mediated processes in somatic hypermutation remains, what about UNG protein levels and activity regulation through the cell cycle in fast dividing germinal center B-cells?

***XRCC1*.** The scaffold protein XRCC1 involved in recruiting DNA ligase III in the last step in BER repair shows some cell cycle-specific interactions via two different domains, BRCT1 and BRCT2. It was shown that BRCT1 domain of XRCC1 interacts with p38 subunit of DNA polymerase α and thus recruits S phase-restricted single-strand break repair pathway (SSBR), involving PARP-1 and PARP-2 enzymes (Lévy et al., 2009). BRCT2 domain in XRCC1 binds DNA ligase III, and mutations in this domain were shown to abolish single strand break repair in G1 but not in S (Taylor et al., 2000).

Mismatch repair factors. Protein levels of MSH2, MSH6, MLH1 and PMS2 did not vary in different phases of the cell cycle in cell culture of non-transformed, non-immortalized human embryonic fibroblasts (Meyers et al., 1997; Szadkowski and Jiricny, 2002). However, mRNA levels of PMS2 remained very low and fluctuated between 50% drop and 50% increase in late G1 and S/G2/M, respectively. MSH6 mRNA increased 2.5-fold at late G1/early S phase, whereas the protein level did not follow this upregulation (Szadkowski and Jiricny, 2002). The rationale for this mRNA increase remains a mystery, since any free MSH6 protein is degraded unless protected within MSH2-MSH6 complex.

Association of monoUb-PCNA and Polη with the MSH2-MSH6 complex, shown to occur in response to oxidative (Zlatanou et al., 2011) and alkylating damage (Peña-Diaz et al., 2012) in cell lines, takes place independently of S-phase. In case of alkylating damage, this non-canonical MMR is confirmed to be activated in G1 phase.

Winners of the competition for cell cycle-restricted mutagenesis. In order to test our hypothesis on cell-cycle coordinated mutagenic repair pathways in SHM, we decided to pick two SHM factors whose activity is to be confined to a single phase of the cell cycle, preferentially the one that would allow us to discriminate its optimal timing in somatic hypermutation. Our search among aforementioned SHM factors for a candidate in cell cycle dependent-restriction concluded with AID and Pol η. Criteria guiding this choice included background on potential cell cycle-dependent gene expression or cell cycle-regulated activity

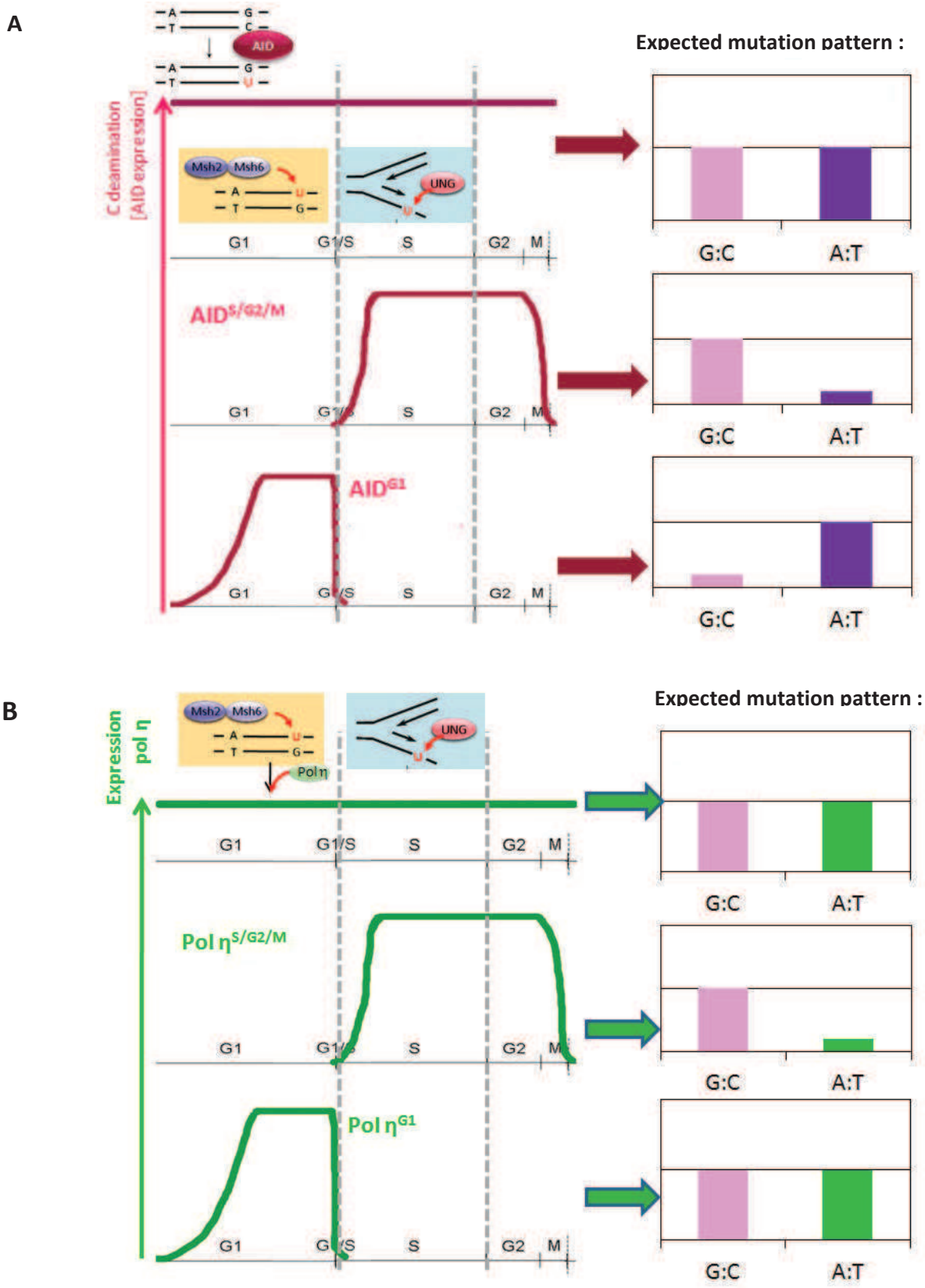


Figure 28. Different outcomes for G1-phase or S-phase restriction of two major SHM factors, AID (A) and Pol η (B). More details in text.

of candidates, as well as clear readouts of their cell-cycle restricted mutagenesis (Figure 28). According to our model, if AID activity is confined to G1 phase, A:T mutations would prevail over G:C; conversely, if AID is to deaminate only in S phase, G:C mutation would dominate. With Pol η restricted to S phase, there would be no MMR's preferred A:T mutator to generate the majority of A:T mutations in G1 phase and their frequency would be reduced.

The main hurdle in testing our cell cycle hypothesis was physically confining these proteins to a precise cell cycle phase. Several promoters were known to confer cell cycle-regulated expression, other peptides could bestow cell cycle-controlled localization to the fusion with fluorescent protein. However, we were looking for fluorescent marker technology that would allow us to rapidly detect the B cell population expressing the confined SHM effector, reliably verify its cell cycle status by measuring cell DNA content and sort out this population from total hypermutating B-cells for the purpose of establishing the mutation pattern. Luckily, an existing technology for tracking cycle in living cells allowed us to resolve all these issues with one bold stroke. The sesame word: Fucci.

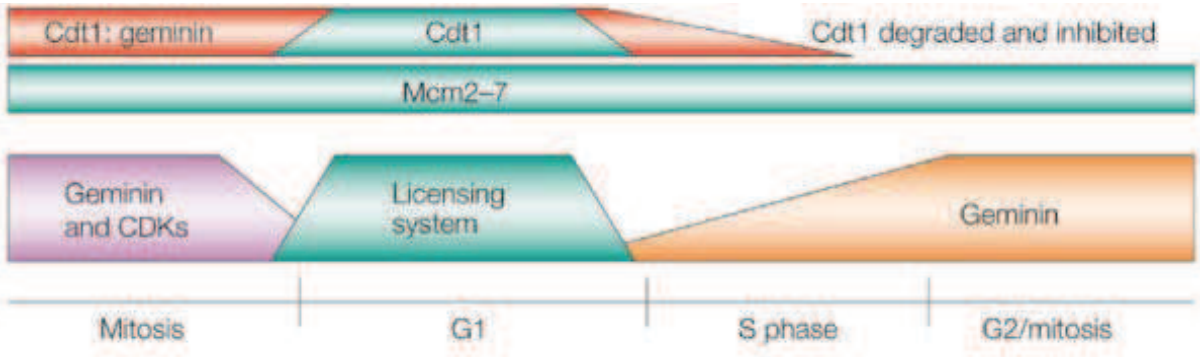
III.4. Cell cycle sensors can be used as protein activity restrictors

In summary, successive transitions between the different cell cycle phases is achieved through a complex network of cell cycle effectors and regulators, themselves regulated at several levels:

- differential expression throughout the cycle (e.g. G1/S-complex specific induction of S phase cyclin (CycA) synthesis)
- dynamic protein-protein interactions (Cyclin-Cdk interactions, Cdk-CKI interactions..);
- post-translational modifications leading to activation, inhibition (phosphorylations) or changes in stability (ubiquitinylation via APC/C and SCF);
- phase-dependent compartmentalization of cell cycle factors (accumulated M-phase cyclin B being translocated from cytoplasm to nucleus in order to initiate mitosis).

It was just the matter of time when the functional elements (enclosed in DNA or protein sequence) that govern these spatiotemporal characteristics of cell cycle actors would be identified and exploited for further characterization of dividing cells within a tissue or at the whole organism scale. With the advent of a variety of fluorescent proteins and improvement of cell visualization techniques (microscopy, cytometry), tracking cell cycle became a matter of imagination and color palette availability. Here we shall expose in more detail the cell cycle-tracking fluorescent reporter system we chose in an effort to manipulate the activity of somatic hypermutation key players in dividing germinal center B-cells.

Figure 29. The activity of components of the origin licensing system and their cell-cycle regulation. Whereas the licensing system is only active in G1, it is inhibited in S and G2 phase by Geminin (orange) and in mitosis by combination of Geminin and Cdk (violet). Mcm2-7: multimer complex of minichromosome maintenance helicase. Adapted from: (Blow and Dutta, 2005)



Fucci. Fluorescent ubiquitination-based cell cycle indicator is based on two ubiquitination-regulated proteins, Cdt1 and Geminin, both phase-specific and with mutually exclusive activity during the cycle. Degrons of these proteins were identified and C-terminally fused to fluorescent proteins, giving rise to G1-restricted fast-folding variant of mKusabira Orange (orange to red mKO2-hCDT1 degron) and monomeric azami green protein confined to S/G2/M period (mAG-Geminin degron) (Sakaue-Sawano et al., 2008).

Cdt1 is a replication origin licensing co-factor participating in the assembly of pre-replicative complex (pre-RC), a multi-subunit complex that gathers on many replication origins on chromosomes prior to DNA synthesis. Its role is to label the replication origin that will soon "fire" – initiate replication – and therefore let the cell know that this portion of DNA was already replicated once. The labeling proceeds by recruiting the minichromosome maintenance (Mcm) multimer helicase onto a replication origin that will further help opening the local DNA structure and ensure replication forks formation. Once other components of replicative machinery have gathered and started off, Mcm helicase will be pushed by the replication fork further away from the replication origin. Thereafter the replication origin will become delicensed for any further replication (Blow and Dutta, 2005).

Replication origin licensing takes place throughout the G1 phase. Once Mcm helicase is loaded, Cdt1 along with some other pre-RC factors is no longer needed – even unwanted because it threatens to re-license the same origin later on in the S phase. The main way of delicensing an origin is through downregulation of Cdt1 activity by means of both phase-specific Cdt1 degradation and inhibitory protein-protein interaction with Geminin.

At the G1/S transition, Cdt1 is targeted for phosphorylation by G1/S complex within the PARRRLRL motif in 90 aa–long peptide sequence. This phosphorylated Cdt1 degron is tracked by Skp2/SCF complex for proteasomal degradation from G1/S transition onwards (Nakayama and Nakayama, 2006). Cdt1 interaction with Geminin has two-fold purpose: it figures as an additional safety-lock, in case any Cdt1 escapes degradation; paradoxically, Geminin-bound Cdt1 is spared from degradation and accumulates in this inactive form during mitosis, when SCF runs out of specific substrates. On the other hand, Geminin is one of the substrates for APC/C complex, being ubiquitinated within its N-terminal 110 aa-long degron and thus targeted for degradation, starting from anaphase until late G1. This APC/C-dependent degradation sets free the accumulated Cdt1 in late mitosis, ready for some *de novo* origin licensing. It is thought that Geminin is re-stabilized in the late G1 after G1-complex-mediated inactivation of APC/C (Figure 29).

Fusing the whole sequence Cdt1 and Geminin to fluorescent proteins risked to perturb the cycle of the cells expressing them in two ways: first, by inevitably overexpressing them if introduced ectopically, and second by potentially behaving as dominant negative regulators of existing whole-length proteins in the cell. Since signal for SCF-mediated degradation is independent from Geminin-binding domain, the latter was removed along with other portions of Cdt1, leaving the minimal Cdt1 degron (amino-acids 30-120) still capable to translocate the

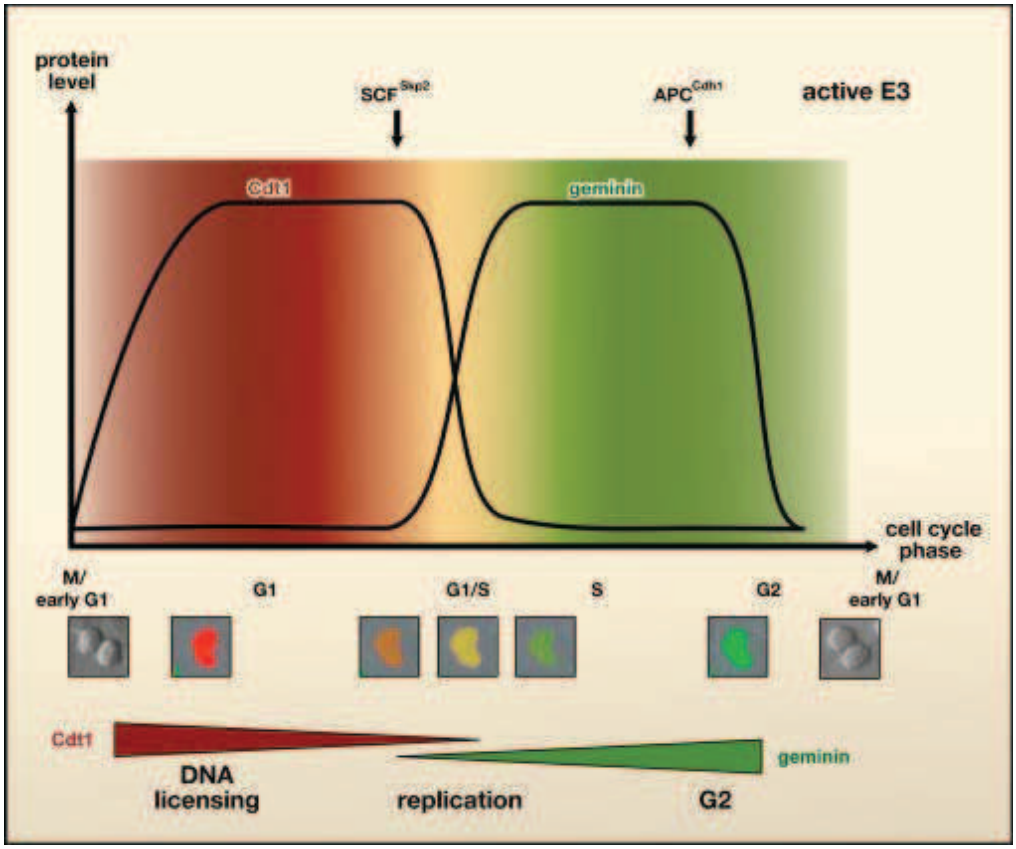


Figure 30. Illustration of the dual Fucci cell cycle sensor. For details, see text. Adapted from (Méchali and Lutzmann, 2008; Sakaue-Sawano et al., 2008).

fluorescent-protein fusion into the nucleus. A minimal Geminin destruction box was devised, leaving out the Cdt1 binding domain. Interestingly, degrons showed to be quite sensitive to the nature of fused fluorescent protein: Cdt1-30/120 maintained the expression confined to G1 phase only when fused to mKO2 or mCherry, while Geminin-1/110 was shown to be permissive to a larger palette of fluorochromes.

These colored cell cycle probes were simultaneously expressed in a variety of cell lines, in germinal center B-cells, in transgenic mice or genetically labeled tumor cells injected in mice, even in zebrafish and fruit fly cells. In all contexts they showed the similar temporal pattern of expression: green fluorescence associated with Geminin degron was limited to S/G2/M status, while red fluorescence associated with truncated CDT1 colored cells that entered G1. An overlap of both colors yielding yellow nuclei was observed at G1/S transition. Therefore, an overlap of remnant APC/C and awaking Skp2/SCF activity can be monitored at G1/S transition by tracking these yellow nuclei. On the contrary, mitotic cells did not express any color, consistent with anaphase-dependent degradation of Geminin and a lag in accumulation of hCDT1 since no Geminin-binding ensured readily available Cdt1 stocks (Figure 30). Both cell-probes were shown not to interfere with the normal cell cycle progression.

The elegant approach of these dual cell cycle probes was principally aimed at cell cycle monitoring and associating other cellular events to temporal order of phase transitions. However, at the time this project was initiated, no attempt was made to fuse these cell-cycle sensors to cellular proteins and turn them into cell cycle "restrictors" with clearly demarcated expression pattern in G1 or S/G2/M phase. With careful AID- and Pol η- fusion design and after many hours spent cloning, transducing cells and cell cycle analyzing, in the end there was nothing lovelier than a phase-synchronized colorful population emerging in the the ideal cellular context, *ex-vivo* cultured splenic B-cells of bone-marrow transplanted chimera mice.

Chapter IV : Mechanism and strand polarity of Pol η synthesis in somatic hypermutation

"The devil is in the details"

Somatic hypermutation is in many forms a skewed process, starting from the very fact of being one million times more intense than spontaneous mutagenesis. From that point of view, nothing guarantees the same randomness of spontaneous events to a mechanism deploying so many versatile actors to achieve such a high level of diversification. In this section we shall focus on strand biases in somatic hypermutation, and especially the one concerning mutagenesis at A:T basepairs and thus implying Pol η misinsertion activity.

Many names for only two strands. Before taking off, we prefer to reduce the ambiguity of the nomenclature to the minimum by establishing the firm terms we shall use throughout the story. We shall deal with only two DNA strands, but these two entities can assume so many disparate functions that a clarification is at need. According to the convention, top DNA strand is always read by the experimenter in 5'→3' direction and will represent the strand that harbors the exact information of RNA transcript sequence coded by a given gene. Therefore, top strand of a displayed DNA sequence is always a coding strand. Transcription complex will synthesize RNA by transcribing the bottom strand, namely the non-coding or transcribed strand.

During the replication, both DNA strands become replication templates but the synthesis on each one of them does not proceed in the same manner: one will be the template of the leading strand synthesis, while the other is the subject of discontinuous synthesis, giving rise to the lagging *de novo* strand. Top or bottom strand will assume their template roles depending on the closest replication origin and therefore, the direction of the replication fork – no convention fixes that as for the transcription.

Finally, all observed mutations, whatever the strand they are originally introduced in, are reported directly as read from the top strand. With these notions in mind, we can start discussing strand biases in somatic hypermutation.

IV.1. Strand bias of mutations at G:C basepairs

First occurrence for a strand bias in detailing somatic hypermutation mechanism came from the fact that not all bases mutate with the same frequency (Chapter II.1). A is more mutated than G and C, whereas T is the least mutated. In the final outcome of SHM, both G and C bases are equally mutated as read from the top strand. G:C mutations tend to

concentrate in so-called hotspot motifs (Rogozin and Kolchanov, 1992) of defined sequence WRCY or inverted RGYW sequence. If there was a strand preference for mutations on G:C basepairs, one motif sequence would be more mutated than the the inverted one within the same readout strand. This postulate, however, assumes the fact that there is equal distribution of both motifs on one strand, i.e. that the base distribution within the read DNA is randomized. This is not the case of V-D-J genes as CDR regions evolved so far as to skew codon usage in order to increase the hotspot concentration and therefore maximize the probability of targeting a precise amino acid replacement, most frequently the one that will come in direct contact with the antigen. In the absence of selective pressure on the sequence of J_H4 intronic region, its sequence is more randomized and the local targeting to these G:C hotspot sequences is still valid, but there is no strand discrimination when targeting them.

All SHM mutations stem from deamination events initiated at C bases by AID. It was thus intriguing to look at the potential strand bias introduced by AID during the initiation phase of somatic hypermutation. This pattern was available for analysis in *Ung*$^{-/-}$ *Msh2* $^{-/-}$ mice which allow for studying of the genuine AID footprint on G:C basepairs (Rada et al., 2004). Once again, only mutagenic outcome in these mice is G:C transition, reflecting high-fidelity DNA polymerases replicating over uracils. Counting unique, non-dynastically related mutations in *Ung*$^{-/-}$ *Msh2* $^{-/-}$ mice turned out to be a daunting task; mutation counts were either over- or underestimated, depending on the adopted analysis strategy. One pattern showed equal C and G mutation frequency, indicating no strand bias; however, other way of analyzing unveiled a slight "penchant" of AID for coding strand (C:G=60:40)(Reynaud et al., 2009).

In either case, AID was confirmed to target cytosines in both strands in spite of long persisting dilemma on how AID gains access to transcribed strand, occupied by RNA Pol II complex or hybridized with RNA within a transcription bubble (Chapter II.5). In S-regions, this is made possible by the action of RNA exosome (Basu et al., 2011).

IV.2. A:T mutagenesis strand polarity

Let us for a moment neglect any hot-spot targeting or base distribution and imagine that A:T mutations were introduced at random with equal frequency on both DNA strands. The frequency at which an A →T mutation occurs at the top strand would be equal to the frequency of T→A mutation in the same strand, the latter mutation reflecting the A →T mutation that occurred in the bottom strand. As mutation databases in different hypermutating substrates were growing larger, a different picture of A:T mutagenesis was becoming increasingly evident in rearranged VDJ genes and flanking region: mutations at A bases are in twofold excess over mutations at T bases, both read off the coding strand. This bias could be explained neither by an imbalance in base distribution of the examined region, nor by the specific sequence context of mutating As or Ts (Milstein et al., 1998), suggesting 1) that there was a targeting mechanism that was directing the A:T mutator preferentially to

the one strand or another, and 2) that the A:T mutator was itself targeting As and Ts non equally.

In vitro Pol η mutation pattern. Even before it was straightforwardly demonstrated that Pol η is the sole contributor of A:T mutagenesis *in vivo*, *in vitro* pattern of Pol η synthesis was already qualifying it as a perfect A:T mutator. When copying undamaged single-strand DNA templates *in vitro*, recombinant human Pol η was incorporating opposite to all four template bases. However, a preference was clearly demarcated (Matsuda et al., 2000): Pol η preferentially inserted dGTP when facing T and dATP when facing A in the template strand. T:dGTP misincorporation was four to five times more frequent than A:dATP misinsertion. This pattern was reliably reproduced with mouse Pol η and with different substrates as Pol η templates (gapped LacZ gene, gapped VκOx gene) (Matsuda et al., 2001; Pavlov et al., 2002).

Therefore, when Pol η has no choice but to copy a single offered template strand outside of any context of being coding/non-coding, its most frequent mutations in the synthesized strand are A→G and to 4-fold lesser extent, T→A.

Rogozin et al. identified a hotspot motif for A:T mutagenesis, W<u>A</u> while analyzing many mutation databases from hypermutating regions in humans, mice, chicken and rabbit (Rogozin et al., 2001) A W<u>A</u> hotspot subset, T<u>A</u> was claimed to accumulate frequencies of A mutations (A→G, A→T and A→C) that were the same in Ig genes and in artificial gapped LacZ substrate copied by Pol η. This argument was used to prove good correspondence between the pattern of Pol η mutagenic synthesis *in vitro* and *in vivo* mutation data on A:T.

What preferentially directs Pol η to the coding strand?

With this intrinsically skewed mutagenic signature, it seemed that Pol η could explain the A:T strand bias as a whole if only correctly targeted to the top strand, in order to elongate it while using the bottom strand as a substrate. Therefore, in the quest to unveil the main genesis factor of A:T mutation bias, two points must be taken into account:

- linking A:T mutations to a strand-specific deamination site (as uracil or abasic site), that would serve as the general trigger for Pol η recruitment. One team suggested that the transcriptional machinery, which drags AID behind it, can eventually deposit mismatch complex and Pol η preferentially onto uracils that were firstly introduced in the non-transcribed top strand (Mayorov et al., 2005). Others reasoned around creating an abasic site in the gapped bottom strand after MMR intervention. This AP site can work as a trigger for polymerase switch while gap filling opposite to it; one of favored polymerases in the switch would be Pol η (Frieder et al., 2009). Even though mutation analysis were in accordance with these hypothesis, none of them provided the satisfactory explanation on how uracils in one strand are more powerful in recruiting Pol η than onto the other strand.
- strand directionality of SHM factors upstream of Pol η. By a mechanism that is yet to be detailed, it is generally accepted that Pol η is recruited in somatic hypermutation as

a part of non-conventional mismatch repair (Chapter II.4.3). Consequently, if one or more mismatch repair factors would have a strand preference, it would explain which is the strand that Pol η copies when filling the gap. This was suggested in (Unniraman and Schatz, 2007) study, whose authors tried to standardize *in vivo* environment for Pol η synthesis by creating transgenic mouse lines containing Igκ transgene with a specific substrate for dissecting Pol η targeting process. This substrate contained stretches of adenine and thymine bases in control version, dubbed TgA; in the test version, TgC, the access of Pol η to this substrate is governed after deamination of a single cytosine inserted in the middle of AT stretches in the top strand. A TgG substrate that contained the complementary AT sequence with C in the bottom strand was created as well. Authors first monitored the mutagenesis on test and control substrates, and concluded that TgG and control substrates accumulated very low number of mutations, thought to be engendered by recruiting Pol η to the neighboring deaminated Cs in the flanking sequence. TgC had significantly more mutations that TgG and TgA, indicating that C in the top strand targets better the A:T mutator machinery. This bias was alleviated when transgenic mice were bred with Msh2-deficient mice, allowing the authors to conclude that MMR machinery is responsible for this targeting bias.

Although one must acknowledge the authors for a considerable effort in devising a model to control the impact of uracil mutagenic potential, many objections can be brought up to counteract the conclusions of this study. The first one concerns the uncontrollable behavior of Ig transgenes that are largely under impact of the integration site; as a consequence, AT-containing substrates did not accumulate more than one mutation per sequence. This resulted in too small mutation database, compromising any statistically significant conclusions about the top strand C or about the impact of Msh2 absence. Many previous studies restrained from commenting A:T bias in Msh2- deficient mice due to the low number of A:T mutations, which did not prevent the authors from concluding about mutational asymmetry disruption due to Msh2 absence on the basis of 2-3 mutations left in the AT substrate. Lastly, authors argued in the light of total mutation numbers that transgenes accumulated depending on the C position, but did not comment the mutational pattern that did not correspond to Pol η basic pattern *in vivo* at all (the most frequently detected mutation being T→C in TgC which accumulated an exploitable number of mutations). When total Igκ transgene mutations in transgenic mice with Msh2-proficient background were analyzed, they did not even present the usual A:T bias!

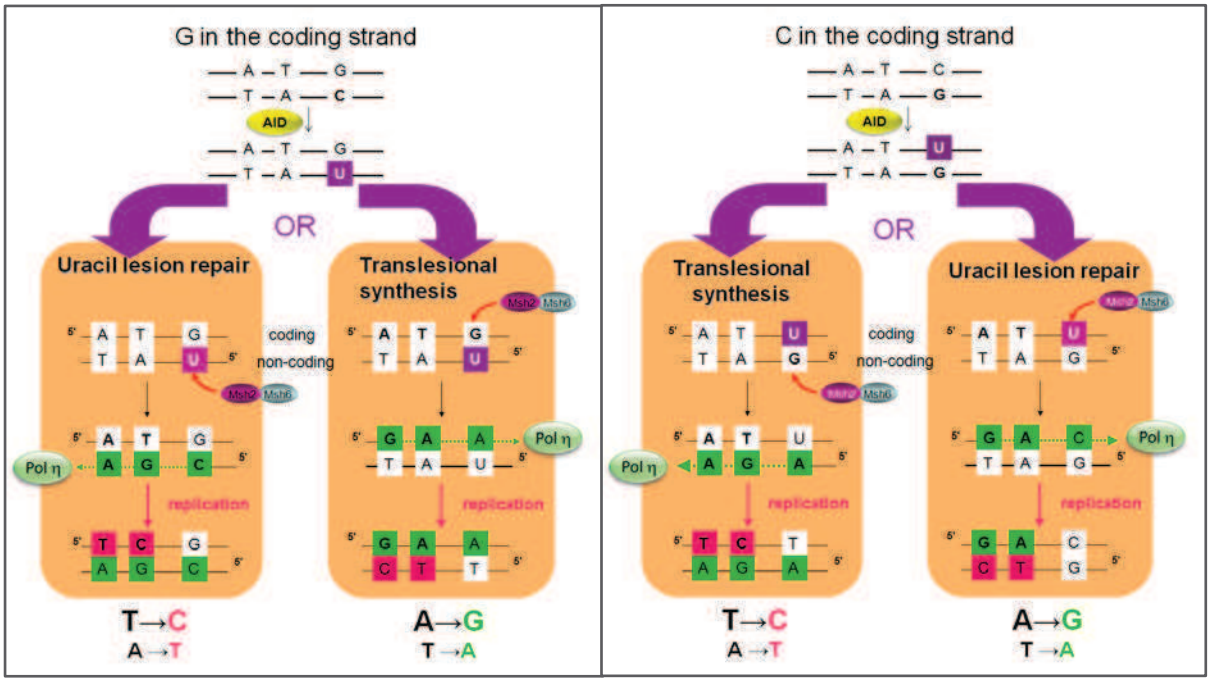


Figure 31. Different mutagenic outcomes in A/T-rich substrates if A :T mutator is targeted to repair the U-containing strand or to perform translesion synthesis on the opposite strand. Left, the case when U is in the non-coding strand; right, when U is in the coding strand. See text for details.

IV.3. What is the mechanism of Pol η synthesis in SHM and can it account for the A:T bias?

We decided to improve the somatic hypermutation substrate from (Unniraman and Schatz, 2007) study and test on it the hypothesis of a dual mechanism of Pol η activity in SHM depending on the orientation of the recruiting deamination site. Recruited by the non-canonical MMR to the degraded strand, Pol η mutates As and Ts while filling the gap in the error-prone manner. However, non-canonical mismatch repair has no constraints to obey the same strand signals as conventional mismatch repair in post-replicative repair. In other words, MMR in SHM can attack either U-containing strand or G-containing strand of the mismatched base pair, depending on the introduction of the nick (Chapter II.4.3 and III.1).

In this context, polymerase η can perform gap filling across the U lesion-containing strand, and therefore carries out its physiological translesion bypass activity. If the U-containing strand is attacked, Pol η performs the reparation of the uracil by elongating the 3' end of the degraded U-containing strand. We wanted to verify which mechanism is favored in the context of somatic hypermutation. The precise mechanism adopted by Pol η can generate different mutational outcomes at A:T basepairs if top or bottom strand U is attacked (Figure 31). In case of MMR attacking the top U-containing strand, Pol η would be recruited to the top strand to perform lesion repair and generate A:T mutations, among which A→G and T→A would be the first and second most abundant mutations, respectively. Conversely, a single bottom strand U would yield a patch containing T→C and minor A→T mutations in the top strand. Mutation readouts would provide the reverse picture for top strand U or bottom strand U if translesional synthesis mechanism was preferred in somatic hypermutation. A controlled hypermutating substrate would allow us to clearly discern between the two proposed mechanisms, if it contained mostly A and T bases and was integrated into normally hypermutating mouse Ig locus. This substrate has to be long enough to minimize the influence of flanking cytosines and has to contain its own deamination sites at central positions. Inserting this substrate into mouse Ig locus would not only allow us to obtain a more precise picture about the length of the Pol η-initiated patch, but if inserted in both orientations, hypermutating substrates would enable us to follow the fate of both top and bottom-strand Us . We chose to introduce not more than three of them into the substrate to mimic the physiological situation in which AID can deaminate on the same strand in a processive manner (Pham et al., 2003), enhancing somatic hypermutation and even providing us an insight into in vivo interactions between mutagenic pathways. We hoped thus to have defined a better hypermutating substrate, which will acquire mutations at A:T basepairs in physiological context of immunoglobulin locus.

Part Two : Research article

AID expression restricted to the G1 phase of the cell cycle recapitulates the mutation spectrum of immunoglobulin genes

Marija Zivojnovic[1], Olivier Albagli-Curiel[2], Annie De Smet[1], François Huetz[3], Jean-Claude Weill[1], Claude-Agnès Reynaud[1] and Sébastien Storck[1]

[1]INSERM U783, Université Paris Descartes, Faculté de Médecine-Site Broussais, Paris, (France)
[2]INSERM U845, Université Paris Descartes, Faculté de Médecine-Site Cochin, Paris (France)
[3]CNRS URA 1961, Institut Pasteur, Paris (France)

Short communication

Summary

Somatic hypermutation in immunoglobulin genes is the result of the processing of uracils generated by AID (activation-induced cytidine deaminase) achieved by two different repair pathways, uracil glycosylase for mutations at G/C bases and mismatch repair, together with DNA polymerase eta, for mutations at A/T bases. The apparent independent behavior of these two pathways revealed by gene inactivation of the different partners involved have led to the formulation of hypermutation models in which each pathway would act at a distinct stage of the cell cycle. We performed transduction of hematopoietic stem cells from AID-deficient animals with cell cycle-restricted variants of AID, and used them to restore Rag2-deficient hosts and analyze mutations in B cells after immunization. We report here that restriction of AID expression to the G1 phase of the cell cycle is able to generate an equal proportion of A/T and G/C mutations at the Ig loci, thus demonstrating that uracils generated in G1 are substrates for both uracil and mismatch repair pathways.

Introduction

Affinity maturation of the immune response is the outcome of a mutagenic process targeted at the immunoglobulin (Ig) heavy and light chain loci (Berek and Milstein, 1987). This process is triggered by AID, a cytidine deaminase that is preferentially induced in fast cycling B cells activated in germinal centers, where the T-dependent immune response takes place (Muramatsu et al., 2000). Mutations that accumulate in Ig V genes become selected upon appropriate T helper signals, during a complex process of cyclic trafficking between the dark zone of germinal centers, where mutagenesis is considered to take place, and the light zone where B cell proliferation slows down and allows for isotype switch and selection of variants with higher affinity (Victora et al., 2010). Isotype switch, the second molecular event targeted by AID at the Ig locus, is a non-homologous recombination process involving the concerted formation and processing of two double-strand breaks, resulting in the exchange of the heavy chain constant regions and subsequent acquisition of new functional capacities linked with each heavy chain isotype.

Unrepaired uracils can generate C to T and G to A transitions if replicated over. However, two different repair pathways appear to contribute to enlarge the mutation spectrum at G/C bases and to spread mutations at distance from the initial deamination site. Uracil glycosylase generates abasic sites that will be substrates for translesional DNA polymerases, generating transversions (and also transitions) at G/C bases. A/T mutations are in contrast introduced during a patch of error-prone synthesis, mobilizing the Msh2/Msh6 mismatch recognition complex together with ExoI, ubiquitinated PCNA, and DNA polymerase eta as the sole contributor of A/T mutagenesis (Di Noia and Neuberger, 2007).

In B cells from Ung-deficient animals, no transversions at G/C bases are observed, but mutations at A/T bases appear relatively unaffected (Rada et al., 2002). In contrast, the impact of Msh2/Msh6 deficiency appears more complex: mutagenesis at A/T bases is markedly reduced, but the overall mutation frequency is also decreased, transition mutations at G/C bases are more prominent and the mutation pattern shows increased focusing at hotspot positions (Delbos et al., 2007). This dissociation lead to the suggestion that both G/C and A/T mutagenesis pathways may work independently, even though it is still unclear to what extent these two pathways may cooperate or compete for uracil recognition (Krijger et al., 2009; Reynaud et al., 2009).

Models based on a distribution of tasks according to the cell cycle have been proposed, in which mismatch repair would preferentially handle uracils generated in the G1 phase, preventing Ung from accessing the lesion, and Ung, the expression of which is increased when DNA replicates, would generate abasic sites in the S/G2 phase (Liu and Schatz, 2009; Weill and Reynaud, 2008).

Experimental data that contradict this model have been reported, in which expression of the Ung-specific UGI inhibitor was restricted to distinct phases of the cell cycle (Sharbeen et al., 2012). The authors used an elegant system of transduction by retroviral expression vectors of naive B cells harboring a transgenic anti-lysozyme B-cell receptor, and performed analysis of somatic mutations *in vivo*, after transfer of the transduced cells in a normal host and immunization with lysozyme. The authors concluded that Ung preferentially excises uracils in G1. Consequently, another cell cycle based model, with opposite distribution of repair activities was recently proposed (Li et al., 2013).

We report here a cell cycle restriction study of AID expression, obtained by retroviral transduction of *Aicda*$^{-/-}$ hematopoietic stem cells and restoration of Rag-2-deficient mice, and show that AID expression restricted to the G1 phase appears to recapitulate mutagenesis produced by the two uracil glycosylase and mismatch repair pathways.

Results and Discussion.

AID expression vectors with G1- or S/G2/M-restriction

The Fucci system (Fluorescent, Ubiquitination-based Cell Cycle Indicators) described by Sakaue-Sawano et al. (2008) consists in fluorescent proteins coupled to peptides that target their degradation in specific phases of the cell cycle (degrons). These degrons comprise a 90 amino acid peptide of the hCDT1 replication licensing factor for restricting expression in G1 (30-120 N-terminal amino acids), and a 110 amino acid peptide from Geminin for S/G2/M restriction.

We chose to fuse mKO2-hCDT1 and mCherry-Geminin reporters at the C-terminus of AID. We also generated mutant versions in which the sites that regulate their

proteasomal degradation were abolished, to control for the functionality of the AID fusion proteins. This was performed by mutating the cyclin box in hCDT1 (RRL to AAA within the ^{65}PARRRLRL^{72} cyclin/Cdk binding consensus site of the protein) that targets its phosphorylation-directed degradation (Liu et al., 2004); and, for Geminin, by deletion of the 9 amino acid destruction box (RRTLKVIQP) that begins at amino acid 33, a mutation that abolishes APC/C-mediated ubiquitination at the M/G1 transition (McGarry and Kirschner, 1998).

Four different retroviral vectors were generated, based on the pMIG plasmid in which expression of constructs is driven by an MSCV (murine stem cell virus) promoter and which includes an EGFP reporter under IRES control. Transduction of NTZ cells (a 3T3 derived cell line harboring a tetracycline-inducible EGFP reporter with a TAG stop codon) (Yoshikawa et al., 2002) was performed as a preliminary test of the functionality of the four constructs, and revealed in all cases an effective mutational targeting of the EGFP reporter (not shown). These four retroviral constructs were used to transduce bone marrow hematopoietic stem cells (HSC) from AID-deficient mice. Three days after transduction, the relative fraction of cells in each phase of the cell cycle was evaluated, based on EGFP, mKO2 and mCherry expression (Fig. 1). For the mutant vectors, transduced cells co-expressed EGFP and either mKO2 or mCherry, with double-expressing cells on the diagonal of the FACS plot. For vectors with cell cycle restriction, two populations can be distinguished: an EGFP-only subset, and a double EGFP/mKO2 (or mCherry)-positive population. Analysis of the cell cycle indicated that approximately 75% of mKO2-positive cells were in G1, while G1 represented 50% of the total population. The non-G1 cells, in agreement with the initial description of this restriction system, were mainly in the early S phase in which expression of hCDT1 and Geminin fused proteins have been shown to partially overlap (Sakaue-Sawano et al., 2008). For mCherry-expressing cells, 85% were in S/G2/M, thus indicating for both vectors that appropriate restriction to the corresponding phase of the cell cycle was achieved in HSC. Conversely, restriction was abolished by the mutations introduced in the hCDT1 and Geminin peptides.

It should be mentioned that these degrons have been shown surprisingly (and without straightforward explanation) to be sensitive to the fluorescent protein to which they are fused (Sakaue-Sawano et al., 2008). The hCDT1 peptide is notably unable to trigger degradation of EGFP or of RFP, but is functional with mKO2, while the Geminin degron has been previously validated with both mKO2 and mAG fusions (Sakaue-Sawano et al., 2008). The use of mCherry reported here constitute a new addition to the Fucci system, with a cell cycle regulation similar to the one described for other reporters.

Mice restored with these different constructs are thereafter referred to as AID^{G1} for AID-mKO2-hCDT1 (and AID$^{G1\text{-}Mut}$ for the one including the mutated degron), and AID$^{S/G2/M}$ for AID-mCherry-geminin (AID$^{S/G2/M\text{-}Mut}$ for the mutated version).

Restoration of Rag-2-deficient mice with transduced Aicda$^{-/-}$ hematopoietic stem cells.

Transduction efficiency ranged from 20-40% of HSC, and peripheral reconstitution, estimated 2-3 months later in the spleen showed an overall 3-fold reduction in EGFP^{+} cells, thus mirroring the initial transduction rate with a moderate counter-selection of the transduced cells, and similar for all constructs tested. EGFP^{+} cells represented 5-15% of total spleen cells, with a similar distribution within the B cell and T cell compartment (Fig. 2).

The cell cycle profile was further assessed on splenic B cells blasts "ex vivo", i.e. obtained from restored animals, after LPS plus IL4 stimulation. This assay was performed to check the cell cycle restriction in the specific conditions of expression of mature B cells differentiated from transduced precursors (Fig. 3). Although the overall mKO2 or mCherry fluorescence intensity was lower than in HSC analyzed early after transduction, positive cells showed correct cell cycle restriction, a restriction completely abolished by the mutation in the degron peptide.

G1-restricted AID generates the complete spectrum of mutations at the Ig locus.

Two months after HSC restoration, mice were immunized by intra-peritoneal SRBC injection, PNA^{high}EGFP^{+} cells were sorted two weeks later and mutations in the J_H4 intronic region were determined for the four types of AID-expressing vectors (Fig. 4).

Mutation frequency induced by the two control AID-fusion constructs was low, with a high proportion of unmutated sequences. The frequency was 0.24 mutations per hundred base pairs for AID$^{G1\text{-}Mut}$ and 0.074 for AID$^{S/G2/M\text{-}Mut}$. In both cases the G/C over A/T targeting ratio was normal, as expected for control AID vectors. Both controls showed a bias in transitions at G/C bases, a profile that appears to correlate with the AID-fusion proteins, but for which we do not have a straightforward explanation.

Surprisingly, restriction of AID expression in G1 yielded similarly a balanced proportion of mutations at G/C and A/T bases, indicating that uracils generated in G1 are substrates for both Ung and Msh2 repair pathways. Importantly, this balanced ratio remains the same, whatever the stringency of the criteria used to eliminate clonal relationships, an important issue when analyzing small germinal center B-cell samples (see discussion in Material and Methods). The mutation frequency observed was three-fold lower than in the non-restricted control, a difference that likely corresponds to the lower accumulation of a protein that undergoes recurrent cell cycle degradation.

In contrast, a negligible number of mutations over background was generated upon restriction of AID to the S/G2/M phase. The relative difference in mutation frequency with the unrestricted control was at least 10-fold, a ratio clearly underestimated since one-third of the mutations observed are probably contributed by the PCR reaction. The much stronger difference between the mutated and unmutated AID$^{S/G2/M}$ compared with the 3-fold difference between the mutated and unmutated AID^{G1} suggests that AID-dependent

mutagenesis is specifically inefficient in S/G2/M. Moreover, since both degrons lead to the presence of the protein to which they are fused in early S and to the absence of both just after mitosis (Sakaue-Sawano et al., 2008), this would restrict the window of AID efficiency to the G1/late G1 phase.

Normal G/C over A/T ratio in patients with compromised DNA polymerase epsilon activity.

Hypermutation is considered to take place in fast cycling centroblasts. Their division time was estimated around 6-7 hrs by the group of MacLennan, in 1988 (Zhang et al., 1988), a figure that was reduced to 10-12 hrs in recent 2-photon microscopy approaches (Allen et al., 2007). It has been proposed that the rate of cell division may impact the A/T mutation frequency, fast cycling Ramos Burkitt's lymphoma B cells showing higher mutagenesis at A/T bases (Kano et al., 2011). We therefore attempted to study human genetic deficiencies in which alterations of the cell cycle have been described. We selected the recently described case of a family with DNA polymerase epsilon hypomorphic mutations, for which fibroblasts and EBV-transformed cell lines showed a protracted G1 phase with a marked delay in S-phase entry (Pachlopnik Schmid et al., 2012) (in collaboration with A. Fischer and G. de Saint-Basile). Whereas the number of $CD27^+$ memory B cells was strongly diminished, the overall Ig gene mutation frequency was only slightly reduced in these cells. No A/T over G/C imbalance was observed, thus suggesting that an increased length of the G1 phase does not favor *in vivo* a specific dissociation of the two arms of mutagenesis (Fig. 5).

AID-induced mutagenesis in G1: restriction or repair?

This work reports two complementary observations. First, AID expression in the G1 phase is able to generate the complete spectrum of mutations observed in normal conditions at the Ig locus, suggesting that both uracil glycosylase and mismatch repair may operate within a similar time frame. Whereas this result contradicts our proposition of differential recruitment of these two pathways according to the cell cycle, it is in good agreement with the similar study performed recently by the group of Jolly, in which cell cycle restriction of Ugi, a specific inhibitor of Ung, was able to interfere with uracil excision only when expressed in G1 (Sharbeen et al., 2012). This observation provides also support to several studies proposing a cooperation between the two repair pathways, notably the involvement of Ung in excising uracil in the single-strand DNA patch exposed during Msh2-driven error-prone synthesis (Krijger et al., 2009).

The second observation is the almost complete absence of mutation induced by AID expressed in S/G2/M. A trivial explanation could be a low efficiency of our S/G2/M-restricted AID, which would fail to induce mutations above the threshold of PCR background. We think this interpretation unlikely, due to the large difference in mutation induction between the S/G2/M construct and its mutated control. Another explanation could be that

AID generates mutations in S or G2, but that these events are correctly repaired, either during the same phase of the cell cycle, or in the next G1, in which error-free repair by Ung is also proposed to take place (Sharbeen et al., 2012). We have in fact previously reported in the BL2 Burkitt's lymphoma cell line, that AID-dependent mutations introduced in G2 are repaired before the next cell division, whereas those occurring in G1 become fixed after replication (Faili et al., 2002). A third possibility could be a decreased efficiency of AID activity in S/G2/M, due to a specific subcellular localization (e.g. increased cytoplasmic retention)(Patenaude et al., 2009) or to an increased nuclear turnover, a proposition documented in another cellular model, the avian DT40 B cell line (Ordinario et al., 2009). However the capacity of Burkitt's lymphomas or DT40 to precisely mimic the physiological mutation process is clearly questionable.

Error-prone and error-free repair appear to coexist during hypermutation at the Ig locus, but their impact differ among AID off-targets, a few genes like *Bcl-6* accumulating mutations while others - more numerous - being targeted and repaired (Liu et al., 2008). Whether cell-cycle regulated gene specific features (e.g. early *vs*. late replication) may explain the differential outcome of AID-mediated off-target deaminations is an issue that deserves further studies.

This report describes for the first time the transduction of HSC to assay the impact of mutant proteins on the somatic hypermutation process. This approach can be similarly applied to shRNA-mediated down-regulation of gene expression, allowing for multiple inactivations through transduction of HSC from gene-targeted mouse lines. Even though the limiting factor will be whether a biologically significant level of expression can be achieved, this approach represents a manageable alternative to classical knock-out or knock-in strategies to study a mechanism for which no satisfactory *in vitro* model exists.

Acknowledgements

We are indebted to the animal care provided by the Laboratoire d'Experimentation Animale et de Transgénèse, Necker and the Metchnikoff mouse facility, Pasteur. We also thank Alain Fischer and Geneviève de Saint-Basile for their collaboration, and the patients with the Pol epsilon hypomorphic mutation for their participation to this study. We thank Jérôme Mégret for cell sorting, Laïla Kossir for her contribution to the construction of phase-restricted vectors, and Maxime Fayon for his contribution to the Ig gene sequencing of the Pol epsilon-deficient patient.

INSERM U783 is supported by the Ligue contre le Cancer (équipe labellisée) and the Fondation Princesse Grace. Marija Zivojnovic was supported by a 3-year allocation from Ecole Normale Supérieure, and from the Ligue Nationale contre le Cancer for her fourth year of PhD.

Materials and Methods

Mouse strains.

C57BL/6 mice were purchased from Charles River. *Aicda* $^{-/-}$ mice, kindly provided by Prof. T. Honjo, had been back-crossed against C57BL/6 background for 9-10 generations at the time of present experiments and maintained in the Laboratoire d'Expérimentation Animale et Transgénèse (Faculté de Médecine Necker/Broussais). RAG2-deficient mice originating from SOPF-conditioned mouse facility of Transgénèse and Archivage d'Animaux Modèles, CNRS, Orléans, were maintained in the Metchnikoff Immunology mouse facility of Institut Pasteur, Paris, during restoration experiments. All experiments and procedures were approved by the National Ethical Committee of the French Ministry of Agriculture.

Construction of retroviral vectors.

Mouse *Aicda* cDNA was amplified using PCR primers including a $(Gly_4\text{-}Ser)_3$ flexible linker sequence in 3'. AID was inserted by EcoRI/HindIII directional cloning in the expression vector pFucci-G1 Orange (MBL International) in 3' fusion with mKO2-hCDT1. Mutant mKO2-hCDT1 was generated by PCR mutagenesis by mutating RRL to AAA in the PARRRLRL motif responsible for its degradation.

The fragments of mCherry and Geminin_1-110 aa degron were obtained from commercially available cloning vectors: pmCherry (Clontech) and pFucci –S/G2/M Green (MBL International), respectively. *Aicda* cDNA was fused to mCherry and Geminin degron in successive steps of PCR assembly reactions using Phusion polymerase (New England Biolabs) and individual fragments as templates with the following cycling conditions: first 10 cycles without assembly primers (10 sec 98°C, 30 sec 55°C , 60-90 sec 72°C), then 20-25 cycles with same conditions except for annealing temperature (60°C) in presence of assembly primers. Mutant mCherry-geminin was made by PCR site-directed mutagenesis as an in-frame deletion of the 9 aa destruction box sequence (RRTLKVIQP) that begins at amino acid 33 in hGeminin_1-110 aa degron.

pMIG and pMIGR1 (MSCV: IRES-GFP, Addgene entry:9044) retroviral vectors were used for cloning of finalized phase-restricted fusions for AID-mKO2-hCDT1 (AID^{G1}) / AID-mKO2-hCDT1 mutant ($AID^{G1\text{-}Mut}$) and AID-mCherry-geminin ($AID^{S/G2/M}$) / AID-mCherry-geminin mutant ($AID^{S/G2/M\text{-}Mut}$), respectively. All constructs were verified by DNA sequencing.

Production and titration of ecotropic retroviral particles.

HEK 293T Neo^R were transiently transfected by X-tremeGene 9 DNA Transfection Reagent (Roche) with the genomic retroviral and structural vectors coding for ecotropic envelope. 48h post-transduction, 293T medium (DMEM F-12 10% fetal calf serum (FCS, (Hyclone)) was harvested, filtered through hydrophylic HighFlow 0.45 μm Minisart syringe filters (Sartorius) and concentrated by centrifugation 30 min at 3900 g at 4°C in Amicon Ultra

15 100K centrifugal filter devices (Millipore Miltenyi). Retroviral particles were titrated by transducing NIH 3T3 fibroblasts plated at $1x10^4$ cells/cm^2 for 24h with pure and serially diluted retroviral supernatants. GFP expression was assessed 72h post-transduction by flow cytometry, yielding titers between between $3x10^7$ and 10^8 pfu/mL.

Hematopoietic stem cell transduction and adoptive transfer experiments.

Eight to twelve week-old *Aicda*$^{-/-}$ mice were euthanized by CO_2. Total bone marrow cells were harvested by flushing femurs and tibias, and labeled according to the instructions of the Mouse Lineage Cell Depletion kit (Miltenyi Biotech). Lineage negative cells (Lin^-) were purified by negative selection after loading labeled bone marrow onto LS columns (Miltenyi Biotech) within strong magnetic field produced by QuadriMACS holder (Miltenyi Biotech). Purified Lin^- cells were checked for purity and cultured for 48h in DMEM with 15% heat-inactivated fetal calf serum (Fetal Clone I, HyClone) supplemented with 1 mM sodium pyruvate and murine cytokines (Peprotech): IL-3 10 ng/mL, IL-6 10 ng/mL, SCF 100 ng/mL, Flt3 ligand 100 ng/mL, and human thrombopoietin at 60 ng/mL. One day before transduction, 6-well plates (BD Falcon) were coated overnight at 4°C with 50 µg/mL retronectin solution in PBS (TaKaRa) at final concentration of 10 µg/cm^2. Virus-free concentrated 293T medium (mock) or retroviral particles were first allowed to bind to retronectin-coated plates in presence of 4 µg/mL polybren solution for 4 hours at 37°C, followed by a gentle wash. Lin^- cells were swiftly dispatched into mock- or virus-bound wells at $1x10^6$ cells/mL density and allowed to transduce during 4 hrs at 37°C. Cells were detached from retronectin-coated wells by Cell Dissociation Buffer (Invitrogen Life Technologies), washed twice and resuspended in cold PBS prior to injection. A small sample of transduced Lin^- cells was kept in culture for additional 72 h after which the percentage of GFP^+ cells and cell cycle-mediated restriction of mCherry/$mKO2^+$ cells were estimated by flow cytometry.

Eight to ten week-old *Rag2*$^{-/-}$ mice were irradiated 8 hours prior to injection with a lethal dose of 4.5 Gy total body irradiation (RS-2000, RadSource Technologies) and injected intravenously with $2x10^6$ transduced or mock-transduced Lin^- cells in 200 µL PBS per mouse on the same day of transduction. A group of 2-3 irradiated recipient mice was injected with mock-transduced Lin^- cells; for each vector coding a phase-restricted fusion protein (4 groups in total), 3-5 siblings were injected and maintained in the same cage. Both sexes were used, but the sex of donor and recipient mice matched in every Lin^- adoptive transfer experiment. Mice were monitored on a daily basis throughout the first week post-injection, and every 2-3 days from the second week on. Four weeks post-injection, peripheral blood blood was analyzed for the presence of GFP^+ cells, B cells, $CD4^+$ and $CD8^+$ T cells.

Cell cycle analysis.

72h post transduction, Lin^- cells were harvested, washed and resuspended in pure DMEM (at $1x10^6$ cell/mL). Diluted Hoechst 33342 (Invitrogen LifeTechnologies) (1mg/mL in water) was added to a final concentration of 2µg/mL and incubated at 37°C for 30 min

away from light. Cell cycle of $mKO2^+$ or $mCherry^+$ cells was assessed after excluding dead cells with SytoxRed (Invitrogen LifeTechnologies) at a 150-200 events/sec rate by exciting Hoechst 33342 with 355 nm-UV laser on BD LSR Fortessa cytometer (Beckton Dickinson). All cell cycle profiles were analyzed with FlowJo software (TriStar, San Carlos, CA, USA). In vitro activated B cells were processed similarly.

Other flow cytometry analysis.

Mouse peripheral blood cells isolated 4 weeks post-injection and splenocytes following the primary immunization were analyzed for the expression of activated B and T-cell markers by the following panel of rat anti-mouse mAb antibodies: APC eFluor780-conjugated B220 (eBiosciences), PerCP-eFluor710-conjugated IgM (eBiosciences), phycoerithrin (PE)-conjugated CD4 (BD Pharmingen), APC-conjugated CD8 (BD Pharmingen), Alexa647-conjugated GL7 activation marker (eBiosciences), biotinylated PNA (Vector) visualized with streptavidin coupled to PE-Cy7 (eBiosciences). Dead cells were excluded by forward and side scatter intensity and by subsequent gating after labeling with Sytox Blue Dead Cell Stain (Invitrogen Life Technologies). All cytometry profiles were acquired at BD LSR Fortessa or BD FACSCantoII cytometers and data were analysed with FACS Diva software (Beckton Dickinson)

Immunization, sorting strategy, genomic DNA extraction and sequence analysis.

Mice were immunized by injecting intra-peritoneally $5x10^9$ sheep red blood cells (SRBC, Eurobio) per mouse, washed three times and resuspended in 500 mL PBS. Mice were sacrificed 14 days after the primary immunization, or 5 days after the secondary immunization performed at a one-month distance. Spleen cells were prepared and labeled for sorting on FACS Aria I (Beckton Dickinson). Sorted $B220^+$ PNA^{hi} $GL7^+$ cells were heat-lysed at 95°C for 10 minutes and proteins were removed after incubation 30 min at 56°C with proteinase K (Roche). Three to four aliquots of this preparation (up to 1000 cells/aliquot) were used as a genomic DNA template for amplifying the rearranged VDJ locus and adjacent J_H4 intron in a 40-cycle direct PCR reaction. Five forward primers at a ratio 6:3:1:1:1 amplifying V_H1, V_H5, V_H3, V_H7 and V_H9 gene families and a reverse primer complementary to the downstream sequence of J_H4 gene (5' GAGTACTTGAAAACCCTCTCAC) were amplified by Phusion polymerase (New England Biolabs) as described previously (Delbos et al., 2005). The gel-purified PCR product was ligated into pCR2.1 cloning vector (Original TA Cloning Kit, Invitrogen Life Technologies). Plasmid DNA was extracted by Millipore 96-well Miniprep system (Millipore Miltenyi). Sequencing was carried out in ABI Prism 3130xl Genetic Analyzer. The diversity of VDJ junctions and mutations within 450bp of J_H4 intron were analyzed with CodonCode Aligner software (Li-COR).

Three different types of mutation analysis were performed. In the first one, allowing the estimate of mutation frequencies (represented in Figure 3), identical sequences were removed. In the second one, mutations were discarded when two mutations or more were

shared by different sequences, even if they do not displayed clonal relationship (to account for possible PCR hybrids). The third one, the most stringent one, takes into account only one type of mutation at each position: this analysis, used by the SHMTool program (Maccarthy et al., 2009), is not strictly scientifically neutral, since it underestimates recurrent base substitutions, notably at hotspot positions.

In vitro B-cell activation.

$5x10^5$ $B220^+EGFP^+GL7^-PNA^{low}$ B cells were collected in parallel from restored SRBC-immunized animals. Cells were resuspended at $1x10^6$ cells/mL in RPMI complete medium (fetal calf serum 15%, 0.1 mM β-mercaptoethanol, 20 mM HEPES, 1 mM sodium pyruvate and non-essential amino-acids) supplemented with *Escherichia coli* lipopolysaccharide (serotype O55:B5 , Sigma Aldrich) and mouse recombinant IL-4 (eBioscience) to a final concentration of 20 µg/mL and 25 ng/mL, respectively. After 4 days of activation, cells were harvested for cell cycle analysis.

Patient studies

A homozygous hypomorphic mutation in the DNA polymerase epsilon gene has been reported within a large French family (Pachlopnik Schmid et al., 2012). 10 ml of blood were obtained from a 13.5 years old affected member after informed consent, and somatic mutations were determined in the J_H4-J_H5 intronic sequence flanking rearranged V_H genes of sorted IgD^-CD27^+ and IgD^+CD27^+ B cells, as described previously (Weller et al., 2008). PCR conditions were as follows: 40 cycles of 98°C for 10 seconds, 60°C for 20 seconds, and 72°C for 20 seconds. The resulting J_H4-J_H5 PCR products were gel purified and cloned with Zero Blunt PCR Cloning kit (Invitrogen). Sequences were run in an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). Mutation frequencies were calculated by comparing the sequences obtained with germline intronic J_H4-J_H5 sequences over 284 bp, starting at the 3' border of the J_H4 gene segment.

References

Allen, C.D., T. Okada, H.L. Tang, and J.G. Cyster. 2007. Imaging of germinal center selection events during affinity maturation. *Science* 315:528-531.

Berek, C., and C. Milstein. 1987. Mutation drift and repertoire shift in the maturation of the immune response. *Immunol Rev* 96:23-41.

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 Noia, J.M., and M.S. Neuberger. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* 76:1-22.

Faili, A., S. Aoufouchi, Q. Gueranger, C. Zober, A. Leon, B. Bertocci, J.C. Weill, and C.A. Reynaud. 2002. AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat Immunol* 3:815-821.

Faili, A., A. Stary, F. Delbos, S. Weller, S. Aoufouchi, A. Sarasin, J.C. Weill, and C.A. Reynaud. 2009. A backup role of DNA polymerase kappa in Ig gene hypermutation only takes place in the complete absence of DNA polymerase eta. *J Immunol* 182:6353-6359.

Kano, C., R. Ouchida, T. Kokubo, and J.Y. Wang. 2011. Rapid cell division contributes to efficient induction of A/T mutations during Ig gene hypermutation. *Mol Immunol* 48:1993-1999.

Krijger, P.H., P. Langerak, P.C. van den Berk, and H. Jacobs. 2009. Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. *J Exp Med* 206:2603-2611.

Li, S., Y. Zhao, and J.Y. Wang. 2013. Analysis of Ig gene hypermutation in Ung(-/-)Polh(-/-) mice suggests that UNG and A:T mutagenesis pathway target different U:G lesions. *Mol Immunol* 53:214-217.

Liu, E., X. Li, F. Yan, Q. Zhao, and X. Wu. 2004. Cyclin-dependent kinases phosphorylate human Cdt1 and induce its degradation. *J Biol Chem* 279:17283-17288.

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

Maccarthy, T., S. Roa, M.D. Scharff, and A. Bergman. 2009. SHMTool: a webserver for comparative analysis of somatic hypermutation datasets. *DNA Repair (Amst)* 8:137-141.

McGarry, T.J., and M.W. Kirschner. 1998. Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* 93:1043-1053.

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.

Ordinario, E.C., M. Yabuki, R.P. Larson, and N. Maizels. 2009. Temporal regulation of Ig gene diversification revealed by single-cell imaging. *J Immunol* 183:4545-4553.

Pachlopnik Schmid, J., R. Lemoine, N. Nehme, V. Cormier-Daire, P. Revy, F. Debeurme, M. Debre, P. Nitschke, C. Bole-Feysot, L. Legeai-Mallet, A. Lim, J.P. de Villartay, C. Picard, A. Durandy, A. Fischer, and G. de Saint Basile. 2012. Polymerase epsilon1 mutation in a human syndrome with facial dysmorphism, immunodeficiency, livedo, and short stature ("FILS syndrome"). *J Exp Med* 209:2323-2330.

Patenaude, A.M., A. Orthwein, Y. Hu, V.A. Campo, B. Kavli, A. Buschiazzo, and J.M. Di Noia. 2009. Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol* 16:517-527.

Rada, C., G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, and M.S. Neuberger. 2002. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* 12:1748-1755.

Reynaud, C.A., F. Delbos, A. Faili, Q. Gueranger, S. Aoufouchi, and J.C. Weill. 2009. Competitive repair pathways in immunoglobulin gene hypermutation. *Philos Trans R Soc Lond B Biol Sci* 364:613-619.

Sakaue-Sawano, A., H. Kurokawa, T. Morimura, A. Hanyu, H. Hama, H. Osawa, S. Kashiwagi, K. Fukami, T. Miyata, H. Miyoshi, T. Imamura, M. Ogawa, H. Masai, and A. Miyawaki. 2008. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* 132:487-498.

Sharbeen, G., C.W. 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. *J Exp Med* 209:965-974.

Victora, 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.

Weill, J.C., and C.A. Reynaud. 2008. DNA polymerases in adaptive immunity. *Nat Rev Immunol* 8:302-312.

Weller, S., M. Mamani-Matsuda, C. Picard, C. Cordier, D. Lecoeuche, F. Gauthier, J.C. Weill, and C.A. Reynaud. 2008. Somatic diversification in the absence of antigen-driven responses is the hallmark of the IgM+ IgD+ CD27+ B cell repertoire in infants. *J Exp Med* 205:1331-1342.

Yoshikawa, K., I.M. Okazaki, T. Eto, K. Kinoshita, M. Muramatsu, H. Nagaoka, and T. Honjo. 2002. AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* 296:2033-2036.

Zhang, J., I.C. MacLennan, Y.J. Liu, and P.J. Lane. 1988. Is rapid proliferation in B centroblasts linked to somatic mutation in memory B cell clones? *Immunol Lett* 18:297-299.

Figure legends.

Figure 1. Cell cycle restricted expression of AID in transduced hematopoietic stem cells.

Retroviral constructs expressing the G1-restricted form of AID fused to the mKO2 fluorescent reporter (AID-mKO2-G1) and its cell cycle control mutant (represented by a red spark), the S/G2/M-restricted from (AID-Cherry-S/G2/M) and its cell cycle control mutant (same symbol), with an IRES EGFP reporter were used to transduced Lin^- hematopoietic stem cells. Cell cycle analysis was performed three days after transduction, using Hoechst staining to estimate of the cell DNA content. The cell cycle profile of $EGFP^+$ cells (bottom row) and of mKO2-EGFP or mCherry-EGFP double positive cells (upper row) is indicated, with distribution within the different phases of the cell cycle estimated by the FlowJo software.

Figure 2. Restoration of peripheral B and T cells by HSC transduced with AID-expressing vectors.

(A) Percentage of transduced $EGFP^+$ cells among *Aicda*$^{-/-}$ Lin^- hematopoietic stem cells used to restore the Rag2-deficent mice analyzed in (B) and (C), and analyzed three days after transduction. (B) Relative B cell and $CD4^+$ and $CD8^+$ T cell distribution among total spleen cells, two to three months after restoration with *Aicda*$^{-/-}$ transduced HSC. Analysis of wt mice and of animals restored in parallel with mock-transduced *Aicda*$^{-/-}$ HSC is shown for comparison. (C) Proportion of $EGFP^+$ cells, transduced with the different AID expression vectors, among total spleen cells (top), $CD4^+$ and $CD8^+$ T cells and B cells (bottom row). AID^{G1} and $AID^{S/G2/M}$: AID restricted in G1 or S/G2/M by fusion with the hCDT1 or Geminin degron, respectively; $AID^{G1\text{-}MUT}$ and $AID^{S/G2/M\text{-}MUT}$: mutated version of the same AID fusion proteins, in which amino acids responsible for cell cycle control have been mutated or deleted.

Figure 3. Cell cycle restricted expression of AID in LPS blasts from reconstituted animals.

Same as Figure 1, but performed on naive splenic B cells from restored animals, three days after LPS + IL4 stimulation. Activations were performed on sorted $EGFP^+B220^+GL7^-PNA^{low}$ B cells, except for mutated AID-mKO2-G1, for which activation was performed on total naive B cells, and gated on $EGFP^+$ cells for cell cycle analysis.

Figure 4. Mutation profile of germinal center B cells from mice reconstituted with cell cycle-restricted AID-expressing HSC. Base substitution pattern (A) and mutation frequency (B) in J_H4 intronic sequences from Rag2-deficient animals restored with *Aicda*$^{-/-}$ HSC transduced with the different AID-expressing vectors. The mutation profile of mice restored with $AID^{S/G2/M}$ is not shown, since it concerns a total of 8 mutations. Representation of mutation

frequencies for the different animals studied (C) and of the relative G/C and A/T ratio among total mutations obtained for the four different AID constructs (D).

Figure 5. Ig gene mutation profile of patients with a homozygous hypomorphic mutation in DNA polymerase epsilon. Base substitution pattern (A) and mutation frequency (B) in J_H4-J_H5 intronic sequences from peripheral memory B cells (IgD^+CD27^+ and IgD^-CD27^+) of a patient presenting an hypomorphic mutation in the DNA epsilon gene. The control sample values are taken from Faili et al. (Faili et al., 2009), with the range of variation among the 6 individuals studied.

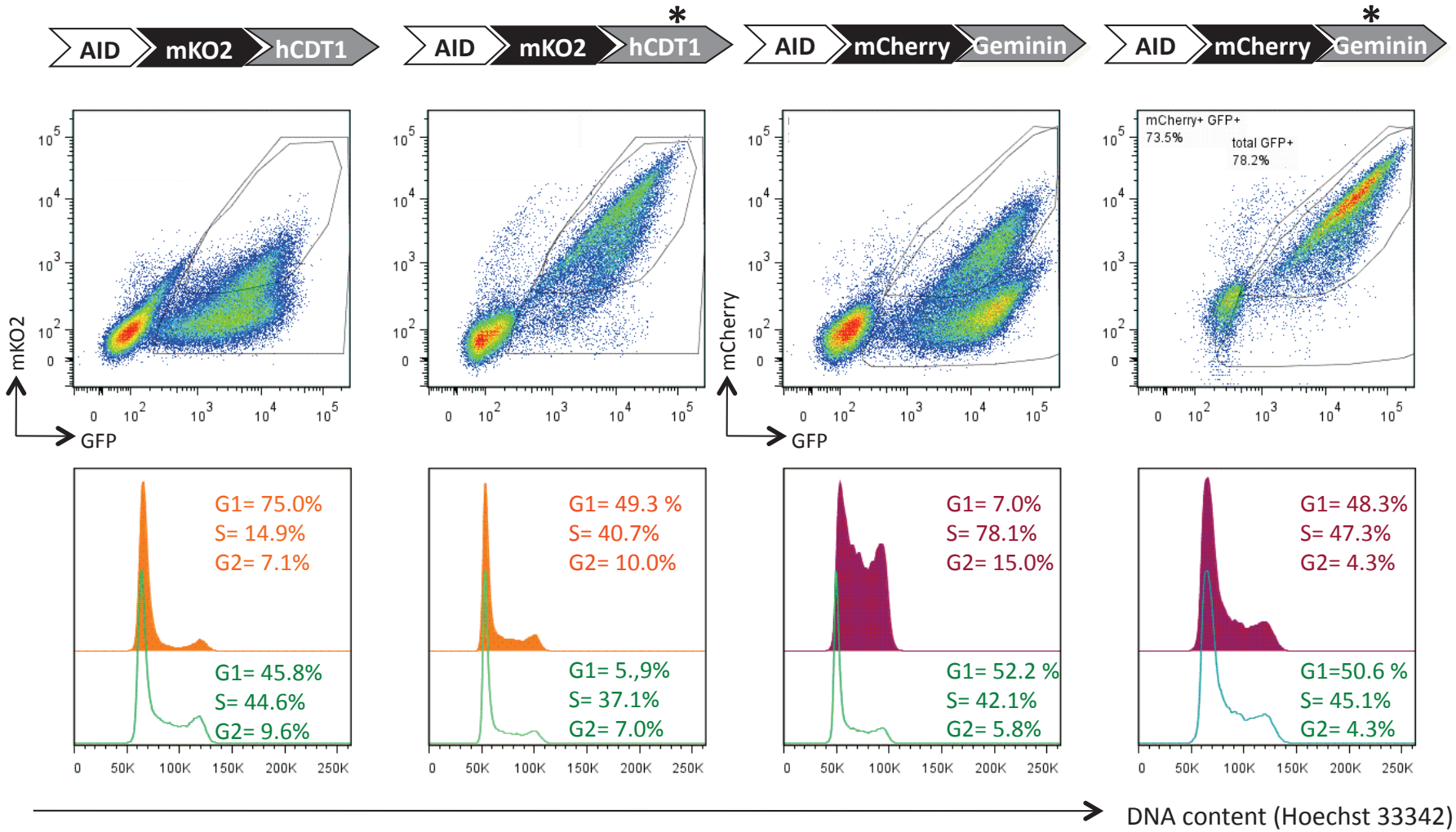

AID
mKO2
hCDT1
*
AID
mKO2
hCDT1
AID
mCherry
Geminin
*
AID
mCherry
Geminin
mKO2
GFP
mCherry
GFP
mCherry+ GFP+
73.5%
total GFP+
78.2%
G1= 75.0%
S= 14.9%
G2= 7.1%
G1= 45.8%
S= 44.6%
G2= 9.6%
G1= 49.3 %
S= 40.7%
G2= 10.0%
G1= 5.,9%
S= 37.1%
G2= 7.0%
G1= 7.0%
S= 78.1%
G2= 15.0%
G1= 52.2 %
S= 42.1%
G2= 5.8%
G1= 48.3%
S= 47.3%
G2= 4.3%
G1=50.6 %
S= 45.1%
G2= 4.3%
0
50K
100K
150K
200K
250K
DNA content (Hoechst 33342)

Figure 1

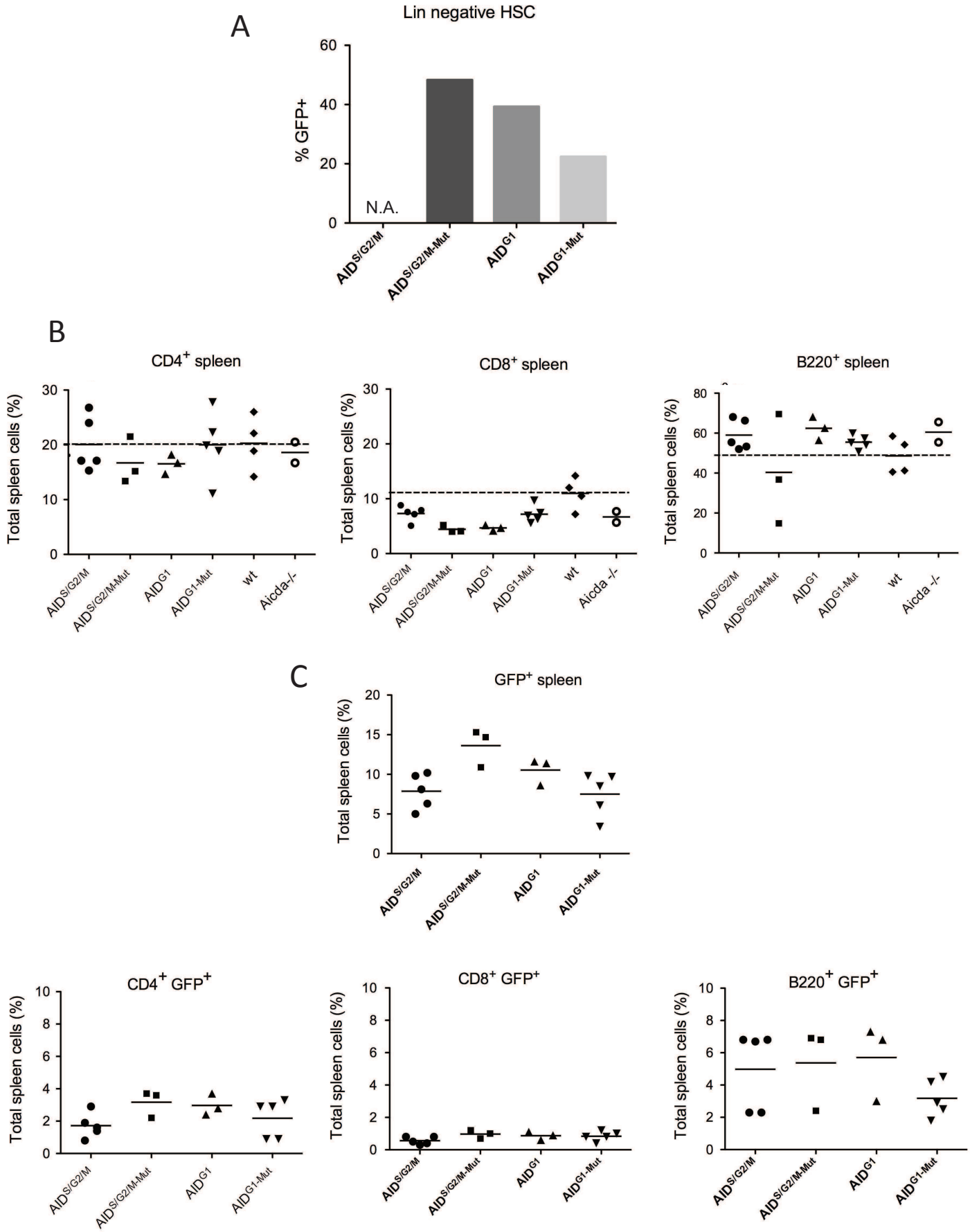

A
Lin negative HSC
% GFP+
60
40
20
0
N.A.
AIDS/G2/M
AIDS/G2/M-Mut
AIDG1
AIDG1-Mut
B
CD4+ spleen
CD8+ spleen
B220+ spleen
Total spleen cells (%)
30
20
10
0
80
60
40
AIDS/G2/M
AIDS/G2/M-Mut
AIDG1
AIDG1-Mut
wt
Aicda-/-
C
GFP+ spleen
15
5
CD4+ GFP+
CD8+ GFP+
B220+ GFP+
8
6
4
2

Figure 2

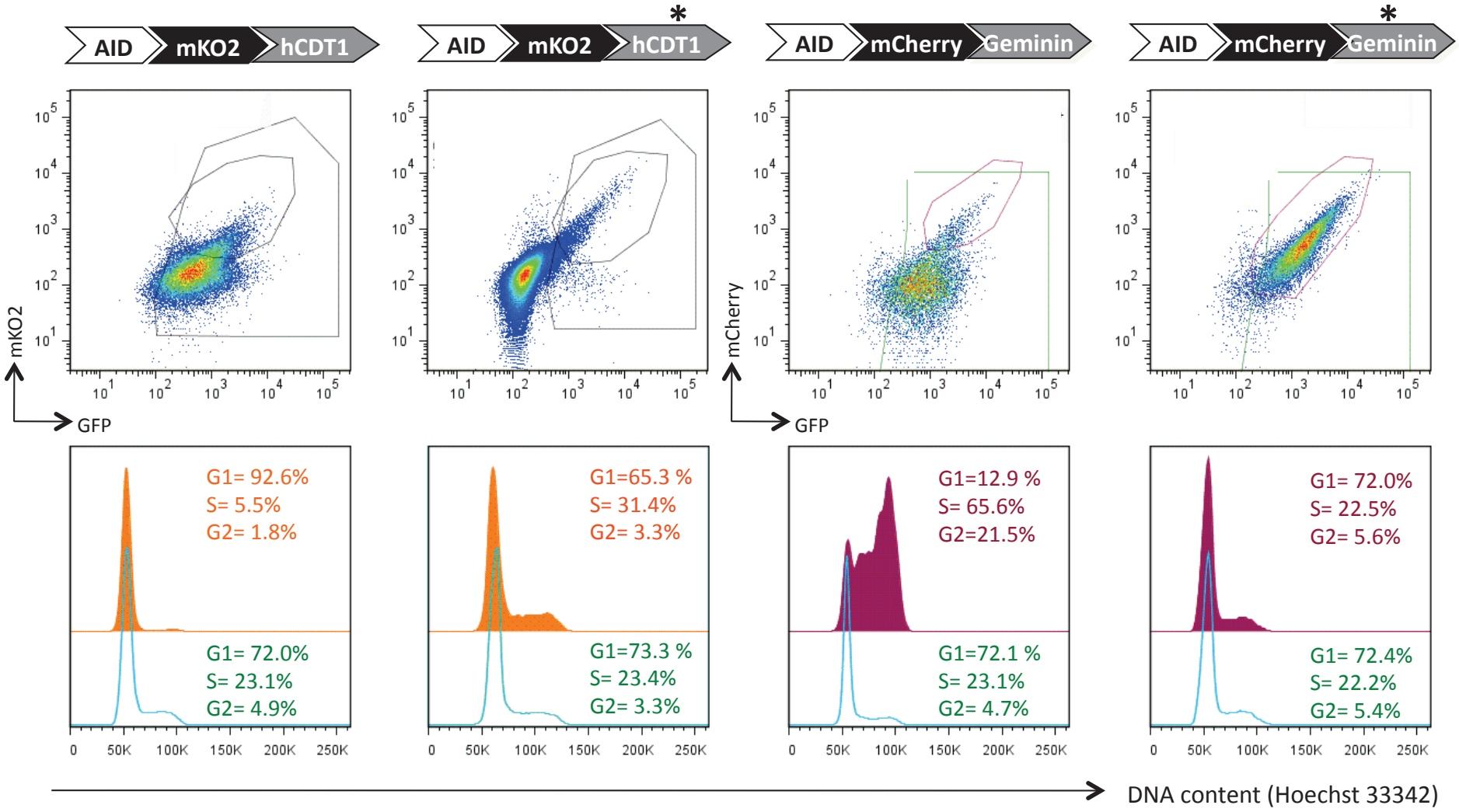

AID
mKO2
hCDT1
*
AID
mKO2
hCDT1
AID
mCherry
Geminin
*
AID
mCherry
Geminin
mKO2
GFP
mCherry
GFP
G1= 92.6%
S= 5.5%
G2= 1.8%
G1= 72.0%
S= 23.1%
G2= 4.9%
G1=65.3 %
S= 31.4%
G2= 3.3%
G1=73.3 %
S= 23.4%
G2= 3.3%
G1=12.9 %
S= 65.6%
G2=21.5%
G1=72.1 %
S= 23.1%
G2= 4.7%
G1= 72.0%
S= 22.5%
G2= 5.6%
G1= 72.4%
S= 22.2%
G2= 5.4%
0
50K
100K
150K
200K
250K
DNA content (Hoechst 33342)

Figure 3

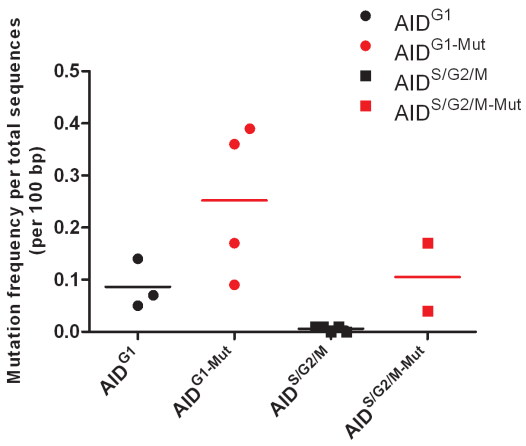
A

	G/C : A/T	transitions : transversions	within G/C			within A/T		
			transitions	transversions		transitions	transversions	
			G to A C to T	G to T C to A	G to C C to G	A to G T to C	A to T T to A	A to C T to G
AID^{G1}	56.9 : 43.1	75.9 : 24.1	76.7	2.0	21.3	75.0	7.8	17.2
$AID^{G1\text{-}Mut}$	58.6 : 41.4	68.0 :32.0	86.6	6.8	6.6	41.6	41.5	16.9
$AID^{S/G2/M}$	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
$AID^{S/G2/M\text{-}Mut}$	50.4 :49.6	63.9 :36.1	78.8	4.7	16.5	48.7	22.0	29.3

B

	AID^{G1}	$AID^{G1\text{-}Mut}$	$AID^{S/G2/M}$	$AID^{S/G2/M\text{-}Mut}$
Number of mice	3	4	5	2
Number of sequences	226	291	232	130
Total length sequenced (bp)*	103,734	133,569	106,720	64,350
Unmutated sequences (percent)	66	77	97	76
Total number of mutations	82	314	8	42
(deletions and insertions)	(3)	(3)	(0)	(2)
Mutation frequency per total sequences (per 100 bp)	0.0819	0.2373	0.0075	0.0737
Mutation frequency per mutated sequences (per 100 bp)	0.3561	0.7047	0.2490	0.3092

C



D

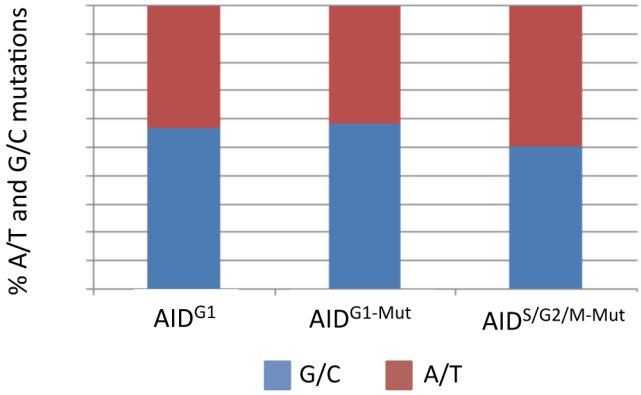


Figure 4

A

	G/C : A/T	transitions : transversions	within G/C			within A/T		
			transitions	transversions		transitions	transversions	
			G to A C to T	G to T C to A	G to C C to G	A to G T to C	A to T T to A	A to C T to G
HC *(range)*	50.0 : 50.0 *(43.6-56.1 : 43.9-56.4)*	51.3 : 48.7 *(47.9-52.4 : 47.6-52.1)*	51.0 *(46.1-54.6)*	16.2 *(12.8-20.4)*	32.8 *(30.4-36.8)*	51.6 *(49.2-57.5)*	21.8 *(17.0-25.1)*	26.6 *(23.2-29.5)*
Pol ε	46.5 : 53.5	47.8 : 52.2	49.3	14.9	35.8	46.3	28.9	24.7

B

	HC IgD^+CD27^+	HC IgD^-CD27^+	Pol ε IgD^+CD27^+	Pol ε IgD^-CD27^+
Number of sequences	137	124	14	12
Total length sequenced (bp)*	38,771	35,092	4760	4080
Unmutated sequences (percent)	4	2	21	8
Total number of mutations	1174	2136	99	182
(deletions and insertions)	(54)	(114)	(4)	(6)
Mutation frequency per total sequences (per 100 bp)	3.03	6.09	2.16	4.61
Mutation frequency per mutated sequences (per 100 bp)	3.14	6.19	2.75	5.03

Figure 5

Part Three: Unpublished research project

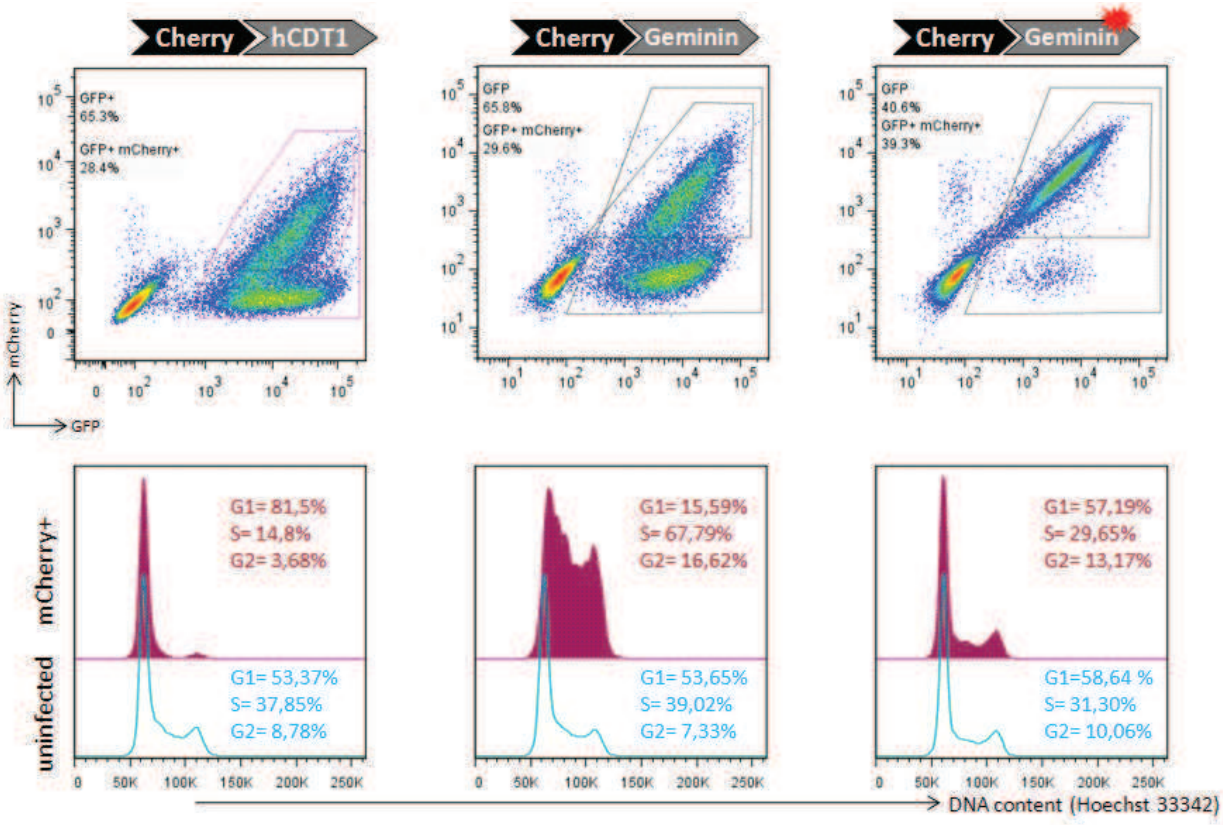
Cell cycle-restricted Polη fusions fail to complement *in vivo* for Pol η activity in somatic hypermutation

Design of Pol η fusions and the quality of restricted Pol η expression

After AID expression was successfully confined to the G1 phase in AID-mKO2-hCDT1:30/120-expressing 3T3 fibroblasts, we set out to replace the fluorochromes from the original constructs (pFucci Orange G1: mKO2-hCDT1:30/120, pFucci-Green S/G2/M: mAG-Geminin:1/110) with mCherry. Since the final fusion protein with mouse polymerase η was to be expressed from pMigR1 construct containing IRES-GFP reporter gene, the need to replace Azami Green (mAG) was obvious. As for pFucci-Orange whose original fluorochrome mKO2 is retained for cloning of AID^{G1} fusion, we sought to replace it with a fluorochrome whose emission signal is better captured with the laser configuration of our flow cytometer. Degradation driven by hCDT1:30/120 degron was shown to be sensitive to the nature of the other fusion partner, whereas Geminin:1/110 was shown to be more permissive (Sakaue-Sawano et al., 2008). However, both degrons governed very successfully the cell-cycle specific degradation of mCherry in NIH 3T3 cells transduced with pMigR1 vectors coding for mCherry fusions (Figure 1). The degradation profile was similar to the one of fusions expressed from the original constructs in HeLa cells. The GFP+mCherry+ population was clearly demarcated from GFP only expressing cells in both cases and was highly enriched in the accumulation-permissive phase. Nonetheless, none of the fusions was spared from a minor leakage of around 15% of $GFP^{+}mCherry^{+}$ cells into phases where the fusion is susceptible to degradation. The majority of "leaky" $GFP^{+}mCherry^{+}$ accumulated at G1/S phase transition, reproducing the pattern of yellow HeLa cells expressing the dual Fucci system in which fluorescence from pFucci Orange G1 and pFucci-Green S/G2/M overlaps at G1/S transition.

Mouse *Polh* cDNA was fused C-terminally to $mCherry^{G1}$ and $mCherry^{S/G2/M}$ fusions. Fusion to Cherry portion was carried out in frame, but lacked the flexible linker that separated AID-C terminus from the fluorochrome in AID constructs. Furthermore, we created control Pol η-mCherry fusions with mutated degrons to verify whether the bulky mCherry-degron tail prevents Pol η from proper folding or full activity. We deleted the Geminin:1/110 portion containing residues for APC/C-mediated ubiquitinylation (degron S/G2/M-Mut) and mutated the PARRRLRL motif of hCDT1:30/120 governing SCF/Skp2 (degron G1-Mut), thus abrogating the cell cycle restriction of corresponding fusions. When the cell cycle profile of SV40T-transformed Pol η deficient mouse embryonic fibroblasts (MEF-T) transduced with vectors expressing $Polη\text{-}Cherry\text{-}degron^{Mut}$ was assessed, the $GFP^{+}mCherry^{+}$ population displayed a cell cycle profile similar to the one of non-transduced cells (Figure 2). For the sake of clarity, Pol η-mCherry-degron/mutated degron fusions will be named from now on Pol $η\text{-}^{G1}$, Pol $η\text{-}^{G1\text{-}Mut}$, Pol $η\text{-}^{S/G2/M}$ and Pol $η\text{-}^{S/G2/M\text{-}Mut}$ in further text.

Figure2. mCherry cell cycle-regulated expression in NIH 3T3 mouse fibroblasts.
pMigR1 retroviral constructs express from the LTR promoter: G1- and S/G2/M-restricted form of mCherry and the cell cycle mutant mCherry-S/G2/M-Mut (represented with a red sparkle) concomitantly with GFP from the pMigR1 IRES. Staining with Hoechst was used to estimate DNA cell content. The cell profile of uninfected cells (bottom row) and of $GFP^{+}mCherry^{+}$ cells (upper row) is indicated, with distributions within different phases of the cell cycle estimated by the Flow Jo software.



Leakiness of MEF-T *Polh*$^{-/-}$cells expressing Pol η-mCherry fused to functional degrons was comparable to percentages of GFP-mCherry double positive cells in NIH 3T3 expressing the respective mCherry-only fusions (around 15% in phases prone to degradation). However, the level of mCherry fluorescence in GFP+ mCherry+ populations of Polη-transduced MEF differed markedly from the one of the restricted mCherry fusions. Despite the fact that the expression of mCherry fusions with or without Pol η was driven by the same retroviral LTR promoter and that in all cases multiplicity of infection (MOI) was in the same order of magnitude, mCherry mean fluorescence intensity (MFI) of $GFP^{+}mCherry^{+}$ population for Pol η fusions was considerably lower compared to fusions with sole mCherry. This was observed for all Pol η fusion in both MEF *Polh*$^{-/-}$and hematopoietic stem cells, irrespective of the restriction profile (Figures 2 and 5).

In vitro activity of Polη in fusion with mutated degrons

To check for a potential alteration of Pol η activity due to its embodiment within a fusion and independently of the cell cycle restriction context, we tested the activity of Pol η fused to mutated degrons in MEF-T cells deficient for Pol η. UV irradiation lead to the formation of repair foci in which Ub-PCNA and Pol eta cluster (Kannouche et al., 2004b) . We set off to test the ability of Pol η-mCherry fusions to accumulate in UV-induced foci in irradiated MEFs. A C-terminal fusion of Pol η with EGFP expressed from a pMigR1 vector lacking IRES-GFP portion was used as a positive control (Figure3). In non-irradiated cells, mCherry was diffusely distributed throughout the nuclei of both Pol η-$^{G1\text{-}Mut}$ and Pol η-$^{S/G2/M\text{-}Mut}$ - transduced cells. Six hours post-irradiation, mCherry fluorescence of both constructs segregated into distinct focal points. The same accumulation pattern was unambiguously reproduced by Pol η-EGFP fusion.

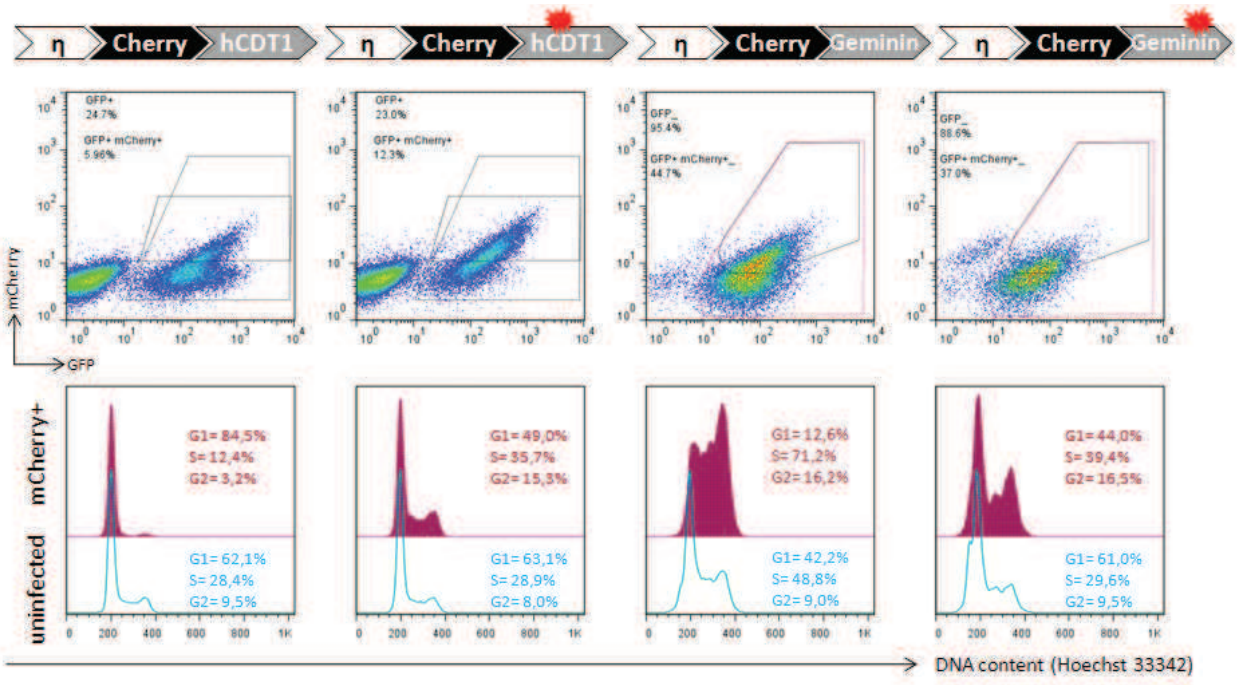


Figure3 Cell cycle restricted expression of Pol η in SV40 T-transformed MEF *Polh*$^{-/-}$ cells.
pMigR1 retroviral constructs expressing the G1-restricted form of Pol η fused to the mCherry fluorescent reporter (Pol η-mCherry-G1) and its cell cycle control mutant (represented by a red spark), the S/G2/M-restricted from (Pol-Cherry-S/G2/M) and its cell cycle control mutant (same symbol), with an IRES EGFP reporter. Cell cycle analysis was performed using Hoechst staining to estimate the cell DNA content. The cell cycle profile of EGFP^{+} cells (bottom row) and of mCherry-EGFP double positive cells (upper row) is indicated, with distribution within the different phases of the cell cycle estimated by the FlowJo software.

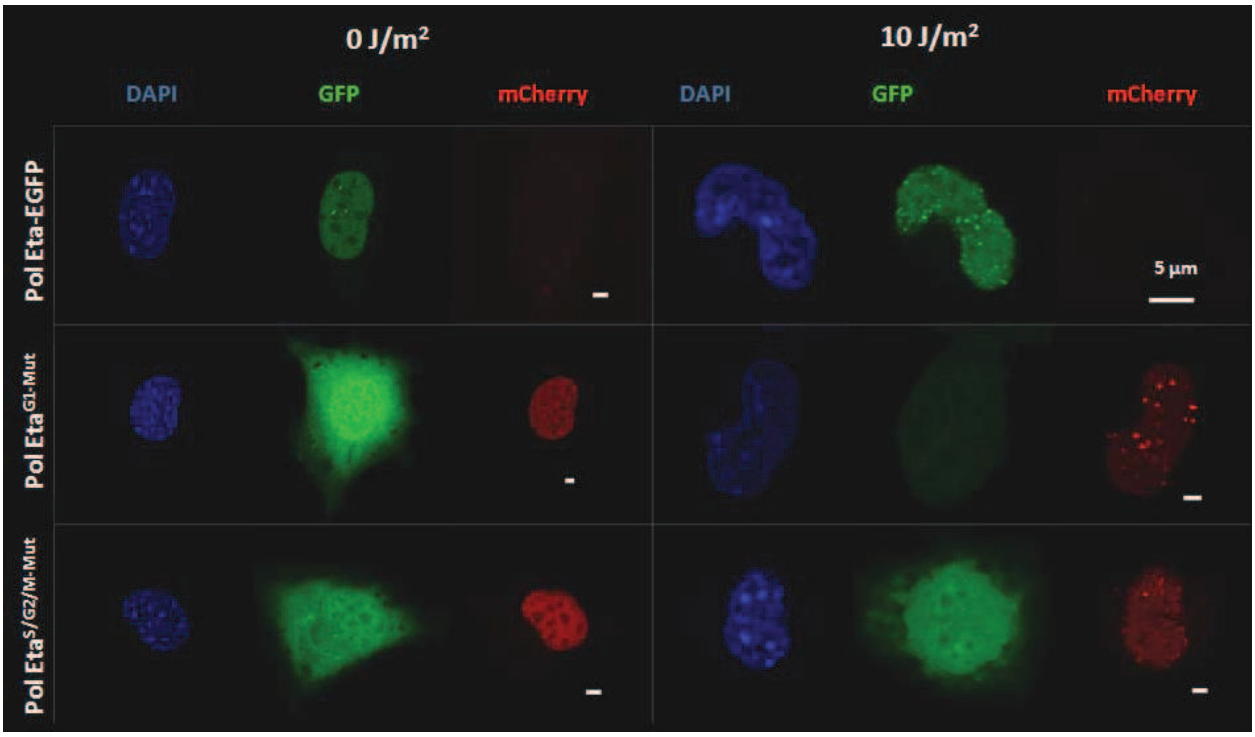


Figure3 Subnuclear organization of Pol η-EGFP fusion and Pol η-mCherry fusions with mutated degrons (Pol η-$^{G1\text{-}Mut}$ and Pol η-$^{S/G2/M\text{-}Mut}$). Transduced and GFP^{+} sorted non-irradiated MEF Poln$^{-/-}$ (left panel including DAPI/GFP/mCherry columns) or UV-irradiated with 10 J/m^2 (right panel) were fixed 6h post-irradiation and mounted on coverslips. Post-acquisition treatment of micrographs was carried out in ImageJ software.

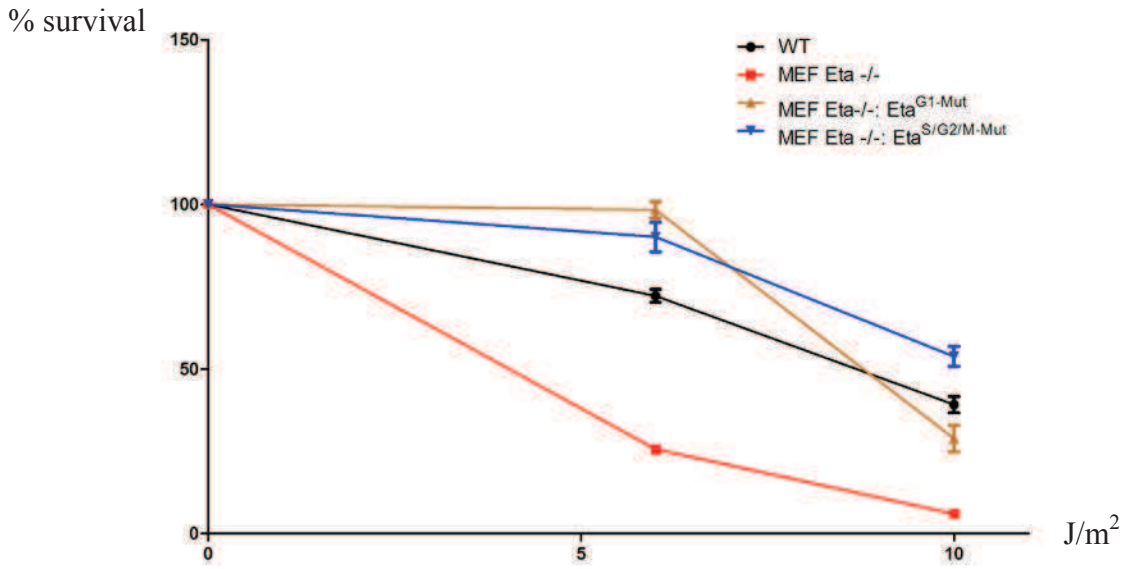


Figure 4 Survival of WT MEFs, Poh$^{-/-}$ MEF-T mock-transduced and transduced with pMigR1 constructs expressing Pol η fusions with abrogated cell cycle restriction.
Survival was measured with a colorimetric assay 72h post-UV irradiation for non-irradiated, 6 J/m^2 and 10 J/m^2-irradiated cells after recovery in culture supplemented with 1mM caffeine. Data points were acquired in triplicates, averaged and subtracted for the values of background emission. Displayed error bars represent SD. Data were analyzed and plotted in GraphPad Prism software.

During the last years, an increasing body of evidence supports a temporal separation between the visible accumulation of TLS polymerases into subnuclear foci and the mere act of loading them onto DNA and performing lesion bypass (Gueranger et al., 2008; Soria et al., 2009). We pursued our testing of Pol η activity in fusion beyond recruitment to the UV-induced foci by measuring the capacity of Pol η-$^{G1\text{-}Mut}$ and Pol η-$^{S/G2/M\text{-}Mut}$ fusions to complement the UV light sensitivity of MEF-T *Polh*$^{-/-}$. To do so, we subjected previously transduced and sorted GFP^{+} cells to two different doses of UV irradiation (6 and 10 J/m^2), allowed them to recover in complete medium supplemented with 1mM caffeine for 72h, followed by the quantification of post-UV irradiation survival. Non-irradiated transduced MEFs were treated simultaneously with irradiated cells (Figure 4).

Pol η-deficient MEFs expressing Pol η-$^{G1\text{-}Mut}$ and Pol η-$^{S/G2/M\text{-}Mut}$ were efficiently rescued from cell death due to the replication fork collapse (observed for *Polh*$^{-/-}$ MEF-T) already at 6 J/m^2 of UV irradiation. This efficiency was close to the success rate of endogenously expressed WT Pol η at 10 J/m^2. These results indicate that stable ectopic expression of Pol η-$^{G1\text{-}Mut}$ and Pol η-$^{S/G2/M\text{-}Mut}$ fusions driven by retroviral promoters is able to complement the UV sensitivity of *Polh*$^{-/-}$ deficient MEFs. However, one should refrain from concluding that a single Pol η-fusion molecule is able to substitute for full WT Pol η activity due to the difficult-to-control Pol η overexpression from LTR promoter and from multiple integration sites.

Adoptive transfer of Polh-/- hematopoietic cells expressing cell-cycle restricted Pol η

We proceeded to generate retrogenic mice with adoptively transferred hematopoietic stem cells (HSC) expressing Pol η-^{G1}, Pol η-$^{G1\text{-}Mut}$, Pol η-$^{S/G2/M}$ and Pol η-$^{S/G2/M\text{-}Mut}$- fusions in the same manner as for cell cycle-restricted AID expression. The phase-restricted cell cycle profiles of GFP^{+}mCherry^{+} Lin^{-} cells expressing Pol η-^{G1} and Pol η-$^{S/G2/M}$ constructs strictly reproduced those of mCherry^{G1} and mCherry$^{S/G2/M}$ fibroblasts, respectively. Transduced HSC cells were more efficient in limiting GFP^{+}mCherry^{+} cells leakage to degradation-prone phases (Figure 5).

Interestingly, if one compares mCherry MFI of the GFP^{+}mCherry^{+} cells in, for example, AID$^{S/G2/M}$ and Pol η$^{S/G2/M}$ Lin^{-} cells for a given value of GFP fluorescence, it becomes clear that Pol η-mCherry fusion starts to visibly accumulate only in GFP^{hi} cells. Values of accumulated AID-mCherry in AID$^{S/G2/M}$ cells closely matched the quantity of GFP per cell. These properties of Pol η-mCherry accumulation and expression underline the existence of additional cellular mechanisms that actively regulate the available Pol η pool in the cell.

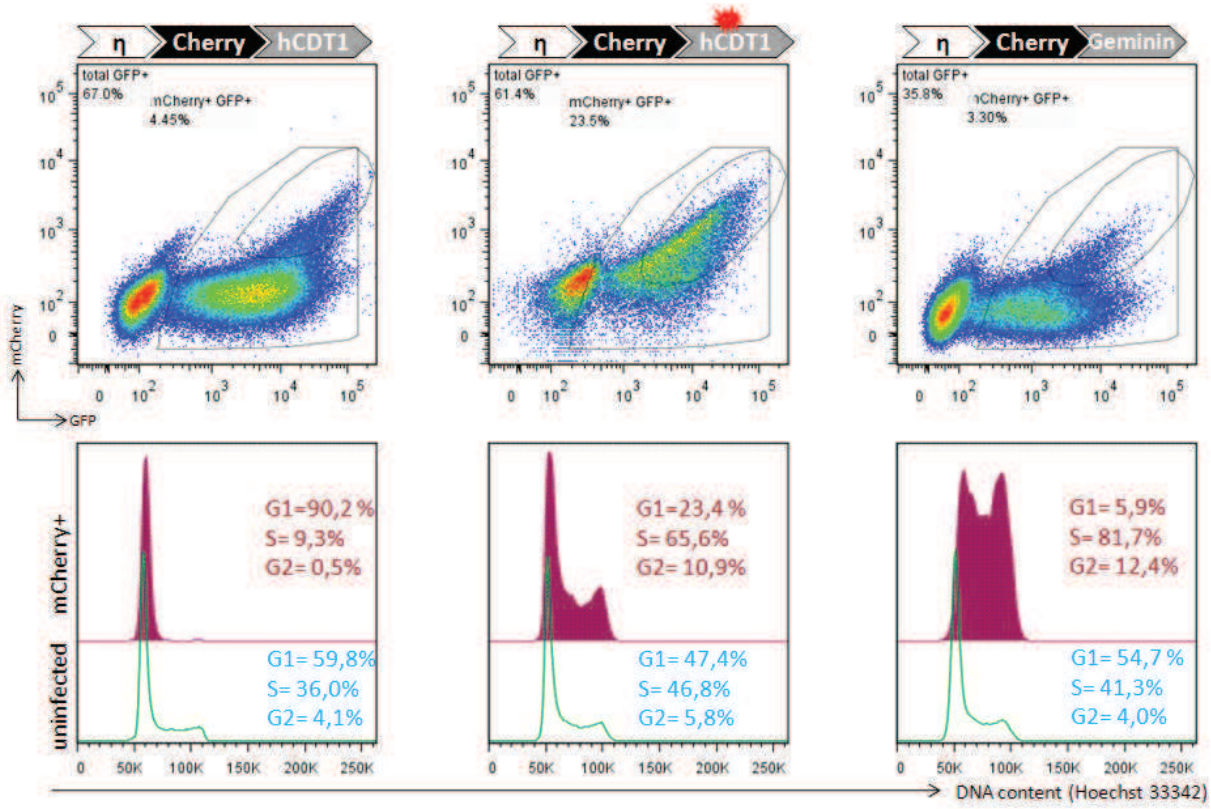


Figure5. Cell cycle restricted expression of Pol η in transduced *Polh*$^{-/-}$ hematopoietic stem cells.
Retroviral constructs expressing G1- and S/G2/M-restricted forms of Pol η along with Pol η-mCherry fused to mutated G1 degron (represented by a red spark), with an IRES EGFP reporter were used to transduced Lin^{-} hematopoietic stem cells. Fusion of Pol η to cell cycle mutant for S/G2/M-conferred restriction were not available for cell cycle analysis. Cell cycle analysis was performed three days afterthe last transduction cycle, using Hoechst staining to estimate the cell DNA content. The cell cycle profile of non-transduced cells (bottom row) and of mCherry-EGFP double positive cells (upper row) is indicated, with distribution within the different phases of the cell cycle estimated by the FlowJo software.

In spite of high transduction levels, Pol η-^{G1}(67% GFP^{+} cells) and Pol η-$^{G1\text{-}Mut}$ (61.4% GFP^{+} cells) poorly reconstituted the B and T-cell pool of irradiated Rag2-deficient receivers. Extremely low percentage of PNA^{hi}GFP^{+} B-cells in spleens of immunized Pol η-^{G1} andPol η-$^{G1\text{-}Mut}$-reconstituted mice prevented us from amplifying the J_H4 intronic region with direct PCR strategy. We temporarily abandoned the plan of restricting of Pol η to G1 phase and pursued the analyses of the Pol η-$^{S/G2/M}$ and Pol η-$^{S/G2/M\text{-}Mut}$-restored mice, which were supposed to yield more informative readout to test our hypothesis of a cell cycle-coordinated SHM process (Figure 28, Chapter III.3).

Pol ηS/G2/M-Mut germinal center B-cells have mutation profile identical to Pol η-deficient mice

B and T-cell pools were successfully reconstituted upon the transfer of Pol η-$^{S/G2/M}$and Pol η-$^{S/G2/M-Mut}$–expressing Lin- cells in irradiated *Rag2*$^{-/-}$ mice (3 mice/ group). Levels of B- and T-cells were comparable to reconstitution with AID-coding constructs (not shown). Immunizations with sheep red blood cells triggered a robust response, allowing us to sort up to 50,000 B220^{+}PNA^{hi} GFP^{+} cells per mouse in the S/G2/M-restricted group (representing a mean of 0.5% of total live spleen cells) and successfully perform V_H-D-J_H4 intron amplification and sequencing. However, only in one Pol η-$^{S/G2/M-Mut}$ mouse could an amplification with sufficient clonal junctional diversity be obtained (Figure 6).

Mutation profile of Pol η$^{S/G2/M}$ mice closely resembled the one of Pol η deficient mice (data taken from (Delbos et al., 2007). Such a profile, biased for A to C and T to G transversions harbors the signature of Polκ, which substitute for Pol η in its absence (as seen in Figure 6D). However, the same held true for the single Pol η-$^{S/G2/M-Mut}$ mouse analyzed. Albeit satisfactory UV sensitivity complementation in *Polh*$^{-/-}$MEFs and successful recruitment to subnuclear foci were observed, Pol η fused with mutated S/G2/M degron and henceforth expressed throughout the cell cycle on a Pol η-deficient background, failed to restore the *in vivo* WT Pol η mutation pattern in germinal center B-cells. We concluded that Pol η fused in C-terminal configuration with mCherry and mutated Geminin:1/110Δ degron severely impaired the enzyme's *in vivo* efficiency to be recruited to Ig genes or to perform A:T mutagenesis in somatic hypermutation and prevents us from concluding about its possible in vivo restriction to a specific phase of the cell cycle.

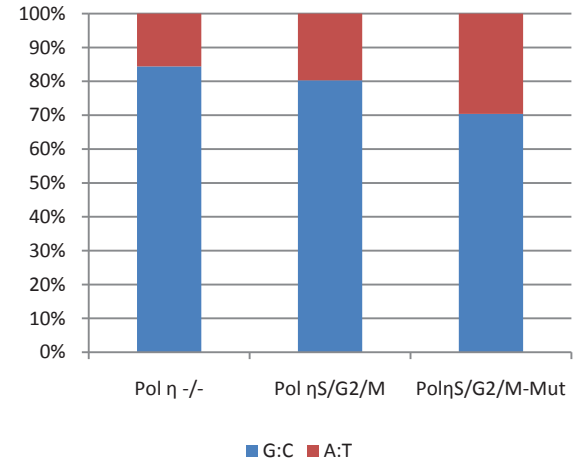
A

	G/C : A/T	transitions : transversions	within G/C			within A/T		
			transitions	transversions		transitions	transversions	
			G to A C to T	G to T C to A	G to C C to G	A to G T to C	A to T T to A	A to C T to G
Pol η $^{-/-}$	84.4 :15.6	53 :47	59.3	11.7	29.0	18.8	24.7	56.5
Pol η$^{S/G2/M}$	80.3 :19.7	49 :51	54.5	10.0	35.5	26.5	20.5	53.0
Polη$^{S/G2/M-Mut}$	70.4 :29.6	53 :47	67.1	11.6	21.4	19.7	29.6	50.7

B

	Pol η$^{S/G2/M}$	Pol η$^{S/G2/M-Mut}$
Number of mice	3	1
Number of sequences	129	56
Total length sequenced (bp)	32,859	35,391
Unmutated sequences (percent)	23	6
Total number of mutations	184	174
(deletions and insertions)	(1)	(0)
Mutation frequency per total sequences (per 100 bp)	0.70	0.58
Mutation frequency per mutated sequences (per 100 bp)	1.18	0.62

C



D

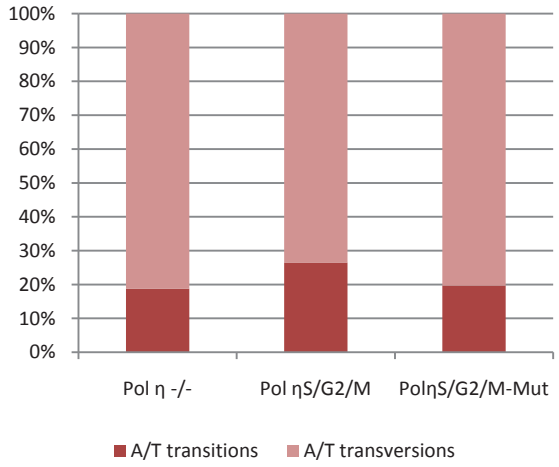


Figure 6. Mutation pattern of J_H4 intron of splenic GFP+ GC B-cells from host mice with adoptively transferred Pol η-$^{G1-Mut}$ and Pol η-$^{S/G2/M-Mut}$-expressing hematopoietic stem cells.
A) Pattern of nucleotide changes in Pol η-deficient mice (dataset taken from (Delbos, 2007)), Pol η-$^{S/G2/M}$ and Pol η-$^{S/G2/M-Mut}$retrogenic host mice. B) Summary of analyzed sequences and pooled mutation rates per reconstituted host group. C) Graphic representation of relative frequencies for A:T and G:C mutations in J_H4 intron of GFP+ GC B cells fromPol η-$^{G1-Mut}$ and Pol η-$^{S/G2/M-Mut}$-expressing retrogenic mice (values from A). D) Graphic representation of relative frequencies for transitions and transversions at A:T base pairs in Pol η-/-, Pol η-$^{G1-Mut}$ and Pol η-$^{S/G2/M-Mut}$,

Materials and methods

Construction of retroviral vectors.

Mouse *Polh* gene, previously amplified from mouse cDNA and cloned into an intermediary vector, was amplified using the following PCR primers: forward (mPol eta BamHI For) and corresponding assembly reverse primers.

In the following course of experiments, we amplified separately the fragments of mCherry, hCDT1_30-120aa and hGeminin_1-110aa degrons from commercially available cloning vectors: pmCherry (Clontech), pFucci-G1 Orange and pFucci –S/G2/M Green (MBL International), respectively. POLH gene were fused to mCherry and degrons in successive steps of assembly PCR reactions using Phusion polymerase (New England Biolabs), individual fragments as templates and following cycling conditions: first 10 cycles without assembly primers : 10 sec 98°C / 30 sec 55°C / 60-90 sec 72°C and 20-25 cycles with same conditions except for annealing temperature (60°C) in the presence of assembly primers.

G1 degron (hCDT1_30-120) in mCherry$^{G1\text{-}Mut}$ moiety was generated by PCR mutagenesis by mutating RRL to AAA in the PARRRLRL motif responsible for its degradation. mCherry$^{S/G2/M\text{-}Mut}$ was generated by PCR site-directed mutagenesis as an in-frame deletion of the destruction box sequence (RRTLKVIQP) that begins at amino acid 33 in hGeminin_1-110aa degron and that abolishes APC-mediated ubiquitination at M/G1 transition.

Pol eta-GFP fusion was generated by successive assembly PCR steps using Ass Pol eta-EGFP For and Ass Pol eta-EGFP Rev.

pMigR1 (MSCV: IRES-EGFP (Pear et al., 1998), courtesy of O.A.-C.) retroviral vector was used as a backbone for cloning of finalized phase-restricted fusions: Pol eta-mCherry^{G1}, Pol eta-mCherry-^{G1Mut}, Pol eta-mCherry-$^{S/G2/M}$, Pol eta-mCherry-$^{S/G2/M\text{-}Mut}$.In order to express Pol eta-GFP fusion, IRES-GFP fragment was subsequently deleted by cutting it out from pMigR1: Pol eta-GFP finalized vector. All constructs were verified by DNA sequencing.

Production and titration of ecotropic retroviral particles.

Cf. Same chapter in Part I, article "AID expression restricted to the G1 phase of the cell cycle recapitulates the mutation spectrum of immunoglobulin genes"

Cell lines and transduction conditions.

NIH 3T3 cells were cultured in DMEM 10% bovine calf serum (HyClone). In-house isolated and SV40T-transformed mouse embryonic fibroblasts (MEF-T) *Polh*$^{-/-}$ were grown in DMEM 10% FCS (Hyclone) with 1mM sodium pyruvate. All mouse cell lines were transduced in the same conditions: briefly, cells plated in suboptimal confluence were incubated with 5 µg/mLpolybren in PBS (Sigma Aldrich) for 1h, supplemented with HEPES 10mM (Invitrogen Life Technologies) and inoculated with retroviral particles at MOI varying from

1-3. Cell plates were centrifuged for 1h30 at 600g 25°C and allowed to incubate with virus for 24h, after which the medium was replaced.

Confocal microscopy.

Adherent cells (MEF-T *Polh*$^{-/-}$) were seeded in 24-well plates containing glass coverslips previously coated with poly-L-lysine 0.01% w/v solution in water (Sigma Aldrich) and allowed to adhere overnight. The next day, medium was carefully removed from the wells and cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences) solution in PBS for 15 min. After washing with PBS, coverslips were mounted in Vectashield DAPI (Vector Laboratories) on glass slides. Micrographs were acquired with Leica TCS SP5 confocal microscope (Leica Microsystems, Necker/Broussais imaging facility). All images were treated with ImageJ 1.46r software (NIH).

Pol η complementation assay for the survival of transduced MEF-T *Polh*$^{-/-}$.

MEF-T *Polh*$^{-/-}$ cells were plated at 2x10^4 cells/cm^2 and transduced with pMigR1: Poleta-expressing vectors as described previously. Seventy-two hours post-transduction, cells were trypsinized, counted, washed and sorted for GFP fluorescence. GFP$^+$ transduced MEF-Ts were grown and checked for GFP expression for a week then plated at 1x10^4 cells per well in 96-well white plates (Nunc) in triplicate. 24 hours after plating, wells were washed with PBS and let dry while irradiating with UVC light (253,7nm)at 0, 6 and 10 J/cm^2 using the UV irradiator at Institut de Recherche Integrée en Cancerologie Gustave Roussy, Villejuif (provided kindly by S. Aoufouchi, UMR CNRS 8200). Immediately after irradiation, MEF-T culturing medium supplemented with 1mM caffeine (Sigma-Aldrich) was added to irradiated wells. Seventy-two hours post-UV irradiation, cells were trypsinized and tested for viability by using the CellTiter-Glo Luminescent Cell Viability Assay (Promega). Colorimetric measurements were carried out with VICTOR X4 2030 Multilabel plate reader (Perkin Elmer), with each plate read three times with 1min interval in between.

Pol η foci formation in transduced MEF-T *Polh*$^{-/-}$.

After transducing and sorting, GFP$^+$ MEF-T *Polh*$^{-/-}$ cells were grown and checked for GFP expression for a week, then plated at 2x10^4 cells/well on sterile round glass coverslips deposed in wells of a 24-well plate and previously coated with polylysine solution in PBS. Cells were grown on coverslips for 24h, and then irradiated with 254 nm UVC light at 0 and 10 J/cm^2 by using UV-sterilization lamps (output: 13.4 W). Six hours post-irradiation, coverslip-grown MEFs were fixed in paraformaldehyde and mounted.

Mouse strains.

Polh$^{-/-}$C57BL/6 mice were maintained in specific pathogen-free conditions in mouse facility of Service de l'Expérimentation Animale et Transgénèse, CNRS Villejuif. RAG2-deficient mice were maintained in specific pathogen-free conditions in Laboratoire d'Expérimentation Animale et Transgénèse, Faculté de Médecine Necker and Broussais mouse facility. All experiments and procedures were approved by the French Ministry of Agriculture.

Hematopoietic stem cell transduction and adoptive transfer experiments.

Cf. Same chapter in Part I, article "AID expression restricted to the G1 phase of the cell cycle recapitulates the mutation spectrum of immunoglobulin genes"

Cell cycle analysis.

Cf. Same chapter in Part I, article "AID expression restricted to the G1 phase of the cell cycle recapitulates the mutation spectrum of immunoglobulin genes"

Other flow cytometry analysis.

Cf. Same chapter in Part I, article "AID expression restricted to the G1 phase of the cell cycle recapitulates the mutation spectrum of immunoglobulin genes"

Immunization, sorting strategy, genomic DNA extraction and sequence analysis.

Cf. Same chapter in Part I, article "AID expression restricted to the G1 phase of the cell cycle recapitulates the mutation spectrum of immunoglobulin genes"

Acknowledgments

We are indebted to the animal care provided by the SEAT (Villejuif) and LEAT (Necker). We kindly thank Said Aoufouchi and Rémi Fritzen for helping us with UV-irradiation of MEF-T and UV sensitivity complementation assays. We would also like to thank Jérôme Mégret for cell sorting, Raphaelle Desvaux and Simon Le Gallou for assistance in micrograph acquisition and Laïla Kossir for her contribution to construction of phase-restricted vectors.

Supplementary Figure 1.Oligonucleotide panel used in this study

V1-FR3	GAGGACTCTGCRGTCTATTWCTGTGC
V5-FR3	GAGGACACRGCCATGTATTACTGTGC
V3-FR3	GAGGACACACCCACATATTACTGTGC
V7-FR3	GAGGACAGTGCCACTTATTACTGTGC
V9-FR3	ATGAGGACATGGCTACATATTTCTGT
Jh4 nested 2	CACCAGACCTCTCTAGACAGC
Jh4 AS 2620	TCCCTCCAGCCATAGGATTG
hCdt1EcoRI-PacI Rev	AAG AAT TCT TAA TTA AAT GGT GTC CTG GTC CTG C
hCdt1mut For	GCC CGC GCG GCA GCG CGG CTG TCG GTG GAC GAG
hCdt1mut Rev	CAG CCG CGC TGC CGC GCG GGC CGG TGG CCT GGC
Ass mCherry-hCdt1 For	CTG TAC AAG GGA TAT CCA TCA CAC TGG CG
Ass mCherry-hCdt1 Rev	GAT GGA TAT CCC TTG TAC AGC TCG TCC ATG C
EGFP-PSLD BglII For	AAA GAT CTG CCG CCA CCA TGG TGA GCA AGG GCG A
EGFP-PSLD BamHI-PacI For	AAG GAT CCT TAA TTA AGT TTC TTG ATT ATC GGT G
mPol eta BamHI For	AAG TCG AGG ATC CGC CAC CAT GGC TCC TGG GCA GAA TCG A
Ass Pol eta-EGFP For	ACT GAC ACA TGT GAG CAA GGG CGA GGA
Ass Pol eta-EGFP Rev	CCT TGC TCA CAT GTG TCA GTG GCT TAA A
hGemEcoRI-PacI Rev	AAG AAT TCT TAA TTA AAG CGC CTT TCT CCG TTT TTC
hGemmut For	TTC TGT CCC ATC TGC ATC TGG ATC TCT TGT TG
hGemmut Rev	CAG ATG CAG ATG GGA CAG AAC TAT TCT TTA TA

References

Delbos, F., Aoufouchi, S., Faili, A., Weill, J.-C., and Reynaud, C.-A. (2007). DNA polymerase η is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. The Journal of Experimental Medicine *204*, 17 –23.

Gueranger, Q., Stary, A., Aoufouchi, S., Faili, A., Sarasin, A., Reynaud, C.-A., and Weill, J.-C. (2008). Role of DNA polymerases eta, iota and zeta in UV resistance and UV-induced mutagenesis in a human cell line. DNA Repair (Amst.) *7*, 1551–1562.

Kannouche, P.L., Wing, J., and Lehmann, A.R. (2004). Interaction of Human DNA Polymerase η with Monoubiquitinated PCNA: A Possible Mechanism for the Polymerase Switch in Response to DNA Damage. Molecular Cell *14*, 491–500.

Pear, W.S., Miller, J.P., Xu, L., Pui, J.C., Soffer, B., Quackenbush, R.C., Pendergast, A.M., Bronson, R., Aster, J.C., Scott, M.L., et al. (1998). Efficient and Rapid Induction of a Chronic Myelogenous Leukemia-Like Myeloproliferative Disease in Mice Receiving P210 bcr/abl-Transduced Bone Marrow. Blood *92*, 3780 –3792.

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., et al. (2008). Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. Cell *132*, 487–498.

Soria, G., Belluscio, L., Cappellen, W.A. van, Kanaar, R., and Gottifredi, J.E. and V. (2009). DNA damage induced Pol η recruitment takes place independently of the cell cycle phase. Cell Cycle *8*, 3340–3348.

Part Four: Research article

Somatic hypermutation at A/T-rich oligonucleotide subtrates show different strand polarities in Ung-deficient or Ung-proficient backgrounds

Marija Zivojnovic[1], Frédéric Delbos[1#], Giulia Girelli Zubani[1], Amélie Julé[1*], Jean-Claude Weill[1], Claude-Agnès Reynaud[1] and Sébastien Storck[1]

[1]INSERM UMR783, Université Paris Descartes, Faculté de Médecine-Site Broussais, Paris, (France)

Present address:

[#]INSERM UMR1064, Institut de Transplantation-Urologie-Néphrologie, Nantes (France)

*INSERM UMR1016 - CNRS UMR8104, Institut Cochin, Université Paris Descartes, Paris (France)

Running title: back-up function of Ung and Pms2 in A-T mutagenesis

Submitted to Molecular and Cellular Biology on October 30 2013

control number: MCB01452-13

Summary

A/T mutations at the immunoglobulin loci are introduced by DNA polymerase eta (Polh), during an Msh2/6-driven repair process resulting in As being two-fold more mutated than Ts. This patch synthesis is initiated by a DNA incision event, the origin of which is still obscure. We report here the analysis of A/T oligonucleotide mutation substrates inserted at the heavy chain locus, including or not internal Cs or Gs. Surprisingly, the template composed of only As and Ts was highly mutated over its entire 90-bp length, displaying a 2-fold decrease in mutation from 5' to 3' and a constant A/T ratio of 4. These features implied that Polh synthesis was initiated from a break in the 5'-flanking region of the substrate and proceeded over its entire length. The A/T bias was strikingly altered in an *Ung*$^{-/-}$ background, thus providing the first experimental evidence supporting the proposition of Schanz *et al.* (1) of a concerted action of Ung and Msh2/6 pathways to generate mutations at A/T bases. New analysis of *Pms2*$^{-/-}$ animals provided a complementary picture, revealing an A/T mutation ratio of 4. We thus propose that Ung and Pms2 provide mutual back-up function for the DNA incision that promotes synthesis by Polh.

Introduction

Somatic hypermutation of immunoglobulin genes is a locus-specific mutagenesis, triggered by an initial deamination event performed by the lymphoid-restricted enzyme, AID (activation-induced cytidine deaminase) (2, 3). Such a process which mostly, but not exclusively, takes place in B cells engaged in an immune response, allows the unique adaptive biological process by which the immune system fighting a pathogen improves the specificity and efficiency of its recognition and, as a consequence, its control and its elimination.

The paradoxical feature of hypermutation is the mutagenic outcome of the processing of this ordinary lesion, uracils generated by cytidine deamination or dUTP incorporation during replication being normally efficiently and faithfully removed from the genome. This mutagenesis appears to result from a combined process of lesion ignorance, incomplete processing or diverted repair, mobilizing as mutagenic partners actors with a major function in genome maintenance and mutation avoidance: uracil glycosylase and mismatch repair (4).

Uracils can remain unprocessed, generating transitions at G/C bases upon the next replication step. They can also be partially processed by uracil glycosylase, giving rise to abasic sites that will be bypassed by translesion DNA polymerases, among which Rev1 is the only one unambiguously identified (5, 6). This process generates transversions, but also transitions at the deamination site. A fraction of uracils can be processed further, resulting in DNA incision. Such nicks constitute a key step in class switch recombination, a second molecular alteration driven by AID at the switch regions

of the Ig locus that involves the rejoining of two double-strand breaks generated at distant sites and that exchanges the constant domain of the heavy chain, allowing the B-cell to acquire new effector capacities. Nicks in DNA also take place in the V region, resulting, in a fraction of events, in classical, error-free base excision repair.

Uracils can also be recognized as mismatches by Msh2-Msh6, which, together with ExoI triggers a long patch error-prone DNA synthesis that relies on DNA polymerase eta (Polη) and ubiquitination of PCNA (7, 8). This pathway generates mainly mutations at A/T bases, at distance from the initial deamination sites, due to the specific misincorporation of Polη that has low fidelity when copying Ts. This patch repair is unusual in that the second part of the mismatch repair complex, MutLa, composed of Pms2 and Mlh1, appears dispensable, and relies on the sole Msh2-Msh6 complex (MutSa), together with Exonuclease I (ExoI) (9, 10). It is thus totally unclear which endonuclease activity provides the DNA incision required to initiate the error-prone synthesis. Moreover, the mutation pattern of Ig genes displays an average two-fold higher targeting at A bases than at T bases, suggesting a preferred Polη synthesis of the coding strand.

To understand the process directing the strand choice, Uniraman and Schatz (2007) designed transgenic light chain substrates in which they inserted a 100-bp oligonucleotide uniquely composed of alternate As and Ts except for one central C or G as AID target. The aim was to correlate the strand bias observed in the mutation pattern with the strand targeted for deamination, either the coding (top) or the transcribed (bottom) strand. In spite of its elegance, this approach had major weaknesses and some inconsistencies: first, the transgene mutation load was very low, 22 mutations being collected in total for the transgene including one C, and almost none for the one containing G. It was thus concluded that only the top strand can be targeted by AID, whereas mice deficient for both uracil glycosylase and mismatch repair (*Ung*$^{-/-}$*Msh2*$^{-/-}$), a configuration that reveals the unprocessed footprint of AID-mediated deaminations, have shown a mutation profile with only a minor bias of C over G targeting (4, 9). Lastly, the mutation pattern observed, biased towards T mutations, was not consistent with the mutagenic outcome of Polη copying the transcribed strand.

Transgenic mutation substrates are notoriously variable in their capacity to become efficiently targeted, making comparisons between mouse lines with unrelated insertion sites difficult. We decided therefore to perform a similar study by using knock-in strategies at the IgH locus, allowing the collection of large mutation databases and the direct comparison of different substrates. Such a strategy gave a rather different picture of Polη performances at the Ig locus, with a long patch length and a different strand bias in Ung-proficient or -deficient backgrounds.

Materials and Methods

In vitro Polη polymerization assay.

The two oligo-C/oligo-G oligonucleotides were annealed and inserted by blunt-end cloning in the EcoRI site of the M13mp18 vector with an orientation resulting in the presence of the oligo-G sequence in the single stranded DNA form. *E. coli* JM109 competent cells (Promega) were used for transformation, production of bacteriophage stocks and mid-size single-stranded DNA preparation. A polymerization initiation primer (ATTTAGGTGACACTATAGTGTAAAACGACGGCCAGT) was designed that includes an SP6 sequence (underlined) and an M13 forward sequence priming 60 bp downstream of the oligo-G cloning site. Purified human Polη (80 nM, Enzymax) was used to elongate this primer in presence of 40 fmol (100 ng) single-stranded M13mp18-oligo-G vector in polymerase:primer:template molar proportions of 50:10:1 in 25 µl of the polymerization mixture containing 250 µg/ml bovine serum albumin (BSA, MP), 40 mM Tris-HCl pH 8.0, 60 mM KCl, 10 mM MgCl2, 10 mM dithiotreitol, 2.5% (v/v) glycerol (Sigma) and 100 µM total dNTP (Eurogentec). Following incubation for 1h at 37°C, Polη was heat-inactivated for 20 min at 65°C, double-stranded polymerization products were denatured at 95°C for 10 min and purified on Nucleospin Gel Extract II column with NTC buffer (Macherey-Nagel). A reaction including a ^{32}P-dATP was run in parallel to check for the efficiency of the polymerase extension by electrophoresis on denaturing gels. Polymerization products that were fully elongated beyond the 5' end of oligo-G were amplified by PCR reaction with Phusion polymerase (New England Biolabs) (10 s 98°C, 30 s 55°C, 75 s 72°C, 25 cycles) using SP6 as a forward primer (ATTTAGGTGACACTATAG) and a reverse primer complementary to the 5' flanking region of oligo-G (GGTTAAAATAAAGACCTGGAG). 250-bp PCR products were gel-purified, cloned into ZeroBlunt cloning vector (Invitrogen) and sequenced with M13 reverse universal primer as described below.

Construction of mutation substrates

A 4.1 kb segment containing the DQ52-J_H locus and extending 510 bp downstream from the end of J_H4 was amplified from E14.1 ES cell DNA and cloned in the pCR-TOPO-XL vector (Invitrogen) and a ClaI site was inserted 41 bp downstream of the J_H4 splicing site by PCR-based mutagenesis. A 2.3 kb fragment adjacent in 3' was amplified similarly, before insertion of both fragments on either side of the floxed Neo^R cassette of the pLNTK vector. The following oligonucleotides: Oligo-G, CG(TATTATTAA)$_4$**G**TATTATTAA**G**TATTATTAA**G(**TATTATTAA)$_4$ and oligo-C, CG(TTAATAATA)$_4$**C**TTAATAATA**C**TTAATAATA**C(**TTAATAATA)$_4$ were annealed before insertion in the *ClaI* restriction site, generating the oligo-G or oligo-C constructs depending upon the orientation of the insertion, named in reference to the presence of G or C in the coding strand. Annealing of the oligo-noG, oligo-noC oligonucleotides (same sequence except for the absence of the three Gs or Cs) and cloning generated similarly the oligo-noG and oligo-noC constructs. Transfection

of E14.1 ES cells, screening of clones and Cre-mediated excision of neo^R gene was performed as described (PCR primers available upon request) (11). Injection of targeted ES clones into C57Bl/6 blastocysts was performed to generate the four mouse lines: Tg-G, Tg-C, Tg-noG, Tg-noC (Service d'Expérimentation Animale et de Transgénèse, Villejuif)

Mouse line and breeding

Screening of knock-in mice was performed using PCR primers flanking the loxP site for the 4 different lines: forward primer, TGGAAGGAGAGCTGTCTTAG and reverse primer, TGAGTACTTGAAAACCCTCTCAC (30 s 94°C, 30 s at 59°C, 90 s 72°C, 35 cycles with GoTaq polymerase, Promega). Knock-in lines were backcrossed with C57Bl/6 mice, and analyzed indifferently in heterozygous or homozygous configuration. Tg-G and Tg-C lines were bred with *Ung*$^{-/-}$ mice (kindly provided by Dr Deborah Barnes) to generate *Ung*$^{-/-}$ Tg-G and Tg-C lines. Pms2-deficient animals were kindly provided by Michael R. Liskay. Animal experiments were done in compliance to guidelines of the Institut National de la Santé et de la Recherche Médicale.

Immunization, cell sorting, genomic DNA extraction and sequence analysis.

For each genotype, 3 to 6 mice were sacrificed for either Peyer's patches or spleen cell isolation. Fourteen days prior to spleen isolation, mice were immunized by intra-peritoneal injection of $5x10^9$ sheep red blood cells (SRBC, Eurobio) resuspended in 500 µl PBS. Spleen and Peyer's patch cells were labeled with an anti-B220 antibody and PNA (phycoerythrin-conjugated anti-B220 (eBiosciences), PNA-FITC (Vector)) for sorting on FACS Aria I (Becton Dickinson). Sorted $B220^+PNA^{high}$ cells were heat-lysed at 95°C for 10 min and proteins were digested 30 min at 56°C with proteinase K (Roche). Three to four aliquots of 10^4 cells were used as a genomic DNA template for amplifying the rearranged VDJ locus and adjacent J_H4 intron. Five forward primers amplifying V_H1, V_H5, V_H3, V_H7 and V_H9 gene families, at a ratio of 6:3:1:1:1, respectively (12), and a reverse primer in the J_H4 intron (CTTGGATATTTGTCCCTGAGGGA) were used with Phusion polymerase (15 sec at 98°C, 30 sec at 64°C, 30 sec at 72°C, 40 cycles). 300 bp-long PCR products were gel purified, A-tailed with Taq polymerase (New England Biolabs) and ligated into pCR2.1 cloning vector (Original TA Cloning Kit, Invitrogen). Plasmid DNA was extracted by Millipore 96-well Miniprep system. Sequencing was carried out in ABI Prism 3130xl Genetic Analyzer. The diversity of VDJ junctions, mutations within 90/93 bp of oligonucleotide transgenes and 64 bp of J_H4 intron were analyzed with CodonCode Aligner software.

Results

An A/T rich mutation substrate inserted at the IgH locus

We designed an oligonucleotide to be used as a hypermutation substrate after its insertion within the heavy chain locus, which includes alternate A and T bases as well as 3 Gs or Cs located in its central part. Since Polη has a preferred misincorporation pattern opposite T bases, the A/T targeting is anticipated to reveal the strand bias of the patch of Polη-mediated DNA synthesis during hypermutation. This 93 base pair long oligonucleotide includes ten repeats of the 9-mer TATTATTAA motif, with three G intercalated within the central repeats ("oligo-G", Fig. 1). Each of these G is thus located within an AID RGYW (R: purine; Y: pyrimidine and W: A or T) mutation hotspot, AGTA. Moreover, since WA (A being the mutated base) is, on one strand or the other, the major Polη mutation hotspot at the Ig locus, each A or T of the oligonucleotide, including those flanking the Gs, are thus embedded in such a motif. The knock-in construct consisted in a 6.4 kb fragment of the mouse J_H4 genomic locus in which a *Cla*I site was generated by mutagenesis, 41 bp downstream of the J_H4 splicing site, a restriction site where the oligonucleotide was inserted. These constructs were used to generate knock-in ES cell line, in which the floxed neo^R gene present 500 bp downstream of the J_H4 intron border was excised. The recombinant mouse line obtained is named "transgene G" (Tg-G).

Three other mouse strains were generated: the "transgene C" (Tg-C) line in which the same oligonucleotide was inserted in reverse polarity, and the "transgene no-C", "transgene no-G" lines, in which the three C or G were absent from the same 90 nucleotide A/T backbone.

In vitro assay of Polη mutagenicity on an A/T rich substrate

Prior to the analysis of mutations generated in B cells harboring the various transgenes, the profile of mutations generated during Polη DNA synthesis was determined in an *in vitro* assay, upon copy of the oligo-G sequence inserted in a single-stranded M13mp18 vector. Since the mutation frequency induced by Polη exceeds by several orders of magnitude the one of Phusion DNA polymerase, we designed a PCR assay of the Polη synthesized products to assess the mutagenicity of the enzyme. This PCR was based on an SP6 sequence included in the DNA synthesis primer, and thus unique to the DNA produced in the reaction, and a primer complementary to complete products that extended in 3' within the M13 vector (Fig. 2A). PCR products of the correct size were gel purified and cloned for sequence analysis.

112 sequences were determined and 54% of them were mutated, harboring between 1 and 8 mutations, thus confirming the correct restriction of the PCR amplification to the newly synthesized product (Fig. 2B). Moreover, no mutations were obtained when Klenow was used instead of Polη as control (not shown). The 239 mutations collected displayed a mutation bias similar to the one described when Polη fidelity was assessed on the beta-galactosidase sequence in the blue/white

complementation assay, or on the Vkox light chain sequence (13, 14). It should be noted that in these two approaches, the mutation pattern was determined after transformation in *E. coli* of mismatch-containing M13 substrates, as opposed to the direct PCR strategy used here. The two most frequent misincorporations resulted in A to G followed by T to A mutations (read in the synthesized strand), in agreement with the preferred G misincorporation of Polη opposite T, and of A opposite A *in vitro*. An overall A/T ratio of 3 was observed, after correction for base composition (the oligo-G including 50 T and 40 A) (Fig. 2C, D). The unusual base composition of the oligonucleotide does not appear therefore to impact the misincorporation specificity of Polη.

The control transgene substrates reveal a DNA polymerase patch synthesis exceeding the oligonucleotide length

Mutation profiles were determined for the two G and C transgenes as well as for the two controls (Tg-noG, Tg-noC). These controls were included to define a core domain within the 90 bp oligonucleotide that would be spared by mutations initiated from either side.

The oligonucleotide transgenes were surprisingly highly mutated, which led us to restrict our analysis to Peyer's patch PNA^{high} B cells from young mice (2 months of age), or, if young mice were not available, to splenic PNA^{high} B cells isolated 14 days after immunization with SRBC (Table 1). Even so, transgenes accumulated up to 17 mutations, a complex situation since such mutations generated additional G and C bases, and thus new and uncontrolled AID targets. The insertion site, at the beginning of the J_H4 intron is indeed located within a highly mutated domain, downstream of the CDR3 region that is a major focus of hypermutation. As control, mutations were determined for each mouse line in 64 bp of flanking sequences (41 upstream and 23 downstream of the *ClaI* insertion site), and this region displayed an average mutation frequency approximately 2-fold lower than the adjacent oligonucleotide transgene for all constructions (Table 1). The high mutation frequency of the different transgenes does not appear therefore abnormal, considering the high density of Polη mutation hotspots they contained, and which might be efficiently targeted once the patch synthesis is initiated.

We divided the sequences analyzed in two categories, total sequences and sequences including 1-3 mutations, which could be considered as a sample resulting from a single (or a small number of) mutation hits (Table 2). Even within the 1-3 mutation sample, control transgenes harbored mutations throughout the 90 bp oligonucleotide, thus precluding the identification of a core region in which the impact of the three Cs or Gs could be assessed (Fig. 3). As discussed in the next chapter, the patch of DNA synthesis performed by Polη appears indeed to exceed the length of the 90 bp A/T substrate.

In vivo mutation profile of DNA polymerase eta.

These control transgenes provide nevertheless a unique opportunity to study the mutagenic properties of Polη *in vivo*, in a context of a balanced and symmetrical number of W<u>A</u> (<u>T</u>W) mutation hotspots that is not encountered in natural DNA templates. 572 mutations were collected for the Tg-

noG line from 8 different mice, with 319 mutations in sequences in the 1-3 mutation range. For the Tg-noC control, no germline transmission was obtained. Two founder mice with high chimerism were therefore studied (Table 2).

Mutation distribution was analyzed along the 90 bp of Tg-noC and Tg-noG templates, and results were grouped in three different segments, in order to represent mutation samples of sufficient size: the 5'-part corresponding to the first three repeats, the 3'-part corresponding to the last three, and the core region corresponding to the internal four repeats (i.e. a region including one repeat on each side of the central Cs and Gs in the Tg-C and Tg-G constructs, see Fig. 1). Mutation frequencies were corrected for the sequence length of each segment, as well as for the base composition of each transgene in case of mutations at As or Ts, the Tg-noC displaying, as mentioned above, 5 A and 4 T in each repeat, and the Tg-noG the reverse base composition.

In the case of Tg-noC, the distribution of mutations in the 1-3 mutation sequence sample was regular throughout the sequence length, whereas a clear decline was observed towards the 3' part of the Tg-noG template, with still a high mutation load up to its 3' end (Fig. 3). The trend was similar when all sequences were analyzed, thus indicating that it is not a difference due to the smaller mutation sample size collected for Tg-noC (not shown). This would therefore suggest that, in spite of their similarity, these A/T sequences present differences, possibly in the secondary structure adopted, which could impact elongation of Polη DNA synthesis.

Interestingly, a similar A/T mutation ratio was observed for both control transgenes in each of the three segments analyzed, indicating a constant strand bias over the 90-nucleotide substrate. The average A/T ratio on the complete sequence was quite high, around 4 for Tg-noC and Tg-noG, respectively (around 3 when all mutated sequences are considered) (Fig. 3 and 4B). This strand bias is approximately twice as high as the one observed at the Ig locus. The mutation pattern was accordingly biased towards A to G, A to T and A to C mutation with decreasing frequency, with A to G being the most prevalent mutation. Complementary mutations at T bases were lower by an approximate 3- to 4-fold ratio (Fig. 4A). This mutation pattern obviously differs from the one observed *in vitro*, a pattern according to which, since A to G predominates, one would expect T to A to be the second most frequent one. Nevertheless, the mutation pattern mirrors the one at the Ig locus *in vivo*, which shows a similar bias, albeit less pronounced, with prevalent mutations at A showing the same decreasing frequency from G, T to C, and a two-fold difference for complement mutations at T positions (from C, A and G).

W<u>A</u> has been described as the main Polη hotspot at the Ig locus *in vivo*. When corrected for its frequency within the oligonucleotide, T<u>A</u> mutates twice more frequently than A<u>A</u>, whether in the T<u>A</u>A or T<u>A</u>T motif (Fig. 5A). As reported by Spencer and Dunn-Walters, the tri-nucleotide context of the mutated base appears to impact the type of mutation generated (15). Our transgene provides only three tri-nucleotide contexts due to its peculiar sequence, A<u>A</u>T, T<u>A</u>A and T<u>A</u>T, when considering mutations at A bases (and the reverse A<u>T</u>T, T<u>T</u>A and A<u>T</u>A for mutations at Ts). A to T mutations

appear specifically disfavored in the TAA context, and the same holds true for the reverse complement motif TTA, an effect apparently contributed by both the 5' and 3' surrounding bases (Fig. 5B).

A reverse complementarity was thus strikingly observed for the overall A and T mutation pattern, including in terms of relative hotspot frequency and nucleotide context. This could be interpreted in two ways: first, Polη could be mutagenic when copying T only (or with several orders of magnitude in nucleotide incorporation infidelity), the 4-fold ratio observed corresponding to the relative proportion of synthesis of each DNA strand, i.e. 80% for the coding strand and 20% for the transcribed strand. Alternatively, Polη could be targeted only to coding strand synthesis, with a four-fold higher mutation frequency when copying Ts than when copying As. The mutation profile of Tg-noC (constant distribution of mutations and A/T ratio along the sequence) does not allow us to discriminate between these two propositions. In contrast, the mutation pattern of Tg-noG is only compatible with the second possibility: the distribution shows indeed a 5' to 3' gradient in mutation frequency while keeping the A/T ratio fairly constant in the whole 90-nucleotide template, thus precluding a contribution in the reverse direction that would markedly bias mutations towards Ts on the 3' side (Fig. 6A, B). This indicates accordingly that our transgenic substrate is targeted on the top strand only, with a patch of Polη synthesis that is initiated by a DNA break in the 5' flank of the oligonucleotide transgene and which can proceed over its total length and even extend beyond its 3' end.

A strand bias reversed by Ung deficiency: a role for the complete mismatch repair complex?

When mutation profiles of Tg-G and Tg-C are confronted with the one of their A/T-only backbone, no major differences are observed, either in terms of mutation distribution along the sequence, A/T ratio or mutation pattern. Whereas Gs and Cs are clearly targeted by AID, their capacity to trigger an Msh2-driven repair process appears thus minor compared to the contribution of sites external to the sequence.

The Tg-G and Tg-C lines were bred into an *Ung*$^{-/-}$ background, in order to more clearly isolate mutations contributed by the Msh2-Msh6 pathway. The most striking alteration was a diminution in the overall A/T bias, to 2.5 and 2.2 for Tg-G and Tg-C, respectively. This reduction was strikingly more pronounced in the 3' part of the oligonucleotide template (Fig. 3 and 4B). It thus appears that Ung deficiency promoted the alternate use of entry points 3' to the oligonucleotide transgene, triggering Polη synthesis in reverse orientation, over a short part only of the oligonucleotide sequence, the 5' segment conserving a high A/T ratio.

Uracil glycosylase can trigger error-free repair of uracils at the Ig locus, at a frequency that is still a debated issue, but which represents a fraction only of the deamination events leading to mutations at G/C bases. Such error-free processing involves the incision of the DNA chain, and has been proposed to provide an entry point for the Mhs2-Msh6 driven long patch DNA synthesis (1). This process, which requires the simultaneous formation of two nearby uracils, is more likely to take

place in regions highly targeted for mutation, and on the coding strand exposed by transcription. In its absence, the Pms2-Mlh1 endonuclease could provide nicks in DNA with a looser strand bias, typically associated with a non-canonical mismatch repair process that lacks directionality (see discussion). We think that our observations provide the first set of data comforting this proposition, even though they pertain to an artificial substrate. We indeed interpret the strand bias observed in our transgenes as being generated by a uracil glycosylase-induced incision in 5' of the oligonucleotide transgene, coincident with recognition by the Msh2-Msh6 complex of another deamination event downstream of it in this highly mutated region of the Ig locus, and promoting a long patch of Polη mutagenic synthesis. In the uracil glycosylase-deficient context, access to the non-coding strand would be provided by Mlh1-Pms2, thus re-equilibrating the A/T ratio towards values observed in Ig genes *in vivo* (Fig. 6D).

An additional difference observed in the Ung-deficient background concerns the targeting of the G and C bases. We evaluated the mutation frequency at these three bases compared to the rest of the transgene, corrected for their representation (3 G/C for 90 A/T, i.e. a ratio of 30). Gs were 3-fold more mutated than the neighboring A/T bases in both the wt and *Ung*$^{-/-}$ context, for sequences with 1-3 mutations. Cs were slightly more mutated than G in the *Ung*$^{-/-}$ background, confirming an efficient targeting of both DNA strands. But, in contrast, mutations at Cs were 5-times less frequent in wt than in *Ung*$^{-/-}$ animals (Fig. 4C). This can be interpreted again as an error-free repair coupled with U/G mismatch recognition, providing entry sites within the transgene and favored, like for external incisions, on the coding strand (Fig. 6C). Their overall number is however too modest to impact the global A/T mutation pattern, thus precluding the assessment that was initially intended in the design of the transgenes of a differential effect between the Tg-C and Tg-G configurations.

Increased strand bias of A/T mutations in Pms2-deficient background.

Pms2 or Mlh1 inactivation has been previously reported to have no discernable impact on somatic hypermutation, suggesting that the MutLa complex is not involved in triggering the MutSa-driven error-prone repair. However these data are either based on small mutation samples, or used tools for their analysis that minor the contribution of recurrent hotspot mutations (see discussion).

We therefore readdressed the question of a possible role of MutLa, by determining a large mutation database from both Pms2 and Ung-deficient backgrounds (Fig. 7). The most obvious difference between the wt and the *Pms2*$^{-/-}$ pattern was the large increase in the A over T mutation ratio, which reached an average value of 4 (Fig. 7C). No such bias was observed in *Ung*$^{-/-}$ mice. Like for the transgene in the Ung-deficient background, this would suggest that the MutLa complex is mainly producing the endonuclease incision required for initiation of Polη synthesis, an incision with a more moderate strand bias. In its absence, Ung would compensate for this function, with a heavier bias towards the non-transcribed strand in which multiple deamination events might be more likely to

occur. The transgene would only be different in that its "default" MutSa pathway would rely on breaks generated by Ung.

Discussion

We report here the analysis of oligonucleotide mutation substrates inserted by knock-in at the IgH locus, to study the strand bias characteristics of the MutSa-driven error-prone repair synthesis taking place during somatic hypermutation in germinal center B cells. This mechanism represents the only developmental process in which the outcome of DNA polymerase eta polymerization on undamaged DNA templates can be assessed.

These oligonucleotide substrates consisted in 10 repeated motifs of the TATTATTAA 9-mer, inserted in both orientations and including or not Gs and Cs in their central part. They were initially designed to determine the impact on the strand choice of the error-prone DNA synthesis process of AID targets embedded in a 90 bp backbone of alternate WA hypermutation hotspots, with substrates lacking such G/C internal entry points as controls. Surprisingly however, both G/C and control substrates were similarly targeted, indicating that the internal sites failed to add significantly to the mutations generated by the vigorous error-prone synthesis promoted by flanking sequences. The substrates were indeed inserted downstream of the J_H4 segment, i.e. in a genomic region flanking CDR3 and highly targeted for hypermutation, which may account for the unexpected inefficiency of the AID hotspots that were designed.

The control substrates allowed us to define the Polη misincorporation profile on targets offering a symmetrical distribution of WA and TW motifs. Mutations were four times more frequent at As than at Ts, with decreased frequency from A to G, A to T and A to C. Reverse complementarity was observed for the mutation pattern at Ts. This stands in marked contrast with the Polη error specificity determined *in vitro*, in which, if G misincorporation opposite T is the most frequent error, the second most frequent one is A opposite A (13, 14). We controlled that this difference was not contributed by the biased nucleotide composition of our transgene, by establishing its mutation profile when copied *in vitro* by Polη, and found the same types of errors as the ones reported in other *in vitro* studies. Polη contribution in hypermutation is obviously taking place in a different context, involving factors like ubiquitinated PCNA whose role is not reproduced in *in vitro* assays, and which may notably affect its elongation after a misincorporated base (8).

The whole 90 bp length of the A/T mutation substrates appeared targeted along a repair patch initiated in its flanking region. Interestingly, if both substrates devoid of G/C (Tg-noG and Tg-noC) behaved slightly differently, Tg-noG displayed a very informative pattern: mutation frequency decreased by two-fold between the 5' and 3' sides of the 90 bp substrate, while the A/T ratio remained constant, around 4. This can only be interpreted as the result of a synthesis initiated exclusively from the 5' flank, since a decreasing progression from 5' to 3' and a backward synthesis initiated from the 3'

side would obviously affect the A/T ratio according to the relative contribution of synthesis proceeding from each side (schematized in Fig. 6). We can therefore conclude that our control transgenes reveal the Polη misincorporation profile upon a synthesis only performed on the coding strand, a configuration possibly resulting from their specific location and specific base composition. This represents accordingly a surprisingly long patch of DNA, exceeding 90 bases, which can be copied by Polη from a single initiation event. The 2-fold A over T bias observed in vivo in Ig genes would thus result from an error-prone synthesis proceeding in an approximate 3 to 1 ratio on the coding and non-coding strand, respectively. Such ratio can either result from a similar frequency of initiation on each strand, or from a difference in the length of the patch produced (or both).

The second observation contributed by these mutation substrates pertains to the question of the still elusive activity that triggers DNA incision after recognition of the U:G mismatch by the MutSa complex. Jiricny and colleagues proposed that uracil glycosylase may provide such nicks, in a configuration in which two nearby uracils (tested within a 300 bp distance) are generated and processed, the 5' one providing an incision mediated by Ung that would allow exonucleolytic processing with 5' to 3' directionality, triggered by Msh2-Msh6 bound to the U:G mismatch in 3' (1). The relevance of this proposition can be questioned, since Ung-deficiency has limited impact on A/T mutagenesis (5). Moreover, if the processive action of AID has been described in the switch regions that harbor a high density of AID hotspots (16), one should mention that hypermutation around the V region is estimated around 10^{-3} per base pair and per cell cycle, making simultaneous deamination events unfrequent, and possibly restricted to the most highly targeted regions, along the transcription bubble that provides accessibility to AID.

Our transgenic substrates provide the first experimental support for this proposition. They indeed appear to be targeted for A/T mutagenesis exclusively from nicks introduced in the 5' flanks of the transgene and generating a four-fold A over T mutation ratio reflecting the performances of Polη. When backcrossed into an Ung-deficient background, these transgenes showed, in contrast, a reduced A/T mutation ratio, with a marked decrease towards their 3' end, reflecting a synthesis initiated from the 3' flanking sequence and proceeding over shorter distance in the reverse orientation. Our interpretation of these striking changes is through the involvement of the endonuclease activity of the MutLa complex, acting instead of Ung, and generating a less pronounced strand bias, as anticipated for a mismatch repair process acting outside replication and lacking the directionality provided by the discontinuities of the replication fork (17).

Like for Ung, Pms2 and Mlh1 deficiencies have been reported to have no impact on A/T mutagenesis (10). More recent data, using catalytic knockin mutants, have further addressed the specific contribution of the endonuclease activity of Pms2 and did not report any major impact of its inactivation on Ig gene hypermutation (18). We however think that the specific analysis performed by the authors may preclude subtle changes to be monitored. The first reason is the multiplicity of the templates analyzed, and the predominance of mutations collected in a functional, selected V_H

sequence. Second, and more importantly, mutations were analyzed using the SHMTool program, which only records a single mutation of each type at each nucleotide position (19). Such an analysis will obviously minor the contribution of recurrent changes observed at mutation hotspots.

A new mutation database from J_H4 intronic sequences from Pms2-deficient B cells was generated and revealed a marked A/T strand bias, with a ratio of 4, suggesting, like for the transgene, that DNA incision promoting A/T mutagenesis may be mainly provided by Ung and targeted to the coding strand. The absence of impact of single Pms2 or Ung deficiencies on the A/T mutation frequency would thus be explained by the reciprocal backup they exert on each other. Whereas the function of the MutLa complex may predominate at the natural Ig locus, Ung appears to be the *primum movens* in the transgene context, possibly due, as mentioned above, to its specific location and/or base composition.

Our transgenic substrates thus allowed a thorough description of the profile of Polη-mediated mutagenesis, showing a long patch size and a mutation profile distinct from the one described in *in vitro* assays. Interestingly, a Polη signature has been recently defined among cancer cell lines, predominantly in B-cell lymphomas, and it would be interesting to confront it with the profile described in this study (20). Together with the analysis of mutation in the Pms2-deficient background, our transgenes also allowed us to propose a mutual back-up function between Ung and MutLa in the DNA incision step promoting A/T mutagenesis, with distinct strand preference. Combined gene inactivations will obviously be required to establish whether other factors may contribute as well in the targeting of this error-prone repair process.

Acknowledgments

We thank the SEAT (Villejuif) for its expertise in generation of ES-derived mouse lines and mouse breeding. We thank Deborah Barnes and Michael Liskay for, respectively, providing *Ung*$^{-/-}$ and *Pms2*$^{-/-}$ mice. We thank Jérôme Mégret for cell sorting. INSERM U783 is supported by the Ligue contre le Cancer (équipe labellisée) and the Fondation Princesse Grace. Marija Zivojnovic was supported by a 3-year allocation from Ecole Normale Supérieure, and by the Ligue Nationale contre le Cancer for her fourth year of PhD.

REFERENCES

1 **Schanz S, Castor D, Fischer F, Jiricny J.** 2009. Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. Proc Natl Acad Sci U S A **106:**5593-5598

2 **Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T.** 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell **102:**553-563

3 **Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Forveille M, Dufourcq-Labelouse R, Gennery A, Tezcan I, Ersoy F, Kayserili H, Ugazio AG, Brousse N, Muramatsu M, Notarangelo LD, Kinoshita K, Honjo T, Fischer A, Durandy A.** 2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell **102:**565-575

4 **Rada C, Di Noia JM, Neuberger MS.** 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

5 **Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS.** 2002. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr Biol **12:**1748-1755

6 **Jansen JG, Langerak P, Tsaalbi-Shtylik A, van den Berk P, Jacobs H, de Wind N.** 2006. Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J Exp Med **203:**319-323

7 **Delbos F, Aoufouchi S, Faili A, Weill JC, Reynaud CA.** 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

8 **Langerak P, Nygren AO, Krijger PH, van den Berk PC, Jacobs H.** 2007. A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. J Exp Med **204:**1989-1998

9 **Weill JC, Reynaud CA.** 2008. DNA polymerases in adaptive immunity. Nat Rev Immunol **8:**302-312

10 **Chahwan R, Edelmann W, Scharff MD, Roa S.** 2011. Mismatch-mediated error prone repair at the immunoglobulin genes. Biomed Pharmacother **65:**529-536

11 **Torres R, Kühn R**. 1997. Laboratory Protocols for Conditional Gene Targeting. Oxford University Press, Oxford, UK

12 **Delbos F, De Smet A, Faili A, Aoufouchi S, Weill JC, Reynaud CA.** 2005. Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. J Exp Med **201:**1191-1196

13 **Matsuda T, Bebenek K, Masutani C, Rogozin IB, Hanaoka F, Kunkel TA.** 2001. Error rate and specificity of human and murine DNA polymerase eta. J Mol Biol **312:**335-346

14 **Pavlov YI, Rogozin IB, Galkin AP, Aksenova AY, Hanaoka F, Rada C, Kunkel TA.** 2002. Correlation of somatic hypermutation specificity and A-T base pair substitution errors by DNA polymerase eta during copying of a mouse immunoglobulin kappa light chain transgene. Proc Natl Acad Sci U S A **99:**9954-9959

15 **Spencer J, Dunn-Walters DK.** 2005. Hypermutation at A-T base pairs: the A nucleotide replacement spectrum is affected by adjacent nucleotides and there is no reverse complementarity of sequences flanking mutated A and T nucleotides. J Immunol **175:**5170-5177

16 **Xue K, Rada C, Neuberger MS.** 2006. The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice. J Exp Med **203:**2085-2094

17 **Pena-Diaz J, Bregenhorn S, Ghodgaonkar M, Follonier C, Artola-Boran M, Castor D, Lopes M, Sartori AA, Jiricny J.** 2012. Noncanonical mismatch repair as a source of genomic instability in human cells. Mol Cell **47:**669-680

18 **van Oers JM, Roa S, Werling U, Liu Y, Genschel J, Hou H, Jr., Sellers RS, Modrich P, Scharff MD, Edelmann W.** 2010. PMS2 endonuclease activity has distinct biological functions and is essential for genome maintenance. Proc Natl Acad Sci U S A **107:**13384-13389

19 **Maccarthy T, Roa S, Scharff MD, Bergman A.** 2009. SHMTool: a webserver for comparative analysis of somatic hypermutation datasets. DNA Repair (Amst) **8:**137-141

20 **Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinsk M, Jager N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, Lopez-Otin C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdes-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR.** 2013. Signatures of mutational processes in human cancer. Nature **500:**415-421

Figure legends

Figure 1. A knock-in strategy for the generation of A/T rich oligonucleotide mutation substrates.
Oligonucleotide mutation substrates were inserted by knock-in at the heavy chain locus, 41 bp downstream of the J_H4 segment. A schematized view of mutation distribution within 490 bp of the J_H4 intron is represented, to highlight the mutation density around the oligonucleotide insertion site. These oligonucleotides consist in 10 repeats of a 9-mer sequence, TATTATTAA, with four repeats flanking on either sides three Gs separated by the same 9-mer sequence (oligo-G). Oligo-C represents the same sequence inserted in reverse orientation. Oligo-noG and oligo-noC consist in the same A-T backbone from which Gs or Cs are omitted. The three segments of the oligonucleotide transgene used in subsequent mutation analysis are represented: 5' segment, core region, and 3' segment.

Figure 2. Pattern of Polη synthesis *in vitro* on an A/T-rich oligonucleotide substrate.
A) Scheme of the strategy used for amplification of *in vitro* Polη DNA synthesis products. Oligo-G was inserted in a single-stranded M13 template and subjected to Polη-mediated DNA synthesis. The primer sequence includes an SP6 sequence that allows the selective amplification of complete *in vitro* synthesized products. B) Total mutations generated during *in vitro* Polη synthesis of the oligo-G template. C) Pattern of mutations at A/T bases generated by Polη in vitro after correction for base composition and reported to the synthesized strand. D) Graphical representation of the mutation profile with indication of the dNTP misincorporated opposite the template base.

Figure 3. Distribution of mutation and A/T ratio along the oligonucleotide sequence.
Mutation frequency (corrected for relative nucleotide length and normalized to 100 for the 5' segment) and A/T ratio (corrected for relative A/T composition) is represented for sequences including 1-3 mutations and analyzed along the three segments of the oligonucleotide transgenes defined in Fig. 1: 5' part, core segment and 3' part.

Figure 4. Mutation profile at A/T and G/C bases for the different oligonucleotide substrates.
A) Distribution of mutations were represented for each transgenic substrate and analyzed in two categories: total sequences and sequences with 1-3 mutations over the 90 or 93 bp oligonucleotide template, and corrected for base composition B) Average A/T mutation ratio for the different transgenes C) Relative mutation frequency at Cs and Gs *vs.* the A/T backbone in the wt and Ung-deficient contexts (corrected for their 30-fold lower representation).

Figure 5. Symmetrical targeting of WA and TW motifs.
(A) Mutations were analyzed for each three-nucleotide context, with reference to the central mutated position. (B) Distribution of mutation outcome at the central nucleotide position is represented for

each three-nucleotide context. Analysis of mutations is restricted to Tg-noG for which a larger database was assembled.

Figure 6. A putative scheme of Ung/Pms2 strand targeting at the A/T oligonucleotide substrate.
A) A schematic impact on A/T mutagenesis of the two possible Polη synthesis profiles accounting for the observed A/T ratio of 4. Scenario 1: an 80/20 A/T mutation profile on top strand synthesis; scenario 2: a 100% A mutation profile with 80% of synthesis on the top strand and 20% on the bottom strand. Only the first proposition corresponds to the pattern observed in Tg-noG, the profile of mutations at Tg-noC (not represented) being compatible with both (see Fig. 3). B) Ung-triggered DNA incision at uracils (represented by orange scissors) is proposed to cooperate with U:G mismatch recognition by Msh2/Msh6 to promote initiation of Polη-mediated error-prone DNA synthesis at the nicks generated (1). Nearby deamination events are more likely to occur on the non-transcribed, exposed strand. C) Initiation of DNA synthesis at internal Cs of the Tg-C substrate triggered by Ung, accounting for the reduced C mutation frequency in the wt *vs*. the Ung-deficient background (Fig. 4C). D) Reduced A/T ratio in the *Ung*$^{-/-}$ context, due to a more moderate strand bias of the non-canonical mismatch repair involving the Pms2-Mlh1 endonuclease activity (blue scissors). The recruitment of the complete mismatch repair complex (MutSa and MutLa) is proposed to be the default pathway at the Ig locus, but Ung would take over the provision of DNA incisions in the specific context of the oligonucleotide substrate, or in a Pms2-deficient context (see discussion).

Figure 7. An A/T bias linked with Pms2 deficiency.
A) Analysis of mutations in J_H4 intronic sequences from wt, *Ung*$^{-/-}$ and *Pms2*$^{-/-}$ mice. B) Mutation profile of wt, *Ung*$^{-/-}$ and *Pms2*$^{-/-}$ mice. C) A over T mutation ratio. *P*-value = 0.012 (two-tailed Mann Whitney test).

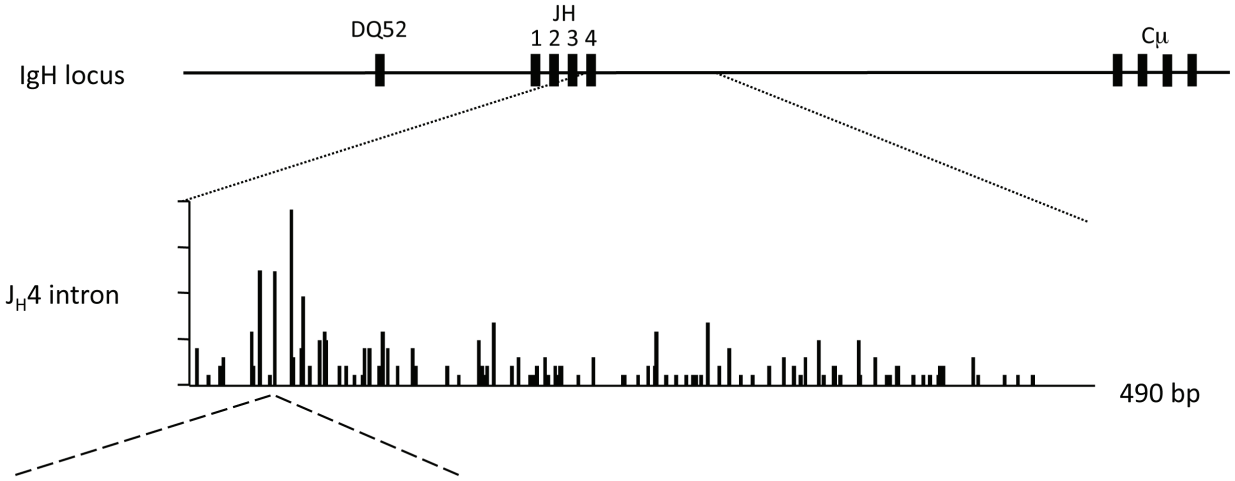
Table 1. Relative mutation frequencies in oligonucleotide transgenes and adjacent J_H4 sequences.

Transgenic mouse	Age (months)	Organ	J_H4 intron mutations Total number	J_H4 intron mutations Frequency (%)	Transgene mutations Total number	Transgene mutations Frequency (%)	Mutation ratio Transgene/ J_H4 intron
Tg-noG	2	IPP	18	0.55	110	1.91	
	2	IPP	43	0.50	92	2.42	
	5	Spleen	25	0.56	58	0.92	
	6	Spleen	31	0.54	127	1.60	
	6	Spleen	15	0.29	74	1.03	
	6	Spleen	18	0.34	64	0.81	
	6	Spleen	21	0.44	54	0.82	
Total			**171**		**579**		**2.41**
Tg-noC	5	Spleen	57	0.84	128	1.38	
	10	Spleen	24	0.64	74	1.43	
Total			**81**		**202**		**1.77**
Tg-G	2	IPP	N.A.	N.A.	20	0.58	
	2	IPP	N.A.	N.A.	43	1.26	
	2	Spleen	N.A.	N.A.	65	1.80	
	2	Spleen	N.A.	N.A.	35	1.06	
	5	Spleen	21	0.69	85*	2.18	
	5	Spleen	20	0.87	47*	1.53	
	5	Spleen	20	0.77	80*	2.32	
Total			**61**		375 (**212***)		**2.54**
Tg-C	2	IPP	63	1.68	250	5.07	
	2	IPP	12	0.32	40	0.84	
	2	Spleen	34	1.67	111	3.98	
	2	Spleen	25	0.75	102	2.38	
	5	Spleen	52	1.56	136	2.98	
	5	Spleen	25	0.69	95	3.10	
Total			**211**		**737**		**2.54**
Tg-G x Ung KO	2.5	IPP	18	0.74	72	2.04	
	2.5	IPP	131	2.56	333	4.49	
	2.5	IPP	125	2.38	422	5.56	
Total			**274**		**827**		**2.08**
Tg-C x Ung KO	2	IPP	12	0.24	98	1.38	
	2	IPP	67	1.56	183	2.91	
	2	IPP	21	0.61	54	1.09	
Total			**100**		**335**		**2.31**

Table 2. Mutation distribution in various transgenic substrates.

	Tg-noC / Tg-noG				Tg-G / Tg-C				Tg-G / Tg-C x Ung KO			
	Tg-noC all	Tg-noG all	Tg-noC 1-3 mut	Tg-noG 1-3 mut	Tg-C all	Tg-G all	Tg-C 1-3 mut	Tg-G 1-3 mut	Tg-C all	Tg-G all	Tg-C 1-3 mut	Tg-G 1-3 mut
A to G	68	185	34	110	239	109	55	48	93	171	45	26
A to T	57	148	30	77	177	96	58	53	62	155	16	29
A to C	34	84	17	48	118	46	33	21	39	107	16	18
Total A mutations	**159**	**417**	**81**	**235**	**534**	**251**	**146**	**122**	**194**	**433**	**77**	**73**
T to C	17	71	4	38	61	44	11	15	35	95	12	15
T to A	14	48	7	28	69	33	18	11	41	76	10	14
T to G	10	36	5	18	41	22	9	11	14	53	6	7
Total T mutations	**41**	**155**	**16**	**84**	**171**	**99**	**38**	**37**	**90**	**224**	**28**	**36**
C to T / G to A					11	8	4	6	32	49	20	12
C to A / G to T					5	5	2	8	0	0	0	0
C to G / G to C					13	9	0	6	0	2	0	1
Total C/G mutations					**29**	**22**	**6**	**20**	**32**	**51**	**20**	**13**
Total mutations*	**200**	**572**	**97**	**319**	**734**	**372**	**190**	**179**	**316**	**708**	**125**	**122**

* Difference with Table 1 concerning total mutation numbers correspond to the elimination of mutations from clonally related sequences.



Hypermutation substrates:

Oligo-G (93bp) $(TATTATTAA)_3$TATTATTAAGTATTATTAAGTATTATTAAGTATTATTAA$(TATTATTAA)_3$

Oligo-NoG (90bp): $(TATTATTAA)_3$TATTATTAA TATTATTAA TATTATTAA TATTATTAA$(TATTATTAA)_3$

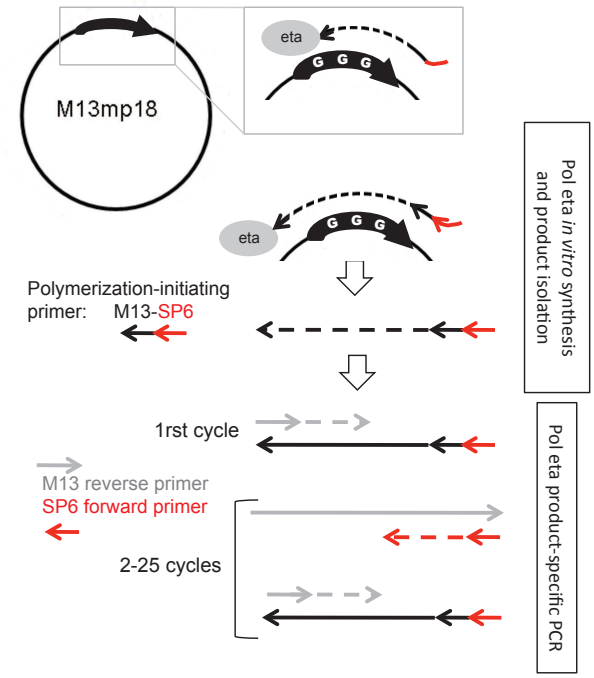
Oligo-C (93bp): $(TTAATAATA)_3$TTAATAATACTTAATAATACTTAATAATACTTAATAATA$(TTAATAATA)_3$

Oligo-NoC (90bp) : $(TTAATAATA)_3$TTAATAATA TTAATAATA TTAATAATA TTAATAATA$(TTAATAATA)_3$

5' segment core segment 3' segment

Figure 1

A



B

Number of sequences	112
Unmutated sequences (percent)	61
Total number of mutation	239
Number of deletions and insertions	11
Mutation frequency/total seq (per 100 bp)	2.40
Mutation frequency/mutated seq (per 100 bp)	4.41

C

From \ To	A	C	G	T	Total
A	-	1.2	70.6	3.2	75.0
T	17.5	1.0	6.5	-	25.0

D

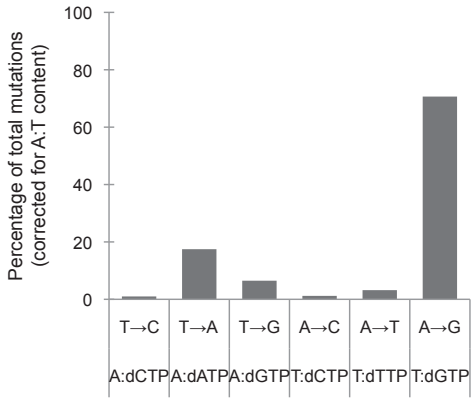


Figure 2

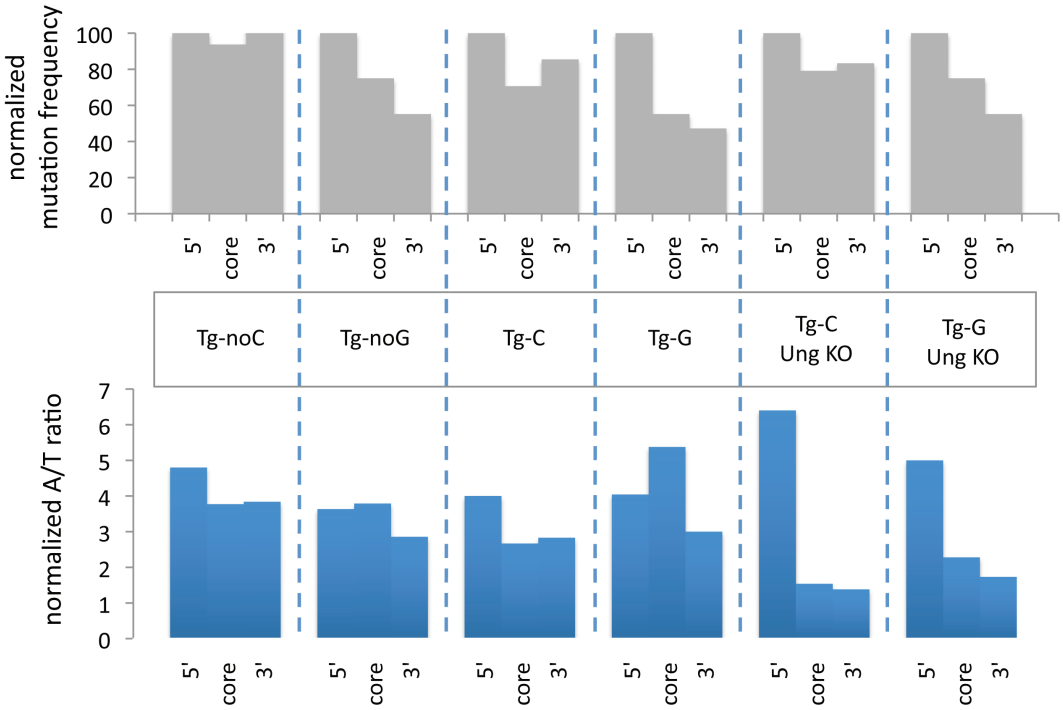

normalized mutation frequency
100
80
60
40
20
0
5'
core
3'
Tg-noC
Tg-noG
Tg-C
Tg-G
Tg-C Ung KO
Tg-G Ung KO
normalized A/T ratio
7
6
5
4
3
2
1
0

Figure 3

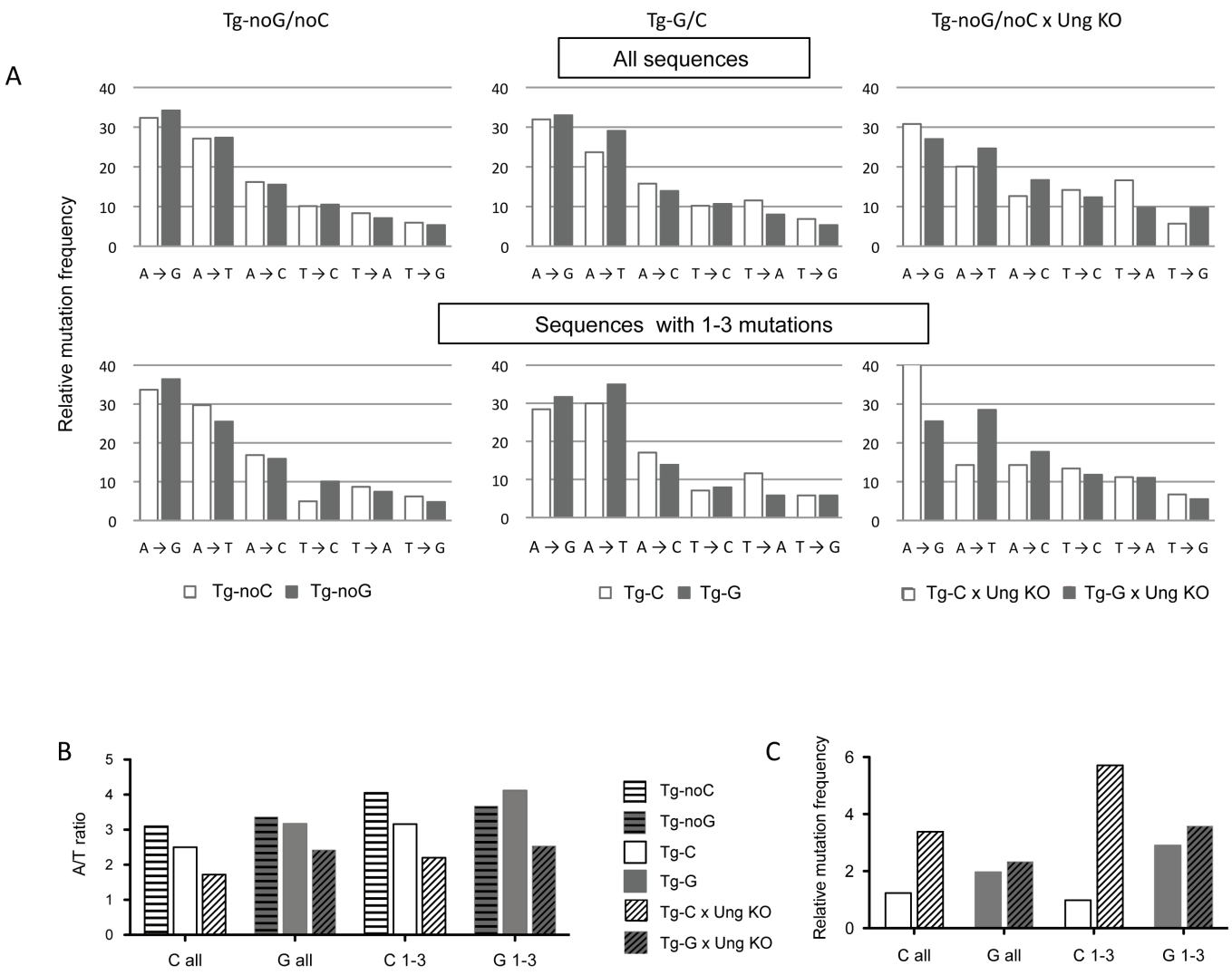
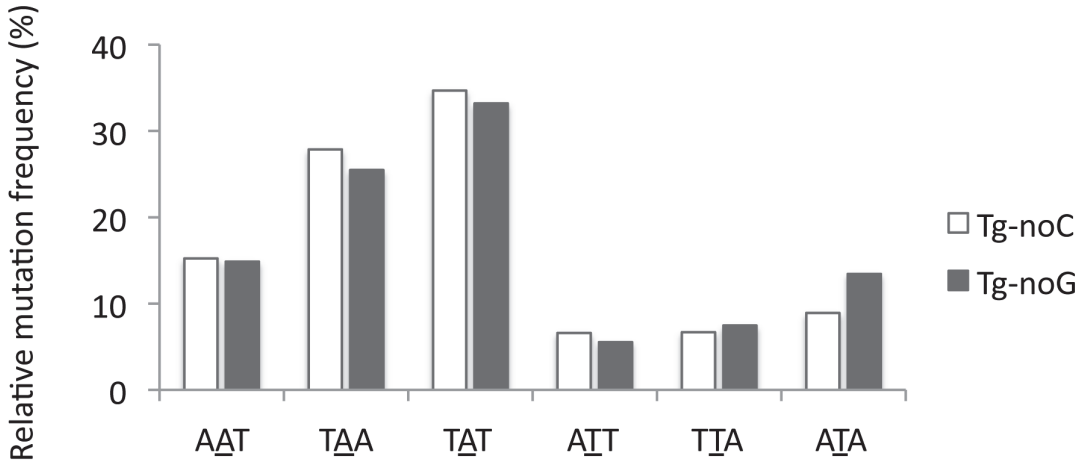

A
Tg-noG/noC
Tg-G/C
Tg-noG/noC x Ung KO
All sequences
Sequences with 1-3 mutations
Relative mutation frequency
A→G A→T A→C T→C T→A T→G
Tg-noC Tg-noG
Tg-C Tg-G
Tg-C x Ung KO Tg-G x Ung KO
B
A/T ratio
C all
G all
C 1-3
G 1-3
C
Relative mutation frequency

Figure 4

A


Relative mutation frequency (%)
40
30
20
10
0
AAT
TAA
TAT
ATT
TTA
ATA
Tg-noC
Tg-noG

B

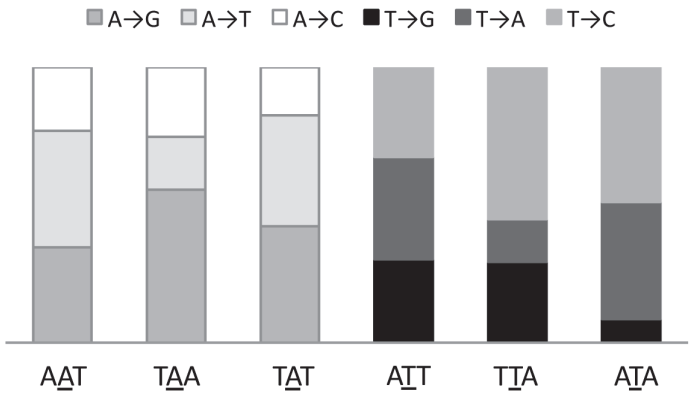

A→G
A→T
A→C
T→G
T→A
T→C
AAT
TAA
TAT
ATT
TTA
ATA

Figure 5

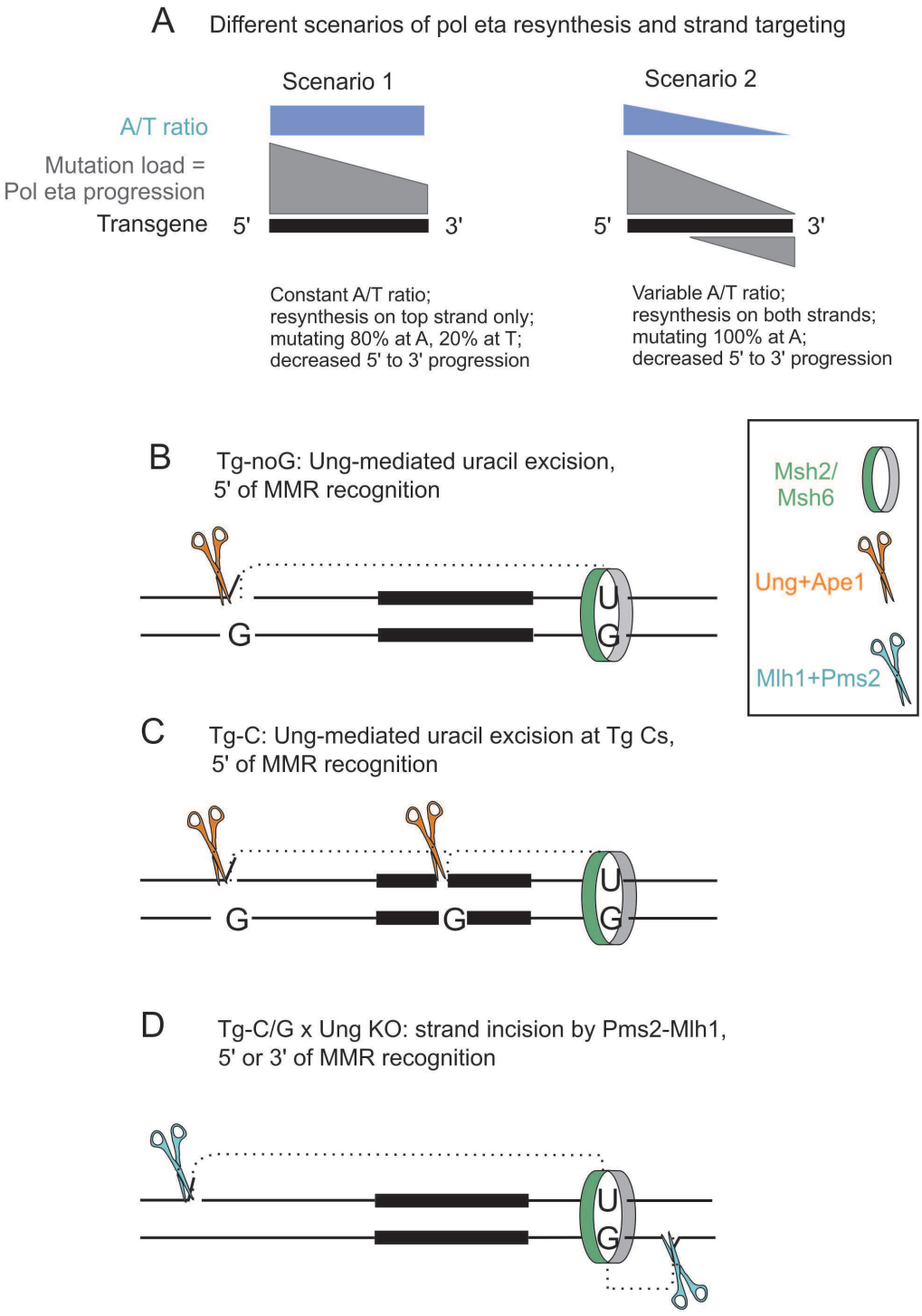

A
Different scenarios of pol eta resynthesis and strand targeting
Scenario 1
Scenario 2
A/T ratio
Mutation load =
Pol eta progression
Transgene
5'
3'
5'
3'
Constant A/T ratio;
resynthesis on top strand only;
mutating 80% at A, 20% at T;
decreased 5' to 3' progression
Variable A/T ratio;
resynthesis on both strands;
mutating 100% at A;
decreased 5' to 3' progression
B
Tg-noG: Ung-mediated uracil excision,
5' of MMR recognition
Msh2/
Msh6
Ung+Ape1
Mlh1+Pms2
U
G
G
C
Tg-C: Ung-mediated uracil excision at Tg Cs,
5' of MMR recognition
U
G
G
G
D
Tg-C/G x Ung KO: strand incision by Pms2-Mlh1,
5' or 3' of MMR recognition
U
G

Figure 6

A

	wt	Ung$^{-/-}$	Pms2$^{-/-}$
Number of mice	4	4	7
Number of sequences	217	323	574
Total length sequenced (bp)	96,131	143,089	254,282
Unmutated sequences (%)	21	38	35
Total number of mutations	1,219	997	1,712
(deletions and insertions)	16	8	15
Mutation frequency per total sequences (per 100 bp)	1.27	0.70	0.67
Mutation frequency per mutated sequences (per 100 bp)	1.60	1.13	1.03

B

	G/C : A/T	transitions : transversions	within G/C			within A/T		
			transitions	transversions		transitions	transversions	
			G to A C to T	G to T C to A	G to C C to G	A to G T to C	A to T T to A	A to C T to G
wt	41.7 : 58.3	51.9 : 46.1	60.1	15.6	24.3	49.9	30.3	19.8
Ung$^{-/-}$	53.1 : 46.9	71.0 : 29.0	92.0	4.3	3.7	47.1	27.7	25.2
Pms2$^{-/-}$	47.9 : 52.1	57.3 : 42.7	67.4	11.0	21.6	47.6	31.5	20.9

C

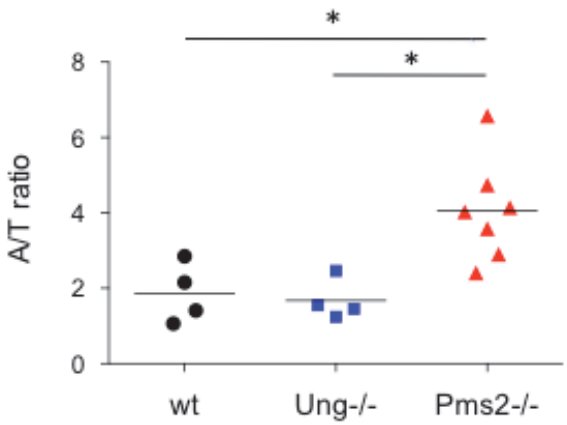


Figure 7

Discussion

PhD project I: To what extent does cell cycle govern different phases of somatic hypermutation?

Cell cycle restriction model : advantages, alternatives and drawbacks

Our "cell cycle restrictor" approach allowed us for the first time to confine the expression of factors directly implicated in somatic hypermutation (AID, Pol η) to defined periods in cell cycle (G1, S/G2/M) and to efficiently track the restriction of fluorescent fusion proteins by flow cytometry. To our knowledge, only Ugi protein, the highly specific inhibitor of uracil-N-glycosylase 2 was subjected to the cell cycle-driven degradation in order to observe the effects of phase-restricted inhibition of UNG on the somatic hypermutation (Sharbeen et al., 2012), a report published during the course of this study. A caveat of this indirect approach is the inherent inertia of a restricted Ugi system, potentially introducing a lag of undefined length between the onset of inhibitor degradation and the release of UNG from inhibition. Authors had to ensure as well whether the phase-restricted degradation of Ugi targeted the UNG trapped in the inhibitory complex with Ugi at the moment of degradation. The approach of directly fusing SHM key players to functional elements of timely degradation (degrons) alleviates these uncertainties.

However, the system of cell cycle-regulated degradation based on "Fucci" degrons is not perfect when it comes to restricting the degradation/accumulation pattern to the entire length of a precise phase. No known degron can confine the expression of a fusion partner in S or G2 phase only. A fusion with Cdt1:30/120 degron has a marked lag in accumulating in G1 phase, as demonstrated by the portion of transduced cells in early G1 phase not expressing the fluorescent fusion with this degron (our own observations, (Sakaue-Sawano et al., 2008; Sharbeen et al., 2012)). This is probably compensated in the normal cell cycle by the timed release of Cdt1 from the inhibitory complex of Geminin due to Geminin degradation at the end of mitosis, this interaction being absent from the "Fucci" system. Finally, the

overlap in function of APC/C and Skp2/SCF ubiquitin ligases in late G1/early S does not allow us to finely dissect any changes in mutation pattern that can be decisively introduced during this time frame. Finally, care must be taken not to overexpress a fluorescent fusion with a degrons (a common issue when using systems with retroviral gene expression...) since the degradation machinery can easily get saturated, causing "leakage" of fluorescent reporter to distant phases and general deregulation of phase-restricted expression. This may be the reason for observing G2/M $GFP^{+}mCherry^{+}$ cells in hCDT1:30/120-governed expression in different cell lines.

Not every protein can be the fusion partner of "Fucci" degrons, be it the fluorescent protein or the effector protein. Fluorescent protein is shown to be required as a "buffer zone" between the effector and the degron, since direct degron fusions with Ugi protein were ineffective (Sharbeen et al., 2012). The following fluorescent reporters were shown to function properly as cell cycle probes in fusion with Cdt1 degron: mCherry (reported here and (Abe et al., 2013)), mKO2 (Sakaue-Sawano et al., 2008), and GFP (Sharbeen et al., 2012) and with Geminin degron: mCherry (reported here), mAG and mKO2 (Sakaue-Sawano et al., 2008) and mVenus (Abe et al., 2013).

Other precautions must be taken into account when choosing the effector partner. Basic fusion of degron and fluorescent protein confers nuclear localization to the fluorescent fusion (Sakaue-Sawano et al., 2008); localization signal sequences of exclusively cytoplasmic proteins may interfere with the ones of the Fucci degrons. Interference of degradation-signaling motifs between the effector partner (e.g. UNG) and degron must be considered when searching the candidate lists. Furthermore, degrons will confer cell-cycle regulated degradation only if its N-terminus is fused to its partners: can the effector protein stand this strict fusion directionality? We have chosen the configuration where both fluorescent reporter and degrons are fused to the C-terminus of our effector proteins. This fusion design was approved for AID protein in many instances; furthermore, we introduced the flexible hinge linker domain between AID and the fluorescent marker, ensuring that AID C-terminal signal sequences are exposed to the AID partners and do not interfere with degron targeting. This might be the reason of our success in restricting AID expression, since similar attempts with AID of other groups failed (D. Schatz, personal communication).

On the contrary, Pol η is mainly fused *via* its N-terminus in many studies (with exception to *S.cerevisiae* RAD30 gene product which supports well the C-terminal fusions (Skoneczna et al., 2007)). We provided evidence that C-terminal fusion of Pol η without flexible linker did not impair its *in vitro* recruitment to UV-induced foci and can fairly carry out UV-lesion bypass in deficient cells, independently of the cell cycle context. Our systematic usage of mutant degrons as a control of the putative alteration of the function of the fusion proteins is extremely important in this context to discriminate between phenotypes

caused by cell-cycle regulation or by side-effects of the fluorescent protein and/or degron peptide.

In contrast with results in cell lines, Pol η C-terminal fusion with Geminin:1/110Δ motif showed compromised mutagenic activity in somatic hypermutation, although the recruitment to UV-foci proved positive. One interpretation of this failure is a hindered interaction with unidentified partner necessary to recruit Pol η downstream of non-canonical MMR, owing to the architecture of the fusion with Geminin:1/110Δ motif. Another one, confirmed by fluorescence expression levels, is a much lower expression of Pol η in lymphocytes derived from transduced HSC than in the fibroblasts in which its complementation capacity was tested.

Other cell cycle control systems can be applied to fluorescent markers in order to create cell cycle sensors, and many of them are in their turn theoretical candidates for cell cycle "restrictors" of effector proteins. Different promoters govern cell cycle-dependent expressions of proteins known to participate in limited periods, e.g. human S-complex cyclin A promoter that induces expression at the G1/S transition or promoters containing FoxM1 binding sites that govern the expression during G2/M (Henderson et al., 2013). However, the choice of time frames for restriction by cell cycle-dependent promoters is quite finite; moreover, this form of cycle-regulated expression is known for its inherent inertia and difficult-to-modulate dynamics.

Other peptides can confer phase-specific localization or pattern to the fusion partner, e.g. if PSLD peptide (*p*hosphorylated *s*ubcellular *l*ocalization *d*omain), C-terminal portion of human DNA helicase B (HDHB) is fused to GFP, it concentrates green fluorescence in nucleus in G1 and swiftly translocates it to cytoplasm in post-G1 period (Gu et al., 2004). As Pol eta can only synthesize DNA when it is present in the nucleus, its activity would be restricted to the G1 phase when fused to EGFP-PSLD. The theoretical advantage of this shuttling system is the permanent stability of the protein throughout the cell cycle, which means potentially more molecules immediately available at the beginning of the G1 phase and a higher accumulation throughout G1 phase. However, it remains to be proven that some degradation of EGFP-PSLD does not accompany its nuclear export.

Other proteins can confer focal organization to the fusion partner in different intervals of S/G2 phases (PCNA: early and mid S-phase, Dnmt1: mid S to late G2, DNA ligase 1: mid S-phase), while they adopt diffuse pattern in cytoplasm outside of these intervals (Kurzawa and Morris, 2010). This form of local activity restriction dependent on the cell cycle phase is applicable to Pol η, while it might severely interfere with AID shuttling necessary for deamination in SHM. Moreover, assessing cell cycle-regulated restriction of activity by microscopy is less straightforward than by flow cytometry.

Lastly, other degrons exist that can govern cell cycle-dependent expression and they are integral part of proteins normally degraded in a specific phase: last 491-527 amino-acids

confer S/G2/M-specific degradation to RAG2 protein, while destruction box (D-box) motif of M complex cyclin B confines its expression from late G2 to anaphase (Chapter III.2). While both degrons worked well within fusion with Ugi-GFP, our own experiments with AID-GFP-RAG2:491/527aa fusion did not provide satisfactory G1-restricted expression of AID.

Aspects of the bone marrow-transplanted retrogenic mice model in SHM: usefulness and downsides

To date, germinal center reaction in not reproduced *in vitro*; the only insight is gained through mouse, human and chicken B-cell lines immortalized or transformed at different stages of B-cell activation. These B-cell lines either constitutively perpetuate somatic hypermutation at low rate (human Ramos, mouse 18-81, chicken DT40) or have inducible AID expression (human Burkitt lymphoma BL2, mouse A20, mouse CH12F3). However, the pattern of *in vivo* somatic hypermutation is not fully reproduced in these cell lines, most of them having intensified mutation bias at G:C over A:T basepairs. Unbalanced mutagenesis may reflect a downregulation of A:T mutagenic factors due to the transformed or immortalized status of the activated B-cell in spite of a proficient MMR pathway in many of these cell lines.

Another plausible model consists in culturing *ex vivo* naive B-cells and triggering their activation. Although these cells readily switch in culture when supplemented with the right cytokine combination and are easy to transduce, they accumulate mutations only in S-regions and spare V(D)J region, although AID is confirmed to be properly induced in these cells. This differential targeting could reflect the huge amount of AID hotspots in S regions relative to that of the variable region of Ig genes; however, this phenomenon cannot be accounted for by this simple explanation. Although they upregulate AID expression upon stimulation, e*x vivo* mouse B cells can not mimic real germinal center B cells, as they for example fail to induce Bcl-6 expression (the master gene of the germinal center transcriptional program); they may therefore be short of unknown factor(s) required for proper targeting of AID to V(D)J regions.

Thus, mice models remain the only reliable playground for studying somatic hypermutation and reporting effects that are closest to the physiological situation. Several *in vivo* models were available at the time we started this study. The most obvious and most frequently practiced in the lab is the knock-in; it guarantees the physiological expression of Pol η and AID-fusions from the endogenous promoter and surrounding regulatory elements. However, generating 8 knock-in lines simultaneously (AID-G1, Pol η-G1, AID-S/G2/M, Pol η-S/G2/M and respective mutants) was somewhat daunting, ever for our expertise in knock-in/knockout mice generation... So we reached out for alternative mice models that will allow us to express our manipulated SHM factors more quickly but reliably:

Adoptive transfer of transduced HSC in irradiated immunocompromised host. This first method of choice is well-known for its success in another Necker team (Lagresle-Peyrou et al., 2006; Yates et al., 2002). A thing to keep in mind is that this model worked well for protein factors that conferred selective advantage to transduced cells in the host environment and served as a proof-of-concept for gene therapy approach. This was certainly not the case for highly mutagenic AID or low-fidelity polymerase η, which could in contrast contribute to the counter-selection of cells overexpressing them. In order to transduce hematopoietic stem cells with a B2-class insert in a P2 facility (the only one available at Necker/Broussais), we had to use retroviral vectors suited for expression in murine stem cells, such as derivatives of MSCV virus (pMIG, pMigR1) (Haas et al., 2003; Van Parijs et al., 1999; Pear et al., 1998; Robbins et al., 1998). Retroviral transduction implicated inducing HSC to proliferate while minimizing the odds for engaging into further differentiation steps. In order to avoid more perturbations, post-transduction HSC were not selected or sorted; they were injected with the total isolated Lin- population, allowing the pool of short-term progenitor cells in this subset to exert a radioprotective function during the establishment of long-term blood cell lineages (Olivier Albagli-Curiel, personal communication). Another prerequisite for successful reconstitution of adaptive immune network of cells in immunocompromised hosts is the minimum number of total Lin- cells injected, irrespective of the transduction rate: this threshold number ensured the survival of transduced progenitors which could in turn give rise to a large number of founder B-cells participating in an anti-SRBC immune response.

By retroviral gene expression in adoptive germinal center B-cell from SWHEL Rag1-/-donor. SW*HEL Rag1*$^{-/-}$donor provides for B-cells which express the same BCR against HEL-antigen due to knock-in of HEL-specific rearranged BCR heavy chain in IgH locus and stable ectopic expression of the low-copy HEL BCR light chain (Phan et al., 2005; Sharbeen et al., 2010). SW*HEL Rag1*$^{-/-}$ splenic B-cells are transduced *ex vivo* and delivered into HEL antigen-primed receiver simultaneously with a novel HEL immunization. Only 6 days post-transfer, GC B-cells differentiated from these transduced cells can be isolated owing to fluorescent HEL binding and expression of the fluorescence marker. These cells undergo class switch recombination and accumulate mutations within the physiological context of cognate antigen encounter and GC reaction. Thus, although this system is extremely faster and directs expression only in the B cell lineage, its application to the type of questions that we asked remains complex: since our donor cells had to be either *Aicda*$^{-/-}$ or *Polh*$^{-/-}$, we should have first established mouse strains bearing at least 6 different mutant alleles (i.e. *Aicda*$^{-/-}$*SW HEL IgH*$^{KI/wt}$*IgL*^{tg}*Rag1*$^{-/-}$*).*

Adoptive transfer of ex vivo transduced fetal liver pre-B cells. Authors of (Rolink et al., 2004) studied transferred these B progenitors into Rag-deficient mice along with normal T-cells in order to restore the fully effective adaptive immune response. This rather complicated procedure can be simplified if the transduced pre-B cells are transferred to μMT

mice, which have fully developed T-cell subsets but lack the entire B-cell lineage (Kitamura et al., 1991). Consequently, expression of AID and Pol η fusions would be strictly limited to the B-cell lineage, which is a great advantage compared to transduced bone marrow transplantation. Moreover, isolated pre-B cells can be transduced far more efficiently (and effortlessly...) than hematopoietic stem cell and are less sensitive to manipulation (can be frozen/thawed, kept for longer times in culture and subsequently selected). Unfortunately, our attempts indicated that, in our hands, this system works occasionally, but not reproducibly for reasons that have not been identified so far [Sebastien Storck and Amélie Julé, unpublished results].

After checking all available options, we set up and successfully carried out HSC transductions with all 8 retroviral constructs, while transplantation worked well for 6 transduction groups (exceptions being Pol η-^{G1} and Pol η-$^{G1\text{-}Mut}$). This procedure can be completed in about 3 months for a mouse cohort reconstituted by one retroviral vector, in contrast to 1 year for the knock-in. Interestingly, this system can be easily applied to knock-down studies using retroviral vectors coding shRNA against a given target.

In order to transduce cells like HSC that are known for being refractory to retroviral infections, we used the retronectin technology to boost the AID or Pol η fusions delivery. This approach, although costly and time-consuming (4 hours for virus attachment, 4 hours for transduction), allowed us to attain 50% of the HSC transduction rate in a single infection cycle. Recently, a new technique was reported using the HIV peptide nanofibrils to bring into proximity a target cell and the viral particle (Yolamanova et al., 2013) in a 5 minute-long procedure, reproducing the efficiency of retronectin-mediated transductions.

However, these highly efficient gene delivery methods were a double-sword in our study: higher numbers of transduced HSC would allow us to sort more GFP^{+} GC B-cells (which was the limiting step for the subsequent PCR amplification) if we managed to avoid high multiplicity of infection due to multiple integration sites and LTR promoter-driven overexpression (hampering the successful phase-restricted expression or even being toxic for the cell). This mouse model was already used to ectopically express different mutated forms of AID, leading to B and T cell tumors (Komeno et al., 2010). It is therefore essential for this procedure to find appropriate transduction conditions that provide the best balance between the percentage of transduced cells and the level of expression.

Additionally, one must take into account the regulatory systems specific for the effector partner and the particular cell type that may impact the overall level of the retrovirally expressed fusion in the cell. This argument may account for the *in vivo* inefficiency of Pol η-$^{S/G2/M\text{-}Mut}$ in B-cells, while high levels of this fusion in the MEF culture were sufficient to exert classical Pol η activity in UV-repair.

Different modes of mutation profile analysis

Amplification of hypermutating regions by PCR became the dominant approach in the analysis of the diversification by somatic hypermutation. However, during the bulk genomic DNA amplification of small GC B-cell sample, one must anticipate the artifacts linked to this technique due to elevated number of amplification cycles, thermostable polymerase misincorporation (minor but not to be ignored if overall mutation frequency is low) and crossovers-like events among related PCR products (Honjo and Alt, 1995). All these events lead to increased extent of clonal relatedness in mutating pools based on the small number of founder cells (defined as a GC B-cell with unique V(D)J junction). A mandatory step in our analyses is purging doubled sequences, followed by the manual rectification of dynastically related ones: if two or more sequences with the identical junction contain different mutations in J_H4 intron, shared mutations are counted only once while independently acquired mutations are scored separately.

This analysis mode allows scoring for many shared mutations that arise as the consequence of mutating a confirmed RGYW or WA hotspot in J_H4 intron *in vivo*, a parameter which is largely neglected by traditional analysis with SHM tool software (MacCarthy et al., 2009). This tool scores as one all mutations of the same type found at an individual nucleotide position, attenuating the accumulating effect of hotspots and underestimating the mutational load of clonally related datasets (Rada et al., 2004).

In a more stringent analysis mode, we applied another criterion concerning clonality of mutations: if two or more sequences have unrelated junctions but share two of more identical mutations, these mutations are considered as PCR crossovers and are counted only one per position. Most importantly, all three analysis modes gave comparable results when applied to mutation profiles of GFP^+ GC B-cells from reconstituted hosts, notably concerning the A/T over G/C ratio of mutations.

New SHM model from combined data

Data from diverse studies associated UNG activity to the S phase in cell cycle, revealing S-phase expression peak or colocalization with nuclear replication foci in S phase cells (Chapter III.3). UNG was found to interact with PCNA *via* its PIP domain (Moldovan et al., 2007; Otterlei et al., 1999), which can be easily explained by the need to have UNG somewhere near the replication fork in case of dUTP misincorporation. More intriguingly, UNG association with RPA protein is greatly enhanced upon UNG phosphorylation away from the residues in phosphodegron, while the same UNG phosphorform showed increased catalytic turnover on the single-stranded uracil *in vitro* (Hagen et al., 2008). While waiting for the validation in other *in vitro* and *in vivo* models, this last piece of puzzle fits well multi-

facetted action of UNG in somatic hypermutation: an activated UNG phosphoform could preferentially attack uracils in single-stranded DNA prior to the passage of replication fork and give rise to mutagenic translesion, justifying its S phase-associated action in our cell cycle hypothesis.

However, many uncertainties linger about UNG-expression in GC B-cells: apart from being upregulated 6 to 20 times upon activation (Di Noia et al., 2006), not much is known about the timing or outcomes of pre-replicative, mid-S phase or post-replicative UNG-mediated recognition of uracils. The last question is indirectly tackled by restricting expression of the UNG inhibitor Ugi to the different parts of the B-cell cycle by fusing it to different degrons and expressing it in adoptive GC SW*HEL Rag1*$^{-/-}$ B-cell model (Sharbeen et al., 2012). All outcomes of UNG intervention in SHM were inhibited if the UNG inhibitor was restricted to G1 phase. No U:G handling, error-prone or error-free, was observed in other phases.

This study predicted in a way the crash of our "SHM-per-cell cycle phase" hypothesis, since AID expressed in S/G2/M yielded almost no mutation. This absence of mutations in S/G2/M in our restriction model can be accounted for in several ways:

- by inefficiency of AID$^{S/G2/M}$ fusion to mutate the Ig locus. Following arguments are against this proposition: firstly, AID fusion with the mutated counterpart successfully induced mutations, both *in vivo* and in NTZ 3T3 cell line (not shown). Secondly, constraining AID to S/G2/M period of NTZ 3T3 cell cycle also yielded a fair number of mutations (0.2% total mutation frequency) and GFP-reversion (up to 0.77% of living cells). As in the case of Pol η$^{S/G2/M\text{-}Mut}$, this *in vitro* efficiency of AID$^{S/G2/M}$ fusion might be due to overexpression in the cell line. In order to test AID$^{S/G2/M}$ efficiency *in vivo*, one would have to express this fusion against *Aicda*$^{-/-}$ *Ung*$^{-/-}$ *Msh2* $^{-/-}$ background, a mouse line whose breeding is underway.
- by the impossibility of S/G2/M-restricted AID to access the Ig locus and perform deaminations: it is plausible that AID may be sequestered away from the Ig locus or even from the nucleus by cytoplasmic retention during this period of the cell cycle. Other possibility is that S/G2/M passage triggers increased AID instability in the nucleus, as already showed for the DT40 chicken cell line.
- by a novel hypothesis that places all SHM mutagenic actors downstream of AID in a common time frame, starting from the beginning of the G1 phase to the late G1/early S phase (Figure D32). This hypothesis assumes another important point, yet to be proved: due to its high transcription rate associated with somatic hypermutation, the Ig locus is susceptible to be early replicating. These premises might explain why most uracils generated in S/G2/M were not further diversified or not even replicated: all events contributing to mutagenesis had already taken place by the time AID$^{S/G2/M}$ arrived on the scene. AID$^{S/G2/M}$-generated uracils could be repaired error-free throughout

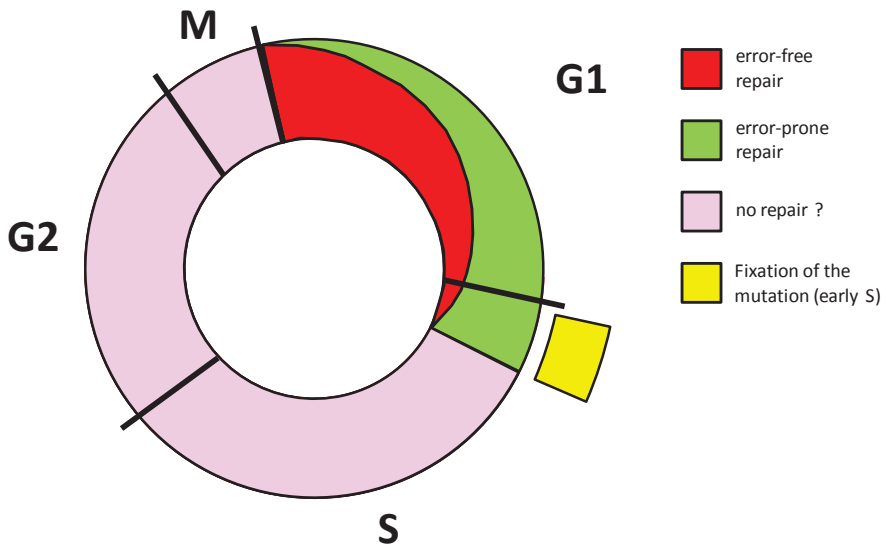


Figure D1. New somatic hypermutation model based on combined data from our study and (Sharbeen, 2012). More details in the text.

S/G2/M; however, this options is contradicted by data from (Sharbeen et al., 2012) in which S/G2/M-restricted inhibition of UNG did not contribute to an increased mutation load in HEL Ig genes. Error-free repair of uracils in S/G2/M can nevertheless be compatible with the observation of (Sharbeen et al., 2012), if one assumes, in their configuration, that S/G2/M-introduced uracils will persist unrepaired up to the next G1 phase where UNG within BER will repair them error-free.

The coexistence of these three different outcomes within the same time frame (repair i.e. U→C reversion/G:C transversions/G:C transitions) having UNG as a denominator can be simply explained by lapses between following decisive events: AID deamination, uracil recognition and abasic site creation (Figure D1). If uracil is introduced early in the G1 phase, its most certain fate is repair by UNG and downstream BER components. The later in G1 it is introduced, the bigger the probability that UNG will manage to generate an abasic site thereafter occupied by early S phase progressing replication fork. This late G1/early S phase configuration of abasic site prevents the recruitment of other BER factors, but promotes involvement of translesional polymerases and generates G:C transversions (and transitions). However, a portion of late G1 uracils ignored by UNG can also persist into early S phase and be replicated, fixing G:C transitions. If events such as mid- or late S-phase AID deaminations exist, these uracils will remain in double-stranded configuration of freshly replicated Ig locus and await next G1 phase until they are handled error-free, or be readily repaired when occurring at a stage of strong Ung activity. Our observation that expression of AID in S/G2/M does not sustain SHM is in accordance with this model.

The contribution of our study to this model is positioning the origins of A:T mutagenesis in the G1 phase as well, since AID-mediated deaminations in G1/early S phase were able to induce mutations at both A:T and G:C basepairs (Figure D1). Even though our approach did not directly imply the involvement of downstream MMR factors in generating A:T mutations in G1, other groups already tracked the activity of non-canonical MMR outside of its conventional post-replication scope and possibly in G1 and early S (Peña-Diaz et al., 2012; Zlatanou et al., 2011). However, results of our study increase the stakes for G1-restricted Pol η: will the G1 restriction allows Pol η to restore the 50% of A:T mutations in transduced Polη-deficient B-cells? An improved reconstitution of new *Rag2*$^{-/-}$ mice cohort with Pol η-$^{G1\text{-}Mut}$ HSC is underway and will allow us to refine and potentially re-apply our G1-restricted approach to Pol η as well.

PhD project II: Does the uracil orientation govern the Pol η synthesis in SHM, or is it directed by the single-stranded nick position?

Unfortunately, we realized that the initial question that animated the design of the A/T-rich hypermutation substrate knock-in was not to be answered by this study: control A/T substrates that lacked core G/C entry sites for AID were heavily mutated and reproduced the mutation pattern of G/C-containing A/T-rich substrates. Recruitment of the error-prone repair to the flanking sequences of highly targeted J_H4 intron portion seemed to literally submerge any positional effect of core cytosines on the mutation pattern even in the central parts of the substrate considered "more protected". However, our control hypermutating substrates combined with their highly targeted proximal environment turned out to be an extremely useful model in the following aspects:

- directly assessing polymerase η activity *in vivo* on undamaged DNA template with controlled sequence context (symmetrical distribution of WA and TW motifs of knock-in substrate) and in ideal conditions of unaltered SHM targeting to Ig locus;
- estimating the length of gap-filling activity of Pol η *in vivo*: our control substrates are targeted exclusively on the top strand and extended over the whole length (90 bp) from the patch initiated within the 5' flank. This specific targeting pattern was deduced from the distribution of A/T mutation ratio in the body of the TgNoG substrate and probably results from the specific location and base content of control substrates;
- providing the first *in vivo* simulation of scenario described in (Schanz et al., 2009) where UNG and mismatch repair collaborate in somatic hypermutation with the latter taking advantage the introduced single stranded break, initiated by the former;
- tweaking the exaggerated A:T mutation bias at the substrate sequence owing to preferential guidance of intrinsically unbiased Msh2/Msh6 recruiter complex to the top strand. On Ung-deficient background, this bias is reversed to the physiological value, candidating UNG as the primary actor of strand directionality associated with A:T mutagenesis in hypermutating substrates.

Some groups consider the absence of MutLα complex as a necessity for error-prone MMR to take place (Chahwan et al., 2011). However, EXO-1 entry in error-prone MMR can be provided by the very same MutLα complex if PMS2 cryptic endonuclease activity is induced (Kadyrov et al., 2006). Although the PMS2 mutant form lacking endonuclease activity did not seem to alter SHM (Oers et al., 2010), the effect of PMS2 absence in that case may be compensated by other "nicking" scenarios, and A:T bias of hypermutating substrates in Ung-deficient knock-in mice strongly suggest Ung as the next one in the waiting list

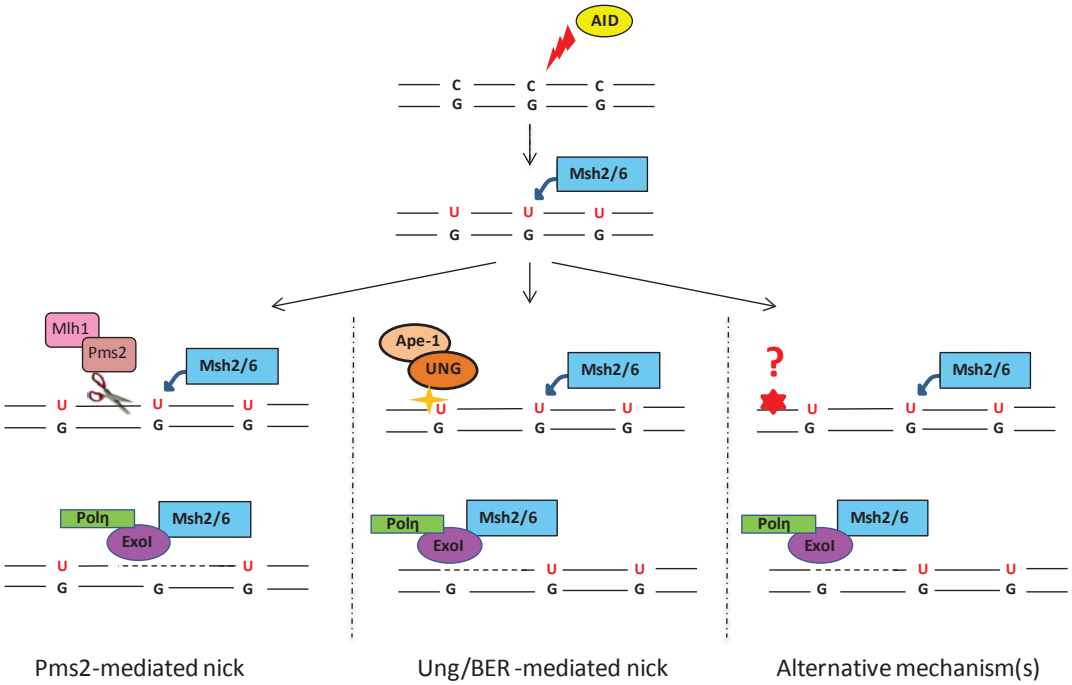


Figure D2. The mechanistic model combining different nicking scenarios inferred from the analysis of AT-rich hypermutating substrates in Ung-proficient and Ung-deficient background.

(Figure D33). Indeed, the marked A:T bias of hypermutating control substrates closely resembles the one of *Pms2*$^{-/-}$ mice. Further crossing of individual genetic knock-outs could help us to discern the exact hierarchy of nick generation mechanism and narrow down the candidates for a third, alternative mechanism, possibly mobilizing additional glycosylases. Our transduction strategy appears as an ideal tool to interrogate these additional partners by shRNA inhibition in *Ung*$^{-/-}$*Pms2*$^{-/-}$ HSC.

We might think at some point that we had identified all the key players of somatic hypermutation on the basis of our knowledge from phenotypes of knock-in or knock-out mice. Nevertheless, one should always be cautious when bringing up such reductionist conclusions because as many other cellular processes, somatic hypermutation will have an ace up its sleeve if the physiological pipeline is altered: as soon as one predominantly recruited effector is removed, there always seems to be another one backing it up. In a B-cell world where the competition is so harsh that no place for jammed molecular machineries, redundancy will confer selective advantage. It is therefore not striking to observe no marked phenotype in somatic hypermutation for *Pms2*$^{-/-}$ mice if UNG is capable to complement the requirements of mutagenic mismatch repair. This seemingly "non-canonical" mismatch repair which is thought to rely solely on MSH2-MSH6 complex according to the knock-out studies is finally not so different from the conventional one in terms of recruiting MutLα and reaching out for a free 3' end.

In conclusion, we are not yet done with short-listing for somatic hypermutation factors. Moreover, the greater challenges lie ahead as identifying interactions between them calls out for more wit and imagination in devising the new integrative models compared with the small puzzle pieces that the classical methods have to offer.

ANNEXE 1: ABBREVIATIONS

AP site	abasic site
AID	activation-induced cytidine deaminase
APC/C	anaphase-promoting complex or cyclosome
APC	antigen presenting cell
BAFF	B-cell activating factor
BCR	B-cell receptor
BER	base excision repair
CIP	Cdk-inhibitory proteins
CDR	complementary determining region
CSF	colon-stimulating factor
CSR	class-switch recombination
Cdk	cyclin-dependent kinases
CPDs	cyclo-butane pyrimidine dimer
DSB	double strand break
Fucci.	fluorescent ubiquitination-based cell cycle indicator
GC	germinal center
HEL	hen egg lysosyme
HEV	high-endothelial venules
HSC	hematopoietic stem cell
HDHB	human DNA helicase B
Ig	immunoglobulin
IPP	Ileal Peyer's Patches
kb	kilobase
Lin-	lineage negative cells
LTR	long terminal repeat
MEF	mouse embryonic fibroblasts
MHC	major histocompatibility complex
mKO2	monomeric Kusabira Orange protein 2
MMR	mismatch-repair
MSCV	mouse stem cell virus
MSH	MutS homologs
MLH	MutL homolog
NHEJ	non-homologous end-joining pathway
PAMPs.	pathogen-associated molecular pattern
PCNA	proliferative cell nuclear antigen
PIP	PCNA-binding peptide domain
PALS	periarterioral lymphoid sheath

PMS2	post-meiotic segregation increased protein 2
PNA	peanut agglutinin
Pol	polymerase
PAD	polymerase-associated domain
PSLD	phosphorylated subcellular localization domain
RAG	recombination-activated genes
RFC	replication factor C
SCF	stem cell factor
SCF	SKP1-CUL1-F box protein complex
SHM	somatic hypermutation
SMUG	single-strand selective monofunctional uracil-DNA glycosylase
SRBC	sheep red blood cells
SSB	single strand break
TCR	T-cell receptor
TDG	thymidine-DNA glycosylase
Tg	transgene
TLR-4	Toll-like receptors
TLS	translesion synthesis
Ub	ubiquitin
UNG	uracil-N-glycosylase
WT	wild type
XPV	xeroderma pigmentosum variant
YFP	yellow fluorescent protein

Bibliographie

Abe, T., Sakaue-Sawano, A., Kiyonari, H., Shioi, G., Inoue, K., Horiuchi, T., Nakao, K., Miyawaki, A., Aizawa, S., and Fujimori, T. (2013). Visualization of cell cycle in mouse embryos with Fucci2 reporter directed by Rosa26 promoter. Development *140*, 237–246.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell.

Antonchuk, J., Sauvageau, G., and Humphries, R.K. (2002). HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. Cell *109*, 39–45.

Aoufouchi, S., Faili, A., Zober, C., D'Orlando, O., Weller, S., Weill, J.-C., and Reynaud, C.-A. (2008a). Proteasomal degradation restricts the nuclear lifespan of AID. J Exp Med *205*, 1357–1368.

Aoufouchi, S., Faili, A., Zober, C., D'Orlando, O., Weller, S., Weill, J.-C., and Reynaud, C.-A. (2008b). Proteasomal degradation restricts the nuclear lifespan of AID. J Exp Med *205*, 1357–1368.

Arakawa, H., Hauschild, J., and Buerstedde, J.-M. (2002). Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science *295*, 1301–1306.

Bachl, J., and Olsson, C. (1999). Hypermutation targets a green fluorescent protein-encoding transgene in the presence of immunoglobulin enhancers. European Journal of Immunology *29*, 1383–1389.

Barreto, V., Reina-San-Martin, B., Ramiro, A.R., McBride, K.M., and Nussenzweig, M.C. (2003). C-Terminal Deletion of AID Uncouples Class Switch Recombination from Somatic Hypermutation and Gene Conversion. Molecular Cell *12*, 501–508.

Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J.P., Manis, J.P., and Alt, F.W. (2005). The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. Nature *438*, 508–511.

Basu, U., Meng, F.-L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., et al. (2011). The RNA Exosome Targets the AID Cytidine Deaminase to Both Strands of Transcribed Duplex DNA Substrates. Cell *144*, 353–363.

Bemark, M., Khamlichi, A.A., Davies, S.L., and Neuberger, M.S. (2000). Disruption of mouse polymerase ζ (Rev3) leads to embryonic lethality and impairs blastocyst development in vitro. Current Biology *10*, 1213–1216.

Berek, C., and Milstein, C. (1987). Mutation drift and repertoire shift in the maturation of the immune response. Immunol. Rev. *96*, 23–41.

Betz, A.G., Milstein, C., González-Fernández, A., Pannell, R., Larson, T., and Neuberger, M.S. (1994). Elements regulating somatic hypermutation of an immunoglobulin κ gene: Critical role for the intron enhancer/matrix attachment region. Cell *77*, 239–248.

Bienko, M., Green, C.M., Sabbioneda, S., Crosetto, N., Matic, I., Hibbert, R.G., Begovic, T., Niimi, A., Mann, M., Lehmann, A.R., et al. (2010). Regulation of Translesion Synthesis DNA Polymerase η by Monoubiquitination. Molecular Cell *37*, 396–407.

Blagodatski, A., Batrak, V., Schmidl, S., Schoetz, U., Caldwell, R.B., Arakawa, H., and Buerstedde, J.-M. (2009). A cis-Acting Diversification Activator Both Necessary and Sufficient for AID-Mediated Hypermutation. PLoS Genet *5*, e1000332.

Blow, J.J., and Dutta, A. (2005). Preventing re-replication of chromosomal DNA. Nat Rev Mol Cell Biol *6*, 476–486.

Brandt, V.L., and Roth, D.B. (2008). G.O.D.'s Holy Grail: Discovery of the RAG Proteins. J Immunol *180*, 3–4.

Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc. Natl. Acad. Sci. U.S.A. *100*, 4102–4107.

Brenner, S., and Milstein, C. (1966). Origin of antibody variation. Nature *211*, 242–243.

Burnet F. M. (1957). A modification of Jerne's theory of antibody formation using the concept of clonal selection. Austr. J. Sci. 20, 67–69.

Chahwan, R., Edelmann, W., Scharff, M.D., and Roa, S. (2011). Mismatch-mediated error prone repair at the immunoglobulin genes. Biomed. Pharmacother. *65*, 529–536.

Chahwan, R., Edelmann, W., Scharff, M.D., and Roa, S. (2012). AIDing antibody diversity by error-prone mismatch repair. Semin. Immunol. *24*, 293–300.

Charles A Janeway, J., Travers, P., Walport, M., and Shlomchik, M.J. (2001). Immunobiology.

Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature *430*, 992–998.

Chen, Y., Cleaver, J.E., Hatahet, Z., Honkanen, R.E., Chang, J.-Y., Yen, Y., and Chou, K. (2008). Human DNA polymerase η activity and translocation is regulated by phosphorylation. PNAS *105*, 16578–16583.

Conticello, S.G., Langlois, M., Yang, Z., and Neuberger, M.S. (2007). DNA Deamination in Immunity: AID in the Context of Its APOBEC Relatives. In Advances in Immunology, Frederick W. Alt and Tasuku Honjo, ed. (Academic Press), pp. 37–73.

Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419–423.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67–79.

Davidson, N.O., and Shelness, G.S. (2000). APOLIPOPROTEIN B: mRNA editing, lipoprotein assembly, and presecretory degradation. Annu. Rev. Nutr. *20*, 169–193.

Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J.-C., and Reynaud, C.-A. (2005). Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. J. Exp. Med. *201*, 1191–1196.

Delbos, F., Aoufouchi, S., Faili, A., Weill, J.-C., and Reynaud, C.-A. (2007). DNA polymerase η is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. The Journal of Experimental Medicine *204*, 17 –23.

Dickerson, S.K., Market, E., Besmer, E., and Papavasiliou, F.N. (2003). AID Mediates Hypermutation by Deaminating Single Stranded DNA. The Journal of Experimental Medicine *197*, 1291 –1296.

Dogan, I., Bertocci, B., Vilmont, V., Delbos, F., Mégret, J., Storck, S., Reynaud, C.-A., and Weill, J.-C. (2009). Multiple layers of B cell memory with different effector functions. Nature Immunology *10*, 1292–1299.

Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.-H., Robbiani, D.F., Di Virgilio, M., San-Martin, B.R., Heidkamp, G., Schwickert, T.A., et al. (2008). MicroRNA-155 Suppresses Activation-Induced Cytidine Deaminase-Mediated Myc-Igh Translocation. Immunity *28*, 630–638.

Esposito, G., Texido, G., Betz, U.A.K., Gu, H., Müller, W., Klein, U., and Rajewsky, K. (2000). Mice reconstituted with DNA polymerase β-deficient fetal liver cells are able to mount a T cell-dependent immune response and mutate their Ig genes normally. PNAS *97*, 1166–1171.

Esposito, G., Godin†, I., Klein, U., Yaspo, M.-L., Cumano, A., and Rajewsky, K. (2000). Disruption of the Rev3l-encoded catalytic subunit of polymerase ζ in mice results in early embryonic lethality. Current Biology *10*, 1221–1224

Fagreus A. (1948). The plasma cellular reaction and its relation to the formation of antibodies in vitro. J. Immunol. 58, 1–14.

Faili, A., Aoufouchi, S., Guéranger, Q., Zober, C., Léon, A., Bertocci, B., Weill, J.-C., and Reynaud, C.-A. (2002). AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. Nature Immunology *3*, 815–821.

Faili, A., Aoufouchi, S., Weller, S., Vuillier, F., Stary, A., Sarasin, A., Reynaud, C.-A., and Weill, J.-C. (2004). DNA Polymerase η Is Involved in Hypermutation Occurring during Immunoglobulin Class Switch Recombination. J Exp Med *199*, 265–270.

Faili, A., Stary, A., Delbos, F., Weller, S., Aoufouchi, S., Sarasin, A., Weill, J.-C., and Reynaud, C.-A. (2009). A Backup Role of DNA Polymerase κ in Ig Gene Hypermutation Only Takes Place in the Complete Absence of DNA Polymerase η. J Immunol *182*, 6353–6359.

Fischer, J.A., Muller-Weeks, S., and Caradonna, S. (2004). Proteolytic degradation of the nuclear isoform of uracil-DNA glycosylase occurs during the S phase of the cell cycle. DNA Repair *3*, 505–513.

Friedberg, E.C., Lehmann, A.R., and Fuchs, R.P.P. (2005). Trading places: how do DNA polymerases switch during translesion DNA synthesis? Mol. Cell *18*, 499–505.

Frieder, D., Larijani, M., Collins, C., Shulman, M., and Martin, A. (2009). The concerted action of Msh2 and UNG stimulates somatic hypermutation at A . T base pairs. Mol. Cell. Biol *29*, 5148–5157.

Fritz, E.L., and Papavasiliou, F.N. (2010). Cytidine deaminases: AIDing DNA demethylation? Genes & Development *24*, 2107 –2114.

Ganesh, K., and Neuberger, M.S. (2011). The relationship between hypothesis and experiment in unveiling the mechanisms of antibody gene diversification. FASEB J *25*, 1123–1132.

Geisberger, R., Rada, C., and Neuberger, M.S. (2009). The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence. PNAS *106*, 6736–6741.

Genschel, J., Bazemore, L.R., and Modrich, P. (2002). Human Exonuclease I Is Required for 5′ and 3′ Mismatch Repair. J. Biol. Chem. *277*, 13302–13311.

Gonzalez, S.F., Degn, S.E., Pitcher, L.A., Woodruff, M., Heesters, B.A., and Carroll, M.C. (2011). Trafficking of B Cell Antigen in Lymph Nodes. Annu. Rev. Immunol. *29*, 215–233.

Goodman, M.F., and Scharff, M.D. (2005). Identifying protein–protein interactions in somatic hypermutation. J Exp Med *201*, 493–496.

Gu, J., Xia, X., Yan, P., Liu, H., Podust, V.N., Reynolds, A.B., and Fanning, E. (2004). Cell cycle-dependent regulation of a human DNA helicase that localizes in DNA damage foci. Mol. Biol. Cell *15*, 3320–3332.

Gueranger, Q., Stary, A., Aoufouchi, S., Faili, A., Sarasin, A., Reynaud, C.-A., and Weill, J.-C. (2008). Role of DNA polymerases eta, iota and zeta in UV resistance and UV-induced mutagenesis in a human cell line. DNA Repair (Amst.) *7*, 1551–1562.

Guo, C., Fischhaber, P.L., Luk-Paszyc, M.J., Masuda, Y., Zhou, J., Kamiya, K., Kisker, C., and Friedberg, E.C. (2003). Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. EMBO J. *22*, 6621–6630.

Haas, D.L., Lutzko, C., Logan, A.C., Cho, G.J., Skelton, D., Jin Yu, X., Pepper, K.A., and Kohn, D.B. (2003). The Moloney Murine Leukemia Virus Repressor Binding Site Represses Expression in Murine and Human Hematopoietic Stem Cells. J. Virol. *77*, 9439–9450.

Hagen, L., Kavli, B., Sousa, M.M., Torseth, K., Liabakk, N.B., Sundheim, O., Peňa-Diaz, J., Otterlei, M., Hørning, O., Jensen, O.N., et al. (2008). Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA. EMBO J *27*, 51–61.

Hardeland, U., Kunz, C., Focke, F., Szadkowski, M., and Schär, P. (2007). Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. Nucl. Acids Res. *35*, 3859–3867.

Hardy, R.R., Kincade, P.W., and Dorshkind, K. (2007). The Protean Nature of Cells in the B Lymphocyte Lineage. Immunity *26*, 703–714.

Hasham, M.G., Donghia, N.M., Coffey, E., Maynard, J., Snow, K.J., Ames, J., Wilpan, R.Y., He, Y., King, B.L., and Mills, K.D. (2010). Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. Nat. Immunol. *11*, 820–826.

Häsler, J., Rada, C., and Neuberger, M.S. (2011). Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1α (eEF1A). PNAS *108*, 18366–18371.

Haurowitz F., and Breinl F. (1930). Chemische untersuchung des prazipitates aus hamoglobin und anti-hamoglobin-serum und bemerkungen ber die natur der antikorper. Z. Physiol. Chem. 192, 42–57.

Hendel, A., Krijger, P.H.L., Diamant, N., Goren, Z., Langerak, P., Kim, J., Reißner, T., Lee, K.-Y., Geacintov, N.E., Carell, T., et al. (2011). PCNA Ubiquitination Is Important, But Not Essential for Translesion DNA Synthesis in Mammalian Cells. PLoS Genet *7*, e1002262.

Henderson, L., Bortone, D.S., Lim, C., and Zambon, A.C. (2013). Classic "broken cell" techniques and newer live cell methods for cell cycle assessment. American Journal of Physiology - Cell Physiology *304*, C927–C938.

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

Holmes, J., Jr, Clark, S., and Modrich, P. (1990). Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines. Proc. Natl. Acad. Sci. U.S.A. *87*, 5837–5841.

Honjo, T., and Alt, F.W. (1995). Immunoglobulin genes (London; San Diego: Academic Press).

Hozumi, N., and Tonegawa, S. (2004). Pillars Article: Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. J Immunol *173*, 4260–4264.

Imai, K., Slupphaug, G., Lee, W.-I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H.E., et al. (2003). Human uracil–DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. Nat Immunol *4*, 1023–1028.

Jansen, J.G., Langerak, P., Tsaalbi-Shtylik, A., Berk, P. van den, Jacobs, H., and Wind, N. de (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J Exp Med 203, 319–32

Jeevan-Raj, B.P., Robert, I., Heyer, V., Page, A., Wang, J.H., Cammas, F., Alt, F.W., Losson, R., and Reina-San-Martin, B. (2011). Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. The Journal of Experimental Medicine *208*, 1649 –1660.

Jerne N. K. (1955). The natural-selection theory of antibody formation. Proc. Natl. Acad. Sci. 41, 849–859.

Johnson, R.E., Kondratick, C.M., Prakash, S., and Prakash, L. (1999). hRAD30 Mutations in the Variant Form of Xeroderma Pigmentosum. Science *285*, 263–265.

Jung, Y.-S., Qian, Y., and Chen, X. (2012). DNA polymerase eta is targeted by Mdm2 for polyubiquitination and proteasomal degradation in response to ultraviolet irradiation. DNA Repair *11*, 177–184.

Kadyrov, F.A., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. Cell *126*, 297–308.

Kadyrov, F.A., Holmes, S.F., Arana, M.E., Lukianova, O.A., O'Donnell, M., Kunkel, T.A., and Modrich, P. (2007). Saccharomyces cerevisiae MutLα Is a Mismatch Repair Endonuclease. J. Biol. Chem. *282*, 37181–37190.

Kannouche, P., Broughton, B.C., Volker, M., Hanaoka, F., Mullenders, L.H.F., and Lehmann, A.R. (2001). Domain structure, localization, and function of DNA polymerase η, defective in xeroderma pigmentosum variant cells. Genes Dev. *15*, 158–172.

Kannouche, P.L., Wing, J., and Lehmann, A.R. (2004a). Interaction of Human DNA Polymerase η with Monoubiquitinated PCNA: A Possible Mechanism for the Polymerase Switch in Response to DNA Damage. Molecular Cell *14*, 491–500.

Kannouche, P.L., Wing, J., and Lehmann, A.R. (2004b). Interaction of Human DNA Polymerase η with Monoubiquitinated PCNA: A Possible Mechanism for the Polymerase Switch in Response to DNA Damage. Molecular Cell *14*, 491–500.

Kavli, B., Andersen, S., Otterlei, M., Liabakk, N.B., Imai, K., Fischer, A., Durandy, A., Krokan, H.E., and Slupphaug, G. (2005). B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. J Exp Med *201*, 2011–2021.

Kavli, B., Otterlei, M., Slupphaug, G., and Krokan, H.E. (2007). Uracil in DNA—General mutagen, but normal intermediate in acquired immunity. DNA Repair *6*, 505–516.

Kenter, A.L. (2012). AID targeting is dependent on RNA polymerase II pausing. Semin. Immunol. *24*, 281–286.

Kitamura, D., Roes, J., Kuhn, R., and Rajewsky, K. (1991). A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin [mu] chain gene. Nature *350*, 423–426.

Kohli, R.M., Maul, R.W., Guminski, A.F., McClure, R.L., Gajula, K.S., Saribasak, H., McMahon, M.A., Siliciano, R.F., Gearhart, P.J., and Stivers, J.T. (2010). Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. J. Biol. Chem. *285*, 40956–40964.

Komeno, Y., Kitaura, J., Watanabe-Okochi, N., Kato, N., Oki, T., Nakahara, F., Harada, Y., Harada, H., Shinkura, R., Nagaoka, H., et al. (2010). AID-induced T-lymphoma or B-leukemia/lymphoma in a mouse BMT model. Leukemia *24*, 1018–1024.

Kondo, M., Wagers, A.J., Manz, M.G., Prohaska, S.S., Scherer, D.C., Beilhack, G.F., Shizuru, J.A., and Weissman, I.L. (2003). BIOLOGY OF HEMATOPOIETIC STEM CELLS AND PROGENITORS : Implications for Clinical Application. Annual Review of Immunology *21*, 759–806.

Kothapalli, N., Norton, D.D., and Fugmann, S.D. (2008). Cutting Edge: A cis-Acting DNA Element Targets AID-Mediated Sequence Diversification to the Chicken Ig Light Chain Gene Locus. J Immunol *180*, 2019–2023.

Krijger, P.H.L., Langerak, P., Berk, P.C.M. van den, and Jacobs, H. (2009). Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. J Exp Med *206*, 2603–2611.

Krijger, P.H.L., van den Berk, P.C.M., Wit, N., Langerak, P., Jansen, J.G., Reynaud, C.-A., de Wind, N., and Jacobs, H. (2011). PCNA ubiquitination-independent activation of polymerase Î· during somatic hypermutation and DNA damage tolerance. DNA Repair *10*, 1051–1059.

Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annu. Rev. Biochem. *74*, 681–710.

Kurzawa, L., and Morris, M.C. (2010). Cell-Cycle Markers and Biosensors. ChemBioChem *11*, 1037–1047.

Lacks, S.A., Dunn, J.J., and Greenberg, B. (1982). Identification of base mismatches recognized by the heteroduplex-DNA-repair system of Streptococcus pneumoniae. Cell *31*, 327–336.

Lagresle-Peyrou, C., Yates, F., Malassis-Seris, M., Hue, C., Morillon, E., Garrigue, A., Liu, A., Hajdari, P., Stockholm, D., Danos, O., et al. (2006). Long-term immune reconstitution in RAG-1-deficient mice treated by retroviral gene therapy: a balance between efficiency and toxicity. Blood *107*, 63–72.

Langerak, P., Nygren, A.O.H., Krijger, P.H.L., Berk, P.C.M. van den, and Jacobs, H. (2007). A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. J Exp Med *204*, 1989–1998.

Längle-Rouault, F., Maenhaut-Michel, G., and Radman, M. (1987). GATC sequences, DNA nicks and the MutH function in Escherichia coli mismatch repair. The EMBO Journal *6*, 1121.

Larson, E.D., Cummings, W.J., Bednarski, D.W., and Maizels, N. (2005). MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. Mol. Cell *20*, 367–375.

Lederberg, J. (1959). Genes and antibodies. Science *129*, 1649–1653.

Lévy, N., Oehlmann, M., Delalande, F., Nasheuer, H.P., Dorsselaer, A.V., Schreiber, V., Murcia, G. de, Murcia, J.M., Maiorano, D., and Bresson, A. (2009). XRCC1 interacts with the p58 subunit of DNA Pol α-primase and may coordinate DNA repair and replication during S phase. Nucl. Acids Res. *37*, 3177–3188.

Lindahl, T. (1982). DNA repair enzymes. Annu. Rev. Biochem. *51*, 61–87.

Lindahl, T., Karran, P., and Wood, R.D. (1997). DNA excision repair pathways. Current Opinion in Genetics & Development *7*, 158–169.

Liu, M., and Schatz, D.G. (2009). Balancing AID and DNA repair during somatic hypermutation. Trends in Immunology *30*, 173–181.

Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. Nature *451*, 841–845.

MacCarthy, T., Roa, S., Scharff, M.D., and Bergman, A. (2009). SHMTool: A webserver for comparative analysis of somatic hypermutation datasets. DNA Repair *8*, 137–141.

Martomo, S.A., Yang, W.W., Wersto, R.P., Ohkumo, T., Kondo, Y., Yokoi, M., Masutani, C., Hanaoka, F., and Gearhart, P.J. (2005). Different mutation signatures in DNA polymerase η- and MSH6-deficient mice suggest separate roles in antibody diversification. PNAS *102*, 8656–8661.

Martomo, S.A., Saribasak, H., Yokoi, M., Hanaoka, F., and Gearhart, P.J. (2008). Reevaluation of the role of DNA polymerase theta in somatic hypermutation of immunoglobulin genes. DNA Repair (Amst.) *7*, 1603–1608.

Masuda, K., Ouchida, R., Hikida, M., Nakayama, M., Ohara, O., Kurosaki, T., and O-Wang, J. (2006). Absence of DNA polymerase theta results in decreased somatic hypermutation frequency and altered mutation patterns in Ig genes. DNA Repair (Amst.) *5*, 1384–1391.

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

Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F., and Kunkel, T.A. (2000). Low fidelity DNA synthesis by human DNA polymerase-[eta]. Nature *404*, 1011–1013.

Matsuda, T., Bebenek, K., Masutani, C., Rogozin, I.B., Hanaoka, F., and Kunkel, T.A. (2001). Error rate and specificity of human and murine DNA polymerase eta. J. Mol. Biol. *312*, 335–346.

Matsumoto, Y., Marusawa, H., Kinoshita, K., Endo, Y., Kou, T., Morisawa, T., Azuma, T., Okazaki, I.-M., Honjo, T., and Chiba, T. (2007). Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med *13*, 470–476.

Mayorov, V.I., Rogozin, I.B., Adkison, L.R., and Gearhart, P.J. (2005). DNA Polymerase η Contributes to Strand Bias of Mutations of A versus T in Immunoglobulin Genes. J Immunol *174*, 7781–7786.

McBride, K.M., Barreto, V., Ramiro, A.R., Stavropoulos, P., and Nussenzweig, M.C. (2004). Somatic Hypermutation Is Limited by CRM1-dependent Nuclear Export of Activation-induced Deaminase. J. Exp. Med. *199*, 1235–1244.

McBride, K.M., Gazumyan, A., Woo, E.M., Schwickert, T.A., Chait, B.T., and Nussenzweig, M.C. (2008). Regulation of class switch recombination and somatic mutation by AID phosphorylation. J Exp Med *205*, 2585–2594.

McDonald, J.P., Levine, A.S., and Woodgate, R. (1997). The Saccharomyces cerevisiae RAD30 gene, a homologue of Escherichia coli dinB and umuC, is DNA damage inducible and functions in a novel error-free postreplication repair mechanism. Genetics *147*, 1557–1568.

McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W., and Weigert, M. (1984). Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. Proc. Natl. Acad. Sci. U.S.A. *81*, 3180–3184.

Meyers, M., Theodosiou, M., Acharya, S., Odegaard, E., Wilson, T., Lewis, J.E., Davis, T.W., Patten, C.W.-V., Fishel, R., and Boothman, D.A. (1997). Cell Cycle Regulation of the Human DNA Mismatch Repair Genes hMSH2, hMLH1, and hPMS2. Cancer Res *57*, 206–208.

Michael, N., Shen, H.M., Longerich, S., Kim, N., Longacre, A., and Storb, U. (2003). The E Box Motif CAGGTG Enhances Somatic Hypermutation without Enhancing Transcription. Immunity *19*, 235–242.

Milstein, C., Neuberger, M.S., and Staden, R. (1998). Both DNA strands of antibody genes are hypermutation targets. PNAS *95*, 8791–8794.

Moldovan, G.-L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. Cell *129*, 665–679.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific Expression of Activation-induced Cytidine Deaminase (AID), a Novel Member of the RNA-editing Deaminase Family in Germinal Center B Cells. J. Biol. Chem. *274*, 18470–18476.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553–563.

Murphy, K., Walport, M., and Janeway, C. (2012). Janeway's immunobiology (New York: Garland Science).

Nakayama, K.I., and Nakayama, K. (2006). Ubiquitin ligases: cell-cycle control and cancer. Nat Rev Cancer *6*, 369–381.

Neuberger, M.S., Harris, R.S., Di Noia, J., and Petersen-Mahrt, S.K. (2003). Immunity through DNA deamination. Trends Biochem. Sci. *28*, 305–312.

Nilsen, H., Haushalter, K.A., Robins, P., Barnes, D.E., Verdine, G.L., and Lindahl, T. (2001). Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil–DNA glycosylase. EMBO J *20*, 4278–4286.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular Mechanisms of Antibody Somatic Hypermutation. Annual Review of Biochemistry *76*, 1–22.

Di Noia, J.M., Rada, C., and Neuberger, M.S. (2006). SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. EMBO J *25*, 585–595.

Nowak, U., Matthews, A.J., Zheng, S., and Chaudhuri, J. (2011). The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. Nat Immunol *12*, 160–166.

Oers, J.M.M. van, Roa, S., Werling, U., Liu, Y., Genschel, J., Hou, H., Sellers, R.S., Modrich, P., Scharff, M.D., and Edelmann, W. (2010). PMS2 endonuclease activity has distinct biological functions and is essential for genome maintenance. PNAS *107*, 13384–13389.

Ohta, H., Sekulovic, S., Bakovic, S., Eaves, C.J., Pineault, N., Gasparetto, M., Smith, C., Sauvageau, G., and Humphries, R.K. (2007). Near-maximal expansions of hematopoietic stem cells in culture using NUP98-HOX fusions. Exp. Hematol. *35*, 817–830.

Okazaki, I., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. J. Exp. Med. *197*, 1173–1181.

Ordinario, E.C., Yabuki, M., Larson, R.P., and Maizels, N. (2009). Temporal Regulation of Ig Gene Diversification Revealed by Single-Cell Imaging. J Immunol *183*, 4545–4553.

Orkin, S.H., and Zon, L.I. (2008). SnapShot: Hematopoiesis. Cell *132*, 712.e1–712.e2.

Orthwein, A., Patenaude, A.-M., Affar, E.B., Lamarre, A., Young, J.C., and Di Noia, J.M. (2010). Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. The Journal of Experimental Medicine *207*, 2751 –2765.

Orthwein, A., Zahn, A., Methot, S.P., Godin, D., Conticello, S.G., Terada, K., and Noia, J.M.D. (2012). Optimal functional levels of activation-induced deaminase specifically require the Hsp40 DnaJa1. The EMBO Journal *31*, 679–691.

Otterlei, M., Warbrick, E., Nagelhus, T.A., Haug, T., Slupphaug, G., Akbari, M., Aas, P.A., Steinsbekk, K., Bakke, O., and Krokan, H.E. (1999). Post-replicative base excision repair in replication foci. The EMBO Journal *18*, 3834–3844.

Van Parijs, L., Refaeli, Y., Lord, J.D., Nelson, B.H., Abbas, A.K., and Baltimore, D. (1999). Uncoupling IL-2 signals that regulate T cell proliferation, survival, and Fas-mediated activation-induced cell death. Immunity *11*, 281–288.

Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., Muramatsu, M., Honjo, T., Morse, H.C., Nussenzweig, M.C., and Dalla-Favera, R. (2008). AID is required for germinal center–derived lymphomagenesis. Nature Genetics *40*, 108–112.

Patenaude, A.-M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzo, A., and Di Noia, J.M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. Nat Struct Mol Biol *16*, 517–527.

Pauling, L. (1940). A Theory of the Structure and Process of Formation of Antibodies*. J. Am. Chem. Soc. *62*, 2643–2657.

Pavlov, Y.I., Rogozin, I.B., Galkin, A.P., Aksenova, A.Y., Hanaoka, F., Rada, C., and Kunkel, T.A. (2002). Correlation of somatic hypermutation specificity and A-T base pair substitution errors by DNA polymerase η during copying of a mouse immunoglobulin κ light chain transgene. PNAS *99*, 9954–9959.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., et al. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell *143*, 122–133.

Pear, W.S., Miller, J.P., Xu, L., Pui, J.C., Soffer, B., Quackenbush, R.C., Pendergast, A.M., Bronson, R., Aster, J.C., Scott, M.L., et al. (1998). Efficient and Rapid Induction of a Chronic Myelogenous Leukemia-Like Myeloproliferative Disease in Mice Receiving P210 bcr/abl-Transduced Bone Marrow. Blood *92*, 3780 –3792.

Peña-Diaz, J., Bregenhorn, S., Ghodgaonkar, M., Follonier, C., Artola-Borán, M., Castor, D., Lopes, M., Sartori, A.A., and Jiricny, J. (2012). Noncanonical mismatch repair as a source of genomic instability in human cells. Mol. Cell *47*, 669–680.

Pérez-Durán, P., Belver, L., Yébenes, V.G. de, Delgado, P., Pisano, D.G., and Ramiro, A.R. (2012). UNG shapes the specificity of AID-induced somatic hypermutation. J Exp Med *209*, 1379–1389.

Peters, A., and Storb, U. (1996). Somatic Hypermutation of Immunoglobulin Genes Is Linked to Transcription Initiation. Immunity *4*, 57–65.

Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature *418*, 99–104.

Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature *424*, 103–107.

Phan, T.G., Gardam, S., Basten, A., and Brink, R. (2005). Altered Migration, Recruitment, and Somatic Hypermutation in the Early Response of Marginal Zone B Cells to T Cell-Dependent Antigen. J Immunol *174*, 4567–4578.

Phillips, L.G., and Sale, J.E. (2009). Be prepared: DNA polymerase η gets ready for action—in G1. Cell Cycle *8*, 3257–3260.

Pillai, S., and Cariappa, A. (2009). The follicular versus marginal zone B lymphocyte cell fate decision. Nature Reviews Immunology *9*, 767–777.

Pluciennik, A., Dzantiev, L., Iyer, R.R., Constantin, N., Kadyrov, F.A., and Modrich, P. (2010). PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. PNAS *107*, 16066–16071.

Poltoratsky, V., Prasad, R., Horton, J.K., and Wilson, S.H. (2007). Down-regulation of DNA polymerase beta accompanies somatic hypermutation in human BL2 cell lines. DNA Repair (Amst.) *6*, 244–253.

Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002a). Immunoglobulin Isotype Switching Is Inhibited and Somatic Hypermutation Perturbed in UNG-Deficient Mice. Current Biology *12*, 1748–1755.

Rada, C., Jarvis, J.M., and Milstein, C. (2002b). AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. PNAS *99*, 7003–7008.

Rada, C., Di Noia, J.M., and Neuberger, M.S. (2004). Mismatch Recognition and Uracil Excision Provide Complementary Paths to Both Ig Switching and the A/T-Focused Phase of Somatic Mutation. Molecular Cell *16*, 163–171.

Radbruch, A., Muehlinghaus, G., Luger, E.O., Inamine, A., Smith, K.G.C., Dörner, T., and Hiepe, F. (2006). Competence and competition: the challenge of becoming a long-lived plasma cell. Nature Reviews Immunology *6*, 741–750.

Radman, M. (1974). Phenomenology of an inducible mutagenic DNA repair pathway in Escherichia coli: SOS repair hypothesis,in: P.L.F. Sherman, M. Miller, C. Lawrence, W.H. Tabor (Eds.) (Springfield, Ill).

Ranjit, S., Khair, L., Linehan, E.K., Ucher, A.J., Chakrabarti, M., Schrader, C.E., and Stavnezer, J. (2011). AID Binds Cooperatively with UNG and Msh2-Msh6 to Ig Switch Regions Dependent upon the AID C Terminus. J Immunol *187*, 2464–2475.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A., et al. (2000). Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2). Cell *102*, 565–575.

Reynaud, C.A., Anquez, V., Grimal, H., and Weill, J.C. (1987). A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell *48*, 379–388.

Reynaud, C.-A., Delbos, F., Faili, A., Guéranger, Q., Aoufouchi, S., and Weill, J.-C. (2009). Competitive repair pathways in immunoglobulin gene hypermutation. Phil. Trans. R. Soc. B *364*, 613–619.

Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R., et al. (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Mol. Cell *36*, 631–641.

Robbins, P.B., Skelton, D.C., Yu, X.-J., Halene, S., Leonard, E.H., and Kohn, D.B. (1998). Consistent, Persistent Expression from Modified Retroviral Vectors in Murine Hematopoietic Stem Cells. PNAS *95*, 10182–10187.

Robinson, S.N., Ng, J., Niu, T., Yang, H., McMannis, J.D., Karandish, S., Kaur, I., Fu, P., Del Angel, M., Messinger, R., et al. (2006). Superior ex vivo cord blood expansion following co-culture with bone marrow-derived mesenchymal stem cells. Bone Marrow Transplant. *37*, 359–366.

Rogozin, I.B., and Kolchanov, N.A. (1992). Somatic hypermutagenesis in immunoglobulin genes: II. Influence of neighbouring base sequences on mutagenesis. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression *1171*, 11–18.

Rogozin, I.B., Pavlov, Y.I., Bebenek, K., Matsuda, T., and Kunkel, T.A. (2001). Somatic mutation hotspots correlate with DNA polymerase η error spectrum. Nature Immunology *2*, 530–536.

Rolink, A.G., Andersson, J., and Melchers, F. (2004). Molecular mechanisms guiding late stages of B-cell development. Immunological Reviews *197*, 41–50.

Rouaud, P., Vincent-Fabert, C., Saintamand, A., Fiancette, R., Marquet, M., Robert, I., Reina-San-Martin, B., Pinaud, E., Cogné, M., and Denizot, Y. (2013). The IgH 3′ regulatory region controls somatic hypermutation in germinal center B cells. J Exp Med.

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., et al. (2008). Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. Cell *132*, 487–498.

Sale, J.E., Calandrini, D.M., Takata, M., Takeda, S., and Neuberger, M.S. (2001). Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. Nature *412*, 921–926.

Saribasak, H., and Gearhart, P.J. (2012). Does DNA repair occur during somatic hypermutation? Seminars in Immunology *24*, 287–292.

Saribasak, H., Maul, R.W., Cao, Z., McClure, R.L., Yang, W., McNeill, D.R., Wilson, D.M., and Gearhart, P.J. (2011). XRCC1 suppresses somatic hypermutation and promotes alternative nonhomologous end joining in Igh genes. J Exp Med *208*, 2209–2216.

Schaetzlein, S., Chahwan, R., Avdievich, E., Roa, S., Wei, K., Eoff, R.L., Sellers, R.S., Clark, A.B., Kunkel, T.A., Scharff, M.D., et al. (2013). Mammalian Exo1 encodes both structural and catalytic functions that play distinct roles in essential biological processes. PNAS *110*, E2470–E2479.

Schanz, S., Castor, D., Fischer, F., and Jiricny, J. (2009). Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. PNAS *106*, 5593–5598.

Schatz, D.G., Oettinger, M.A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell *59*, 1035–1048.

Schenten, D., Kracker, S., Esposito, G., Franco, S., Klein, U., Murphy, M., Alt, F.W., and Rajewsky, K. (2009). Polζ ablation in B cells impairs the germinal center reaction, class switch recombination, DNA break repair, and genome stability. J Exp Med *206*, 477–490.

Schrader, C.E., Guikema, J.E.J., Linehan, E.K., Selsing, E., and Stavnezer, J. (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. J Immunol *179*, 6064–6071.

Shapiro-Shelef, M., and Calame, K. (2005). Regulation of plasma-cell development. Nature Reviews Immunology *5*, 230–242.

Sharbeen, G., Cook, A.J.L., Lau, K.K.E., Raftery, J., Yee, C.W.Y., and Jolly, C.J. (2010). Incorporation of dUTP does not mediate mutation of A:T base pairs in Ig genes in vivo. Nucl. Acids Res. *38*, 8120–8130.

Sharbeen, G., Yee, C.W.Y., Smith, A.L., and Jolly, C.J. (2012). Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase. J Exp Med *209*, 965–974.

Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 Gene in Normal B Cells by the Process of Somatic Hypermutation of Ig Genes. Science *280*, 1750–1752.

Schenten, D., Gerlach, V.L., Guo, C., Velasco-Miguel, S., Hladik, C.L., White, C.L., Friedberg, E.C., Rajewsky, K., and Esposito, G. (2002). DNA polymerase kappa deficiency does not affect somatic hypermutation in mice. Eur. J. Immunol. *32*, 3152–316

Shibahara, K., and Stillman, B. (1999). Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell *96*, 575–585.

Shinkai, Y., Rathbun, 0Gary, Lam, K.-P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., Stall, A.M., et al. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell *68*, 855–867.

Shinkura, R., Ito, S., Begum, N.A., Nagaoka, H., Muramatsu, M., Kinoshita, K., Sakakibara, Y., Hijikata, H., and Honjo, T. (2004). Separate domains of AID are required for somatic hypermutation and class-switch recombination. Nature Immunology *5*, 707–712.

Skoneczna, A., McIntyre, J., Skoneczny, M., Policinska, Z., and Sledziewska-Gojska, E. (2007). Polymerase eta Is a Short-lived, Proteasomally Degraded Protein that Is Temporarily Stabilized Following UV Irradiation in Saccharomyces cerevisiae. Journal of Molecular Biology *366*, 1074–1086.

Soria, G., Belluscio, L., Cappellen, W.A. van, Kanaar, R., and Gottifredi, J.E. and V. (2009). DNA damage induced Pol η recruitment takes place independently of the cell cycle phase. Cell Cycle *8*, 3340–3348.

Sorrentino, B.P. (2004). Clinical strategies for expansion of haematopoietic stem cells. Nature Reviews Immunology *4*, 878–888.

Stanlie, A., Aida, M., Muramatsu, M., Honjo, T., and Begum, N.A. (2010). Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. PNAS *107*, 22190–22195.

Stavnezer, J., and Schrader, C.E. (2006). Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. Trends in Genetics *22*, 23–28.

Stavnezer, J., Guikema, J.E.J., and Schrader, C.E. (2008). Mechanism and Regulation of Class Switch Recombination. Annual Review of Immunology *26*, 261–292.

Van der Stoep, N., Gorman, J.R., and Alt, F.W. (1998). Reevaluation of 3′Eκ Function in Stage- and Lineage-Specific Rearrangement and Somatic Hypermutation. Immunity *8*, 743–750.

Szadkowski, M., and Jiricny, J. (2002). Identification and functional characterization of the promoter region of the human MSH6 gene. Genes Chromosomes Cancer *33*, 36–46.

Talmage D. W. (1959). Immunological specificity. Unique combinations of selected natural globulins provide and alternative to the classical concept. Science 129, 1643–1648.

Taylor, R.M., Moore, D.J., Whitehouse, J., Johnson, P., and Caldecott, K.W. (2000). A Cell Cycle-Specific Requirement for the XRCC1 BRCT II Domain during Mammalian DNA Strand Break Repair. Mol. Cell. Biol. *20*, 735–740.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 Is a Negative Regulator of Activation-Induced Cytidine Deaminase. Immunity *28*, 621–629.

Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2010). B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. Nature Immunology *11*, 148–154.

Uchimura, Y., Barton, L.F., Rada, C., and Neuberger, M.S. (2011). REG-γ associates with and modulates the abundance of nuclear activation-induced deaminase. J Exp Med *208*, 2385–2391.

Umar, A., Buermeyer, A.B., Simon, J.A., Thomas, D.C., Clark, A.B., Liskay, R.M., and Kunkel, T.A. (1996). Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. Cell *87*, 65–73.

Unniraman, S., and Schatz, D.G. (2007). Strand-Biased Spreading of Mutations During Somatic Hypermutation. Science *317*, 1227–1230.

Victora, G.D., and Nussenzweig, M.C. (2012). Germinal Centers. Annual Review of Immunology *30*, 429–457.

Vincent-Fabert, C., Fiancette, R., Pinaud, E., Truffinet, V., Cogné, N., Cogné, M., and Denizot, Y. (2010). Genomic deletion of the whole IgH 3′ regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. Blood *116*, 1895–1898.

Wang, M., Rada, C., and Neuberger, M.S. (2010). Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. J Exp Med *207*, 141–153.

Washington, M.T., Johnson, R.E., Prakash, L., and Prakash, S. (2001). Accuracy of lesion bypass by yeast and human DNA polymerase η. PNAS *98*, 8355–8360.

Weigert, M.G., Cesari, I.M., Yonkovich, S.J., and Cohn, M. (1970). Variability in the lambda light chain sequences of mouse antibody. Nature *228*, 1045–1047.

Weill, J.-C., and Reynaud, C.-A. (2008). DNA polymerases in adaptive immunity. Nat Rev Immunol *8*, 302–312.

Welner, R.S., Pelayo, R., and Kincade, P.W. (2008). Evolving views on the genealogy of B cells. Nature Reviews Immunology *8*, 95–106.

Wilson, T.M., Vaisman, A., Martomo, S.A., Sullivan, P., Lan, L., Hanaoka, F., Yasui, A., Woodgate, R., and Gearhart, P.J. (2005). MSH2–MSH6 stimulates DNA polymerase η, suggesting a role for A:T mutations in antibody genes. J Exp Med *201*, 637–645.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol *11*, 607–620.

Xu, Z., Fulop, Z., Wu, G., Pone, E.J., Zhang, J., Mai, T., Thomas, L.M., Al-Qahtani, A., White, C.A., Park, S.-R., et al. (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., Rada, C., and Neuberger, M.S. (2006). The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2−/− ung−/− mice. J Exp Med *203*, 2085–2094.

Yabuki, M., Fujii, M.M., and Maizels, N. (2005). The MRE11-RAD50-NBS1 complex accelerates somatic hypermutation and gene conversion of immunoglobulin variable regions. Nature Immunology *6*, 730–736.

Yabuki, M., Ordinario, E.C., Cummings, W.J., Fujii, M.M., and Maizels, N. (2009). E2A Acts in cis in G1 Phase of Cell Cycle to Promote Ig Gene Diversification. J Immunol *182*, 408–415.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat Immunol *12*, 62–69.

Yates, F., Malassis-Séris, M., Stockholm, D., Bouneaud, C., Larousserie, F., Noguiez-Hellin, P., Danos, O., Kohn, D.B., Fischer, A., De Villartay, J.-P., et al. (2002). Gene Therapy of RAG-2−/− Mice: Sustained Correction of the Immunodeficiency. Blood *100*, 3942–3949.

De Yébenes, V.G., Belver, L., Pisano, D.G., González, S., Villasante, A., Croce, C., He, L., and Ramiro, A.R. (2008). miR-181b negatively regulates activation-induced cytidine deaminase in B cells. The Journal of Experimental Medicine *205*, 2199 –2206.

Yolamanova, M., Meier, C., Shaytan, A.K., Vas, V., Bertoncini, C.W., Arnold, F., Zirafi, O., Usmani, S.M., Müller, J.A., Sauter, D., et al. (2013). Peptide nanofibrils boost retroviral gene transfer and provide a rapid means for concentrating viruses. Nat Nanotechnol *8*, 130–136.

Yoshikawa, K., Okazaki, I., Eto, T., Kinoshita, K., Muramatsu, M., Nagaoka, H., and Honjo, T. (2002). AID Enzyme-Induced Hypermutation in an Actively Transcribed Gene in Fibroblasts. Science *296*, 2033 –2036.

Zan, H., Shima, N., Xu, Z., Al-Qahtani, A., Evinger Iii, A.J., Zhong, Y., Schimenti, J.C., and Casali, P. (2005). The translesion DNA polymerase |[theta]| plays a dominant role in immunoglobulin gene somatic hypermutation. EMBO J *24*, 3757–3769.

Zeng, X., Winter, D.B., Kasmer, C., Kraemer, K.H., Lehmann, A.R., and Gearhart, P.J. (2001). DNA polymerase η is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. Nat Immunol *2*, 537–541.

Zhang, Q., Iida, R., Shimazu, T., and Kincade, P.W. (2012). Replenishing B lymphocytes in health and disease. Current Opinion in Immunology *24*, 196–203.

Zheng, B., Han, S., Spanopoulou, E., and Kelsoe, G. (1998). Immunoglobulin gene hypermutation in germinal centers is independent of the RAG-1 V(D)J recombinase. Immunol. Rev. *162*, 133–141.

Zlatanou, A., Despras, E., Braz-Petta, T., Boubakour-Azzouz, I., Pouvelle, C., Stewart, G.S., Nakajima, S., Yasui, A., Ishchenko, A.A., and Kannouche, P.L. (2011). The hMsh2-hMsh6 Complex Acts in Concert with Monoubiquitinated PCNA and Pol η in Response to Oxidative DNA Damage in Human Cells. Molecular Cell *43*, 649–662.

Bibliography from table T2

Bemark, M., Khamlichi, A.A., Davies, S.L., and Neuberger, M.S. (2000). Disruption of mouse polymerase ζ (Rev3) leads to embryonic lethality and impairs blastocyst development in vitro. Current Biology *10*, 1213–1216.

Casali, P., Pal, Z., Xu, Z., and Zan, H. (2006). DNA repair in antibody somatic hypermutation. Trends in Immunology *27*, 313–321.

Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J.-C., and Reynaud, C.-A. (2005). Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. J. Exp. Med. *201*, 1191–1196.

Esposito, G., Godin†, I., Klein, U., Yaspo, M.-L., Cumano, A., and Rajewsky, K. (2000). Disruption of the Rev3l-encoded catalytic subunit of polymerase ζ in mice results in early embryonic lethality. Current Biology *10*, 1221–1224.

Foti, J.J., and Walker, G.C. (2010). SnapShot: DNA polymerases II mammals. Cell *141*, 370–370.e1.

Goodman, M.F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. Annu. Rev. Biochem. *71*, 17–50.

Jansen, J.G., Langerak, P., Tsaalbi-Shtylik, A., Berk, P. van den, Jacobs, H., and Wind, N. de (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J Exp Med *203*, 319–323.

Kunkel, T.A. (2003). Considering the cancer consequences of altered DNA polymerase function. Cancer Cell *3*, 105–110.

Martomo, S.A., Yang, W.W., Wersto, R.P., Ohkumo, T., Kondo, Y., Yokoi, M., Masutani, C., Hanaoka, F., and Gearhart, P.J. (2005). Different mutation signatures in DNA polymerase η- and MSH6-deficient mice suggest separate roles in antibody diversification. PNAS *102*, 8656–8661.

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

Schenten, D., Gerlach, V.L., Guo, C., Velasco-Miguel, S., Hladik, C.L., White, C.L., Friedberg, E.C., Rajewsky, K., and Esposito, G. (2002). DNA polymerase kappa deficiency does not affect somatic hypermutation in mice. Eur. J. Immunol. *32*, 3152–3160.

Weill, J.-C., and Reynaud, C.-A. (2008). DNA polymerases in adaptive immunity. Nat Rev Immunol *8*, 302–312.

Bibliography from table T3

Bardwell, P.D., Martin, A., Wong, E., Li, Z., Edelmann, W., and Scharff, M.D. (2003). Cutting Edge: The G-U Mismatch Glycosylase Methyl-CpG Binding Domain 4 Is Dispensable for Somatic Hypermutation and Class Switch Recombination. J Immunol *170*, 1620–1624.

Bardwell, P.D., Woo, C.J., Wei, K., Li, Z., Martin, A., Sack, S.Z., Parris, T., Edelmann, W., and Scharff, M.D. (2004). Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1–mutant mice. Nature Immunology *5*, 224–229.

Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419–423.

Ehrenstein, M.R., and Neuberger, M.S. (1999). Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. The EMBO Journal *18*, 3484–3490.

Jansen, J.G., Langerak, P., Tsaalbi-Shtylik, A., Berk, P. van den, Jacobs, H., and Wind, N. de (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J Exp Med *203*, 319–323.

Langerak, P., Krijger, P.H.L., Heideman, M.R., Berk, P.C.M. van den, and Jacobs, H. (2009). Somatic hypermutation of immunoglobulin genes: lessons from proliferating cell nuclear antigenK164R mutant mice. Phil. Trans. R. Soc. B *364*, 621–629.

Di Noia, J.M., Rada, C., and Neuberger, M.S. (2006). SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. EMBO J *25*, 585–595.

Phung, Q.H., Winter, D.B., Alrefai, R., and Gearhart, P.J. (1999). Cutting Edge: Hypermutation in Ig V Genes from Mice Deficient in the MLH1 Mismatch Repair Protein. J Immunol *162*, 3121–3124.

Rada, C., Ehrenstein, M.R., Neuberger, M.S., and Milstein, C. (1998). Hot Spot Focusing of Somatic Hypermutation in MSH2-Deficient Mice Suggests Two Stages of Mutational Targeting. Immunity *9*, 135–141.

Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002). Immunoglobulin Isotype Switching Is Inhibited and Somatic Hypermutation Perturbed in UNG-Deficient Mice. Current Biology *12*, 1748–1755.

Rada, C., Di Noia, J.M., and Neuberger, M.S. (2004). Mismatch Recognition and Uracil Excision Provide Complementary Paths to Both Ig Switching and the A/T-Focused Phase of Somatic Mutation. Molecular Cell *16*, 163–171.

Shen, H.M., Tanaka, A., Bozek, G., Nicolae, D., and Storb, U. (2006). Somatic Hypermutation and Class Switch Recombination in Msh6−/−Ung−/− Double-Knockout Mice. J Immunol *177*, 5386–5392.

Wiesendanger, M., Kneitz, B., Edelmann, W., and Scharff, M.D. (2000). Somatic Hypermutation in Muts Homologue (Msh)3-, Msh6-, and Msh3/Msh6-Deficient Mice Reveals a Role for the Msh2–Msh6 Heterodimer in Modulating the Base Substitution Pattern. J Exp Med *191*, 579–584.

Winter, D.B., Phung, Q.H., Umar, A., Baker, S.M., Tarone, R.E., Tanaka, K., Liskay, R.M., Kunkel, T.A., Bohr, V.A., and Gearhart, P.J. (1998). Altered spectra of hypermutation in antibodies from mice deficient for the DNA mismatch repair protein PMS2. PNAS *95*, 6953–6958.

Xanthoudakis, S., Smeyne, R.J., Wallace, J.D., and Curran, T. (1996). The redox/DNA repair protein, Ref-1, is essential for early embryonic development in mice. PNAS *93*, 8919–8923.