DISS. ETH NO. 17949

INVOLVEMENT OF MISMATCH REPAIR PROTEINS IN ACTIVATION-INDUCED DEAMINASE-MEDIATED ANTIBODY DIVERSIFICATION

A dissertation submitted to

ETH ZURICH

for the degree of
Dr. sc. ETH Zurich

Presented by

SILVIA SCHANZ

Dipl. Biol. Universität Freiburg im Breisgau

03.04.1978

Citizen of

GERMANY

Accepted on the recommendation of

Prof. Josef Jiricny
Prof. Hans Hengartner
Prof. Ruedi Aebersold
Prof. Matthias Wabl

2008

TABLE OF CONTENTS

1   SUMMARY

Antibody diversification in vertebrates is mediated by three metabolic processes: V(D)J recombination, somatic hypermutation (SHM) and class switch recombination (CSR). While V(D)J recombination occurs in a very early stage in the bone marrow and leads to the production of a primary repertoire of low-affinity antibodies, SHM and CSR are part of a maturation process resulting in high-affinity binding antibodies. This affinity maturation takes place in germinal centers, a special microenvironment within secondary lymphoid organs, upon activation of the B cells through antigen binding and T cell interaction. Both SHM and CSR are triggered by a single enzyme, activation-induced cytidine deaminase (AID), which deaminates cytidines in immunoglobulin gene loci and thus gives rise to U/G mispairs in DNA. AID is essential for the initiation of both SHM and CSR, but both processes are believed to be accomplished by the downstream processing of the deamination damage, rather than by AID itself. Recent genetic studies implicated base excision repair (BER) and mismatch repair (MMR) proteins in this processing. As canonical post-replicative MMR is responsible for the removal of biosynthetic errors from newly synthesized DNA, its involvement in a metabolic process leading to a high number of mutations (SHM) and double strand breaks and recombination (CSR) appears to be counterintuitive, unless the process involves a non-canonical BER and MMR, most likely in conjunction with novel interacting partners.

In order to gain new insights into the processes of SHM and CSR, we set out to search for AID interacting proteins, making use of several protein purification and interaction approaches followed by mass spectrometric analysis. While we did not observe a direct interaction between AID and MMR proteins, we identified a number of other proteins potentially interacting with AID. Among these, the helicases RuvB-like1/2 were identified and thus became a subject of further studies in the lab.

In order to investigate the interference of MMR and BER in the processing of AID-mediated U/G mispairs, we exploited an *in vitro* MMR assay system. Using heteroduplex DNA substrates containing several U/G mismatches, such as possibly arise through the processive deamination of cytosines by AID, we could show that a subset of uracil residues can act as initiation sites for MMR-dependent repair of

neighboring U/G mispairs. We furthermore showed that this uracil-directed repair is dependent on MutSα, while MutLα seems to be of minor importance.

Our data provide a novel insight into the SHM process and help explain some important genetic findings that have been puzzling the field for several years.

1.1 Zusammenfassung

Drei metabolische Prozesse sind in Vertebraten massgeblich an der Diversifizierung von Antikörpern beteiligt: V(D)J Rekombination, somatische Hypermutation (SHM) und Klassenwechsel-Rekombination ("class switch recombination", CSR). Während die V(D)J Rekombination bereits sehr früh im Knochenmark stattfindet und dadurch ein erstes Repertoire an schwach affinen Antikörpern erzeugt wird, sind SHM und CSR Schritte eines Reifeprozesses, der in hoch affinen Antikörpern resultiert. Diese Affinitätsreifung findet nach Aktivierung der B Zellen durch Antigenbindung und Interaktion mit T Zellen in Keimzentren, speziellen Mikrostrukturen innerhalb sekundärer Lymphorgane, statt. Sowohl SHM als auch CSR werden von einem einzigen Enzym initiiert, der "activation-induced cytidine deaminase" (AID), welches Cytidine innerhalb von Immunoglobulin (Ig) Genloci deaminiert und folglich U/G Fehlpaarungen in der DNA generiert. Obwohl AID sowohl für die Initiation von SHM als auch von CSR essentiell ist, sind auch die nachfolgend ("downstream") agierenden Proteine für die Prozessierung des Deaminationschadens massgeblich. Die Ergebnisse genetischer Studien ergaben kürzlich Hinweise auf eine Beteiligung zweier Reparaturmechanismen, "base excision repair" (BER) und "mismatch repair" (MMR) an SHM und CSR. Die konventionelle "Mismatch"-Reparatur ist verantwortlich für die Behebung von biosynthetischen Schäden in neu-synthetisierter DNA. Eine Beteiligung dieses Reparaturweges an einem metabolischen Prozess, der einerseits eine grosse Anzahl an Mutationen einführt (SHM) und andererseits zu Doppelstrangbrüchen und Rekombination (CSR) führt, erscheint deshalb nicht eingängig. Es sei denn der Prozess involviert BER und MMR Mechanismen, die auf neue Interaktionspartner zurückgreifen und damit nicht der Norm entsprechen.

Um neue Erkenntnisse über SHM und CSR zu gewinnen, haben wir uns in der vorliegenden Arbeit mit der Suche nach neuen Interaktionspartnern von AID

beschäftigt. Zur Anwendung kamen dabei unterschiedliche Proteinreinigungs-protokolle und Interaktionsssstudien-Ansätze, gefolgt von massenspektrometrischen Analysen. Wir identifizierten eine Reihe neuer, potentieller Interaktionspartner von AID, wobei direkte Interaktionen zwischen AID und MMR-Proteinen nicht beobachtet werden konnten. Unter den identifizierten Proteinen waren unter anderen die Helikasen RuvB-like1 und 2, welche daraufhin genauer untersucht wurden.

Um das Zusammenspiel zwischen MMR und BER bei der Prozessierung von AID-generierten U/G Fehlpaarungen genauer zu untersuchen, machten wir uns ein *in vitro* "MMR-Assay" System zu Nutze. Durch die prozessive Deamination durch AID entstehen mehrere U/G Fehlpaarungen im Ig Locus. Durch Einsatz eines DNA Substrates, das zwei solcher Fehlpaarungen enthält, konnten wir zeigen, dass einzelne Uracile als Eintrittsstelle für eine MMR-abhängige Reparatur benachbarter U/G Basenpaare fungieren können. Darüberhinaus konnten wir zeigen, dass diese von Uracil initiierte Reparatur abhängig ist von MutSα, wohingegen MutLα von geringerer Bedeutung ist.

Unsere Ergebnisse liefern damit neue Einblicke in den Prozess der SHM und helfen bedeutende genetische Studien zu deuten, die das AID-Gebiet seit mehreren Jahren beschäftigen.

2 INTRODUCTION

2.1 The immune system

The human body is surrounded by microorganisms and thus constantly defends itself against infection. Parasitic and infectious diseases currently cause about one third of all cases of death in the world, a number comprising more than all deaths of all forms of cancer combined.

More than 200 years ago, Edward Jenner, the originator of the concept of adaptive immunity, demonstrated that a patient that was inoculated with cowpox was subsequently protected against smallpox. This experiment is considered the origin of vaccination and has paved the way for successful elimination of the disease two centuries later. However, while some long known diseases could be eradicated, new infectious diseases are continually emerging.

Yet, despite the constant exposure to pathogens, humans hardly become ill. So how does the body defend against these microbes? What are the underlying mechanisms of successful elimination of the invaders, including bacteria, parasites, viruses and fungi?

The answer to all those questions lies in the fact that vertebrates are endowed with a elaborate collection of mechanisms known as the immune system. This system consists of many different types of tissues, cells and proteins interacting in a highly complex and dynamic network. To survive the challenge of pathogen invasion, sophisticated ways evolved to recognize, neutralize and erase the invaders. A crucial task that has to be carried out is to distinguish between self and non-self molecules. While self molecules are considered to be part of the organism's own body, non-self molecules are recognized as foreign by the immune system and are called antigens (*anti*body *gen*erators). Antigens bind to specific immune receptors and therefore initiate an immune response.

This points out the importance of a tight regulation of the immune system's activity. Failure to distinguish between the body's own and foreign substances, as well as between an invader's pathogenicity or harmlessness, will lead to diseases such as allergy and autoimmune diseases[1, 2].
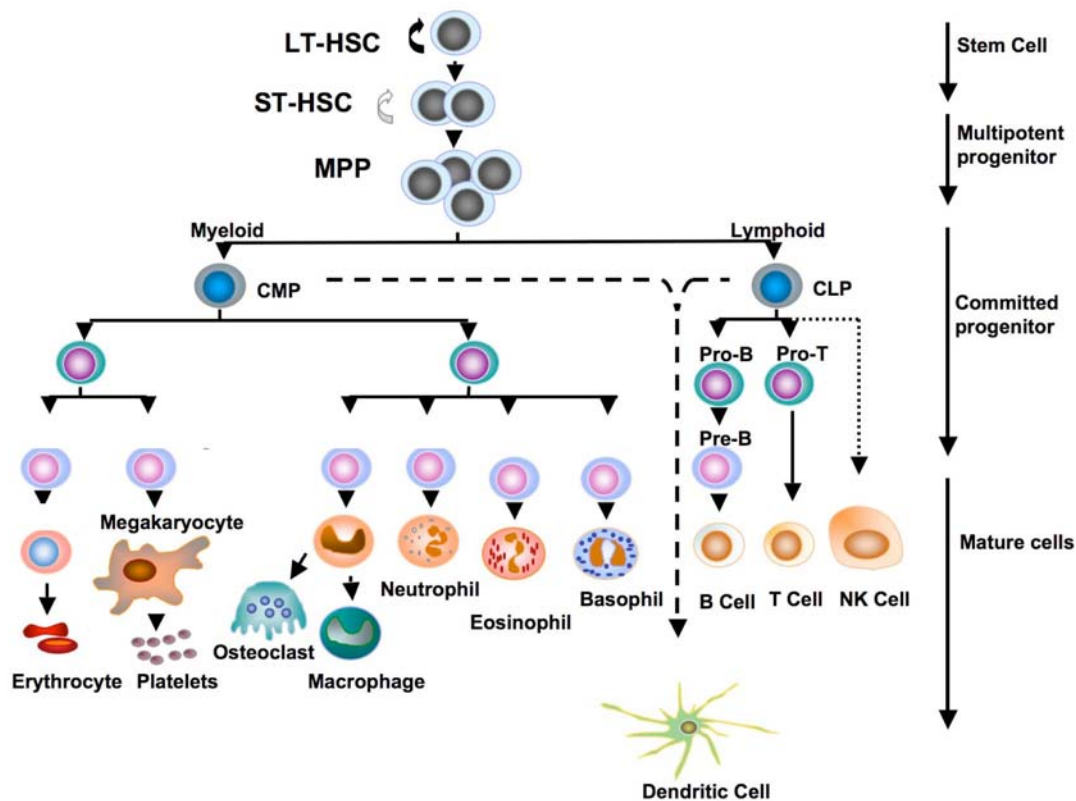
Fig. 2.1 The hematopoietic system. All cells of the blood arise from pluripotent hematopoietic stem cells (HSC) in the bone marrow. These cells exist in a long-term (LT) and short-term (ST) state and develop into multipotent progenitors (MPP). Those progenitors give rise to two types of stem cells, namely common myeloid progenitors (CMP) and common lymphoid progenitors (CLP). CMPs give rise to the myeloid lineage, which comprises the following leukocytes: erythrocytes, megakaryocytes, which generate platelets important in blood clotting, and granulocytes. Granulocytes include neutrophils, eosinophils, basophils and monocytes that develop into osteoclasts and macrophages. Dendritic cells can arise from both CMPs and CLPs. CLPs generally give rise to the lymphoid lineage of white blood cells or leukocytes – the B and T lymphocytes and the natural killer cells (NK). Adapted from[3].

A characteristic of the adaptive immune system is its immunological memory. An individual who has been exposed to a certain infectious agent once will immediately and efficiently create an immune response during future encounters with this pathogen. It possesses protective immunity against it. This process of acquired immunity forms the basis of vaccination.

The immune system is divided into the innate and adaptive. The innate immune system serves as a first, unspecific line of defense against pathogens. In contrast, the adaptive immune system is endowed with cells carrying highly diverse antigen-specific receptors enabling the system to recognize virtually any foreign antigen. Both parts of the immune system depend upon the activity of white blood cells or leukocytes, derived from the hematopoietic stem cells of the bone marrow (Fig. 2.1).

Early in development, stem cells give rise to all types of blood cells, such as red blood cells, platelets and the two main categories of white blood cells, the lymphoid and myeloid lineages. While the myeloid lineage comprises most of the cells of the innate immune system, the adaptive immune system is based on cells derived from the lymphoid lineage[1, 2].

2.1.1   Innate immunity

Most microorganisms that enter the body of a healthy individual are recognized and destroyed within minutes or hours by the non-specific defense mechanisms. The immune response occurs in a generic way and does not depend on the host's prior exposure to the pathogen. This immediate reply can be found in all multicellular organisms, be it in the plant or animal world and is referred to as the innate immune response.

One of the first reactions of the immune system to irritation or infection is inflammation. This process is initiated through the release of chemical factors such as cytokines and eicosanoids (prostaglandins, leukotrienes) by injured or infected cells. Prostaglandins produce fever and dilation of blood vessels, while leukotrienes attract certain white blood cells. The typical symptoms of an infection are swelling and redness that are caused by increased blood flow into a tissue.

2.1.1.1 Epithelial surfaces: the front line of host defense

Disease causing agents can enter the body at multiple sites and produce a variety of symptoms. As a first means of defense, the body exhibits physical barriers between the external world and the internal milieu through its surface epithelia. Epithelia span the skin and all linings of the body's tubular structures, namely the respiratory, gastrointestinal and urogenital tracts. Tight junctions hold together the epithelial cells and prevent easy entry by potential invaders. Thus, infection only occurs if pathogens can colonize and cross those barriers. While the skin with its dry surface provides a strong obstacle, microorganism uptake occurs most often via internal epithelial surfaces. However, interior epithelium is covered with a mucus layer providing a

certain protection against microbial and chemical insults. The slimy mucus contains secreted mucin and other glycoproteins that contribute to prevention of pathogen adherence. In addition, antimicrobial peptides are associated with microorganism killing or growth inhibition. Amongst the most prominent are the defensins, short peptides of 12-50 amino acids than can be found in all animals and plants. They work most likely through specific disruption of the pathogen's membrane integrity.

The importance of mucus flow and beating of epithelial cilia in infection clearing becomes obvious in individuals with disturbed ciliary movement and defective mucus secretion. This is the case in cystic fibrosis, where patients frequently suffer from lung infections due to the lack of pathogen clearance.

Apart from providing physical barriers, surface epithelia also produce chemical substances that are microbicidal. For example, the antibacterial enzymes phospholipase A and lysozyme are secreted in tears and saliva. Other examples of chemical barriers against infection are the acidic pH of the stomach and the digestive enzymes.

2.1.1.2 Cellular barriers of the innate system

Leukocytes are considered to be the second arm of the innate immune system and act like single-celled organisms. The most prominent innate leukocytes are phagocytes that include macrophages, neutrophils and dendritic cells. Other important components are mast cells, eosinophils, basophils and natural killer cells. While the latter cells belong to the lymphoid lineage, all aforementioned leukocytes are derived from the myeloid lineage of white blood cells.

Innate leukocytes are responsible for the initial sensing of the pathogens, their clearance through contact or engulfing and eventual killing.

Macrophages are the most efficient phagocytes and reside in tissues throughout the body, mainly in areas where infections are most likely to arise, including the lungs and the gut. Prominent places are also the liver, the spleen and the connective tissue. Macrophages literally mean "large eating cells" (Greek) and they are differentiated from phagocytic cells present in the blood called monocytes. Macrophages are long-lived cells that constantly circulate through the tissues of the body. Upon

encountering pathogens, bacterial molecules are bound to receptors on the surface of macrophages. This triggers the cell to actively engulf the invader through internalization in a membrane-enclosed vesicle known as "phagosome" or endocytic vacuole. Those phagosomes then fuse with so-called lysosomes. This leads to the destruction of the microbes through the generation of a "respiratory burst" involving the release of reactive oxygen species and enzymes from the lysosome. In addition, chemokines are produced and distributed by the macrophages that in turn attract other immune cells to help clear the infection.

Neutrophils, the second major family of phagocytes, are, in contrast to macrophages, short-lived and are abundant in the blood. They cannot be found in normal, healthy tissue. Neutrophils are recruited to sites of infection by molecules such as formyl-methionine-containing peptides released by the microbes or by activated macrophages. The mechanism of pathogen elimination is similar to that of macrophages.

Dendritic cells (DCs) are a third class of phagocytes and are so named due to their resemblance to neuronal dendrites, even though they have no connection to the nervous system. They are mainly present in the skin (here they are called Langerhans cells) and the inner mucosal lining of the stomach, intestines, lungs and nose. Dendritic cells are very important in antigen presentation and therefore serve as a link between the adaptive and the innate immune system. Conventional dendritic cells take up antigens in peripheral tissues where they get activated. They then travel to secondary lymphoid organs where they are the most potent stimulators of T-cell responses. However, follicular dendritic cells present antigens to B-cells in lymphoid follicles.

Eosinophils and basophils are related to neutrophils. These cell types are referred to as granulocytes, due to the presence of granules in their cytoplasm. They are also called polymorphonuclear cells (PMNs), due to their distinctive formed nuclei. Their granules contain a variety of toxic substances and free radicals that inhibit growth and eventually kill invading microbes. Eosinophils and basophils are primarily involved in the defense against parasitic infections and play a role in allergic reactions where they

might also cause tissue damage. The activation and toxin release of granulocytes is therefore tightly regulated, in order to prevent inappropriate tissue destruction.

Mast cells emerge throughout the connective tissue, mainly in the dermis and the submucosal tissues. Similarly to eosinophils and basophils, they possess large granules containing various mediator molecules, most prominently the vasoactive amine histamine. This explains also their crucial role in allergic reactions. Mast cells display high-affinity Fcε receptors that mediate binding to IgE monomers. Upon antigen-binding to this IgE, mast cell degranulation and activation is triggered, leading to a systemic and immediate hyper-sensitivity reaction.

Natural killer cells (NKs) are large granular, non-B, non-T lymphocytes belonging to the innate immune system. Their main task is to kill virus- and other intracellular pathogen-infected cells by inducing the infected cell to undergo apoptosis. Furthermore, they can attack and destroy tumor cells. To distinguish between healthy and non-healthy cells, they follow a process called "missing-self". This describes the situation arising in host cells after viral infection, where lower levels of the cell-surface marker MHC (major histocompatibility complex) are displayed. Natural killer cells are so named because they kill their target cells without preceding activation. Apart from contributing to innate immunity to viruses, they also play a role in antibody-dependent cell-mediated cytotoxicity (ADCC).

2.1.1.3 Pattern recognition receptors

When microorganisms penetrate the epithelial barricades, cells of the innate immune system enter the stage to mount a first line of defense. To successfully distinguish self from non-self, the innate immune system relies on the recognition of particular types of molecules that are normally displayed on the pathogen's surface but are absent in the host. These pathogen-associated molecular patterns (PAMPs) often occur as repeats on the pathogen surface and induce two types of immune responses: phagocytosis and inflammatory responses. They both occur immediately and without prior exposure to the specific antigen. The receptors dedicated to recognize those

11

pathogen-associated molecules are collectively called pattern recognition receptors (PRR). They appear either as membrane-bound receptors on the surface of host cells (as discussed below) or as soluble receptors in the blood (e.g. components of the complement system, see 2.1.1.4).

Pathogen-associated immunostimulants can be of various types. One of the most prominent examples is formylated methionine. Prokaryotic translation initiation differs from the one in eukaryotes in that formylated rather than unmodified methionine is used as the initiating amino acid. Thus, the appearance of formyl methionine-containing peptides in a host is a clear sign of microbe invasion and evokes a primary immune response. Also the composition of the outer surface of many microorganisms is clearly different to the one of their multicellular hosts. It includes flagella of bacteria, the peptidoglycan cell wall, as well as teichoic acids on Gram-positive and lipopolysaccharide (LPS) on Gram-negative bacteria. In addition, short sequences of bacterial DNA containing unmethylated repeats of the dinucleotide CpG can serve to discriminate between invading prokaryotes and their eukaryotic hosts.

Membrane-bound pattern recognition receptors arise in different versions. The macrophage mannose receptor is one example of receptors that recognize pathogen surfaces directly. It is a cell-bound C-type (calcium-dependent) lectin that binds to certain bacterial and viral surface sugars, including those found on the human immunodeficiency virus (HIV). Scavenger receptors are also membrane-bound, but are specialized in recognition of anionic polymers and acetylated low-density lipoproteins.

Other prominent pattern recognition receptors are the Toll-like receptors (TLRs). They belong to a family of evolutionarily conserved transmembrane receptors and function exclusively as signaling receptors. Toll-like receptors were originally discovered in the fruitfly *Drosophila melanogaster* for their role in embryonic development. In humans and mice there are only about 10 expressed TLR genes. But due to the fact that the TLRs have low specificity, they are able to detect most pathogenic microbes. Some TLRs are located on the cell surface, while others act intracellularly in the fight against microbes that have been taken up via endocytosis. Binding of pathogen-associated molecules to the TLRs initiates an intracellular signaling cascade that eventually activates the transcription factor NF-κB. This, on

one hand, stimulates the induction of various important mediators of innate immunity, e.g. cytokines and chemokines. On the other hand, it activates the expression of costimulatory-molecules that are essential for the induction of the adaptive immune response, such as B7.1 (CD80) and B7.2 (CD86) produced by tissue dendritic cells and macrophages in response to LPS-signaling through TLR-4.

In addition to membrane-bound receptors, the repertoire of pattern recognition receptors comprises also soluble receptors that circulate in the blood as part of the complement system.

2.1.1.4 The complement system

Complement was originally discovered as a heat-labile constituent of normal plasma. The complement system is a biochemical cascade comprised of about 20 soluble proteins that has been named after its capability to supplement the ability of antibodies to clear pathogens or label them for destruction by other cells. Although first discovered as an effector arm of antibody response, it is now widely accepted that it originated as part of the innate immunity system, where it can be activated early in infection even in the absence of antibodies. During innate immune response, pathogens are coated by complement that eventually facilitates their destruction.

Complement is activated either directly by pathogens or indirectly by pathogen-bound antibodies. It can be classified into three distinct pathways: the classical pathway, the lectin pathway and the alternative pathway. Pathogen-bound IgG or IgM antibodies activate the classical pathway, while the second pathway is initiated through interaction with mannan-binding lectin, a serum protein. The alternative pathway provides an amplification loop for the other two. Complement activation leads to a cascade of reactions that emerges on the surface of microbes and produces active components with multiple effector functions. The early events consist of a sequence of cleavage reactions where the cleaved products bind to the pathogen surface and activate the next component. The pathways converge at the formation of a C3 convertase enzyme. The resulting active C3b fragment attaches to pathogens, facilitates their recognition by specific complement receptors on phagocytic cells and finally leads to the engulfment of the pathogens opsonized by C3b.

The inflammatory, self-amplifying and destructive properties of the complement cascade necessitate that its activated key components are rapidly inactivated after they have done their job, so as to prevent it from spreading to nearby host cells. Deactivation is modulated by a system of regulatory proteins that terminate the activity of complement proteins either through direct binding or cleavage. In addition, complement factors are highly unstable and are rapidly degraded, unless they bind immediately to an appropriate component in the cascade or a nearby membrane[1, 4].

2.1.2 Adaptive immunity

The adaptive immune system is the branch of the immune system that protects against severe infection from parasites, fungi, viruses and bacteria. A human being born with a defective adaptive immune system will soon die, unless extraordinary measures are taken. This is e.g. the case for "scid" (severe combined immunodeficiency) patients that lack B-and T cells due to mutations in the RAG genes, necessary for the generation of T cell receptors and immunoglobulins (see chapter 2.2). The highly sophisticated adaptive immune response is called into play by the primary innate immune system.

Unlike the innate immune response, the adaptive one is highly specific to the particular antigen that induced it. It is based on the clonal selection from a repertoire of lymphocytes displaying highly-diverse antigen-specific receptors that allow the immune system to detect any pathogen. Besides eliminating the antigen, the adaptive immune system is also able to create long-lasting specific protective immunity that prevents reinfection. This is achieved through the generation of increased numbers of differentiated memory lymphocytes, allowing a more rapid and effective response should the same pathogen reinvade the host.

2.1.2.1 Lymphocytes and antigen-presenting cells

The main effector cells of the adaptive immune system are a type of leukocyte, called lymphocytes. B-cells and T-cells are the two major forms of lymphocytes, which are both derived from the same pluripotent hematopoietic stem cells (Fig. 2.1). They

continuously circulate through peripheral lymphoid organs, such as lymph node or the spleen, where lymphocytes are activated. Before their activation through antigen encounter, B-cells and T-cells are not distinguishable from one another. At this stage, they are known as naïve lymphocytes. Those that have bound their antigen differentiate further into fully-functional lymphocytes called effector lymphocytes. Overall, the human body has about 2 trillion (2×10^{12}) lymphocytes, depicting about 20 – 40% of all white blood cells. Their total mass is similar to that of liver or brain. Despite their abundance, the central role of lymphocytes was not revealed until the late 1950s. This is when experiments with irradiated mice showed that only lymphocytes could restore the adaptive immune response[5].

There are in principle two classes of adaptive immune responses: the antibody response carried out by the B-cells and the cell-mediated immune response conducted by the T-cells. Both will be described in more detail below. Common to both is the first step of adaptive immunity induction, which requires the activation of specialized, professional antigen-presenting cells (APC). These accessory cells are very efficient at internalizing antigens, either by receptor-mediated endocytosis or phagocytosis. The antigens are then fragmented and displayed, most likely complexed with major histocompatibility complex (MHC) molecules, on their surface, where it will be recognized by the effector cells.

Dendritic cells have the broadest range of antigen presentation and are probably the most efficient APCs. Other important antigen-presenting cells are macrophages and B-cells. In contrast to DCs and macrophages, B-cells do not present their antigens in conjunction with MHC molecules. Instead, they express surface antibody molecules and therefore they are able to very efficiently present the antigen to which their antibody is directed. But they are inefficient in presenting most other antigens. B-cells and dendritic cells (and to a lesser extent macrophages) are equipped with particular immunostimulatory receptors permitting an enhanced activation of T-cells. In addition to the most prominent APCs, similar specialized cells can be found in certain organs, e.g. Kupffer cells in the liver or microglia in the brain. They are derived from macrophages and are also able to effectively present antigens.

Antigens can be divided into exogenous and endogenous. Dendritic cells usually take care of exogenous pathogens, such as parasites, bacteria and toxins in tissue. Following internalization of the foreign particle, dendritic cells migrate to T-cell

containing lymph nodes with the help of chemotactic signaling. On their way, dendritic cells undergo a process of maturation, where they develop the ability to communicate with T-cells. The engulfed pathogen is converted to smaller fragments called "non-self" antigens and displayed on the cell surface through coupling to "self"-receptors called major histocompatibility complex (in humans also known as human leukocyte antigen (HLA)). Exogenous antigens are generally exhibited on MHC class II molecules. The MHC-II:antigen complex is then recognized by and activates CD4+ helper T-cells that are passing through the lymph node.

In contrast to exogenous antigens, endogenous antigens are produced inside the host cell after viral infection. Host enzymes digest the virally-associated proteins and the proteolytic peptides are displayed on the cell surface on MHC class I molecules. MHC-I:antigen complexes are usually recognized by CD8+ cytotoxic T-cells, leading to their activation.

2.1.2.2 T-cells and antigen recognition

In contrast to peripheral or secondary lymphoid organs where mature naïve lymphocytes are maintained and adaptive immune responses are initiated, central or primary lymphoid organs are the places where lymphocytes are generated. Both T and B lymphocytes descend from the bone marrow. While this is the location where B cells mature, T cells travel to the thymus, an organ in the upper chest after which they are named, and mature there. After successful completion of their maturation, lymphocytes enter the blood stream and circulate through the peripheral lymphoid tissue until they get activated by their specific antigen.

In contrast to B cells that recognize intact antigens, T cell activation depends on the interaction with fragments of the original pathogen that has been partially degraded inside the APC. A second major difference to B cells, which secrete antibodies and can reach their targets far away from their site of activation, is that T cells only work at short range. They either perform their duty directly within a secondary lymphoid organ or after they have migrated to the site of infection. The limited action sphere is a result of the fact that T cells need direct interaction with their target cells to either signal or kill them.

Antigen recognition receptors made by T cells, unlike antibodies made by B cells, exist only in membrane bound forms. This is what made them difficult to isolate and it was not until the late 1980s that they were biochemically described and identified. T cell antigen receptors are composed of two disulfide-linked polypeptide chains, called TCRα and TCRβ. Each chain contains two immunoglobulin like domains, one variable and one constant. Their three-dimensional structure determined by X-ray diffraction is very reminiscent of one arm of a Y-shaped antibody molecule.

The pools of gene segments encoding the TCR chains are located on different chromosomes and are brought together by site-specific recombination during T cell development in the thymus. This process is similar to that used by B cells to generate antibody diversity and will be described later (see 2.2.1).

While most T cell receptors are made up of α and β chains, a small minority of T cells produces a different but related type of receptor heterodimer. It is composed of γ and δ chains and can be found in T cells that arise early in development. Those cells occur mainly in epithelia (e.g. in gut and skin) and their function is unclear.

Even though the variable antigen-binding chains of T cell receptors have exquisite specificity for their respective antigen they do not possess intrinsic signaling capacity. In the fully functional antigen-receptor complex they are thus tightly associated with a number of variant membrane-bound accessory proteins that initiate signaling upon antigen binding to the cell surface receptor. The binding is thus transferred to a signal that is passed from the cell surface to the cell interior. The α:β TCR heterodimer recognizes and binds its peptide:MHC ligand, whereas four other chains (δ, two ε, γ) are responsible for signaling to the nucleus. They are collectively called CD3 co-receptor complex and are additionally required for cell-surface expression of the antigen-binding chains. Besides CD3 and the α:β heterodimer, a homodimer of ζ chains that is mostly located within the cell completes the T cell receptor complex.

There are two major classes of T cells: cytotoxic T cells and helper T cells. Already at the end of their development they are distinguished by individual cell-surface markers. Those markers are not random, but define the T cell's function and determine their interactions with other immune cells. Cytotoxic cells that directly kill their target cells carry CD8 molecules on their surface. The class of T cells that helps activate their targets rather than directly attacking them displays CD4 receptors.

CD4 and CD8 are accessory receptors that have a direct role in activating T cells by generating their own intracellular signals and are therefore called co-receptors. They are both single-pass transmembrane proteins with extracellular Ig-domains. Like TCRs they perceive MHC molecules, however, unlike TCRs, they bind to nonvariant parts of the MHC proteins, far away from the peptide-binding groove.

T cells only recognize their antigen in association with self-MHC proteins, but not in association with foreign-MHC molecules. That means that T cells show MHC restriction. This traces back on a process of positive selection during T cell development in the thymus. That is where T cells capable of detecting foreign peptides presented on self-MHC molecules are selected, whereas others that would be of no use are eliminated.

Cytotoxic T cells, also known as killer T cells, are a subgroup of T cells that provide protection against intracellular pathogens such as viruses, bacteria and parasites that multiply within the host where they are sheltered from attack by antibodies. Protection is guaranteed through direct killing of the infected cell before the microbes have a chance to proliferate and spread to neighboring cells. The attack is achieved with great precision, sparing adjacent cells and thus minimizing tissue damage. Cytotoxic cells operate through different killing strategies, all of which function by inducing the target cell to kill itself by undergoing apoptosis.

Three types of preformed cytotoxic proteins are involved. One involves perforin, a pore-forming protein that is homologous to the complement component C9 and stored in secretory vesicles. Perforin is released by local exocytosis at the point of contact with the target cell, whereupon it polymerizes in its plasma membrane to form transmembrane channels. This allows access of granulysin and serine proteases, which are stored in the secretory vesicles, to the infected cell. One example is granzyme B that cleaves and thereby activates members of the caspase family of proteins that mediate apoptosis.

Another killing strategy also activates the death-inducing caspase cascade in the target cell, but does it less directly. This pathway involves the Fas ligand, a homotrimeric protein on the cell surface of cytotoxic cells. Binding of Fas ligand to Fas, a transmembrane protein on target cells, triggers the cascade and eventually leads to the killing of the cell.

Cytotoxic T cells also act by releasing cytokines, such as TNF-α, LT-α and IFN-γ. IFN-γ directly inhibits viral replication and activates macrophages. It also induces the expression of MHC class I molecules and therefore provides a positive feedback loop for eradication of the infected cells.

Helper T cells, or CD4 lymphocytes, are immune mediators crucial for controlling the adaptive immune system. They have no direct killing function and have no phagocytic or cytotoxic activity. Instead, they "manage" the immune response by directing other cells to perform these tasks. In contrast to cytotoxic cells, they protect us against both intracellular and extracellular pathogens.

When activated by an antigen-presenting cell, T cells secrete a variety of cytokines and display co-stimulatory proteins on their surface. Naïve T cells require at least two signals for activation that are both delivered by an antigen-presenting cell, usually a dendritic cell. The first signal is provided by the binding of the MHC-peptide complex to T cell receptors, while the second signal is essentially delivered by the B7 co-stimulatory protein binding to CD28 on the T cell surface. A naïve lymphocyte differentiates into either of two distinct types of effector helper T cells, namely T_H1 and T_H2. T_H1 effector cells help activate macrophages to eliminate intracellular pathogens proliferating within phagosomes and trigger cytotoxic T cells to kill infected cells. In contrast, T_H2 cells help stimulate B cells to produce antibodies that eliminate extracellular pathogens and their toxic products.

T helper cells activate their target cells by a combination of membrane-bound and secreted signal proteins. The membrane-bound signal is usually the CD40 ligand. B cells, like T cells, depend on two simultaneous signals for full activation. Signal 1 is provided by antigen-binding to the B cell antigen receptor, while signal 2 is delivered by T_H2 cells in the form of CD40 ligand and various cytokines.

A third, recently-identified subset of CD4 effector cells are the T_H17 cells and a fourth and last, minor subclass of T lymphocytes is constituted by the regulatory T cells (T_{reg}) that suppress and limit the immune system. They are able to control aberrant immune responses to self-antigens and therefore crucially participate in controlling the development of autoimmune diseases[1,2].

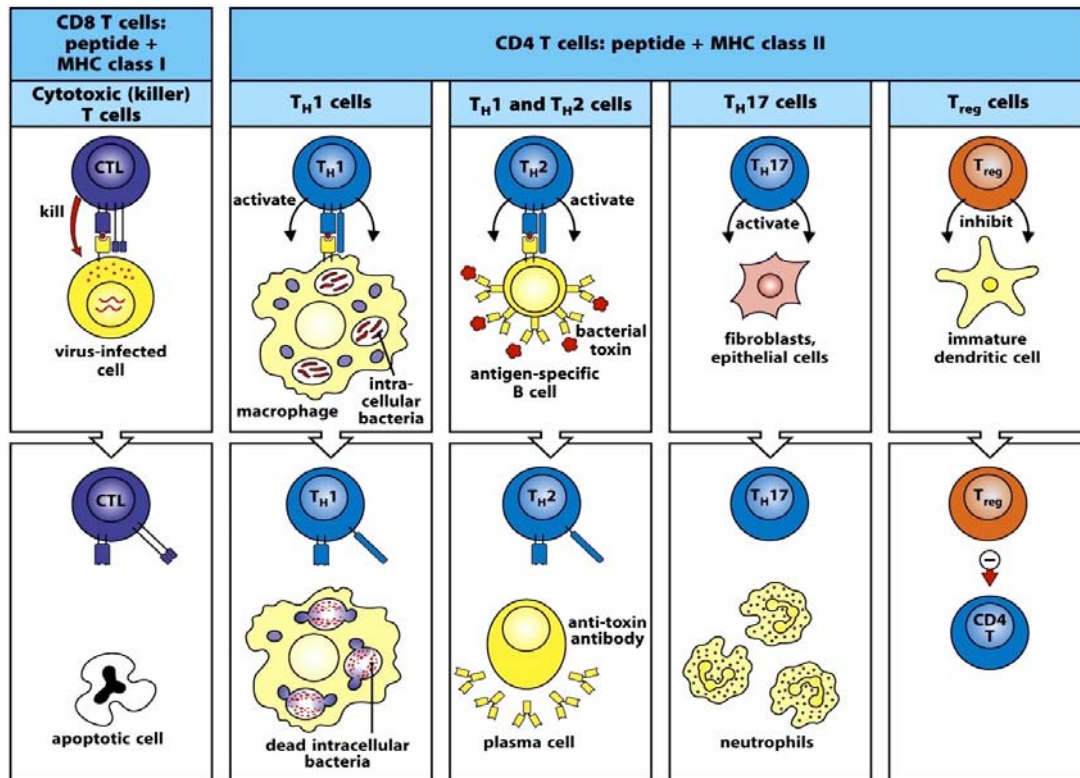For an overview of cytotoxic and helper T cells see Fig. 2.2.

Fig. 2.2 Scheme illustrating the characteristics of CD8 cytotoxic T cells and T_H1, T_H2 and T_H17 CD4 effector T cells. CD8 cytotoxic cells (left panels) initiate killing of virus-infected cells, which display peptide fragments of cytosolic pathogens bound to MHC class I molecules on their cell surface. T_H1 cells (second panels) and T_H2 cells (third panels) both express the CD4 co-receptor and recognize fragments of antigens degraded within intracellular vesicles, displayed at the cell surface by MHC class II molecules. T_H1 cells activate macrophages through cytokines, enabling them to destroy intracellular microorganisms more efficiently. They can also activate B cells to produce strongly opsonizing antibodies belonging to certain IgG subclasses. T_H2 cells, in contrast, produce cytokines that drive B cells to differentiate and produce immunoglobulins of other types, especially IgE, and are responsible for initiating B-cell responses by activating naïve B cells to proliferate and secrete IgM. T_H17 cells (fourth panels) are a recently recognized subset of CD4 effector T cells. They induce local epithelial and stromal cells to produce chemokines that recruit neutrophils to site of infection early in the adaptive immune response. The remaining subset of effector T cells are the regulatory T cells (right panels), a heterogeneous class of cells that suppress T cell activity and help prevent the development of autoimmunity during immune responses. Adapted from[1].

2.1.2.3 B-cells and antibodies

In the absence of fully-functional immune responses, vertebrates would die of infectious diseases. Many pathogenic bacteria proliferate in extracellular spaces of the body and many intracellular pathogens use the extracellular fluids to spread from one cell to another. Those extracellular spaces are protected by the humoral immune response. Antibodies, exclusively synthesized by B cells are the main effectors of the humoral response. They are responsible for preventing the spread of intracellular infections and carry out the destruction of extracellular microorganisms.

B cells originate from the lymphoid lineage derived from pluripotent hematopoietic stem cells. At a very early stage in the bone marrow they undergo a maturation process from pro-B cells to pre-B cells and then to immature B cells that display full antibody molecules on their cell surface. Further maturation occurs after migration to peripheral lymphoid organs. The activation of a naïve B cell is then initiated through antigen binding and T helper cell interaction, as described earlier (see 2.1.2.2). Subsequently, the activated B cell differentiates into an antibody-secreting plasma cell or a memory B cell (Fig. 2.3).
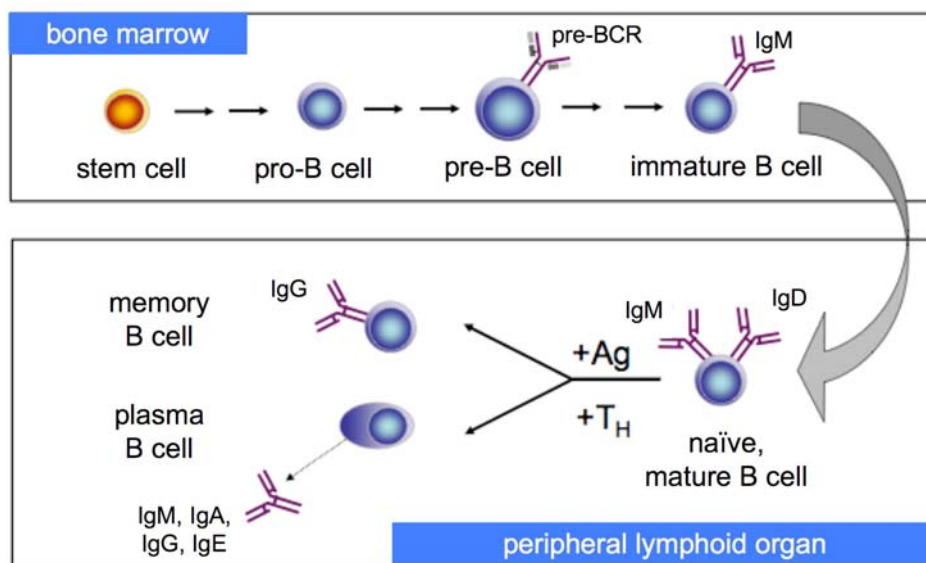

Fig. 2.3 The maturation pathway of B cells. B cells arise from the lymphoid lineage derived from hematopoietic stem cells. In a very early stage in the bone marrow these cells start to differentiate into pro-B cells and pre-B cells, expressing a pre-B cell receptor (pre-BCR). Eventually the cell develops into an immature B cell that exhibits a functional IgM molecule on its surface. After migrating to peripheral lymphoid organs further maturation occurs and the B cell gets activated. Activated B cells undergo rounds of mutation and selection for higher-affinity mutants, resulting in high-affinity antibody-secreting plasma cells and high-affinity memory B cells. Adapted from[6].

As predicted by the clonal selection theory[7], all antibodies produced by one individual plasma B cell have the same antigen specificity. The first antibodies that are created are not secreted, but rather inserted into the plasma membrane, where they serve as antigen receptors. A B cell generally has about 10^5 such receptors in its plasma membrane. After encountering its specific antigen, the B cell starts to secrete antibodies while it is still a small lymphocyte. At the end stage of its maturation pathway however, it is a large plasma cell that continuously secretes antibodies at a rate of 2'000 per second[8].

The activation of B cells through their interaction with helper T cells initially occurs at the border of T cell and B cell areas of secondary lymphoid tissues, such as lymph nodes. That is where both types of lymphocytes are trapped following binding to their specific antigens. They then migrate into the B cell zone or follicle and form germinal centers, where further interactions take place (Fig. 2.4).
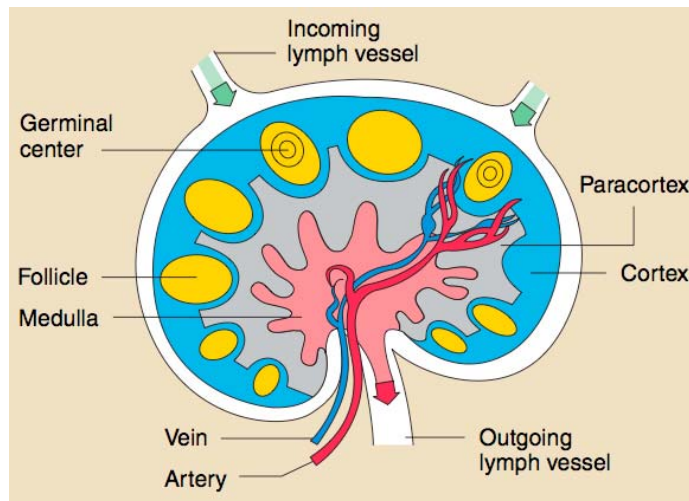

Fig. 2.4 Schematic representation of a lymph node. T cells concentrate in the paracortex, B cells in the cortex and plasma cells in the medulla. B cells are first activated at the border of T- and B cell zones when they encounter both, their specific antigen and activated T cells. They migrate to germinal centers, where an affinity maturation occurs. Mutations that result in low-affinity binding B cell receptors lead to the elimination of the B cell. Thus the germinal center is a site of massive cell death. Some mutations, however, will improve the ability of the B cell receptor to bind antigen. These cells survive and undergo repeated cycles of mutation and selection during which some of the progeny B cells undergo differentiation to either memory B cells or plasma cells and leave the germinal center. Adapted from[9].

Successful activation of B cells requires binding of the antigen to the surface immunoglobulin, the B cell receptor, and, as for T cells, to a co-receptor. The B cell co-receptor complex is composed of three proteins: CD19, CD21 and CD81. Interactions with co-receptor proteins activate intracellular signaling pathways and amplify the antibody production. In addition co-stimulatory molecules are increasingly expressed on the B cell surface, therefore facilitating and eliciting T cell help. Those cells recognize antigen fragments displayed by the B cell as peptide:MHC class II complex. Stimulation then occurs through interaction between CD40 on the B cell and CD40 ligand on the T cell. Further increase of activation is achieved through other TNF/TNF-receptor family ligand pairs and by the targeted release of cytokines. In turn, B cells also provide signals to T cells, e.g. through B7-proteins, to obtain continuous support by the T helper cells.

The class of cytokines released by the T cells determines not only the differentiation of a naïve B cell into an antibody-secreting B cell or memory cell. Cytokines also impair and bias the antibody class and its antigen specificity through affecting class switch recombination and somatic hypermutation, respectively. Those two processes will be described in more detail later (2.2.3, 2.2.4).

Antibodies are produced in billions of variations, each endowed with a different amino acid sequence and different antigen-binding site. They are collectively called immunoglobulins (abbreviated as Ig) and are amongst the most abundant protein components in the blood. They account for about 20% of the total protein in plasma by weight. Antibodies contribute to immunity in several ways. One of these is a process called neutralization, in which the binding of pathogens to specific molecules on the target cell surface is inhibited by the pathogen/antibody interaction. Antibodies also protect against bacteria that multiply outside of cells by facilitating the uptake of the pathogen by phagocytes. The process of coating the surface of a microbe with antibodies is called opsonization. Coating of pathogens with antibodies can additionally activate proteins of the complement system by the classical pathway. Complement receptors on phagocytic cells bind to complement proteins bound to the pathogen that eventually leads to its uptake and internalization.

Antibodies are Y-shaped molecules that carry two identical antigen binding sites at the tips of the Y (Fig. 2.5)[10]. Binding sites for various cell-surface receptors and complement components are located at the tail of the Y. An antibody, or immunoglobulin, consists of four polypeptide chains, two identical light and two identical heavy chains. Each chain is organized in a number of immunoglobulin (Ig) domains, consisting of β sheet structures of about 110 amino acids. A light chain possesses one variable (V_L) and one constant (C_L) domain, while a heavy chain exhibits one variable (V_H) and three constant (C_H) domains. The variable parts of all four chains contain small hypervariable regions (complementarity-determining regions, CDRs) that are brought together in the three dimensional structure to form the antigen-binding site.
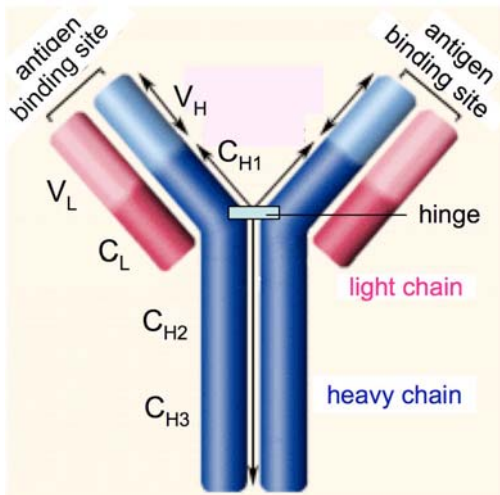
Fig. 2.5 Structure of an antibody molecule (Ab). Each molecule consists of four chains, two identical heavy chains and two identical light chains. A light chain comprises one variable domain (V_L) and one single constant domain (C_L). In contrast, heavy chains comprise one variable domain (VH), but three constant domains (C_H1, C_H2, C_H3). The variable domains of all four peptide chains form the antigen binding site that is located at the N-terminus of the Ab. The C-terminus is composed of C_H2 and C_H3 of the heavy chains and defines the effector function of the Ab. The peptide chains are connected with each other *via* a flexible hinge region.

Antibodies are produced in two different forms: they are either membrane-bound, where they serve as B cell receptors, or soluble and secreted when they act as mediators of the adaptive immune response. The sole difference is the C-terminus of the heavy chain. The membrane-bound molecules contain a hydrophobic domain that facilitates anchoring in the plasma membrane, whereas secreted antibodies lack the transmembrane domain due to alternative splicing, allowing the escape from the cell.

In humans, there are five antibody classes (IgM, IgD, IgG, IgA, IgE), all endowed with a distinct heavy chain (µ, δ, γ, α, ε) and mediating a different characteristic biological response. The heavy chains form the antibody tails (F_c region) that define the biological effector functions of the antibodies and determine which other proteins will interact with the different antibodies. Light chains exist in two forms (κ and λ) that can associate to either type of heavy chain. In contrast to heavy chains, light chains do not influence the antibody's effector function, but determine the antigen specificity of the antibody molecule.

2.1.2.4 Immunological memory

We are able to develop lifelong immunity to various common infectious diseases after initial exposure to the antigen. This is due to the fact that the immune system can remember prior infections. When we are first exposed to a specific antigen, a primary immune response is triggered. This is characterized by a delay of several days, a rapid

and exponential rise of antibodies followed by a gradual decline. If the same pathogen reinvades the body after weeks, months or years, a secondary immune response is triggered, which is faster and the response is more intense, which is indicative of an antigen-specific immunological memory.

When naïve B cells first encounter their specific antigens, they differentiate into either effector or memory cells. Effector B cells are actively engaged in the secretion of antibodies and the clearance of extracellular pathogens. In contrast, memory cells are not engaged in the response, but are endowed with the ability to be more easily and more quickly induced to become effector B cells when they later encounter the same antigen. Similarly to naïve B cells, memory cells can then differentiate into effector cells or give rise to more memory cells.

Immunological memory is considered to be adaptive, because the immune system prepares itself for future challenges. It can be divided into passive and active memory. Passive memory is generally short-term and lasts for only up to a few weeks. One example is the transfer of the mother's antibodies across the placenta to the fetus. Another one is the artificial transfer of antibody-rich serum from one individual to another. Active memory however is typically long-term or life-long and can be naturally acquired by infection followed by B cell or T cell activation, or artificially through vaccination[2, 4].

2.2 Antibody diversification

Even in the absence of antigens, a human body can produce more than 10^{11} different antibody molecules. This is referred to as the pre-immune antibody repertoire. It is large enough to comprise antigen binding sites that recognize almost all antigenic determinants, albeit with low affinity. After repeated stimulation through the same antigen, affinity maturation takes place, leading to an immunoglobulin repertoire of antibodies with much higher affinity for the antigen.

How do lymphocytes manage to generate an almost infinite number of antigen binding receptors from a limited number of genes? One possibility is through "combinatorial diversity", which is achieved by different combinations of heavy and light chain. Given the fact that light and heavy chains are each encoded by 1'000 genes, this results theoretically in 1'000 x 1'000 (10^6) possibilities to create antibodies. Furthermore, the human body has evolved unique genetic mechanisms to produce an almost unlimited number of specific light and heavy chains through a remarkably economical way. Separate gene segments encoding the variable region of the antibody chains are joined together in many different combinations before their transcription ("recombinatorial diversity").

First evidence for the existence of DNA rearrangements during B cell development came in the 1970s from experiments in mice. While the sequences for the V-region and the C-region were still present on two different DNA restriction fragments in an embryonic cell, they were localized on the same DNA fragment in an adult tumor cell, indicating that the DNA has been rearranged[11].

2.2.1 V(D)J recombination

V(D)J recombination is a specialized double-strand break and rejoining pathway that generates the exon encoding the variable domain of T cell receptors and immunoglobulins. These joining events in pre-T and pre-B cells form the basis of the antigen-specific adaptive immune system in vertebrates.

During V(D)J recombination, the DNA sequences encoding V regions are generated by rearrangements of a relatively small number of inherited gene segments.

The V domain of the light chain is encoded by two separate DNA segments, where the first one is called variable (V) (first 95-101 amino acids) and exists in multiple copies. The second one is referred to as the joining (J) segment (encoding up to 13 amino acids) and exists only in a few copies. The heavy chain comprises a third gene segment named diversity (D), which lies between the V and J segments.

As for the light chain, the V segment that is initially located relatively far away from the C region is joined to the J segment. Since the J segments are located in close proximity to the C region, this rearrangement brings together the V, J and C regions. The process in heavy chain rearrangement is very similar, except that the D and J segments are combined before the V segment joins them to form a complete V-region exon. For both, light and heavy chain genes, RNA splicing is responsible for joining the assembled V-region gene to the now neighboring C-region sequence (Fig. 2.6).
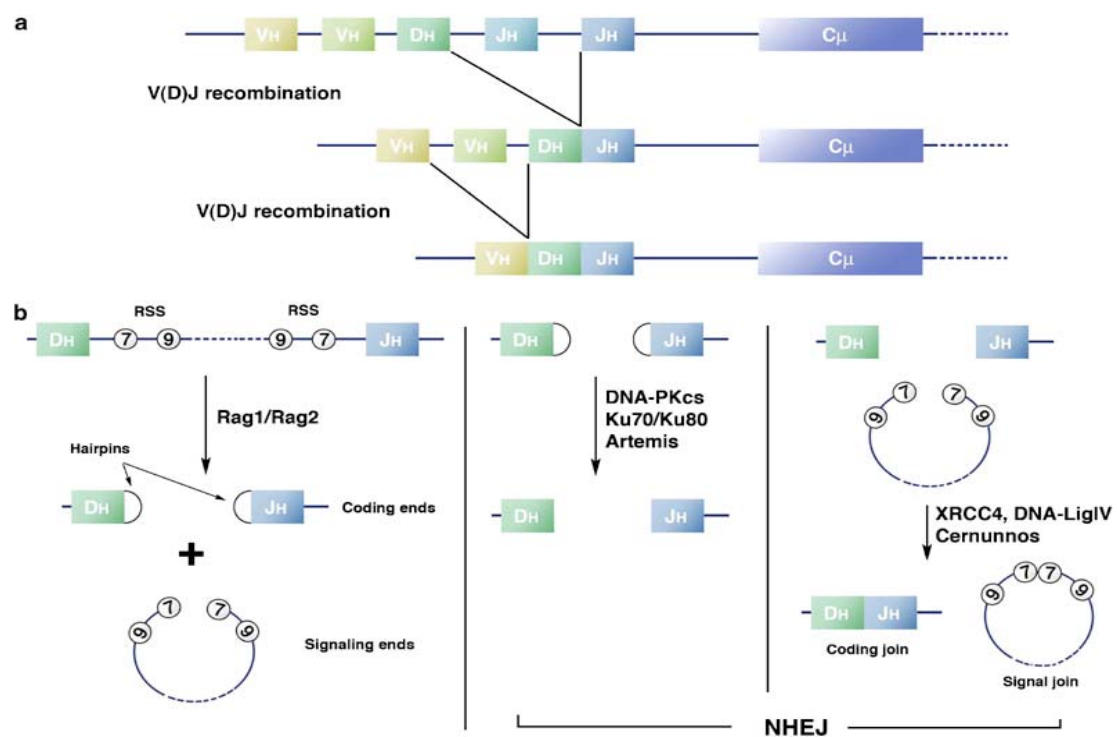
Fig 2.6 V(D)J recombination. **a** Schematic representation of the IgH chain locus and V(D)J recombination process. **b** The V(D)J reaction can be divided into three steps. The Rag1/2 complex introduces a DNA double strand break at the border between D_H and J_H segments and their respective recombination signal sequences (RSS), creating hairpin-sealed coding ends and blunt signal ends. Artemis, which is recruited and phosphorylated by the Ku/DNA-PK complex, opens the hairpins through its endonuclease activity. The XRCC4/Cernunnos/DNA-LigaseIV complex finally seals coding and signal joins. NHEJ, non-homologous end-joining; IgH, immunoglobulin H. Adapted from[12].

The rearrangement of V, D and J segments is locally controlled by the use of conserved noncoding DNA sequences flanking the sites at which recombination takes place. They are called recombination signal sequences (RSSs) and are composed of conserved heptamer and nonamer sequences separated by 12 or 23 base pair (bp) spacers. The 12/23 rule implies that gene segments with 12 bp spacers can only be joined to those with 23 bp spacers thus inhibiting the joining of two segments of the same kind.

The process of somatic V(D)J recombination involves a complex of enzymes that is composed of both lymphocyte-specific and ubiquitous DNA-modifying proteins. The lymphocyte-specific proteins of the so-called V(D)J recombinase enzyme complex are encoded by two closely-related genes named *RAG-1* and *RAG-2* (*rag* = recombination activating genes). Expression of these genes is exclusively restricted to developing lymphocytes while they are assembling their antigen-receptors. RAG-1 and RAG-2 introduce double-strand breaks at the flanking DNA sequences. This is followed by a rejoining process mediated by the RAG proteins themselves and enzymes involved in general non-homologous DNA end joining repair (NHEJ). Amongst these are the heterodimer Ku (Ku70:Ku80) that forms a ring around DNA, and DNA-dependent protein kinase (DNA-PK). DNA-PK is known to be involved in the repair of double-stranded DNA and has a serine/threonine protein kinase activity when bound to DNA ends. It recruits and phosphorylates several proteins, such as Artemis, which possesses nuclease activity and cleaves the hairpin loop initially formed by the RAG proteins. The DNA ends are then finally joined together by the XRCC4/DNA ligase IV complex. The intervening DNA sequences are looped out, leading to a loss of genetic information.

While site-specific recombination processes are usually precise, this is not the case during the joining of antibody gene segments. Instead, nucleotides are often lost from the ends of the recombining gene fragments or random nucleotides are inserted by the lymphoid-specific enzyme terminal deoxynucleotidyl transferase (TdT). This is called junctional diversification and, together with the recombinatorial diversity achieved through differential rearrangement of various gene segments, it enormously increases the diversity of V-region coding sequences[1, 12].

2.2.2 Activation-induced cytidine deaminase (AID)

The encounter with specific antigens triggers a secondary antibody diversification process in B cells. Somatic hypermutation (SHM), gene conversion (as it occurs in chicken and rabbit[13, 14]) and class switch recombination (CSR) are the mechanisms that are mainly involved in this affinity maturation cascade[15-18].

All three processes are initiated through targeted deamination of deoxycytidine residues, leading to deoxyuridine and thus transforming C/G pairs into U/G mispairs. The catalyst of this first step is activation-induced cytidine deaminase (AID), an enzyme that is exclusively expressed in activated B lymphocytes[17].

AID was first identified in 1999 by Honjo and colleagues in a subtractive hybridization screen for cDNAs that were induced during cytokine-mediated activation of switch recombination in a mouse B cell line[17]. Due to its striking homology to APOBEC1, the catalytic component of a tissue-specific RNA editing complex, AID was first proposed to work as an RNA-editing enzyme on mRNA molecules[16, 17, 19]. However, later studies revealed that AID acts directly by deaminating dC to dU on single-stranded DNA[20-26]. Its role in SHM and CSR was confirmed by studies in AID deficient mice and humans that are viable, but suffer from type II hyper-IgM immunodeficiency syndrome (HIGM-2) and are unable to carry out SHM and CSR[16, 27, 28].

The regulation of AID gene expression by CD40 ligand stimulation and IL-4 was studied in human and mouse B cells[29]. CD40 ligand stimulation induces NFκB binding to two sites in the 5' upstream region of the AID gene, while IL-4 induces STAT6 binding to one site in that region. The two stimuli strongly synergize in inducing AID mRNA and protein expression.

AID was shown to act as a single-strand specific deaminase. Evidence therefore came from biochemical studies that proved that AID deaminates dC to dU on ssDNA, but not on double-stranded DNA (dsDNA), DNA:RNA hybrids or RNA in any form[30-32]. Deamination on dsDNA will only occur during ongoing transcription or when single-stranded bubbles are otherwise introduced[32-38].

AID was described to act as a dimer carrying out processive C-to-U-deaminations prior to dissociation from DNA[31, 32, 39]. Preferred sites of mutations lie within a WRCY (W=A/T, R=A/G, Y=C/T) consensus sequence with AGCT being the preferred DNA motif. Biochemical data suggest that AID binds randomly to ssDNA

in search of C residues by jumping and sliding along the DNA backbone (Fig. 2.7). However, it is not clear whether AID acts processively also *in vivo*, where directional deamination might be imposed by the 5'-3'-directionality of RNA polymerase II (RNA Pol II) that was shown to associate with AID[40].
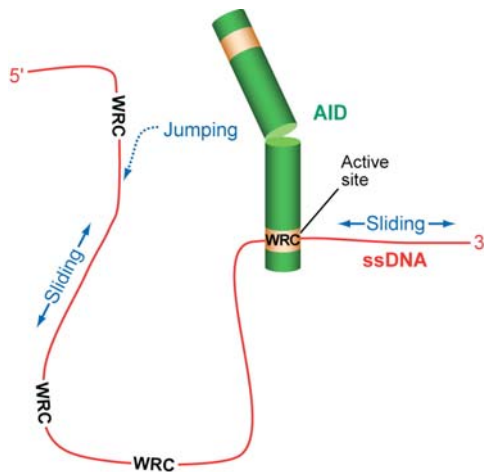

Fig 2.7 Model describing processive C-to-U deaminations. AID is depicted as a dimer on ssDNA. Current biochemical data suggest that AID binding occurs randomly, and enzyme motion, for example sliding and jumping, occurs in either direction along the ssDNA substrate. Deamination by AID occurs processively (making multiple deaminations per substrate molecule) and equally in 5' and 3' directions, with preferential targeting to WRC motifs. Notably, there is no external energy source present, for example ATP or GTP hydrolysis. Adapted from[41].

Elucidation of the three-dimensional structure of AID has been complicated by the difficulty to produce active protein in high quantity. However, sequence comparisons with other deaminases have revealed it to be part of the APOBEC family of polynucleotide cytidine deaminases, with AID exhibiting the greatest homology to APOBEC1. Both proteins contain a canonical cytidine deaminase motif, with key histidine and cysteine residues used for zinc coordination and catalysis[42, 43]. The N-terminal part of AID is positively charged, enabling tight binding to the negatively-charged DNA, and carries a putative bipartite nuclear localization signal (NLS)[42] (Fig. 2.8). In contrast, the C-terminus harbors a leucine-rich nuclear export signal (NES) that is mainly responsible for the predominantly cytoplasmic distribution of AID[42-44]. Mutational studies identified further functional differences between the two termini: N-terminal AID deletion mutants have been observed to be poor hypermutators while retaining CSR capability, while C-terminal deletion mutants are SHM-competent, but lack the ability to perform CSR[45, 46]. The latter observation was first made when mutations in the last 30 amino acids of the C-terminus where identified in patients suffering from the hyper-IgM II syndrome[45-47]. Taken together, this led to the speculation that the N-terminus of AID engages in interactions with proteins that are essential for SHM, while the C-terminus recruits cofactors that are indispensable for

CSR. However, it is currently unclear whether the interactions with MDM2 (a protein that controls p53) and DNA-PK (involved in DNA double-strand break repair) with the C-terminus of AID are of physiological relevance[48-50].
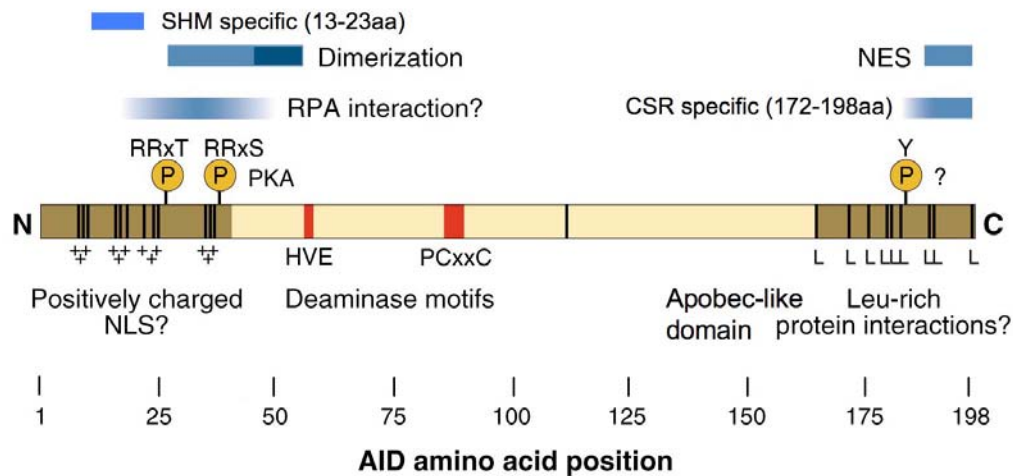

Fig. 2.8 Organization of activation-induced deaminase (AID) functional domains. AID is drawn to scale with the N- and C-terminal ends indicated by N and C, respectively. The positions of positively charged amino acids in the N terminus and leucine (Leu) residues in the C terminus are indicated by black vertical lines with + and L symbols, respectively. Both regions are in brown with the suggested function indicated (NLS, nuclear localization signal). Red boxes indicate the deaminase motifs that are likely implicated in Zn coordination, with the conserved amino acids typed below (H, histidine; V, valine; E, glutamic acid; P, proline; C, cysteine). In the upper part, blue bars indicate the approximate position of known AID functional domains. Abbreviations: SHM, somatic hypermutation; RPA, replication protein A; NES, nuclear export signal; CSR, class-switch recombination. Gold circles indicate consensus phosphorylation sites for protein kinase A (PKA) as well as a tyrosine (Y) that has been found to be phosphorylated *in vivo* (R, arginine; T, threonine; S, serine). Adapted from[21].

One of the most important questions concerning AID is its targeting to the Ig locus. Most researchers support the idea that this process partially depends on proteins specifically interacting with AID. Two proteins, replication protein A (RPA) and protein kinase A (PKA) alpha-regulatory subunit (PKAr1α), have been shown to directly interact with AID and might provide key insights into the mechanism by which AID accesses its ssDNA substrate[51-53]. Upon PKAr1α binding to AID in the cytoplasm, the deaminase becomes phosphorylated by the kinase on two residues within its N-terminus (Ser38 and Thr27 in human AID). A third phosphorylation site was located to the C-terminus (Y184). Although unphosphorylated AID is still enzymatically active to a certain extent, this posttranslational modification is considered to be important for the biological activity of the enzyme. The phosphorylation of AID seems to facilitate its association with RPA, which

dramatically increases the efficiency of AID-catalyzed deamination, possibly through RPA-mediated stabilization of ssDNA and thereby allowing AID access to its substrate[54]. Furthermore, mice with constitutively activated PKA exhibit enhanced CSR, while addition of PKA inhibitors block class switching *in vitro*. However, uncertainties remain concerning the role of phosphorylation of AID and thus the modification is widely believed to be of more regulatory than activating nature.

Epigenetic changes, such as changes in DNA methylation and modifications of histones, can regulate transcription, replication and repair. These modifications change the accessibility of the genes of interest to the proteins involved in DNA metabolism. One example is a recent study in transgenic mice, which implies that demethylation of cytosines early in B cell development might be important for targeting the active kappa light chain allele for SHM[55]. Moreover, studies in primary B cells suggest that histones associated with AID-targeted switch regions undergoing CSR are hyperacetylated, whereas inactive switch regions are not[40, 56, 57].

By analogy to transcription, where specific DNA motifs recruit transcription factors, *cis*-acting sequences might also play a role in recruiting AID to the IgV regions in centroblast B cells. Thus, it was observed that it is not the V region exon itself but rather sequence motifs within the promoter and enhancer regions that are necessary for AID targeting[58, 59]. However, there is controversial data on this topic. Thus, it has been shown that deletion of the core Eμ intronic H chain enhancer had only a minor effect on SHM, implying that it does not play a major role in the targeting of SHM[60]. Moreover, deletion of the core Eμ enhancer in an ectopic H chain gene in Ramos B cells led to a twofold reduction in SHM. In another study, however, the frequency of hypermutation actually appeared lower in the presence of the Ig enhancer[61]. Other *cis*-acting sequences that are possibly involved in AID targeting are ssDNA structures such as G-loops, which have been detected in transcribed switch regions. G-loop structures are likely to provide AID a stable ssDNA target in the non-transcribed strand[62], whereas collapsed R-loops may offer the same in the transcribed strand[63].

Besides *cis*-acting sequences, *trans* factor-binding sites also seem to be involved in SHM targeting. E.g. E-box motifs for E2A, PU.1 and NF-EM5 obviously modify the rate of SHM in V and H chain regions[64, 65].

Another control mechanism for AID action in SHM and CSR seems to be based on the cell-cycle. While error-free repair is essential during the replicative S phase of the

cell cycle, error-prone base excision repair (BER) and mismatch repair (MMR) are involved in SHM and CSR that might be restricted to the non-replicative phases of the cell cycle. A study in Burkitt's lymphoma BL2 cells, where induced mutations were detected only in G_1 and G_2 but not during S phase, supported this hypothesis[66]. Furthermore, during CSR, the IgH gene colocalizes with proteins of the nonhomologous end joining (NHEJ) in G_1[67, 68]. AID might also be targeted through subnuclear co-localization with the genes undergoing SHM. It was shown for example that the subnuclear localization of the H chain gene changes around the time of V(D)J rearrangement[69] and that it colocalizes with NHEJ proteins during CSR[67, 68]. Another aspect of AID is that it exists in various isoforms, with five different transcripts having been described so far. Besides the most prevalent full-length AID transcript, four transcripts that carry deletions in/of exon 3 and/or 4 have been identified. The alternative transcripts of AID were reported in human B cell non-Hodgkin's lymphoma (B-NHL)[70], chronic lymphocyte leukemia[71-74], asthmatic patients[75] and normal B cells stimulated with IL-4 and CD40L[71, 72].

2.2.3 Somatic hypermutation (SHM)

Somatic hypermutation (SHM) is a mechanism that introduces point mutations into the V region of immunoglobulins. It is based on the clonal selection theory, originally proposed by Burnet (1957)[7]: "The theory requires at some stage in early embryonic development a genetic process for which there is no available precedent. In some way we have to picture a 'randomization' of the coding responsible for part of the specification of gamma globulin molecules."

Somatic hypermutation occurs after the primary repertoire of antibodies has been created by V(D)J recombination. Thus, SHM takes place both in a different phase of the immune response as well as in a different part of the body. This process of secondary diversification is largely driven by antigen and only happens in activated B cells. When a B cell that expresses surface IgM antibodies binds to its specific antigen, it undergoes rapid proliferation, differentiates and migrates to specialized structures within secondary lymphoid organs named germinal centers (see Fig. 2.4). That is the microenvironment where B cells become centroblasts and start to express

AID, therefore initiating somatic hypermutation of the antibody V region encoding the antigen-binding sites.

The frequency of mutation at the immunoglobulin locus is approximately 10^{-3} /basepair/generation. Most mutations are single base changes. They begin accumulating 100-200 bp from the transcription start and end 1.5-2 kb downstream, with the mutation frequency decreasing with increasing distance from the promotor (Fig. 2.9). In humans and mice, both C/G and A/T pairs can be targeted with similar frequency. While strand polarity is observed for A/T basepairs, where mutations preferably take place on the transcribed strand, C/G pairs seem to mutate at equal frequencies regardless of strand placement.
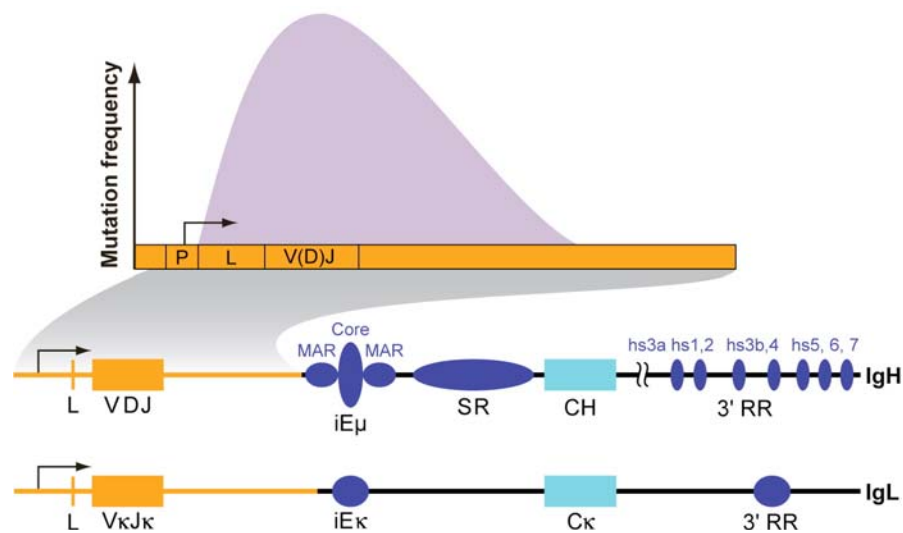

Fig. 2.9 Somatic hypermutation in Ig genes. Mutation frequency along the IgH and IgL locus is depicted. Representative IgH and Igκ genes are depicted below the graph that include the following elements: Leader (L), V region [V(D)J], intronic enhancer (iEμ and iEκ), C region (CH or Cκ), switch region (SR), and the 3' regulatory region (3' RR). The arrow indicates the transcription start site. The V region (orange) experiences significant SHM, whereas the C region (blue) does not, except for the switch region (not shown). SHM is sharply delimited by the V promoter (P) at the 5' end and starts about 100 - 200 bp from the transcription start site. Mutation frequency is maximal over the V(D)J coding exon and exponentially decays at the 3' end at 1.5–2 kb downstream from the transcription start site. Adapted from[41].

Most mutations occur in certain hotspots on DNA encoding Ig loci. Those motifs are composed of a WRCY consensus sequence, with AGCT being the preferred embodiment (as described earlier). However, exceptionally, mutations occur outside these hotspots, or even in genes outside the immunoglobulin locus, such as *Bcl6*, *CD79* and *CD95* [76-78]. Most genes that attract SHM possess binding sites for the E2A transcription factor[64, 79]. In malignant B cells, several other loci including *Myc*, *Pax5*, *Pim1*, *RhoH/TFF* have been identified to undergo erroneous mutation[80, 81]. Therefore,
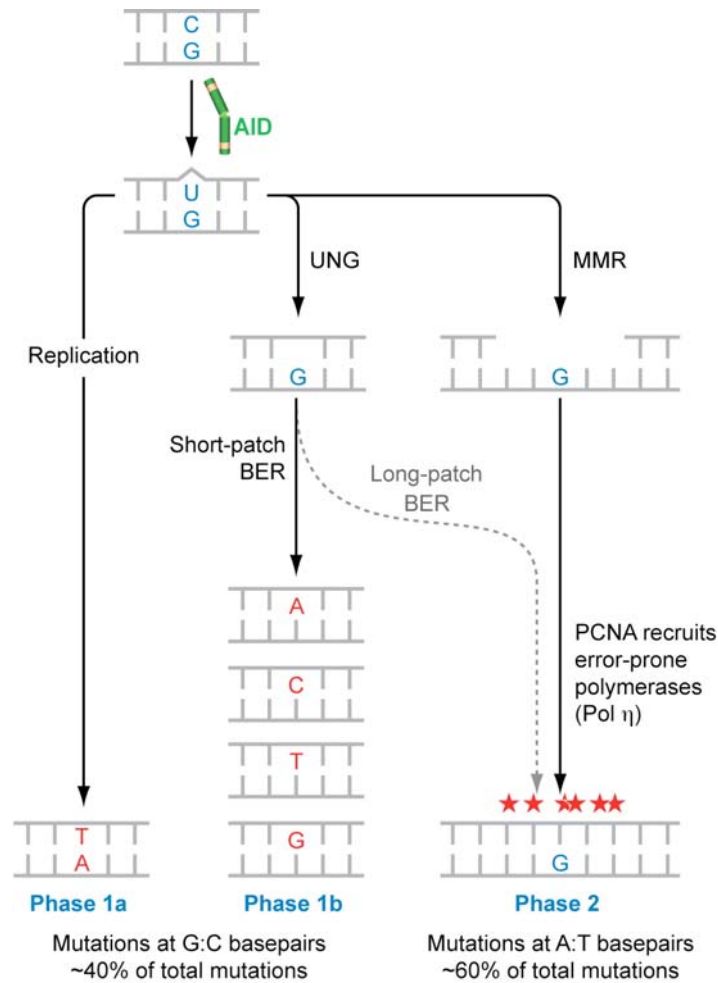
AID targeting to non-immunoglobulin targets is of pathological concern, because it is a feature of many B cell tumors, where it likely supports the proto-oncogene mutations and the chromosomal translocations that are typically detected in many Burkitt and follicular B cell lymphomas[82].

Somatic hypermutation can be divided into two phases. This proposition[83] was first made after the discovery that MSH2 knock-out mice exhibit an altered somatic hypermutation spectrum with mutations at A/T pairs being markedly reduced[84-86].
The first phase depends on the mutagenic activity of AID whereas the second phase is based on the error-prone repair of the AID-induced U/G mismatches (Fig. 2.10). The first phase is again divided into two parts: phase Ia mutations are transitions at the initial C/G sites occurring after the propagation of the AID induced C-to-U mutation through replication. This is how C-to-T transitions on one strand, or G-to-A transitions on the opposite strand, are introduced. If the U/G lesion is resolved by enzymes of the (short patch) BER pathway transversion mutations are created. In this scenario, uracil DNA glycosylase (UDG) removes the uracil from the DNA backbone and thus gives rise to a non-instructive abasic site. If this site is not processed by BER and is allowed to persist until replication, DNA polymerases capable of by-passing AP-sites may insert any of the four nucleotides (phase Ib mutations)[87].
An alternative mechanism of DNA cleavage, an intermediate step in SHM, is proposed to involve the MRN complex. This complex is constituted of the MRE11, Rad51 and NBS1 (MRN) proteins and is generally known for its function in DNA break sensing and repair. It was recently shown that MRE11 is enriched at the IgV locus and that MRE11/RAD51 cleave abasic sites to form ssDNA breaks[88]. Another study suggests that NBS1 overexpression in a hypermutating B cell line significantly increases the mutation rate implying that the MRN complex accelerates the processing of AID-generated DNA lesions[89].

In phase II, members of the MMR machinery and long patch BER are believed to be responsible for the introduction of mutations at neighboring A/T base pairs. Although MMR and BER generally function to maintain genomic stability, they appear to become error prone in the germinal center centroblast B cells, where they act on the Ig genes.

Fig. 2.10 Model of somatic hypermutation. AID deaminates a cytidine residue, creating a uridine:guanosine (U/G) mismatch that is resolved by several pathways that may compete with one another. AID deaminates single-stranded DNA formed during transcription of both strands of the DNA (not shown). The subsequent steps, however, might not occur equally on both strands. (Left) The general replication machinery can interpret the U as if it were a deoxythymidine (T). One of the daughter cells will acquire a C-to-T transition mutation. (Center) UNG can remove the uracil, leaving behind an abasic site. Short-patch base excision repair (BER) can fill the gap with error-prone polymerases, which can insert any nucleotide in place of the U, leading to transitions and transversions at G/C bases. (Right) Mismatch repair (MMR) can recognize the U/G mismatch. The U-bearing strand is excised and, at loci that undergo SHM, monoubiquitylated PCNA (proliferating cell nuclear antigen) recruits error-prone polymerases to fill the gap, leading to transition and transversion mutations at A/T bases as well as at neighboring G/C bases. (Dashed line) Long-patch BER can also be a source of mutations at A/T bases and may compete with MMR. Adapted from[41].

The MMR MutSα heterodimer (MSH2/MSH6) recognizes and binds to the original U/G mismatch[90], and was proposed to be involved in the subsequent gap formation and error-prone DNA synthesis. This would result in the generation of mutations outside of the initial U/G mismatch, e.g. at A/T base pairs. Exonuclease I is a known partner of MSH2 and was observed to physically associate with the V region of immunoglobulins. It therefore became an attractive candidate for strand degradation following U/G recognition by MutSα[91]. The single-stranded gap is then believed to undergo low-fidelity regeneration involving error-prone polymerases such as polymerases θ, ι, ζ, λ, REV1 and polymerase η that seems to play a critical role in this process during which mutations are also introduced at more distal sites[92, 93].

Most mutations at A/T base pairs are dependent on components of the MMR machinery, since deficiencies in MSH2 and MSH6 lead to a significant reduction of mutations at A/T base pairs[83, 84, 94, 95]. However, A/T mutations are ablated entirely only in mice that are simultaneously deficient in both UDG and MSH2, indicating that both pathways provide alternative backup mechanisms for each other[96].

Multiple studies show that somatic hypermutation is dependent on transcription. As described earlier (see 2.2.2), AID preferably acts on single-stranded DNA, such as occurs in a transcription bubble. Another hint comes from the fact that the promotor and transcription-regulatory elements of the Ig locus play a crucial role in the activation of hypermutation[59, 97, 98]. The Ig enhancers (but not the IgV gene itself) are very important for the generation of mutations, whereas the IgV promotor defines the 5' border of the mutation domain.

Further evidence comes from the implication of the transcription machinery *per se*. Storb and colleagues set up a model, where they postulate an association of a mutator factor with a RNA polymerase II (RNA Pol II) transcriptional complex, allowing the deposition of mutations while moving along with the transcription bubble[99, 100]. That the 5'-proximal ~ 150 nucleotides are not mutated could be explained either by the RNA Pol II complex first having to reach the elongation phase before the mutator can dock on, or/and by the need to make the DNA substrate accessible for the mutator by moving the transcriptional complex a certain distance into the gene. An *in vitro* study provided further arguments to support the hypothesis of correlation between SHM and transcription: RNA polymerase II has been shown to immunoprecipitate with transfected AID in *in vitro*-activated B cells, thus suggesting a direct physical interaction between the two proteins[40].

Furthermore, transcription was observed to enhance the *in vitro* AID-mediated deamination of synthetic substrates, as well as AID-mediated mutation of *Escherichia coli* [26, 32, 33, 36, 38, 101].

2.2.4   Class switch recombination (CSR)

While somatic hypermutation is a genetic mechanism that modifies the antigen-binding site of an antibody, class switch recombination (CSR) is a process that defines the biological property of an antibody molecule. During this procedure, the

constant region of the antibody heavy chain is changed, while the variable part stays the same, thus retaining the antibody's specificity while changing its effector function[102]. Several different C-region isotypes can therefore be expressed in the B cell's progeny as the cells mature and proliferate in the course of an immune response.

Plasma B cells begin to synthesize IgM molecules that are inserted into the plasma membrane and serve as antigen receptors. After the cells have left the bone marrow, but before they encounter their specific antigen, they start to coproduce IgD molecules with the same antigen specificity that are also expressed on the cell's surface[103, 104]. Upon stimulation by antigen and helper T cells, B cells start to secrete IgM molecules that dominate the primary antibody response. The switch to other isotypes, such as IgG, IgE or IgA is then induced through the combination of antigen and cytokines that are released by T helper cells later in the immune response. Crucial signaling for CSR is also provided by the specific interaction of the CD40 receptor on B cells and CD154 (CD40L) or, specifically for mouse B cells, through the engagement of the Toll-like receptor 4 (TLR4) by lipopolysaccharide (LPS)[16, 17, 105].

Immunoglobulin isotype switching is achieved through class switch recombination, a type of non-homologous DNA recombination or intrachromosomal deletional recombination event (Fig. 2.11). The process is guided by switch regions comprised of stretches of repetitive DNA (multiples of the GAGCT and GGGGGT sequences) that lie in the intron between the J_H gene and the C_μ gene. Equivalent sites are located upstream of the other genes encoding heavy-chain isotypes, except for the δ gene whose expression is independent of DNA rearrangement. When a B cell switches e.g. from IgM to IgG, the DNA recombination occurs between the switch regions preceding the exons encoding the heavy μ and γ genes (S_μ and S_γ). This leads to the deletion of the C_δ coding region and the rest of the intervening DNA[106, 107]. In contrast to V(D)J recombination, all switching events are generally productive and lead to a functional antibody molecule, since the switch regions lie in introns and thus frameshift mutations are of no importance.

Like somatic hypermutation, class switch recombination also depends on transcription. Cytokines produced by T helper cells and dendritic cells induce transcription from germline (GL) promoters situated upstream from each acceptor switch region. Since these GL transcripts do not encode proteins, they are also called

sterile transcripts[108]. Their main purpose is to direct AID to a specific switch region and to render the DNA accessible to the deaminase. First, the act of transcription provides short regions of ssDNA exhibiting an ideal substrate for AID. Second, transcription through switch regions leads to R-loop structures that are formed when the transcribed strand displaces the non-template strand of the DNA double helix[109, 110]. In addition, stretches of palindromic AGCT sequences may allow AID to act on both strands concurrently. That is how multiple single strand breaks may be introduced on both strands and thus give rise to double strand breaks.
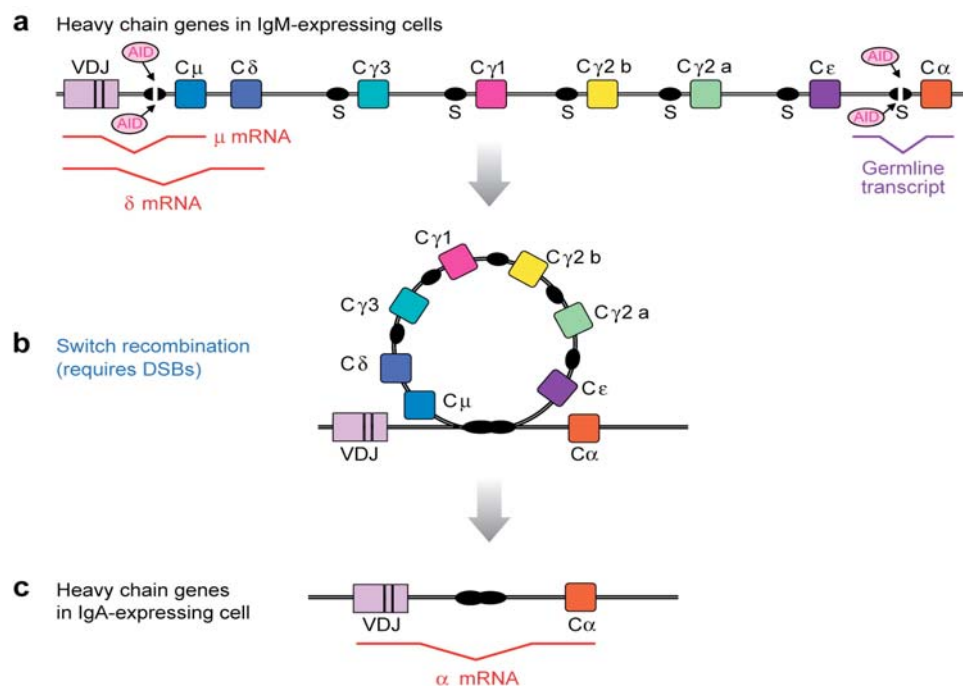
Fig. 2.11 Diagram of Ig class switch recombination (CSR) to IgA. **a** The mouse IgH locus in B cells expressing IgM and IgD (by alternative RNA transcription/processing). During CSR, activation-induced cytidine deaminase (AID) deaminates dC residues in the top and bottom strands of transcriptionally active S regions (Sμ and Sα in the diagram shown), initiating a process that results in double-strand DNA breaks (DSBs) in both S regions and CSR by intrachromosomal deletion (**b**). **c** The IgH locus after CSR to IgA. Splicing diagrams of the μ, δ mRNAs and the germline α transcript are indicated below the diagram of the locus. Similar germline transcripts are induced from unrearranged Cγ, C, and Cα genes, depending on the cytokine stimulation received by the B cell. Adapted from[111].

The CSR process is initiated by AID-mediated deamination of dC nucleotides within S regions. As in SHM, proteins of the BER pathway are required to deal with the initial U/G mismatch. UDG removes the uracil and subsequently APE starts the repair of the UDG-created abasic site by incising the phosphate backbone of DNA, thus leading to single-strand breaks[112]. However, in cells undergoing CSR, polymerase β,

the BER enzyme that is usually responsible for faithfully repairing the lesions, might be downregulated, specifically prevented from accessing switch regions or its activity might be inhibited[111].

That is where MMR, the second repair pathway involved in CSR, is most likely entering the stage: its role seems to be in the conversion of the remaining single-strand breaks (SSB) to double-strand breaks (DSB), which are necessary for CSR[113, 114]. If the SSBs are in close proximity to each other on opposite DNA strands, DSBs can form spontaneously. If not, SSBs are not destabilizing the duplex and in most cases they are simply repaired [90]. Since switch regions are large and breaks can occur anywhere, it seems unlikely that they frequently occur close enough to be spontaneously converted to DSBs[103-105]. A model how MMR proteins could be involved in this conversion will be described in chapter 2.4.1.

After DSBs are created, the donor and acceptor switch regions are recombined. While DSBs created during DNA replication or during the G2 phase of the cell cycle are usually repaired by homologous recombination, DSBs in B cells created during CSR are introduced and resolved during G1. And since S regions lack sufficient homology to undergo homologous recombination, ubiquitously expressed proteins from the NHEJ pathway are recruited for the repair[90, 113].

Four proteins that are essential for NHEJ have been shown to be important for CSR: Ku70, Ku80 and the ligase-complex XRCC4-ligase IV[68, 115-120]. The Ku70-Ku80 heterodimer binds to the ends at DSBs by forming a ring that encircles DNA. Ku then slides away from the ends, allowing the catalytic subunit of DNA-PK (DNA-PKcs) to access the DNA ends. DNA-PKcs is transphosphorylated and phosphorylates Ku70, Ku80 and XRCC4 and thus appears to be an activator as well as a scaffold protein during the ligation event. The binding efficiency of XRCC4-ligase IV is improved by Ku70-Ku80[121-125].

Furthermore, proteins of the MRN complex (MRE11, RAD50, NBS1) scan the DNA and bind DSBs during CSR within seconds of their formation[126, 127]. The kinase ataxia telangiectasia mutated (ATM) then binds the complex via NBS1. Following activation, it phosphorylates multiple substrates, such as NBS1 itself, CHK2, MDC1, p53, 53BP1 and H2AX. This leads to a further recruitment of MRN and other repair factors to the DSB and finally leads to the activation of the cell-cycle checkpoints[128-130]. To conclude, all these proteins are likely to be involved in holding the AID-mediated DSBs in a three-dimensional structure to support exact synapsis between S_μ

and the downstream acceptor S region, and to inhibit recombination with DSBs on other chromosomes.

The activity of AID is crucial for the initiation of CSR. Hence, a lack of AID results in a complete block of class switching. AID$^{-/-}$ cells were shown to exhibit very few S region DSBs, typically only about 10% of wild-type cells[113]. A deficiency in this enzyme in humans leads to an immunodeficiency syndrome, known as hyper IgM type 2, which is characterized by the absence of immunoglobulins other than IgM[27]. The involvement of double-strand break repair, as described above, can be observed in mice lacking Ku proteins that show a significantly reduced number of switched antibodies. Furthermore, it was shown that deficiencies of other DNA repair proteins such as DNA-PKcs also impair CSR, since they are involved in DNA pairing and the end-joining processes[111].

2.3   DNA repair mechanisms

Since the main cause of cancer are mutations in genes critical for cell growth, the maintenance of genome stability is crucial for preventing oncogenesis. DNA is constantly exposed to various exogenous and endogenous sources of damage that, if not circumvented, eventually lead to the activation of proto-oncogenes, inactivation of tumor-suppressor genes and the formation of cancer. Furthermore, the fidelity of the replicative polymerases substantially influences the integrity of the genetic information.

There are three main causes for introducing lesions into DNA: The first one arises from byproducts of normal cellular metabolism such as reactive oxygen species derived from oxidative respiration, or products derived from lipid peroxidation. Another source of endogenous damage comes from hydrolysis within the cell. This is linked to protonation of purines on N^7, which weakens the glycosidic bond. A striking example is the hydrolysis of the glycosidic bond of about 18'000 purine residues per cell per day. More lesions are created by deamination of cytosines, guanines, adenines or 5-methyl-cytosines resulting in uracil, hypoxanthine, xanthine and thymine, respectively, all of which have to be replaced in order to avoid persistent alteration of the genetic material.

The mutational burden imposed by endogenous damage is elevated by various sources of exogenous damage. These comprise environmental agents such as the ultraviolet component of sunlight, leading e.g. to the formation of pyrimidine-dimers. Ionizing radiation readily introduces a great variety of damage including single and double strand breaks in DNA and oxidative damage. Furthermore, various genotoxic chemicals, such as those contained in tobacco smoke, cause mutations that considerably enhance cancer risk[131-134].
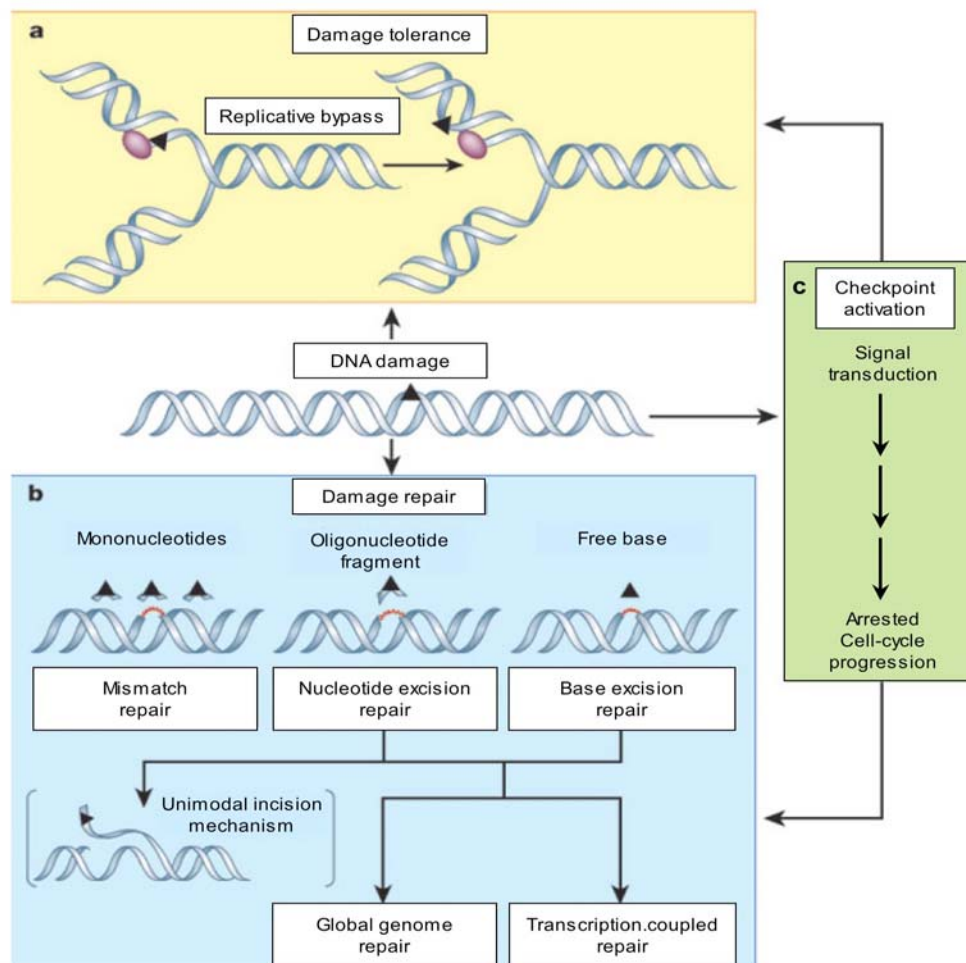
Fig. 2.12 DNA damage (illustrated as a black triangle) results in either repair or tolerance. **a** During damage tolerance, damaged sites are recognized by the replication machinery before they can be repaired, resulting in an arrest that can be relieved by replicative bypass (translesion DNA synthesis). **b** DNA repair involves the excision of bases and DNA synthesis (red lines), which requires double-stranded DNA. Mispaired bases, usually generated by mistakes during DNA replication, are excised as single nucleotides during mismatch repair. A damaged base is excised as a single free base (base excision repair) or as an oligonucleotide fragment (nucleotide excision repair). Such fragments are generated by incisions flanking either side of the damaged base. Nucleotide excision repair can also transpire in some organisms by a distinct biochemical mechanism involving only a single incision next to a site of damage (unimodal incision). **c** The cell has a network of complex signaling pathways that arrest the cell cycle and may ultimately lead to programmed cell death. Adapted from[134].

To deal with and fight against these permanent threats, nature has evolved sophisticated systems to counteract this time-dependent genetic degeneration (Fig. 2.12). These mechanisms either allow tolerance of DNA damage or lead to the activation of checkpoints that provoke attenuation or arrest of the cell cycle in order to permit repair, or, if no longer possible, to induce apoptosis. A third branch comprises a collection of mechanisms that repair or replace the damaged bases and hence restore the integrity of the DNA backbone. Those repair systems are generally divided into four classes that will be discussed further in the following sections: I) direct repair, II) nucleotide excision repair, III) base excision repair and IV) mismatch repair. Focus will be placed on the latter two, given that they are most relevant to the studies discussed in this thesis.

2.3.1   Direct repair

Direct repair, also referred to as damage reversal, is the simplest way of correcting damage in DNA. Rather than excising and replacing the whole base, the modification is directly repaired in a single-step reaction. This mechanism involves specific enzymes that recognize base modifications and promote restoration of the initial state of DNA. Due to its simplicity, the substrate spectrum of these enzymes is very limited, but repair is highly accurate and carried out with great efficiency.

UV irradiation causes the formation of cyclobutane pyrimidine dimers between adjacent thymine residues. This leads to a local distortion of DNA, hindering transcription and replication, and thus necessitating repair of the lesion. One way in which this can be achieved is by a photolyase-mediated process. These enzymes use photons as their energy source and take advantage of chromophore cofactors to split the aberrant cyclobutyl ring and thus restore the original sequence[135-137].

Another example is the repair of the highly cytotoxic and mutagenic lesion O^6-methylguanine. This modification is a direct consequence of DNA methylation, for instance through environmental carcinogens or endogenous metabolic compounds. The lesion is repaired by the suicide enzyme O^6-methylguanine-DNA methyltransferase (MGMT) that transfers the methyl group to an internal cysteine, thus restoring the integrity of the base at the expense of the enzyme[138, 139]. The AlkB

class of proteins is another example of damage reversal-factors that repair 1-methyladenine and 3-methylcytosine in DNA by oxidative demethylation[140].

2.3.2   Nucleotide excision repair (NER)

In terms of lesion recognition, nucleotide excision repair is the most versatile of all repair systems. It readily recognizes distortions as diverse as intra-strand crosslinks, UV induced cyclobutane pyrimidine dimers, protein-DNA crosslinks and a broad range of bulky chemical adducts. This is due to the fact that NER does not attempt to ascertain the exact nature of a DNA lesion, but rather senses the distortion in the structure of the double helix[141].

NER comprises two subpathways with partially distinct substrate specificity: global genome repair (GGR) that surveys the whole genome for DNA lesions and transcription-coupled repair (TCR) that focuses on distortions blocking the elongating RNA polymerases[142].

TCR and GGR require the same set of proteins, with exception of XPC. XPC is dispensable for TCR, since RNA polymerase II recognizes the damage in this process. On the other hand, TCR necessitates other factors that are not required for GGR, such as CSA, CSB and XAB2[143].

Local lesions are initially recognized by the XPC-HR23B heterodimer upon scanning the DNA double helix for distortions. Subsequently, a protein complex is formed, including the 10-subunit factor TFIIH, XPG, XPA and RPA. This complex unwinds the DNA close to the lesion, whereby a stretch of about 30 nucleotides becomes accessible. The single-stranded DNA is stabilized by RPA and the damaged strand is then incised on either side of the lesion: while the 3' cut is carried out by XPG, the 5' cleavage is performed by the heterodimer of ERCC1 and XPF. This allows for excision of the oligonucleotide containing the lesion. The resulting gap is then filled-in by the replicative polymerases δ or ε with the help of RPA, RFC and PCNA. The remaining nick is then sealed by DNA ligase I[133] (Fig. 2.13).

Since the NER machinery comprises a much higher number of components than all the other repair pathways, the above description is very simplified.
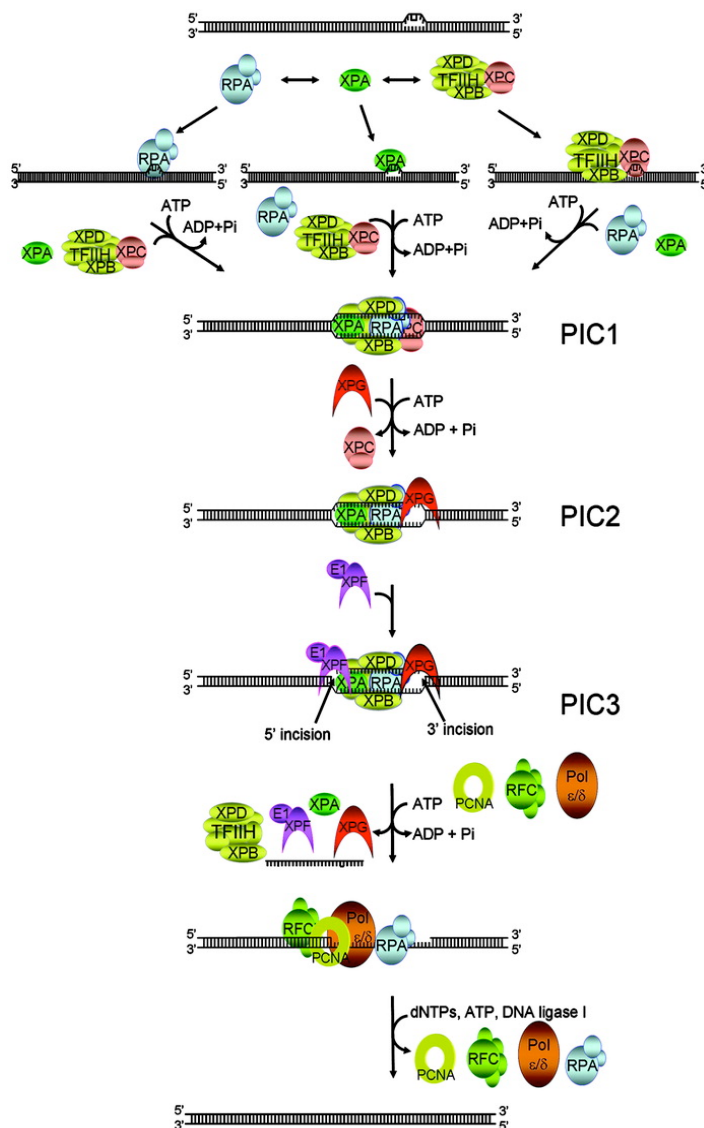
Fig. 2.13 Nucleotide excision repair in human cells. The DNA damage is recognized by the cooperative binding of RPA, XPA, and XPC-TFIIH, which assemble at the damage site in a random order. The four repair factors form a complex at the binding site, and if the binding site is damage-free, ATP hydrolysis by the XPB and XPD helicases dissociates the complex (kinetic proofreading). If the site contains a lesion, ATP hydrolysis unwinds the duplex by about 25 bp around the lesion, making a stable preincision complex 1 (PIC1) at the damage site. XPG then replaces XPC in the complex to form a more stable preincision complex 2 (PIC2). Finally, XPF·ERCC1 is recruited to the damage site to form preincision complex 3 (PIC3). The damaged strand is incised at the 6th ± 3 phosphodiester bond, 3' to the damage by XPG, and the 20th ± 5 phosphodiester bond 5' to the damage by XPF·ERCC1. The resulting 24–32 oligomer is released, and the gap is filled by Pol δ/ε with the aid of replication accessory proteins PCNA and RFC. Adapted from[144].

2.3.3 Base excision repair (BER)

Base excision repair (BER) is the major pathway mediating the removal of corrupt DNA bases. These include ionizing radiation, alkylating agents and reactive oxygen species, methylation, deamination and hydroxylation, respectively[145, 146]. Uracil in DNA is the most frequent lesion addressed by BER, thus making BER the most prevalent repair mechanism in the cell. The pathway is highly conserved throughout evolution and could successfully be reconstituted *in vitro*[147, 148] (Fig. 2.14).
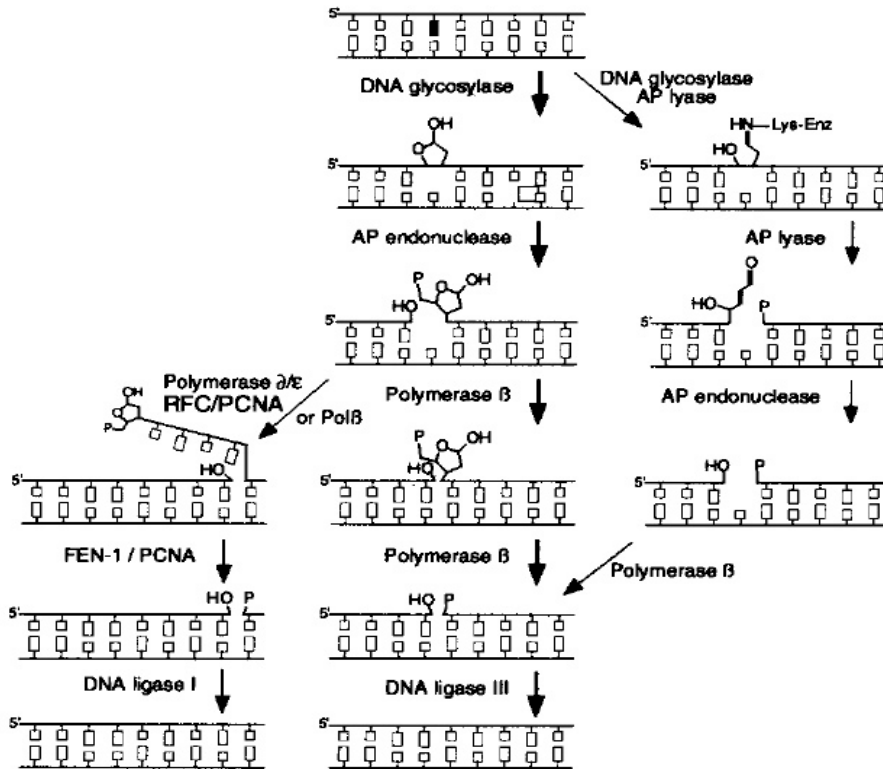
Fig. 2.14 Overview of base excision repair pathways: The short-patch BER pathway shown in the center is predominant. Excision of a damaged base by a DNA glycosylase or spontaneous base loss generates an abasic site, which is processed to a single-strand break by an AP endonuclease. The repair synthesis of one nucleotide and the excision of the abasic site is brought about by polymerase β, and DNA ligase III seals the remaining nick to restore the original DNA sequence. The short-patch pathway involving a bifunctional DNA glycoylase/AP lyase is outlined on the right. This pathway also requires APE1, Pol β and DNA ligase III, with the difference that Pol β is not involved in the removal of the baseless sugar-phosphate. In the long-patch repair pathway, Pol β or Pol δ/ε supported by the replication factors RFC/PCNA, makes a repair tract of 2±6 bases and FEN-1, supported by PCNA, excises the overhanging oligonucleotide. DNA ligase I then seals the nick. Adapted from[149].

Excision repair of the damaged base is initiated by hydrolytic removal of the base through breakage of the glycosidic bond, which links the base to the sugar phosphate backbone. Depending on the type of lesion, this step is carried out by one of several DNA glycosylases. While some of these enzymes are highly-specific for a certain lesion, others possess a broader substrate specificity. The essential role of BER in maintaining genome integrity is reflected by the ubiquity and wide utility of the glycosylase family. It is divided into two mechanistic classes: mono-functional glycosylases, such as uracil-DNA-glycosylase (UDG) and thymidine-DNA-glycosylase (TDG), which contain only the glycosylase activity, and bi-functional glycosylases, e.g. human 7,8-dihydro-8-oxoguanine-DNA glycosylase (hOGG1), which additionally contain a β-lyase activity.

46

Since UDG and TDG are the glycosylases most relevant to the studies presented in this thesis, they will be described more closely in the following section.

In 1974, uracil-DNA-glycosylase (UDG) was identified by Thomas Lindahl in *E. coli*[150]. Since then, homologous proteins were detected in extracts from yeast, mammals and large eukaryotic viruses[151-155]. Their main function is to protect cells from cytotoxic and mutagenic effects of uracil in DNA. While U/G lesions are introduced through deamination of cytosines, U/A mismatches are resulting from misincorporation of dUTP during replication.

The determination of the three-dimensional structure of UDG shed light on the molecular mechanism of catalysis[156-159]. The enzyme was shown to remove uracils *in vitro* in the order of preference ssU > U/G > U/A[160]. Furthermore, it is able to excise, albeit with low efficiency, 5-hydroxyuracil, isodialuric acid and alloxan[161]. Once a modified base is sensed, it is flipped out of the double helix and into the active site of the enzyme. If the substrate specificity is correct, hydrolysis of the glycosidic bond is initiated[156, 157].

In mammalian cells, UNG2, the nuclear form of UDG, was suggested to play a major role in postreplicative removal of incorporated uracil, due to its interaction with RPA and PCNA and its colocalization with these proteins in replicaton foci[162]. The relevance of these observations was emphasized and confirmed by studies in UDG[-/-] mice[163].

Thymine-DNA-glycosylase (TDG) was identified in 1990 by Jiricny and Wiebauer as a G/T binding and processing protein. G/T mismatches occur through spontaneous deamination of 5-methylcytosine, leading to a thymine residue opposing a guanine. TDG was observed to correct the mismatch by replacing the T with a C residue[164-167]. Although thymidine is the best known substrate for human TDG, glycosylases from the same family but of different origin were shown to have a wider substrate spectrum, with uracil being the most efficiently removed residue[168-170]. Moreover, modified uracils, such as 5-fluorouracil and 5-bromouracil, were also found to be very efficiently processes by TDG[169].

TDG was shown to bind AP-sites opposite G residues with an affinity higher than any other substrate, including G/T and G/U mismatches[171, 172]. Experimental data suggest that a conformational change, involving the N-terminal domain of TDG, provides

unspecific DNA interactions, facilitating the processing of a broader spectrum of substrates at the expense of enzymatic turnover. This structural change in the product-bound TDG is then reversed by SUMOylation[173]. Despite the presence of TDG and other glycosylases, e.g. MBD4 and SMUG1, UNG$^{-/-}$ cells accumulate significant amounts of dUMP in their DNA. This is in line with the finding that U/A mispairs, arising during replication, are poor substrates for TDG[163, 170]. However, TDG was demonstrated to be highly active on U/G substrates, implicating a role for TDG in the repair of deaminated cytosines[174].

Following the action of glycosylases, the resulting abasic sites are further processed by AP endonucleases. APE1 is the major human endonuclease responsible for 95% of cellular AP site incision activity[175]. The phosphodiester bond is cleaved 5' to the AP site, creating a SSB with a 5'-sugar phosphate. At this stage of the reaction, DNA polymerase β removes the abasic sugar-phosphate and mediates repair synthesis. Repair is completed by DNA ligase IIIα-XRCC1 heterodimer.

During canonical base-excision repair, called 'short-patch BER', pol β adds one nucleotide to the 3'-hydroxyl end of the strand break and concurrently removes the baseless sugar-phosphate. A second, minor BER pathway is mainly involved in the repair of AP sites at which a sugar residue has been modified, e.g. oxidized or reduced[176, 177]. The repaired stretch comprises two to ten nucleotides, therefore this repair pathway is referred to as 'long-patch BER'. Besides DNA glycosylase and AP endonuclease, it requires other factors such as flap endonuclease (FEN1), proliferating-nuclear antigen (PCNA), replication factor C (RFC), DNA polymerase δ/ε and DNA ligase I.

Since the AP lyase activity of pol β appears to be inactive in this scenario, pol β first adds one nucleotide to the 3' end of the nick and then hands it over to DNA polymerase δ or ε that add further nucleotides. Hence a single-stranded flap structure with a 5' sugar phosphate is generated[178]. FEN1 tracks down the length of the flap to cleave at the single-strand/double-strand junction. To terminate the reaction, the remaining nick is sealed by DNA ligase I. During long-patch repair, FEN1 activity is stimulated by PCNA[176]. Alternatively, the reaction can be catalyzed by Pol δ and PCNA.

2.3.4 Mismatch repair (MMR)

Besides NER and BER, the mismatch repair (MMR) system is the third branch of excision repair. MMR is mainly responsible for the removal of biosynthetic errors from newly synthesized DNA. Firstly, those errors occur through the insufficient fidelity of replicating DNA polymerases, leading to base-base mismatches. Even though only the geometry of Watson-Crick base pairs fits perfectly into the active site of the polymerase and non-Watson-Crick base pairs do not fit, in rare cases non-matching nucleotides are incorporated into the nascent chain. The replication polymerases possess proofreading activity that corrects these misincorporations. However, with a rate of 10^{-6} – 10^{-8}, misincorporated nucleotides escape the proofreading. Considering that the human genome comprises 3×10^9 base pairs, this corresponds to about 100 – 1'000 mutations per cell and cell cycle[179-181]. This mutational burden is pathogenic, as MMR defects predispose to cancer. This was first identified in patients afflicted with the cancer predisposition syndrome hereditary non-polyposis colon cancer (HNPCC), whose tumors are deficient in mismatch correction[182-184].

The second source of biosynthetic errors is the incorrect replication of microsatellites, repeated-sequence motifs such as $[A]_n$ or $[CA]_n$ that are present throughout the genome. During DNA synthesis primer and template strand occasionally dissociate and re-anneal incorrectly. This results in extrahelical nucleotides that are known as insertion/deletion loops (IDLs). If these loops are not repaired, they give rise to frameshift mutations in the next replication cycle. In most such cases, frameshift mutations lead to repeat shortening. This phenotype is known as microsatellite instability and was described in many cancers of the colon, endometrium and other organs[185, 186].

The main task of the MMR system is to address base-base mismatches and IDLs that escape the proofreading activity of the replicative polymerases. In doing so, it has to satisfy two special requirements. Firstly, the recognition is hampered through the fact that the bases are neither damaged nor modified, as in the case for BER. Secondly, a strand discrimination mechanism is required to efficiently distinguish between the error-containing daughter strand and the parental strand carrying the correct genetic information[187].

Besides biosynthetic errors occurring during DNA replication, mismatches also arise as a natural consequence of genetic recombination and are likewise corrected by the MMR system[187, 188]. Furthermore, MMR proteins are implicated in other DNA metabolic pathways such as DNA-damage signaling and mutagenic processes. In addition, other lesions such as UV photoproducts, DNA crosslinks or alkylated, oxidized and methylated bases may - under certain circumstances – be addressed by MMR[188-191].

Taken together, the activity of the MMR system reduces the overall amount of mutations by up to three orders of magnitude. In other words, the spontaneous mutability is increased 50-1'000 fold when MMR proteins are genetically inactivated[192-195]. Deficiencies in the MMR system result in elevated levels of illegitimate recombination between quasi-homologous sequences, increase the number of base substitutions and frameshift mutations and confer resistance to the cytotoxic effects of several classes of DNA damaging agents[193, 196-198]. MMR deficiency was not only described in conjunction with HNPCC, but with a number of sporadic tumors that present in a variety of tissues[199-202]. This points out the importance of active MMR and renders this pathway an attractive candidate for a prognostic marker in cancer diagnosis.

2.3.4.1 Mismatch repair in *E.coli*

Most of our understanding of the MMR pathway comes from studies of *E. coli* mutator strains that ultimately resulted in the complete reconstitution of the MMR system from individual purified proteins[203]. However, the first evidence that mismatches can provoke their own repair goes back to Meselson and colleagues who transfected *E. coli* with phage λ heteroduplex DNA containing several mispaired bases and found that the mispairs were corrected prior to replication[204, 205]. Due to the evolutionary conservation of the MMR pathway, these studies later also helped to shed light on the mechanism in the mammalian system.

The mismatch repair reaction can generally be divided into three steps: initiation, excision and resynthesis. The process requires a total of eleven activities. Four of these, namely MutS, MutL, MutH and UvrD, are specific for MMR and will hence be described in more detail in this section[203, 206-208].

In the initial step, the homodimeric MutS detects a base-base mismatch or IDL and then recruits the homodimeric MutL in an ATP dependent manner. Upon formation of this ternary complex, the MutH endonuclease is activated and incises the newly-synthesized strand of the mispair-containing duplex. During replication, the newly-synthesized strand is transiently undermethylated at GATC sites (because the modifying enzyme deoxyadenine methylase lags behind the replication fork by ~ 2 minutes) and can thus be distinguished from the parental, methylated strand. Hence, DNA methylation is used as a strand discrimination signal in *E. coli*, directing MutH to introduce the nick into the strand carrying the erroneous genetic information. A hallmark of MMR is that the process is bidirectional. Mismatch correction can be initiated irrespective of whether the nick lies within a distance of up to 1 kb 5' or 3' of the mismatch[209-213]. While the single-strand specific exonucleases RecJ and ExoVII support 5'-3'- repair, 3'-5'-repair is assisted by ExoI, ExoX and ExoVII.

There currently exist three models of MutS and MutL cooperation in recruiting MutH. The first comes from electron microscopic data of Allen and colleagues that support an "ATP-driven translocation model": The MutS/MutL complex actively searches for the nearest hemimethyated GATC sequence in an ATP-dependent reaction. In the course of this reaction, intermediate DNA stretches are looped out and structures are formed that could be visualized by electron microscopy[214, 215]. The "sliding clamp model" implies that MutS/MutL bind to the mismatch in an ADP-dependent manner. Mismatch binding triggers the exchange of ADP for ATP, subsequently leading to a conformational change of the MutS/MutL complex. As a result, the affinity for the mismatch is reduced and the complex freely diffuses along the DNA in an ATP hydrolysis-independent fashion. Upon recruitment of MutH at a hemimethylated GATC site, the endonuclease becomes activated and downstream events are initiated[216]. The third model supposes an "induced fit" of the mismatch into the active site of MutS. MutS remains bound at or close to the mismatch. It induces DNA bending and structural changes, thus facilitating the interaction with MutL and MutH[217].

After the erroneous strand has been incised, the methyl-directed excision step is launched. This involves unwinding of the DNA double helix through UvrD (DNA helicase II) starting at the incised DNA strand. However, the reaction is not dependent on MutH incision, that only occurs in Gram-negative bacteria, but can rather be

initiated at any strand break within ~ 1 kb of the mismatch. This evoked the hypothesis that in Gram-positive bacteria, the 3'end of the leading strand as well as both termini of the Okazaki fragments of the lagging strand can serve as initiation signals[218-220]. Following unwinding of the duplex, the error-containing strand is degraded by the aforementioned exonucleases, until resection stops at ~ 100 bp past the mismatch. Intermediate ssDNA is stabilized by single-strand DNA binding protein (SBB) until DNA polymerase III holoenzyme resynthesizes the DNA and DNA ligase seals the remaining nick as the final step in the restoration of genomic integrity[203, 209] (Fig. 2.15).
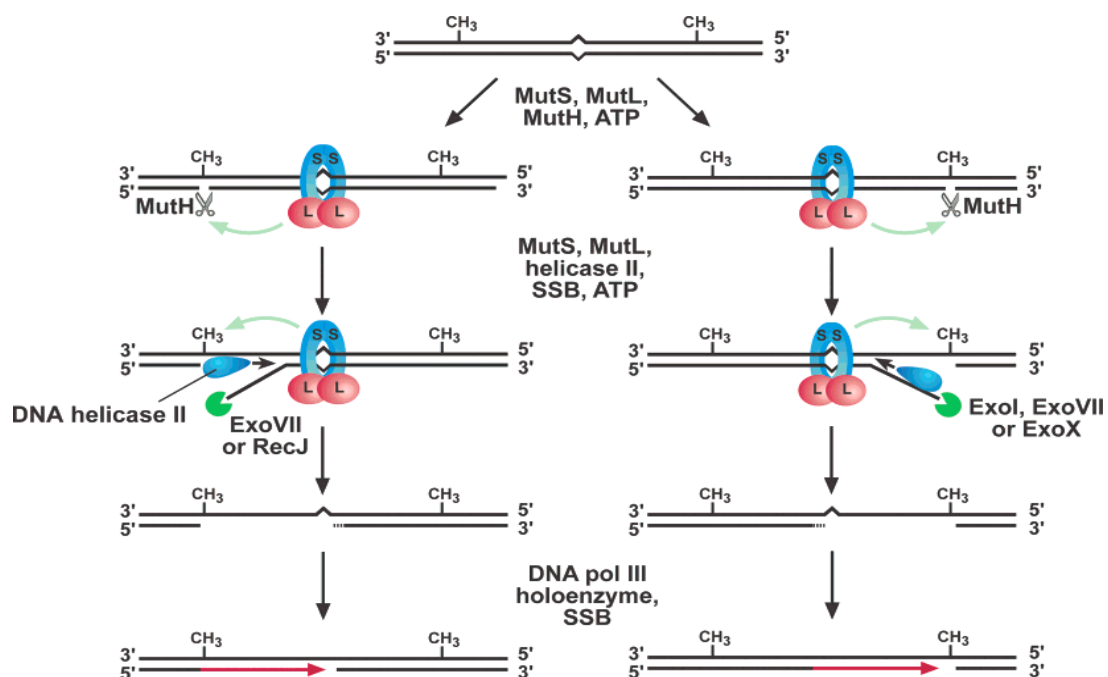
Fig. 2.15 Mechanism of methyl-directed MMR in *E. coli*. Details of the reaction are explained in the text. Although not shown, the integrity of the newly synthesized strand is restored by DNA ligase. The green arrows indicate signaling by MutS/MutL from the site of the mismatch to the nearest hemimethylated (GATC) where MutH incises the strand with the erroneous information. Adapted from[221].

E. coli MutS protein

The MutS protein is responsible for initiation of the mismatch repair reaction through direct binding to the mismatch *per se*. The 95 kDa polypeptide exists in an equilibrium of dimers and tetramers and recognizes IDLs of up to four nucleotides, as well as seven out of eight possible mismatches, with a C/C mismatch being the only non-repairable lesion[209, 222, 223].

The resolution of the MutS crystal structure yielded insight into the mode of action of the enzyme. It was described as a "pair of praying hands", where the MutS homodimer encircles the mismatch-containing DNA with its fingertips and thumbs inducing a bending of the DNA of about 60° (Fig. 2.16). The protein comprises five domains, each endowed with a specific function. Domain I harbors the Phe-X-Glu motif, where the phenylalanine residue anchors the protein to its substrate. This dock-on process is only carried out by one MutS molecule, rendering MutS a functional heterodimer[224-227]. Domain II and III are referred to as "connector" and "core" domains, respectively, and constitute the palm of the hand. They transmit allosteric information from mismatch binding to the ATPase domain. The fourth domain, similar to the first, forms a DNA clamp depicting the fingers of the preying hands. The "ATPase" domain localized in the fifth segment comprises the conserved Walker A and B motifs that are frequently found in DNA repair proteins. The ATPase domains of the two subunits are intertwined thereby exhibiting a platform to modulate affinity to DNA heteroduplexes and coordinating conformational changes[217, 225, 228-231].
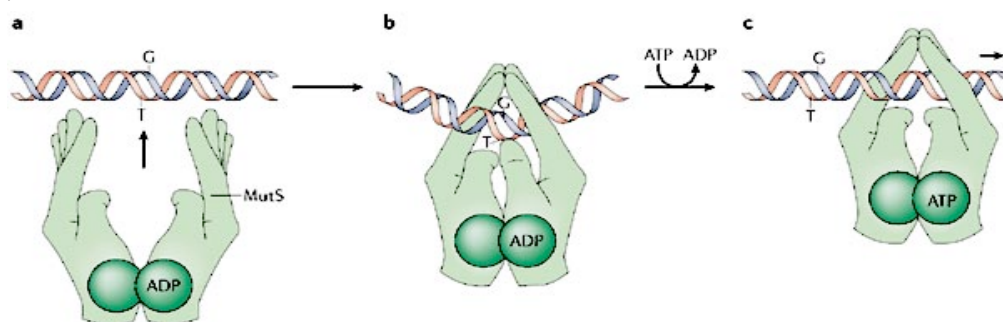
Fig. 2.16 The *E. coli* MutS sliding clamp and its activation. **a** In the free ADP-bound MutS homodimer, the "fingers" have an open conformation and dimerisation is mediated by the "palms" containing the ATPase domains. **b** Upon DNA binding and recognition of a mismatch, in this case a G/T mismatch, the "fingers" wrap around the DNA and form a structure reminiscent of a pair of "praying hands". In this conformation, only one of the two "thumb"-domains interacts with the misincorporated nucleotide *via* the conserved Phe-X-Glu motif, rendering MutS a functional heterodimer. **c** Exchange of ADP by ATP releases the "thumb" from the minor groove of the DNA and allows free diffusion along the duplex in both directions. Adapted from[187].

E. coli MutL protein

Due to its role in mediating the interaction between MutS and MutH, the MutL protein was described as the 'molecular matchmaker' of the mismatch repair process[232]. As shown in gel-shift experiments, the efficiency of mismatch binding by MutS is significantly increased in the presence of MutL[233].

The homodimeric MutL proteins are 68 kDa polypeptides that belong to the GHKL (gyrase/Hsp90/histidine-kinase/MutL) ATPase family. The N-terminus of these proteins is highly conserved and harbors the ATPase domain, whereas the C-terminal part is much less conserved and comprises a linker region and the dimerization domain[234]. It was shown that the ATPase domain is necessary for both mismatch-provoked excision and MutH activation[235, 236]. Furthermore some studies report the *E. coli* MutL ATPase to be strongly activated by binding to single-stranded DNA, but less so by duplex DNA. Although other groups question the DNA-binding capacity of MutL, recent studies with MutL mutants that show impaired DNA binding support the relevance of DNA binding to the MMR process[216, 237-239].

When the crystal structure of MutL was solved, both in the presence of ATP and the non-hydrolysable ATP analogue AMP-PNP, new insights into the molecular mechanisms of the MutL activity were obtained. ATP binding is used to modulate protein conformation and thus to control the interactions with downstream MMR components. MutH binding and activation of its endonucleolytic activity by MutL was described to be dependent on ATP hydrolysis. Similar observations were made with respect to UvrD interaction and activation of its helicase activity by ATP-bound MutL. Based on the structural data of the C-terminus, a working hypothesis of MutL dimerization and interaction with UvrD was postulated[240, 241]. Whereas DNA binding of MutL seems to be involved in the activation of the UvrD helicase, it seems to be dispensable for mismatch-triggered activation of MutH.

E. coli MutH protein

MutH is a monomeric endonuclease that is unique to Gram-negative bacteria. Its crystal structure revealed a clamp-like structure with two arms that are wound around the DNA. The endonuclease provides the strand discrimination signal necessary for downstream MMR components: MutH recognizes the newly-synthesized DNA strand at transiently hemi-methylated GATC sites and cleaves it 5' to the G. Its activity is greatly stimulated by the interaction with MutS and MutL in the presence of ATP. Depending on whether incision takes place 5' or 3' from the mismatch, different exonucleases are then recruited to degrade the error-containing strand[206, 208, 242].

E. coli UvrD protein

UvrD belongs to a superfamily of *E. coli* helicases and is also referred to as DNA helicase II. It is responsible for unwinding of the DNA starting from the nick introduced by MutH in the preceding step of MMR. Thus, DNA becomes accessible for the exonucleases. However, since the processivity of UvrD is very low, only 40-50 nucleotides can be simultaneously unwound. In order to unwind longer DNA stretches, UvrD was proposed to be iteratively loaded onto its substrate by MutL in an ATP-dependent manner[243].

2.3.4.2 Mismatch repair in yeast

Much of our knowledge about the eukaryotic mismatch repair system comes from studies in yeast, in particular *Saccharomyces cerevisiae*. Due to its genetic simplicity, it is an ideal model organism and the high conservation of the amino acid sequences from yeast to man facilitated the identification of human MMR genes[244-247].

In contrast to prokaryotic MMR, the system in yeast comprises several MutS and MutL homologs, namely MSH1-6, MLH1-3 and PMS1.

Similarly to their *E. coli* counterparts, MSH proteins exhibit weak ATPase activities, harbored in the C-terminal domains and their function is strictly dependent on ATP binding and hydrolysis[248, 249]. Besides carrying out repair of nuclear DNA, the proteins are also implicated in other biological processes. MSH1, that is exclusively found in yeast and does not require an additional partner, is implicated in the maintenance and repair of mitochondrial DNA[250-252]. Supported by a recent phylogenetic analysis, MSH1 is thought to constitute the founding member of the eukaryotic MutS homologs[253]. MSH4 and MSH5 are lacking the conserved N-terminal domain found in their prokaryotic homologs and, accordingly, they have not been found to contribute to MMR to date. Instead, these proteins appear to be involved in meiotic recombination[254, 255]. MSH2 turned out to be the leading actor in mismatch repair initiation. It engages in the formation of two different MutS-like heterodimers. In conjunction with MSH6, it binds to base-base mismatches or insertion/deletion loops, whereas, in complex with MSH3, only DNA duplexes with IDLs are recognized[256-259].

The MutL homologs MLH1-3 were identified due to their striking sequence homology with the prokaryotic MutL[260, 261]. PMS1 was initially described in a yeast mutant strain that showed increased levels of postmeiotic segregation, implicating the protein in homologous recombination processes[262]. PMS1, in complex with MLH1, exerts the vast majority of MMR activities, whereas the MLH1-MLH2 and MLH1-MLH3 heterodimers deal only with a subset of biosynthetic errors characterized by specific mutational intermediates[260, 263].

Another remarkable difference between the prokaryotic and the eukaryotic mismatch repair system is the discrimination of the newly-synthesized and the parental strands. This will be discussed on the basis of the human MMR system in the following chapter.

2.3.4.3 Mismatch repair in higher eukaryotes

Despite many similarities between the eukaryotic and the prokaryotic mismatch repair systems, there are some obvious differences. While the bacterial MutS and MutL proteins work as homodimers, the mammalian MutS and MutL homologs appear to be functional only in heterodimeric complexes. Human cells express five MutS homologs, namely MSH2, MSH3, MSH4, MSH5 and MSH6. With MSH2 as the common subunit, two MutS activities are formed that actively participate in MMR: MutSα, composed of MSH2 and MSH6, and MutSβ, a heterodimer of MSH2 and MSH3. MutSα is able to repair all eight base-base mismatches including C-C mismatches and is highly active on IDLs comprising 1-2 nucleotides, with a reduced affinity for larger heterologies. In contrast, MutSβ is specialized for the repair of IDLs of 2 to 10 nucleotides and can bind also single-nucleotide ID loops, albeit weakly. It is inactive on base-base mismatches[264-266]. Both MSH2 and MSH6 have been implicated in cancer development, such as in the aforementioned disease HNPCC. While mutations in *MSH2* strongly contribute to this type of cancer (40%), mutations in *MSH6* play a minor role (10%). Thus, inactivation of *MSH6* leads to a less severe tumorigenic phenotype. In contrast, *MSH3* is less involved in tumor development[267-269].

As mentioned earlier, no MSH1 homolog could be identified in mammals, whereas mammalian MSH4 and MSH5 function in meiotic recombination and play an important role in crossing over events, similar to their role in yeast[270, 271].

Three eukaryotic MutL activities could be identified to date. They are all heterodimers containing MLH1, the key player. MutLα is composed of MLH1 and PMS2 (human PMS2 = yeast PMS1) and supports repair initiated by either MutSα or MutSβ. It is the primary MutL activity in human mitotic cells. Other low abundant complexes composed of human MutL homologs are MutLβ (MHL1/PMS1) and MutLγ (MLH1/MLH3). While no or only marginal MMR activity was reported for MutLβ, MutLγ seems to contribute, albeit to a small degree, to base-base and single nucleotide repair *in vitro*[272-274]. This observation can be explained by recent findings from Modrich and coworkers, who identified an endonuclease activity in the PMS2 subunit of MutLα with the consensus sequence conserved in MLH3, but not in PMS1[275]. This endonuclease acitivity is essential for 3'-to-5' MMR (see below).

While genetic inactivation of MLH1 or PMS2 confers cancer predisposition, mutations in PMS1 do not and the involvement of MLH3 in tumor formation in humans remains to be elucidated[267-269].

So what is the molecular mechanism of mismatch detection and repair in human cells? Repair is believed to be triggered by MutSα binding to the mismatch. Following ATP-binding, the heterodimer undergoes a conformational change that commences to freely diffuse along the DNA in the form of a 'sliding clamp'. The ternary complex composed of ATP-activated mismatch-bound MutSα then recruits MutLα and a cascade of downstream events is launched. Depending on whether the repair complex encounters a strand discontinuity 3' or 5' from the original mismatch, excision of the erroneous strand has different factor requirements. Our knowledge about these processes was profoundly enriched by reconstitution of the MMR system *in vitro*[276, 277] (Fig. 2.17). If the nick that serves as a strand discrimination signal lies 5' to the mismatch the repair reaction can be carried out by only three protein components, namely MutSα, single-strand DNA binding protein RPA and the human exonuclease EXO1, with the loading efficiency of the exonuclease being increased upon the addition of the non-histone chromatin protein HMGB[221, 277-280]. As long as the mismatch is present, further stimulation of EXO1 comes from MutSα. Upon removal of the mispaired bases, MutLα exerts an inhibitory effect by actively

displacing EXO1 from DNA. Pol δ loads at the 3' terminus of the original discontinuity, which carries a bound homotrimeric proliferating cell nuclear antigen (PCNA) molecule. This complex fills the gap and DNA ligase I seals the remaining nick to complete the repair process.
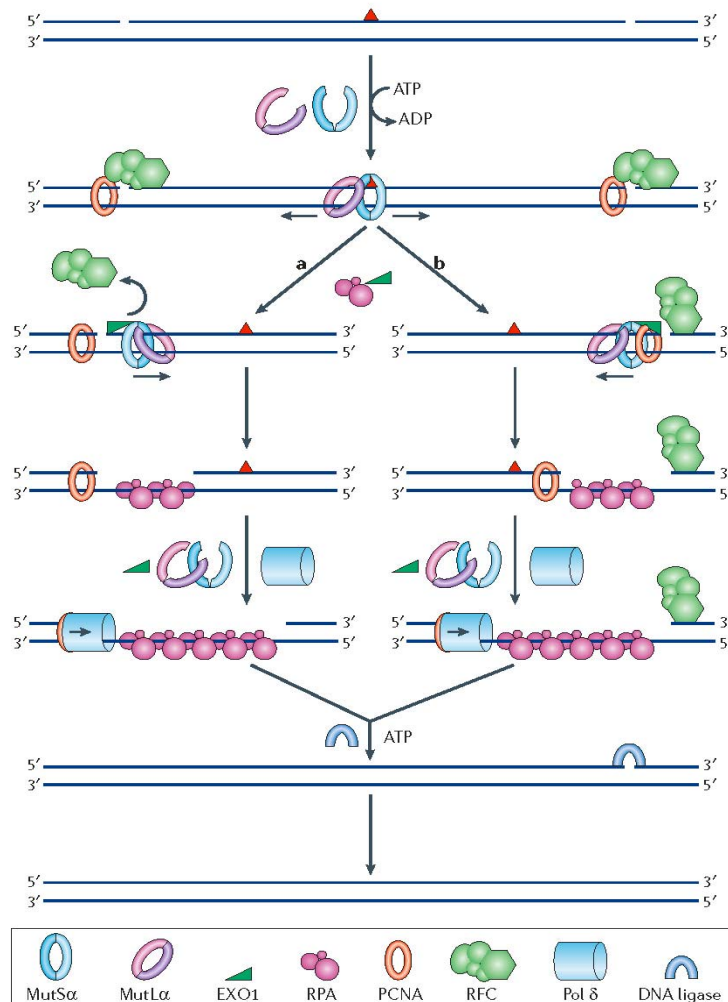

Fig. 2.17 The mismatch repair (MMR) process was recently reconstituted from either MutSα or MutSβ, MutLα, replication protein A (RPA), exonuclease-1 (EXO1), proliferating cell nuclear antigen (PCNA), replication factor C (RFC), DNA polymerase δ (Pol δ) and DNA ligase I. The following is proposed to take place. The mismatch (red triangle)-bound MutSα (or MutSβ) recruits MutLα. The ternary complex undergoes an ATP-driven conformational switch, which releases the sliding clamp from the mismatch site. **a** Clamps that diffuse upstream encounter RFC that is bound at the 5' terminus of the strand break, and will displace it and load EXO1. The activated exonuclease commences the degradation of the strand in a 5'-3' direction. The single-stranded gap is stabilized by RPA. When the mismatch is removed, EXO1 activity is no longer stimulated by MutSα, and is actively inhibited by MutLα. Pol δ loads at the 3' terminus of the original discontinuity, which carries a bound PCNA molecule. This complex fills the gap and DNA ligase I seals the remaining nick to complete the repair process. **b** Clamps that migrate downstream encounter a PCNA molecule that is bound at the 3' terminus of the strand break. The recruitment and the activation of EXO1 results in the degradation of the region between the original discontinuity and the mismatch, possibly through several iterative EXO1-loading events. A recent study reveals a endonuclease activity in MutLα and suggests it to be responsible for introducing the nicks (Kadyrov *et al.* 2006). RFC that is bound at the 5' terminus of the discontinuity prevents degradation in the 5'-3' direction (away from the mismatch). Once the mismatch is removed and the EXO1 activity is inhibited by bound RPA and MutLα, the gap is filled by Pol δ. DNA ligase I seals the remaining nick to complete the repair process. Adapted from[187].

In contrast to 5' repair, 3' nick-directed repair necessitates additional factors such as PCNA and its clamp loader replication factor C (RFC). Together with MutSα, they stimulate a latent endonuclease activity in MutLα that allows introduction of nicks

both 5' and 3' from the mismatch. This recent finding of Modrich and colleagues nicely explains the fact that EXO1 that possesses only a 5'-3'-polarity is sufficient not only for 5'-, but also for 3'-directed repair. After a long search for a cryptic 3'-5' exonucleolytic activity it is now clear that MutLα converts the 3' substrate into multiple 5' substrates that are gradually degraded by EXO1 (Fig. 2.18). Once the degraded stretch of DNA has passed the mismatch site, the EXO1 activity is no longer stimulated and the gap can be faithfully repaired[275].

Although shown to be essential for MMR *in vitro*, EXO1 seems to be partially dispensable *in vivo*, since cells from Exo1[-/-] mice retain substantial MMR activity, indicating that there must be redundant activities *in vivo* that remain to be identified[281]. To eventually complete the repair reaction after excision of the mismatch-carrying strand, DNA polymerase δ and DNA ligase I refill the resulting gap and finally seal the remaining nick in the DNA backbone.
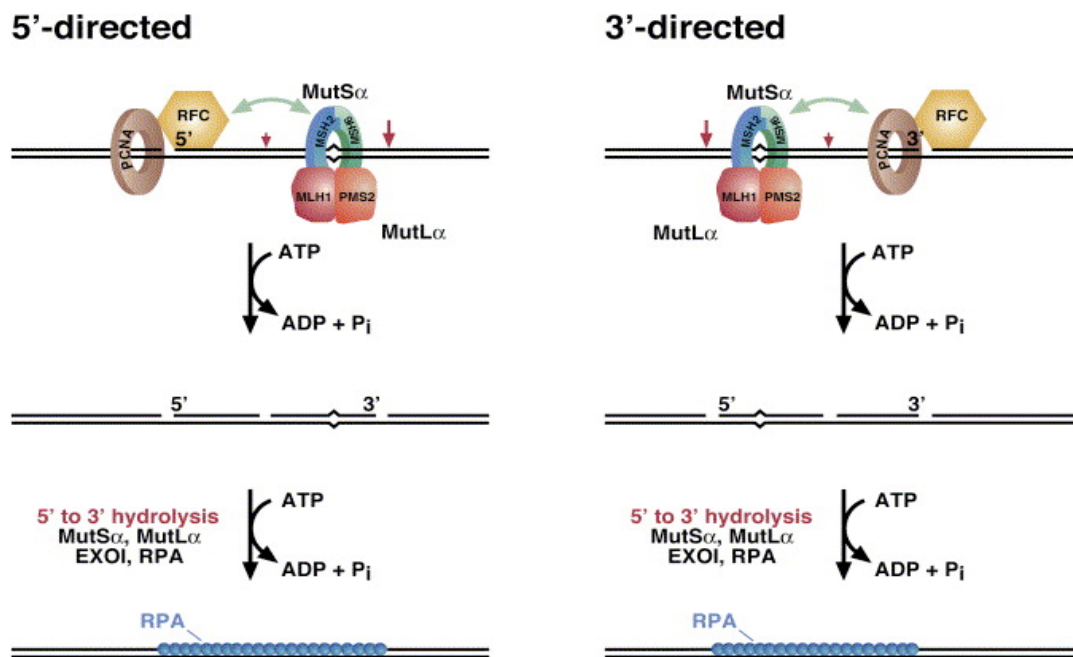
Fig. 2.18 Incision of the discontinuous heteroduplex strand in human mismatch repair. MutSα, PCNA, and RFC activate a latent MutLα endonuclease, which incises the discontinuous strand of 5' or 3' heteroduplex DNAs in an ATP-dependent reaction. Incision displays a bias for occurrence on the distal side of the mismatch relative to the location of the original strand break (large red arrows) but can also occur proximal to the mispair (small red arrows). For a 3' heteroduplex, this yields a new 5' terminus on the distal side of the mismatch that serves as an entry site for MutSα-activated EXOI, which removes the mismatch in a 5'-to-3' hydrolytic reaction controlled by RPA (Genschel and Modrich, 2003). The strong bias for incision of the discontinuous strand implies signaling along the helix contour, which may involve ATP-promoted movement of MutSα or the MutSα•MutLα complex along the helix. This feature of the mechanism is not illustrated in the diagram shown. Adapted from[275].

Human MutS homologs

Twenty years ago, a factor binding with high affinity to G/T mismatches on oligonucleotide substrates was first described and named G/T binding protein (GTBP) according to its substrate[282]. Later it was renamed to human MSH6 due to its close sequence homology with the *S. cerevisiae* MSH6 protein. When first purified to near homogeneity, it turned out that the 160 kDa polypeptide representing MSH6 constantly coeluted with a second one of 100 kDa molecular mass. This was identified as the gene product of MSH2[283, 284]. MSH2 and MSH6 are tightly associated and form the MutSα heterodimer that represents 80 - 90% of the cellular MSH2 pool. Mutation experiments showed that mismatch binding occurred *via* MSH6. This observation was confirmed by DNA cross-linking experiments that only succeeded with the MSH6 protein but not with MSH2[226, 231]. Due to their ATPase domains located in the C-terminal domain of the protein, MutS homologs are considered part of the ABC transporter family[285]. Mutants carrying deletions in that region showed impaired MMR activity, while the mismatch-binding capacity was retained. This suggests that while ATP hydrolysis is dispensable for initiating the repair reaction, it is crucial for dissociation from the mismatch and triggering of downstream events[230, 231, 265, 266, 286, 287]. MutSα contains two functionally interdependent, highly homologous ATP-binding sites. However, not even the recent determination of the crystal structure of MutSα bound to a variety of mismatches could elucidate the exact role of nucleotides in the function of MutSα[288]. Even though prokaryotic and eukaryotic MutS proteins are highly similar, there are differences concerning the mode of nucleotide binding. While bacterial MutS simultaneously binds two ATP molecules, human MutSα has one high-affinity ADP-binding site and one high-affinity ATP-binding site[289].

The observation that the phenotype of cells mutated in the *MSH2* gene, namely high levels of microsatellite instability, is significantly more dramatic than in *MSH6* mutants lead to the assumption that MSH2 might coexist with a partner other than MSH6. The identification of the protein MSH3 supported this hypothesis and evidence was obtained that the MSH2/MSH3 complex, also referred to as MutSβ, recognizes IDLs of up to ten nucleotides. Thus, MutSβ is mainly responsible for loop repair, whereas MutSα is the main actor in mispaired-base restoration. Interestingly,

MSH3 overexpression correlates with a decrease in base-base mismatch repair due to depletion of MutSα. Generally MutSα contains the major part of MSH2 in the cells with MutSβ accounting only for about 10 - 20%[266, 290, 291].

Human MutL homologs

MutLα is a heterodimer composed of MLH1 and PMS2. It is the major MutL activity in human mitotic cells, constituting about 90% of the cellular MLH1. Together with MutSα or MutSβ and ATP, it forms ternary complexes on DNA and supports processing of lesions recognized by the aforementioned molecules. The fact that substitutions in the ATP binding site of MutLα do not affect DNA binding, but disrupt downstream events, emphasizes the role of MutLα as the 'molecular matchmaker' in MMR[272, 292].

Cells that lack either MLH1 or PMS2 show mutator phenotypes and microsatellite instability similar to MSH2$^{-/-}$ cells. However, although *Pms2*$^{-/-}$ mice are cancer-prone, they do not develop gastrointestinal tumors as do *Mlh1*$^{-/-}$ mice, suggesting additional partners of MLH1 in MMR. Hence a backup pathway for PMS2 was proposed and found when PMS1 was discovered. In conjunction with MLH1, it forms the heterodimer MutLβ. However, its role in mismatch repair remains a matter of debate. Although its yeast homolog was reported to contribute to the repair of a subset of IDLs, no such activity could be found to date in human MutLβ. There is also uncertainty with respect to the contribution of PMS1 deficiency to tumor formation. While some studies show that *Pms1*$^{-/-}$ mice are not cancer prone, others report that in some HNPCC patients Pms1 appears to be mutated[247, 260, 293, 294].

The third MutL activity present in human cells is the MutLγ heterodimer, made up of MLH1 and MLH3. Its abundance in the cell is very low compared to MutLα and it is mainly implicated in meiotic recombination[295, 296]. *Mlh3*$^{-/-}$ mice are tumor prone and demonstrate weak mononucleotide microsatellite instability[297]. Its involvement in mismatch repair was recently demonstrated *in vitro*, where a contribution of MutLγ to base-base mismatch and single nucleotide IDL repair, albeit with low efficiency, could be detected[274].

2.3.4.4 Special functions of MMR proteins

Besides their function in mismatch correction, MMR proteins are also involved in a variety of other cellular processes. The most relevant are listed below.

The role of MMR in homologous recombination

MMR proteins are known to modulate the fidelity and efficiency of both mitotic and meiotic recombination[298, 299]. The best understood mechanism is the function of MutS and MutL homologs in the suppression of recombination between quasi-homologous, but non-identical DNA sequences, so called homeologous sequences. Even though present in multiple copies throughout the genome, chromosome rearrangements are infrequent, implying that recombination between such sequences is rare. This suggested the existence of a mechanism to suppress such events and the solution was provided by studies in *E. coli*, describing that deficiency in MutS or MutL significantly increases the frequency of homeologous exchanges[300-302]. Similar results were later obtained for the eukaryotic MutS and MutL counterparts in yeast and mammals[303]. Mismatches arising in recombination intermediates are likely to be recognized by a mechanism related to the postreplicative repair pathway. Depending on the degree of divergence, mismatches are either corrected or the recombination event is disrupted. In this scenario, strand exchange is probably prevented by blocking branch migration upon mismatch detection[304].

The role of MMR in meiosis

In contrast to mitotic recombination that involves the same set of proteins as postreplicative repair, meiotic recombination uses a different combination of MMR proteins. Data obtained from knockout mouse models and studies in yeast indicated that the MSH4-MSH5 heterodimer rather than MutSα or MutSβ is required[305, 306]. Mice lacking either Msh4 or Msh5 are sterile, but not cancer prone. Furthermore, involvement of MutLγ in meiosis was suggested when it was observed that *Mlh1*[-/-] and *Mlh3*[-/-] mice are also sterile[293, 307]. The tetrameric complex composed of MSH4-MSH5 and MLH1-MLH3 was subsequently hypothesized to promote the formation of Holliday junctions and their resolution into cross-overs rather than noncross-overs[308].

Furthermore, the discovery that PMS2 deficiency only renders male mice sterile led to the assumption that the MLH1-PMS2 heterodimer is implicated only in specific events of meiosis[293, 309].

Remarkably, a role for MMR proteins in signaling DNA damage and promoting cell death was postulated when it was observed that abnormal meiosis activates a checkpoint leading to apoptosis in a MSH4-MSH5 and MLH1-MLH3 dependent manner[187, 308, 310, 311].

The role of MMR in triplet repeat instability

Neurodegenerative diseases such as Huntington's disease, myotonic dystrophy, Friedreich's ataxia and fragile-X syndrome are often caused by trinucleotide repeat expansion[312]. During DNA replication, recombination and repair these repeats are likely to form unusual secondary structures, such as mismatches, triplexes and hairpins that may be aberrantly recognized by MutSβ. Accordingly, the role of MutSβ in triplet-repeat destabilization was shown in *Msh3*$^{-/-}$ animals that exhibit increased repeat stability compared to *Msh6*$^{-/-}$ mice[313]. However, although several models implicating MMR in triplet repeat destabilization have been proposed, the exact molecular mechanisms remain to be identified.

2.4   DNA repair proteins in immunoglobulin diversification

Due to its relevance to this work, a special focus will be put on the role of DNA repair proteins in the generation of immunoglobulin diversification. Generally, DNA repair proteins such as those involved in mismatch repair (MMR) and base excision repair (BER) are thought to maintain the integrity of the genome through faithful mutation correction. However, there are some scenarios in which the proper task of those repair proteins is actually to induce and fix mutations rather than correct them. Examples are the processes of antibody diversification comprising somatic hypermutation (SHM) and class switch recombination (CSR), both of which are necessary for the efficient functioning of the immune system.

2.4.1 Contribution of mismatch repair proteins

Mutations that arise spontaneously in the genome are generally eliminated by MMR. Hence, MMR-deficient cell lines display an increased frequency of spontaneous mutations[314]. In contrast, mice deficient for *Msh2* and *Msh6* (MutSα) exhibit decreased levels of SHM. Interestingly, the spectrum of mutations is different to wild-type mice in that the knock-out mice show only few mutations at A/T pairs with the majority of mutations displaying G-to-A or C-to-T transitions at the original AID-induced U/G site within WRCY motifs. This finding led to the proposition of two phases of SHM, discussed earlier[83-85, 94, 95, 315, 316] (see Fig. 2.10). An attractive candidate for gap formation preceding mutagenic repair is EXO1, a known partner of MSH2. Interestingly, it was observed in physical association with the V region of Ig genes and *Exo1*[-/-] mice demonstrate the same skewed mutation spectrum as *Msh2*[-/-] and *MSH6*[-/-] mice[91].

Mismatch repair proteins are mainly involved in the second phase of SHM, with MutSα demonstrating the leading role, while MutSβ, seems not to be involved. Indeed *Msh3*[-/-] mice do not have a detectable change in the frequency of variable (V) region mutations[95]. Phase II is initiated through U/G mismatch recognition that is proposed to trigger a patch of DNA synthesis in which additional mutations are introduced. While MutSα mainly detects base-base mismatches and small IDL, MutSβ is specialized on recognizing larger IDLs. This could explain the crucial role of MutSα in SHM, whereas MutSβ seems to be dispensable[22]. In addition, MutSα was observed to stimulate polymerase η (Pol η) catalysis *in vitro*, thus further contributing to SHM[90]. Pol η is one of several and probably the major translesion synthesis polymerase that appears to be involved in this phase of SHM, as discussed in the next chapter.

Besides its enzymatic function in the second phase of SHM, MSH6 was suggested to play a scaffolding role in the first phase of SHM. Studies with Msh6 mutant mice, lacking MMR activity but expressing wild-type levels of the protein, were defective for phase II, but retained the ability to produce phase I mutations. However, the mutations often appeared to be non-targeted, outside the usual hot spot motifs, thus implying a role for MSH6 in the active targeting of AID to the Ig locus[317].

In contrast to MutSα, MutL homologs seem to play a minor role in SHM. Although the frequency of mutations in Mlh1-deficient mice was twofold lower than in wt mice, the pattern of mutation in *Mlh1*[-/-] clones was similar to wild-type clones. These findings suggested no direct role for Mlh1 in SHM[318]. Similar results were obtained for Pms2 deficiency with respect to SHM. However, a role for Pms2 in CSR was postulated. A recent study identified increased donor/acceptor homology at switch junctions of Pms2-deficient mice and proposed that class switching can occur by microhomology-mediated end-joining. The fact that the lack of Pms2 only affects CSR, while SHM remains unaffected reflects differences in the pathways of break resolution. This is in contrast to the findings for MSH2, whose deficiency equally alters the patterns of SHM and CSR, suggesting parallels in the two mechanisms[319].

As opposed to MLH1 and PMS2, MLH3 deficiency leads to an increased frequency of mutations in the immunoglobulin variable regions, compared to wild type littermates. Hence, MLH3 is the only DNA repair protein whose loss has a positive impact on the rate of SHM, implicating MLH3 in inhibiting the accumulation of mutations in SHM[80].

Besides its role in SHM, MutSα is involved in the process of class switch recombination. It was shown to bind to and promote synapsis of transcriptionaly activated immunoglobulin switch regions. Transcription through those sequence stretches causes formation of characteristic structures that have been identified as extended G-loops[317, 320, 321] and high affinity binding of MutSα to G4 DNA was observed[322]. A knock-out of *Msh2* leads to a 2-10 fold reduction in istoype switching[323, 324]. Studies with mice carrying a mutation in the ATPase domain of Msh2 also revealed disturbed CSR, thus suggesting that the ATPase activity is essential for efficient CSR switching[325]. Similar CSR impairment was observed for Msh6-deficient mice, but not in animals lacking Msh3, indicating that, as for SHM, it is indeed MutSα and not MutSβ that is essential for efficient CSR[316]. MutLα seems to also contribute to CSR, however, its impact is less distinct, since Pms2- and Mlh1-deficient mice exhibit only a 2-4 fold reduction of isotype switching[319, 324].

An attractive model for MMR protein involvement in CSR is to convert SSBs that are not in close proximity to each other on opposite DNA strands to DSBs[113, 114] (Fig. 2.19; also includes model of DSB generation in switch regions by AID-UNG-APE). If

SSBs are near each other on opposite strands, DSBs can form spontaneously, due to thermodynamic lability. However, as S-regions are large and AID-initiated breaks can occur anywhere within that region, it seems unlikely that they are mostly close enough to be spontaneously converted to DSBs. So in most cases, distal SSBs would necessitate conversion to DSBs. A likely scenario is the following: MutSα recognizes and binds to AID-mediated U/G mismatches, MutLα is recruited and EXO1 excises from the nearest 5' SSB created by AID-UDG-APE activity toward the mispaired U/G. EXO1 is then hypothesized to continue the excision past the mismatch until a SSB is reached on the opposing strand, thus forming a DSB[103-105, 114].
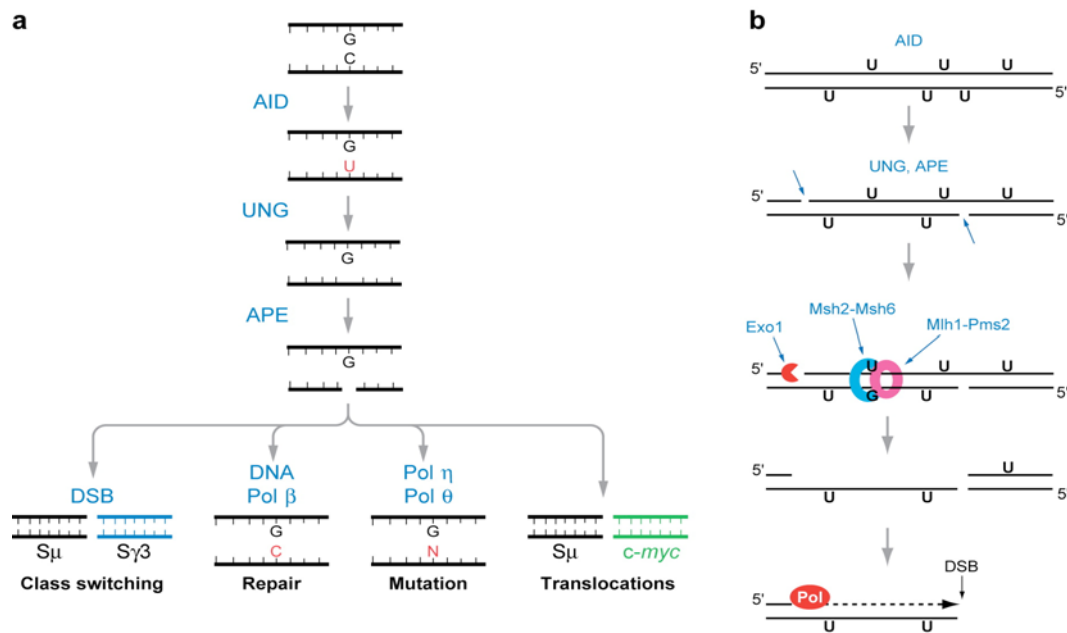

Fig. 2.19 Diagrams of (**a**) model for generation of DNA breaks, mutations, and translocations in IgS regions by AID-UNG-APE and (**b**) model for conversion of SSBs to DSBs by mismatch repair. **a** AID deaminates dC, resulting in dU bases, which are excised by one of the uracil DNA glycosylases, UNG. Abasic sites are cut by AP-endonuclease (APE1 and APE2), creating SSBs that can spontaneously form DSBs if they are near each other on opposite DNA strands, or, if not, mismatch repair activity converts them toDSBs (see **b**). Alternatively, DNA Pol β can correctly repair the nick, preventing CSR, or error-prone translesion polymerases can repair the nick but introduce mutations. Finally, the DNA breaks can lead to aberrant recombinations/translocations. **b** AID is hypothesized to introduce several dU residues in S regions during one cell cycle. Some of the dU residues are excised by UNG, and some of the abasic sites are nicked by APE. The U/G mismatches that remain would be substrates for Msh2-Msh6. Msh2-Msh6, along with Mlh1-Pms2, recruit Exo1 (and accessory proteins) to a nearby 5' nick, from where Exo1 begins to excise toward the mismatch, creating a DSB with a 5' single-strand overhang, which can be filled in by DNA polymerase. Fill-in synthesis is probably performed by translesion polymerases owing to the presence of abasic sites. Alternatively, the overhang is removed by a 5' flap endonuclease (Fen1) or by Exo1. Adapted from[111].

2.4.2   Involvement of glycosylases

There are four mammalian uracil DNA glycosylases linked to the BER pathway: Uracil-DNA-glycosylase (UDG), single-strand selective monofunctional uracil-DNA glycosylase 1 (SMUG1), thymidine-DNA glycosylase (TDG) and methyl-CpG binding domain protein 4 (MBD4)[326]. Together they are responsible for the removal of uracil from the deoxyribose-phosphate backbone.

UDG provides the main uracil excision activity measured in cell extracts and has also been shown to be primarily involved in initiating BER of the Ig genes in centroblast B cells. It participates in the major pathways leading to transversion mutations at G/C base pairs[327, 328], Ig gene conversion[329, 330] and CSR[328, 331]. The genetic inactivation of UDG causes a profound effect on SHM and CSR in humans and mice[327, 331], whereas a deficiency in other glycosylases has little or no effect[21, 332, 333] (TDG has not been tested). In humans, a mutation in UDG is associated with the severe human immunodeficiency syndrome, HIGM ("Hyper IgM"), characterized by elevated levels of IgM and also characteristic of AID$^{-/-}$ patients[331]. Comparable to $Udg^{-/-}$ mice[328], deficient isotype switching in this human genetic disease is accompanied by a shift in the hypermutation spectrum, increasing the frequency of transition mutations at C/G base pairs. Transitions at C/G pairs can occur when a U/G mismatch created upon deamination is fixed during replication.

CSR is reduced by 95% in B cells from Udg-deficient mice and in patients that have deleterious mutations in *UDG*. Besides, S region DSBs are significantly reduced in B cells induced to undergo CSR[105, 328, 331]. None of the other glycosylases, TDG, SMUG1, nor MBD4, can substitute for UDG in SHM and CSR[96, 328, 332, 333], suggesting that it is not just uracil removal *per se* that is important, but specifically excision by UDG. The possible involvement of SMUG1 was studied more intensively and it was observed that when overexpressed, SMUG1 can partially compensate for UDG when UDG is absent. However, in normal B cells, in the presence of UDG, SMUG1 actually seems to diminish the frequency of Ig diversification rather than supporting it, probably by favoring faithful repair of the uracil[333]. The dominance of UDG over the other glycosylases reflects a common theme in antibody diversification. In spite of considerable redundancy, one particular enzyme and its downstream partners appear to be hijacked by the centroblast B cell to process the Ig V region in an error-prone manner.

The observation that UDG-deficiency strongly inhibits switch recombination *in vitro*, but does not abolish isotype switching *in vivo*, led to the assumption that there must be a UDG-independent path of initiating U/G mismatch recognition to CSR[328, 331]. Indeed, subsequent experiments showed that MSH2/MSH6 and UDG provide alternative backup pathways for processing the U/G lesion. Thus, it was shown that only when UDG and MSH2 are simultaneously deleted[334] is CSR entirely ablated, implying that MutSα is also involved in the original U/G mismatch recognition and processing. Similarly, although phase II SHM at A/T base pairs is significantly reduced in Msh2/Msh6 deficient mice, it is completely absent only in Msh2/Udg and Msh6/Udg double-deficient mice. This suggests that a UDG-dependent, BER pathway also contributes to A/T mutagenesis[96, 328, 335]. However, since short-patch BER only acts on the original U/G mismatch, it is likely that long-patch BER and/or MMR are responsible for introducing mutations at neighboring A/T base pairs, although the exact enzymatic mechanism is not yet elucidated.

To summarize, in mice deficient for Udg and Msh2/Msh6, Aid-triggered DNA deamination at IgV and switch regions, leads solely to the generation of G-to-A and C-to-T transition substitutions[96]. Hence, it seems that in the absence of both Udg and MutSα the U/G mispair is scarcely detected as a lesion but is simply replicated.

2.4.3 The role of translesion synthesis polymerases

While phase I mutations at G/C base pairs are a direct consequence of the C→U deamination catalyzed by AID, mutations at A/T pairs require a second mutagenic process that occurs during error-prone patch repair of the AID-generated U/G mismatch. Mounting evidence suggests the involvement of a group of error-prone DNA polymerases called translesion synthesis (TLS) DNA polymerases, known to bypass DNA lesions. Most DNA lesions cannot be accommodated in the active site of replicative DNA polymerases and thus block the replication machinery. In principal there are two ways to circumvent this block: "damage tolerance", using recombinational mechanisms to copy genetic information from the undamaged sister duplex, or translesion synthesis, a process where nucleotides are incorporated opposite the damage or, after excision of the damaged base, opposite the abasic site[93].

In 1999, a new family of DNA polymerases was discovered that was named Y-family and subsequently four mammalian members of that family where shown to be involved in TLS: Pol η, ι, κ and Rev1. Together with Pol ζ, a B-family polymerase, these TLS polymerases are attractive candidates for error-prone patch repair in SHM. TLS polymerases clearly differ from replicative polymerases such as Pol δ in that they exhibit relatively poor processivity and low fidelity, due to the lack of a 3'-5' proofreading exonuclease activity. They are ubiquitously expressed and possess a shallower and less constrained binding pocket than the high-fidelity polymerases[92].

Accumulating genetic evidence[336-342] pointed out a pivotal role for Pol η and thus declared it the prime candidate for SHM at A/T base pairs. Pol η is responsible for bypassing cyclobutane pyrimidine dimers, the major UV photoproduct, and mutations in its gene cause the disease Xeroderma pigmentosum-variant (XP-V)[343]. This disorder is a consequence of the disability to correctly bypass UV-induced DNA lesions[344] and patients suffering from XP-V were observed to exhibit an altered SHM spectrum, with fewer mutations at A/T pairs, suggesting that the absence of Pol η has an effect on the targeting and type of mutations that occur. Mutations at A/T base pairs can thus be considered a direct result of misincorporated nucleotides by this polymerase during patch DNA synthesis[336]. As in humans, Pol η-deficient mice were observed to exhibit a diminished frequency of mutations at A/T base pairs[337, 340]. However, although A/T mutations are greatly reduced in *Pol η*$^{-/-}$ mice, they are essentially only totally abolished in *Pol η*$^{-/-}$, *Msh2*$^{-/-}$ double knock-out cells[342]. A likely explanation of these results is that MSH2-MSH6-mediated recruitment of Pol η is the major pathway to mutations at A/T pairs. However, in the absence of Pol η, a backup polymerase is recruited, providing a low background of mutations at A/T. Further support for the MutSα-dependent recruitment of Pol η comes from the recent finding that MSH2/MSH6 both interact with and stimulate DNA polymerase η[90].

The DNA polymerase ι that colocalizes with Pol η at replication foci[345], has also been implicated in the Ig gene mutagenesis induced in a Burkitt's lymphoma cell line[346]. After co-culture with T cells and surface IgM crosslinking, BL2 cells showed a 4-fold increase in Pol ι mRNA expression. Consistent with SHM-created C-to-T transition, Pol ι introduces T/G mispairs at a frequency of 10^{-2} [347]. However, inactivation of the Pol ι gene failed to alter the pattern of hypermutation[348].

Furthermore, B cell lines and mice that display low levels of the Rev3 transcript, encoding the catalytic subunit of the mispair extender Pol ζ, revealed a decreased mutation frequency at all base pairs. Moreover, impaired affinity maturation of B cells was observed in the mouse study[349, 350]. Yet another study claims the involvement of the translesion polymerase θ, which possesses a dual nucleotide mispair inserter-extender function[351]. Rev1, although structurally a *bona fide* member of the Y-family of TLS polymerases, is in fact a deoxycytidyl transferase, capable of inserting dCMP opposite either Gs or abasic sites in template DNA[352]. Rev1 has a strong preference for inserting deoxycytidine residues into the newly-synthesized strand, which would result in C-to-G and G-to-C transversion mutations[352, 353]. Hence, such mutations are significantly reduced during IgV hypermutation in chicken DT40 and mouse B cells that are deficient for Rev1[354, 355]. Thus, there is good evidence for a role of Rev1 in the first phase of SHM.

Finally the question that needs to be solved is how the IgV region is temporally and spatially targeted for error-prone repair, while under normal circumstances faithful repair is crucial for the maintenance of genome stability. A recent study, which showed that Rad6/Rad18-mediated detection of AID-initiated DNA lesions triggers monoubiquitination of proliferating cell nuclear antigen (PCNA), suggested that this process may be key to the recruitment of error-prone polymerases[356]. This nicely fits the model of a switch from high-fidelity to error-prone DNA synthesis mediated by monoubiquitination of PCNA[356-359]. Modified PCNA may be required to recruit REV1 to repair the abasic sites or the ssDNA breaks arising in short-patch BER. In addition, Pol η was shown to interact with monoubiquitinated PCNA[360], which could nicely explain the initiation of phase II SHM. Several recent studies support the hypothesis of a role for monoubiquitylated PCNA in SHM through facilitating A/T mutagenesis in the Ig gene *via* both MMR and long-patch BER[356-359]. Buerstedde and colleagues investigated the role of PCNA ubiquitination in variants of the DT40 B cell line mutated in K164 of PCNA or in Rad18, involved in PCNA ubiquitination. Interestingly, they not only detected an increased sensitivity of the cells to DNA-damaging agents but also a significant reduction in AID-dependent single-nucleotide substitutions in the immunoglobulin light-chain locus[358]. This demonstrated the first evidence that PCNA-monoubiquitination is exploited for Ig SHM, most likely through the recruitment of error-prone translesion synthesis polymerases.

3   AIM OF MY STUDIES

One of the major challenges of the field is to identify the factors that target AID to its sites of action at the immunoglobulin locus. We therefore set out to identify proteins interacting with AID, in the hope that they could provide us with novel insights into this phenomenon.

The second open question concerns the role of the MMR proteins in SHM and CSR. This we decided to address by using our well-established *in vitro* mismatch repair assay system specifically modified for the purpose by incorporating U/G mispairs in critical sites of the heteroduplex DNA substrate used in the assay.

4   RESULTS

4.1   Identification of new interaction partners of AID

Activation-induced deaminase (AID) was recently shown to trigger somatic hypermutation (SHM) and class switch recombination (CSR)[17]. However, despite extensive research over the last decade, little is known about the partners of AID that are involved in the process of antibody diversification, in particular its specific targeting to the immunoglobulin locus. In an attempt to gain novel insights into this process, we searched for proteins interacting with AID by means of various protein purification and interaction approaches in combination with mass-spectrometry (MS) based analysis.

Since genetic evidence suggests the involvement of MMR proteins in SHM and CSR[21-23] we were particularly interested in uncovering both, direct and indirect interactions between AID and known players of the mismatch repair (MMR) proteins.

4.1.1 Tandem-Affinity-Purification (TAP)

As a first approach, we set out to establish a Tandem-Affinity Purification (TAP) protocol. Several reports have demonstrated the potential of this method, originally set up in *S. cerevisae*[361, 362], for the identification of interacting partners of known proteins in various host cell lines[362, 363]. TAP allows identification of interacting proteins under native conditions *in vivo*, and, due to the double-tag conferred on the protein of interest, with a high degree of purity[363]. The TAP-tag consists of two IgG binding domains of the S. *aureus* protein A and of the calmodulin binding peptide, which are separated by a Tobacco Etch Virus (TEV) protease cleavage site. After successful expression of the recombinant proteins, the AID-mediated protein complex is isolated from the cells *via* TAP, on IgG sepharose and calmodulin affinity resin columns. Subsequent MS-based analysis then reveals the identity of the isolated proteins and bioinformatic annotation enables setting up of protein pathway maps, as shown in Fig. 4.1. In order to avoid competition for interaction partners, TAP-based studies are best carried out using a cell line that lacks the corresponding endogenous protein.
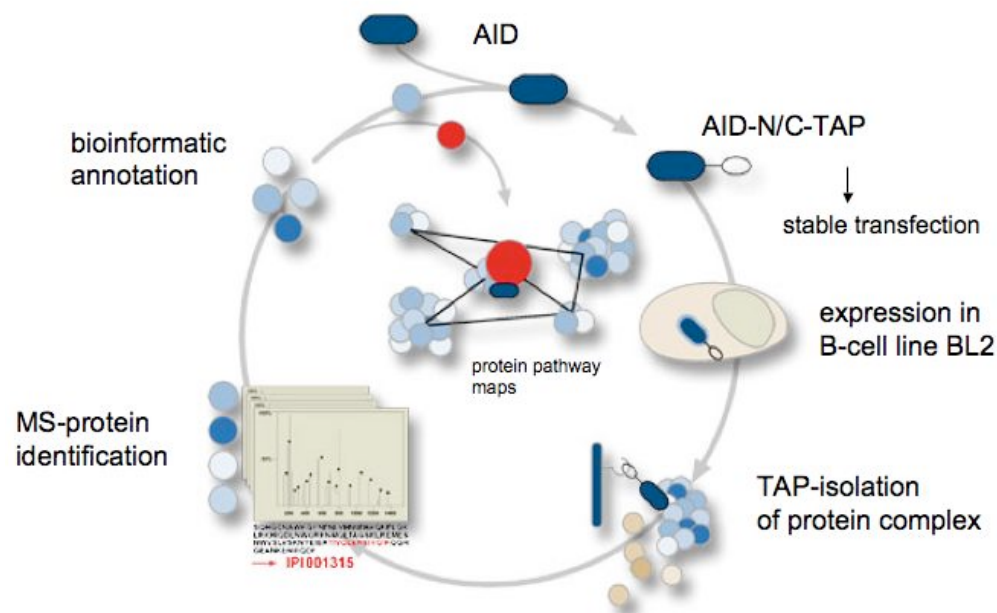
Fig. 4.1 Schematic illustration of a tandem-affinity purification (TAP) approach to isolate activation-induced deaminase (AID) in complex with its interacting proteins. In a first step, a TAP-tag is joined to either the N-or C-terminus of the AID protein. The construct is then stably transfected into BL2 cells, a Burkitt's lymphoma cell line. After expression of the recombinant protein, the protein complex is isolated from the cell extracts. Subsequent mass spectrometric analysis reveals the protein identities and bioinformatic annotation allows to set up protein pathway maps.

We first created DNA vectors for the expression of recombinant AID protein with a N- or C-terminal TAP-tag (pZome-1-N-AID, pZome1-C-AID). As a control, we used the empty vector (pZome-1-N). All constructs were then stably transfected into the Burkitt's lymphoma cell line BL2 by electroporation.

Human Burkitt's lymphoma cell lines have previously been used as *in vitro* models for studying the hypermutation mechanism[364-366]. These cell lines represent the transformed counterparts of germinal center (GC) centroblast B cells, which have already experienced hypermutation *in vivo*. While some Burkitt's lymphoma cell lines exhibit constitutive hypermutation, the process has to be induced in others by surface receptor cross-linking. We chose the Burkitt's lymphoma cell line BL2[365] as a model system. Based on this cell line, Faili and colleagues created an AID-deficient cell line through homologous recombination[66], which we stably transfected with our TAP-AID constructs.

As shown in Fig. 4.2 BL2 cells express all relevant mismatch repair proteins (MSH2/6/3 and MLH1/PMS2) in amounts comparable to HeLa, our reference cell line. The targeted inactivation of AID through homologous recombination affected neither the overall expression of MMR proteins (Fig. 4.2, lane BL2 AID$^{-/-}$), nor the MMR efficiency of the cell extracts, as observed in *in vitro* mismatch repair assays (data not shown).
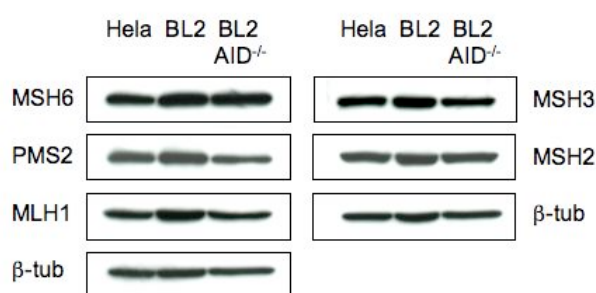
Fig. 4.2 Western blots showing mismatch repair proteins in HeLa and Burkitt's lymphoma BL2 cells. All relevant mismatch repair proteins (MSH2/3/6 and MLH1/PMS2) are expressed in similar amounts in all three cell lines. No obvious differences were observed between parental BL2 and BL2 AID$^{-/-}$ cells. β-tubulin served as a loading control.

Stable transfectants were isolated by puromycin resistance and single clones were analyzed by Western blot for the expression of the TAP-tagged AID protein. Unfortunately, no AID protein could be detected in any of the clones (data not shown). To check whether the AID gene was integrated into the genome of the isolated clones, we performed RT-PCR (Fig. 4.3). During PCR the increase of measured fluorescence is directly proportional to the amount of double-strand DNA

generated. As illustrated in figure 4.3a (upper panel), we detected a fluorescence signal in both AID-N-TAP- and AID-C-TAP single BL2 cell clones, as well as in the positive control sample (AID-proficient BL2 cells), when AID primers were used. This indicated successful integration and transcription of the AID gene. As negative controls we used RNA isolated from BL2 cells that were transfected with the empty pZome-N-TAP vector, and RNA from colon mucosa, that does not express AID. In none of the negative control samples a fluorescence signal was observed. In contrast, all samples contained transcripts for the internal control gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (Fig. 4.3a, lower panel). The RT-PCR results were confirmed by agarose gel electrophoresis (Fig. 4.3b).
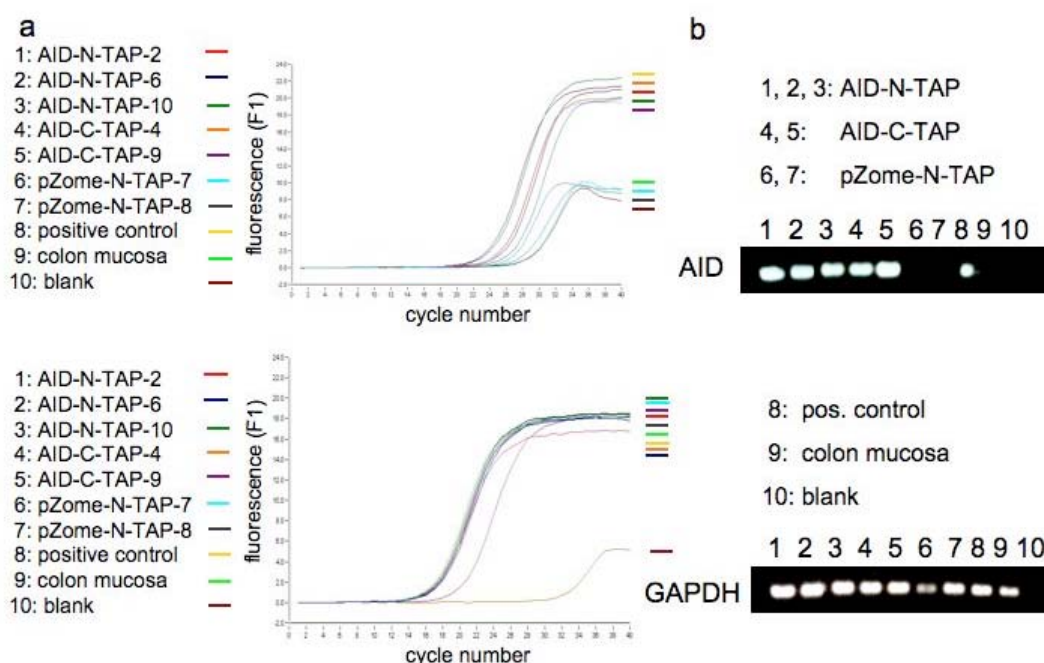
Fig. 4.3 RT-PCR analysis of BL2-TAP-AID single cell clones. **a** RT-PCR plots showing cycle number versus fluorescence intensity for the indicated gene transcripts. **b** Agarose gel depicting gene transcripts amplified by RT-PCR. Strong signals were obtained for AID-TAP cell clones (lanes 1-5, upper panel) and the positive control (lane 8, upper panel) when AID-primers were used. Negative control fractions and blank (lane 6, 7, 9, 10, upper panel) showed no signal, indicating the absence of gene transcripts. The internal control gene GAPDH was expressed in similar amounts in all samples analyzed (lower panel).

To compare the expression of TAP-tagged AID in B cells versus non-B cells, we transiently transfected pZome-1-N-AID, pZome1-C-AID as well as the empty vector pZome-1-N into the human epithelial kidney cell line 293T. Analyzing the resulting single cell clones by Western blot, we were able to detect TAP-tagged AID protein in

293T cells, whereas no signal was obtained in AID-transfected BL2 cells (data not shown). Subsequent RT-PCR studies comparing RNA from the 293T- and BL2-AID cell clones revealed that the level of AID-transcripts measured in transfected 293T cells was substantially greater than in AID-expressing BL2 transfectants (Fig. 4.4 a,b; upper panels). We selected AID-N-TAP-expressing clones derived from 293T and BL2 cells and performed tandem affinity purification from 60 mg of whole cell extracts. The isolated protein complexes were analyzed by western blotting. While small amounts of highly pure AID protein could be detected in the experiment performed with 293T cells, no AID protein was detectable in the elution fraction from BL2 cells (data not shown). Although AID was successfully purified from transiently transfected 293T cells, the amount of total protein was insufficient for MS-based protein identification.
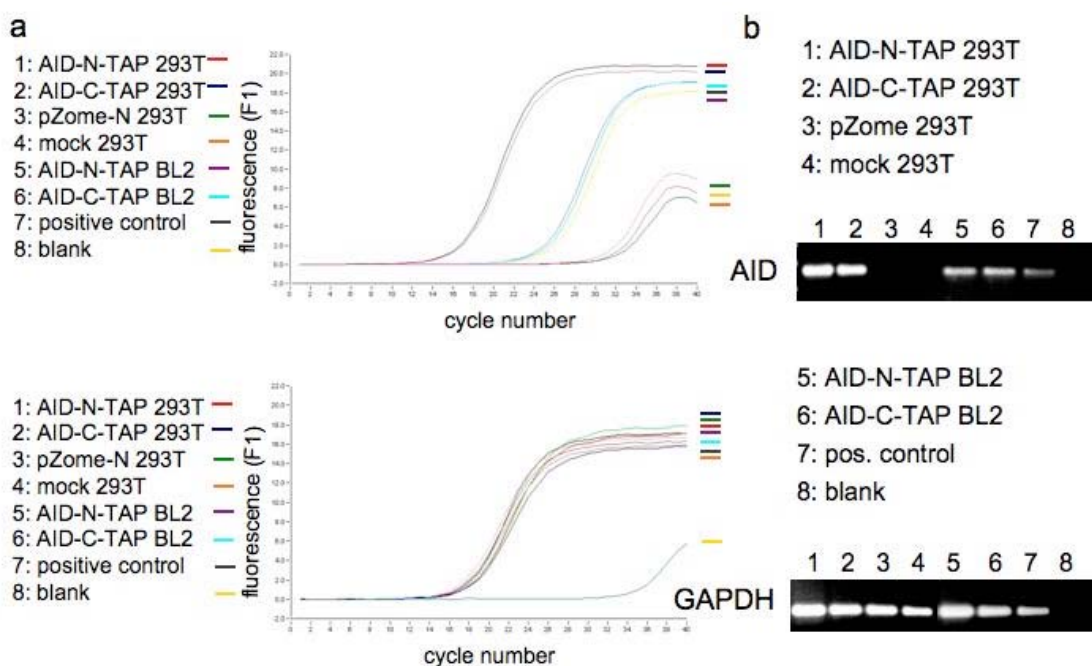
Fig. 4.4 RT-PCR analysis comparing AID-TAP clones from a 293T and BL2 cell background. **a** RT-PCR plots showing cycle number versus fluorescence intensity. **b** Agarose gels comparing the AID transcript levels of AID-TAP-293T and AID-TAP-BL2 cell clones. Whereas the expression levels of the AID transcripts was higher in the samples derived from 293T cell clones (lane 1, 2, upper panel) compared to BL2 cell clones (lane 5, 6), the expression level of the GAPDH transcript was comparable in all samples analyzed (lane 1-7, lower panel).

4.1.2 GST-AID pulldown approach

On the basis of our observations from the TAP-experiment, we concluded that, using mammalian cells, we cannot express recombinant TAP-AID protein in quantities necessary for the subsequent MS analysis. We therefore decided to change the expression system and attempted to overexpress our protein of interest in *E. coli* BL21 cells.

In order to purify the full-length human AID protein from the bacterial cell extracts, we decided to exploit the GST-gene fusion system and constructed an inducible prokaryotic GST-AID expression vector. Glutathione-S-Transferase (GST) binds with high affinity to glutathione-coated agarose beads that are used to isolate the recombinant protein of interest from bacterial cell extracts. The bound GST-hAID protein was then incubated with whole cell extracts from the human B cell line Ramos to allow the AID protein to form complexes with its potential partners.

To distinguish between proteins that bind unspecifically to the GST-tag and those that bind specifically to AID, we performed a second 'pseudo-purification' experiment, using an empty GST-expression vector that was transformed into *E. coli* BL21. The GST and GST-AID protein concentrations were estimated based on Coomassie-staining (gel not shown). Equal amounts of recovered proteins from the GST- and the GST-AID purifications were incubated with Ramos cell extracts and interacting proteins were separated by electrophoresis. Silver-staining of the gel revealed the bait proteins used for the pulldown, GST-only and GST-AID, respectively, as well as additional bands, representing the potential AID interacting proteins (Fig. 4.5 a). The results were highly reproducible, as judged by comparing silver-stained gels from independent experiments. The gel was then cut into ten blocks containing either protein bands from the bait or the unknown proteins. For each block containing a potential interaction partner of GST-AID, the corresponding gel piece in the GST-only fraction was cut out as a control, allowing the elimination of unspecifically bound proteins. The gel pieces were destained, the contained proteins in-gel trypsin digested and subjected to MS analysis. The origin of the identified peptides was determined by Mascot database search.
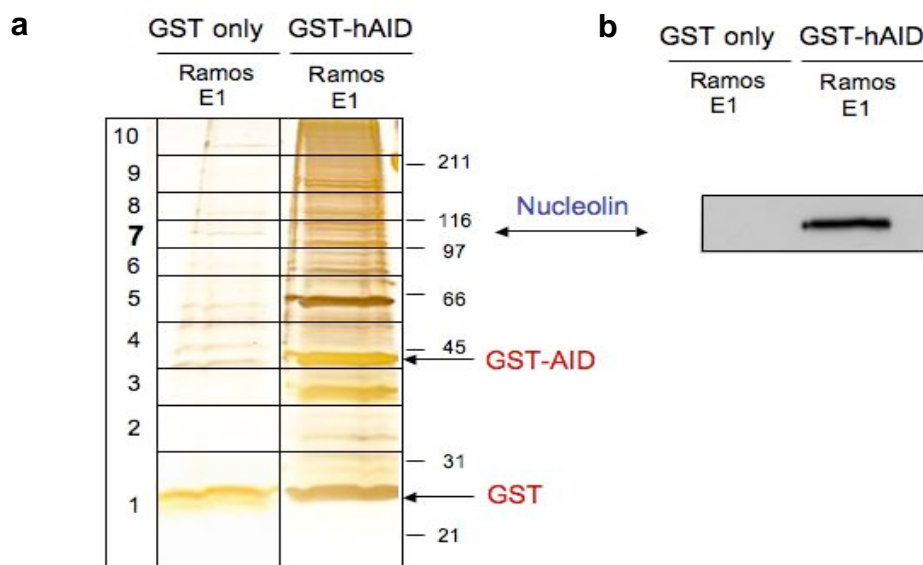
Fig. 4.5 Pull-down of recombinant GST-hAID from *E. coli* incubated with human Ramos B cell extracts. **a** Silver-stained 12% SDS-polyacrylamide gel showing proteins pulled down with the GST-AID construct (right lane) and the corresponding control fraction from the GST-only pulldown (left lane). The respective bait proteins are highlighted in red. Arrows indicate the corresponding silver-stained protein bands. Numbers (1-10) specify the gel slices analyzed by MS. **b** Western blot revealing the presence of nucleolin protein in the GST-AID pull-down fraction (right lane). The complete absence of nucleolin in the control fraction (left lane) suggests that the AID-nucleolin interaction is specific for GST-AID.

In both cases, the bait proteins GST-only and GST-AID were among the most prominent bands. Their identity was verified by MS analysis and western blotting (WB data not shown). Table 4.1 shows the MS data for the AID-bait and a selection of AID interactors. Among other proteins, the MS analysis led to the identification of DNA-PK (see Table 4.1). DNA-PK was previously shown to interact with AID[50], thus validating our GST-pulldown approach as a suitable tool to search for new interaction partners of AID. Nucleolin was detected as the most abundant protein after AID, the bait itself. Nucleolin was shown to be a component of the B cell-specific transcription factor and switch region binding protein LR1[367]. LR1 contains two polypeptides of 106 kDa and 45 kDa. The 106 kDa component is nucleolin that was identified in our screen at the corresponding molecular weight. Transcriptional activation by LR1 was observed at the Eμ enhancer and switch region sites[368, 369]. These data implicated nucleolin in class switch recombination and possibly also in somatic hypermutation and thus convinced us to further examine its potential interaction with AID. As shown by western blotting in Fig. 4.5b, nucleolin was specifically detected in the GST-AID

pulldown experiment, whereas it was completely absent from the GST-only control, indicating that the interaction of AID with nucleolin was specific. In an attempt to investigate this interaction *in vivo*, we constructed a N-terminal haemaglutinin (HA)-AID-fusion construct, as well as a wild-type AID (WT) construct. The corresponding human HA-nucleolin and WT-nucleolin expression vectors were kindly provided by Nancy Maizels. We transfected 293T cells with either HA-nucleolin and WT-AID or HA-AID and WT-nucleolin DNA vectors. The respective HA-tagged proteins in complex with interacting partners were immunoprecipitated from whole cell extracts. The immunoprecipitate was analyzed by SDS-PAGE and by western blotting for the presence of AID and nucleolin. Although HA-nucleolin and HA-AID were present in the respective samples, we could not detect the AID protein in the HA-nucleolin precipitate, nor could we observe nucleolin in the HA-AID precipitate (data not shown). Thus, the interaction between AID and nucleolin, as suggested from the initial GST-AID pulldown experiment, could not be confirmed upon overexpression of the proteins in human 293T cells.

GST-hAID pulldown		
Protein	Protein score	Protein mass
Activation-induced cytidine deaminase (AID)	2705	24337
Nucleolin (Protein C23)	2455	76494
Ras GTPase-activating-like protein (p195)	2181	189761
Poly(ADP-ribose)polymerase 1 (PARP-1)	2056	113680
Heterogeneous nuclear ribonucleoprotein U (hnRNP U)	1753	91033
HNRPA1 protein	1029	29482
Interleukin enhancer-binding factor 3	914	95678
DNA Topoisomerase 1	906	91125
Gamma-interferon-inducible protein	677	88618
Replication factor C (large subunit) RF-C 140 kDa subunit	629	128688
DNA-dependent protein kinase catalytic subunit (DNA-PK)	622	473749
Cell-division cycle 5-like protein (Cdc5-like protein)	553	92422
Chromodomain helicase-DNA-binding protein 4	407	219393

Table 4.1 MS data obtained from GST-hAID pull-down after incubation with Ramos B cell extracts. The table shows a selection of potential AID-interacting proteins with their corresponding Mascot protein scores and protein mass. The bait protein AID highlighted in red was the most abundant protein in the complex fraction analyzed. Nucleolin (blue) was chosen for follow-up studies as described in the text. The other proteins will be targets of further studies.

4.1.3 AID-(phospho)peptide pull-down studies

As a third approach to identify new binding partners of AID, we used AID specific peptides, rather than the full-length protein, as baits. In order to distinguish between interactions depending on phosphorylation of AID and proteins that bind unmodified AID, we designed peptides that were either phosphorylated or unmodified.

The feasibility of a (phospho)peptide pull-down and its validation as means of identifying new biological interaction partners of known proteins was recently demonstrated[370]. Based on this study, we set out to establish our own protocol for a phospho-peptide pull-down. The initial step towards realizing this approach was the design of peptides consisting of 18 amino acids, derived from the wild-type AID sequence (Fig. 4.6a, hypothetical AID structure, Fig. 4.6b). Mass spectrometric analysis revealed that nuclear AID is phosphorylated on serine 38 (S38) and tyrosine 184 (Y184) in activated B cells[51]. Notably, the S38 residue is located within a cyclic AMP-dependent PKA consensus phosphorylation site that is conserved in AID across all species that undergo CSR. Thus, we designed four AID-specific peptides comprising either phosphorylated or non-phosphorylated S38, or phosphorylated or non-phosphorylated Y184, respectively (Fig. 4.6c). Each peptide was biotinylated at the N-terminus, allowing binding to streptavidin-coated magnetic beads. The bead-bound peptides were incubated with Ramos cell extracts to allow binding of interacting proteins. Non-specific interactors were washed off and the remaining proteins eluted. The protein mixture was separated on a SDS gel and the gel was silver-stained (gel not shown). Gel blocks or distinct bands containing high amounts of protein were in-gel trypsin digested and the protein identity established by MS.

Fig. 4.7a shows a list of selected proteins that were identified in the MS screen of proteins pulled-down with the S38 peptides. As for the GST-AID pull-down experiment, DNA-PK was among the most abundant proteins in the eluates from non-phosphorylated and phosphorylated S38 peptide pull-downs. The DNA repair protein RAD50 was specifically pulled-down with the non-phosphorylated S38 peptide, as was the helicase RuvB-like2. The protein RuvB-like1 was mainly found in the unmodified S38 pull-down fraction, but was also detected in the sample pull-down with the phosphorylated S38 peptide.
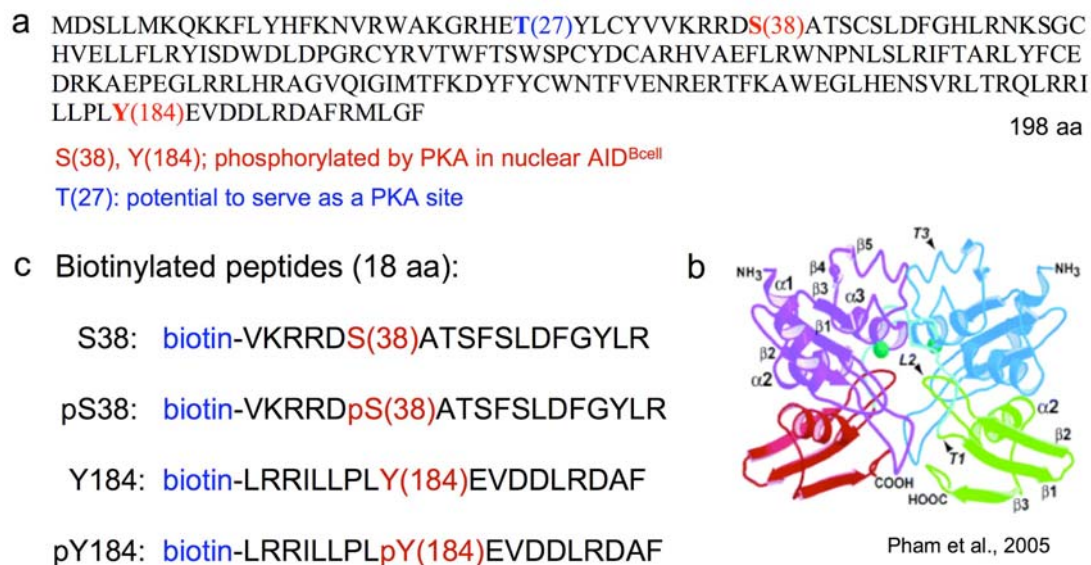
a MDSLLMKQKKFLYHFKNVRWAKGRHE**T(27)**YLCYVVKRRD**S(38)**ATSCSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCE
DRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRI
LLPL**Y(184)**EVDDLRDAFRMLGF

198 aa

S(38), Y(184); phosphorylated by PKA in nuclear AID[Bcell]

T(27): potential to serve as a PKA site

c Biotinylated peptides (18 aa):

S38:   biotin-VKRRDS(38)ATSFSLDFGYLR

pS38:   biotin-VKRRDpS(38)ATSFSLDFGYLR

Y184:   biotin-LRRILLPLY(184)EVDDLRDAF

pY184:   biotin-LRRILLPLpY(184)EVDDLRDAF

b

Pham et al., 2005

Fig. 4.6 Sequence of the AID-protein and biotinylated (phospho)AID peptides. **a** AID is a 198 amino acid (aa) protein that is phosphorylated on serine 38 (S38) and tyrosine 184 (Y184) (red). Threonine 27 (T27) potentially serves as a PKA site (blue). **b** Hypothetical AID structure based on structural alignment with yeast CDD1[371]. **c** AID-specific 18 aa peptides, containing either the S38 or Y184 phosphorylation site. All peptides are biotinylated at the N-terminus, allowing binding to streptavidin-coated beads.

The results of the MS analysis were subsequently confirmed by western blotting (Fig. 4.7b). *In vivo*, RAD50 exists in a complex with MRE11 and NBS1, forming the MRN complex. Thus MRE11 and NBS1 were expected to be pulled-down with the S38 fragment, just like RAD50. Although no peptides were identified via MS, MRE11 and NBS1 could be detected by western blot. The MRN complex has previously been linked to immunoglobulin diversification and thus became a first focus of our studies. Nancy Maizels and colleagues postulated that the MRN complex accelerates somatic hypermutation and gene conversion[89]. Furthermore, they claimed that MRE11/RAD50 possesses a DNA cleavage activity in the AID/UNG-dependent pathway of antibody diversification[88].

To closer investigate the suggested interaction between AID and the MRN complex, we performed several immunoprecipitation (IP) experiments. We transfected human 293T cells with either HA-AID and NBS1-WT, or HA-NBS1 and AID-WT expression constructs, and used anti-HA antibodies to pull-down the AID-NBS1 complex, possibly in conjunction with RAD50 and MRE11. However, neither of the approaches succeeded in revealing a direct interaction between AID and NBS1. We then tested the possibility that the interaction of AID and the MRN complex was

mediated by RAD50 or MRE11. We transfected 293T cells with the AID-WT expression vector, performed an AID-IP and immunoblotted the eluate for the presence of RAD50 and MRE11. Unfortunately, no direct interaction between AID and RAD50 or between AID and MRE11, could be observed. Thus the interaction between AID and RAD50/MRN, indicated by the initial AID peptide pull-down, could not be confirmed using full-length AID protein expressed in human 293T cells.
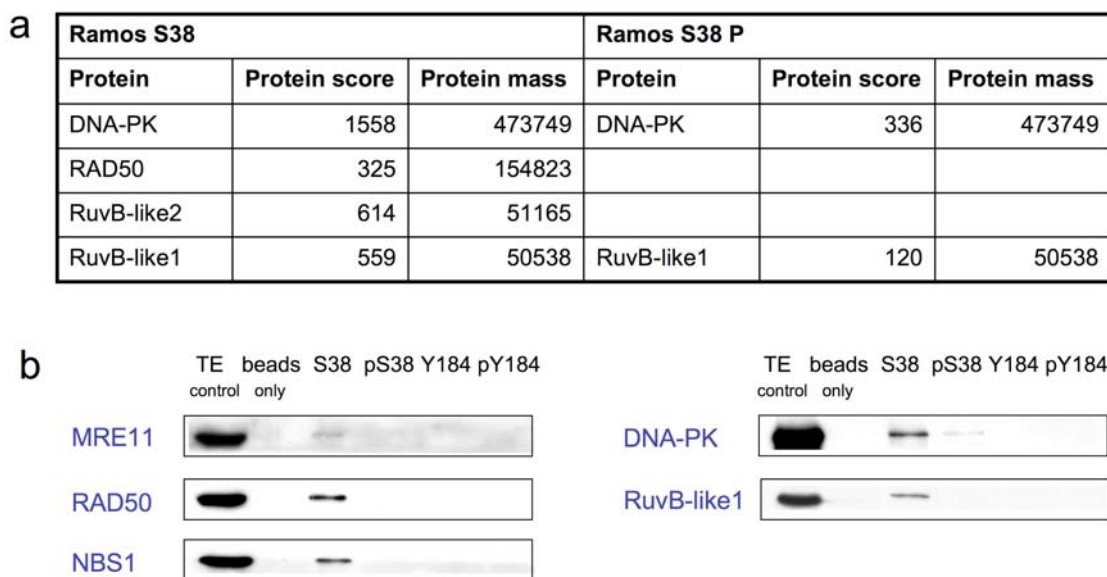

a

Ramos S38			Ramos S38 P		
Protein	Protein score	Protein mass	Protein	Protein score	Protein mass
DNA-PK	1558	473749	DNA-PK	336	473749
RAD50	325	154823			
RuvB-like2	614	51165			
RuvB-like1	559	50538	RuvB-like1	120	50538

b

Fig. 4.7 AID-(phospho)-peptide pulldown *via* streptavidin-conjugated magnetic beads. **a** Table showing MS data of selected proteins pulled down with the non-phosphorylated and phosphorylated S38 peptides (S38/S38P). Protein name, score and mass are listed. **b** Western blots confirming the MS data. Besides RAD50, also MRE11 and NBS1, the other components of the MRN complex, were detected in the pull-down fraction of the S38 peptide. DNA-PK, as a known interactor of AID, was confirmed and RuvB-like1 was revealed as a novel protein interacting with AID.

The next targets we investigated more closely were the RuvB-like1/2 proteins. These helicases are implicated in diverse cellular processes, such as the control of transcription and chromatin remodeling[372]. An interaction was shown with RNA PolII and RPA (p14), thus emphasizing their role in transcription[373]. These data provided a possible link to AID, which was observed to interact with RNA PolII and RPA (p32) as well[40, 54]. The *in vivo* interaction between full-length AID and RuvB-like1 or RuvB-like2, could be confirmed by co-immunoprecipitation experiments in human 293T cells (data not shown). These and subsequent experiments revealed a novel function of RuvB-like1/2 in cell division (Castor *et al.*, manuscript in preparation).

In an attempt to investigate whether the AID-RuvB-like1 interaction is direct, we exploited an ELISA (enzyme-linked immunosorbent assay)-based protein interaction assay. This approach necessitates highly pure protein. Thus we expressed recombinant GST-AID and His-tagged RuvB-like1 protein in *E. coli* and purified it *via* glutathione sepharose- and Ni-NTA-columns, respectively (Fig. 4.8a). Constant amounts of GST-AID were used to coat the ELISA plate and were subsequently incubated with increasing amounts of the His-RuvB-like1 protein. The absolute absorbance increased linearly with increasing amounts of RuvB-like1, indicating specific binding of RuvB-like1 to GST-AID. In contrast, when GST-AID was incubated with various controls (GST-only, carbonate assay buffer, BSA) no specific interaction with GST-AID was detectable (Fig. 4.8b).
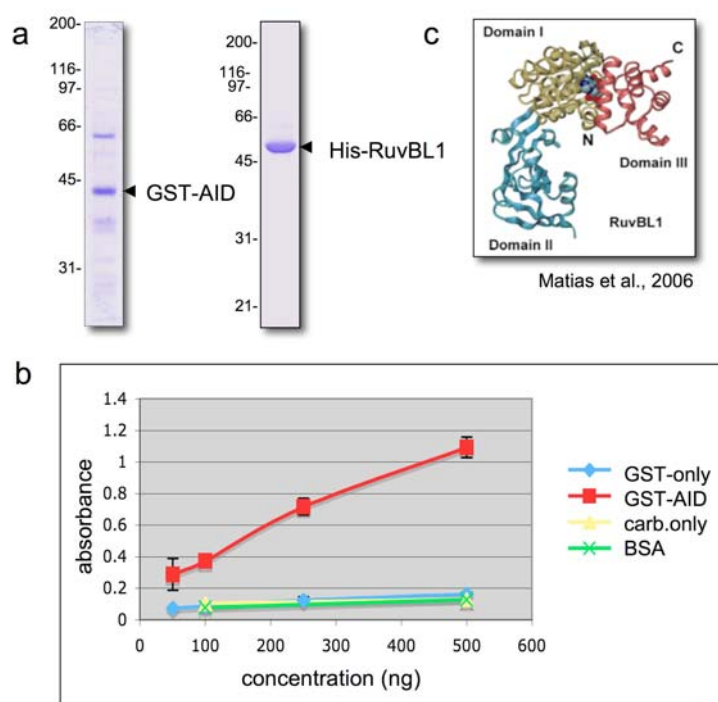

Fig. 4.8 Interaction between AID and RuvB-like1. **a** Coomassie-stained gels showing GST-AID and His-RuvB-like1 purified by glutathione-sepharose- and Ni-NTA-columns. **b** ELISA of RuvBL1 binding to GST-AID. The plate was coated with either GST-AID (red) or controls [GST-only (blue), carbonate buffer (yellow), BSA (green)] and incubated with increasing amounts of RuvBL1 (50, 100, 250, 500ng). The increased absorbance measured for GST-AID indicated specific binding to RuvBL1. **c** Scheme illustrating the crystal structure of RuvBL1 with its three distinct protein domains I (yellow), II (blue) and III (red)[374].

Recently, the crystal structure of RuvB-like1 was published, revealing three distinct protein domains (Fig. 4.8c). To further map the AID-interaction domain on RuvB-like1, we took advantage of the available structural information and created several RuvB-like1 deletion mutants comprising distinct protein domains (domain I, II or III). These protein domains were expressed individually in *E. coli* and analyzed for their binding to full-length AID by ELISA as described. The interaction was mainly mediated by amino acids in domain I of RuvB-like1. Domain II also seems to contribute to the interaction with AID, whereas domain III does not interact with full-length AID (data not shown).

4.2 Mechanistic insights into the role of MMR in somatic hypermutation

4.2.1 Interference of mismatch repair in the processing of U/G mispairs arising through cytosine deamination (manuscript submitted).

Silvia Schanz, Dennis Castor, Franziska Fischer and Josef Jiricny

In this paper, we studied the interference between mismatch repair (MMR) and base excision repair (BER) in the processing of AID-generated U/G mispairs, using an *in vitro* MMR assay system. We incubated whole BL2 cell extracts with heteroduplex DNA substrates containing two U/G mispairs, such as might arise during processive deamination of cytosines by AID. We could show that uracil residues can act as initiation sites for MMR-dependent repair of neighboring mispairs. Using cell lines lacking either MLH1 or MSH2 (293TLα- cells and LoVo, respectively), we could furthermore show that this uracil-directed repair is dependent on the MSH2/MSH6 heterodimer, MutSα, while the MLH1/PMS2 heterodimer, MutLα, seems to be of minor importance. These findings are in line with observations made in studies of nick-directed MMR and help explain results of genetic studies that revealed a dominant role for MutSα in SHM.

Interference of mismatch repair in the processing of U/G mispairs arising through cytosine deamination.

Silvia Schanz[1,2], Dennis Castor[1,2], Franziska Fischer[1] and Josef Jiricny[1,2]*

[1]*Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190,*
CH-8057 Zurich, Switzerland

[2]*Department of Biology, Swiss Federal Institutes of Technology, Sonneggstrasse 23,*
CH-8092 Zurich, Switzerland

Abstract: 258 words

*Corresponding author
Tel.:   +41-44-635 3450
Fax.:   +41-44-635 3484
E-mail: jiricny@imcr.uzh.ch

Antibody diversity in vertebrates is mediated by three metabolic processes: VDJ recombination, somatic hypermutation (SHM) and class switch recombination (CSR). The two latter processes are initiated by activation-induced deaminase (AID), which converts selected cytosines downstream from transcription initiation sites in the immunoglobulin locus to uracils[16, 17, 21, 375]. Although uracil processing is generally error-free, metabolism of AID-induced U/G mispairs in antigen-stimulated B-cells gives rise to numerous mutations, both at the deaminated C/G base pairs and at distal A/T sites. While the former mutations could be linked to the *UNG* gene encoding uracil DNA glycosylase, the latter were linked to the mismatch repair genes *MSH2*, *MSH6* and *EXO1*, as well as the gene encoding polymerase-η[376]. We now show that in substrates containing several U/G mispairs, such as might arise through the processive action of AID, a subset of uracil residues can act as initiation sites for a MSH2/MSH6-dependent repair of neighbouring U/G mispairs. As immunoglobulin gene rearrangements require ubiquitylation of proliferating cell nuclear antigen (PCNA)[358], and as ubiquitylated PCNA preferentially interacts with polymerase-η[360] it is likely that the PCNA-dependent resynthesis of the long mismatch repair tracts will in some cases be catalysed by polymerase-η, which has a propensity towards mutations at A/T base pairs. Our work provides the first mechanistic insight into somatic hypermutation and helps explain how two error-free processes could give rise to a high number of mutations. Our data also imply that MMR and BER interference in the processing of uracil residues on opposite strands could give rise to double strand breaks, which are postulated to trigger CSR.

AID is believed to be targeted to its sites of action by transcription[40, 59, 62]. A single deamination event in a transcription bubble would generate a uracil residue in one of the strands, which would give rise to a U/G mispair once the bubble had moved on and the DNA had reannealed. Under normal circumstances, this lesion is addressed by base excision repair (BER), where the uracil is removed by the uracil DNA glycosylase UNG2, giving rise to an AP-site[21, 22]. The subsequent cleavage of the sugar-phosphate backbone by an AP-endonuclease provides an entry point for polymerase-β, which extends the 3'-terminus of the nick by inserting a dCMP residue. Concurrently, the N-terminal domain of pol-β cleaves off the baseless sugar-

phosphate residue to leave a nick, which is sealed by DNA ligase III/XRCC1[377]. We wanted to test whether this canonical BER process takes place also in extracts of B-cells, and whether we could find any evidence of the involvement of MMR in U/G processing as suggested by the genetic studies.

We constructed supercoiled phagemid substrates containing a single U/G or G/T mispair in the recognition sequence of the *Acl*I endonuclease. The presence of these mispairs makes the DNA refractory to cleavage by this enzyme at this site, but correction of U/G to C/G, or of G/T to A/T, will regenerate the cleavage site and digestion of the heteroduplex DNA with *Acl*I will give rise to a different restriction fragment pattern (Fig. 1a). As shown in Fig. 1b, the U/G repair reaction was very efficient (lane 1). Inhibition of the two major uracil glycosylases, TDG (by immunodepletion) and UNG2 (by the uracil glycosylase inhibitor, Ugi) made the substrate refractory to cleavage (lane 2), which confirmed that the U/G mispair was processed by BER in this extract. The G/T mismatch in the supercoiled substrate was also not repaired (lane 5). This confirmed that circular heteroduplexes lacking nicks or gaps are refractory to MMR, because the latter strand discontinuities represent essential entry points for exonuclease 1 (EXO1), which catalyses the degradation of the discontinuous strand of the heteroduplex up to and ~150 nucleotides past the mispair[187, 378]. Correspondingly, the U/G and G/T substrates were very efficiently processed by MMR in the same substrates into which a nick was introduced ~350 nucleotides 5' from the U or the G, respectively (lanes 4 and 6, respectively).

The above experiments confirmed that a single U/G mispair in a covalently-closed circular substrate is not subject to MMR-dependent processing, in spite of being efficiently recognised by the mismatch recognition factor MutSα (a heterodimer of MSH2 and MSH6)[96], which is highly abundant in proliferating B-cells[379]. Interestingly, available biochemical evidence suggests that AID acts processively to generate several uracil residues in close proximity[31, 32, 39]. In this scenario, MutSα might interfere with U/G processing, due to its mode of action. Upon mismatch recognition, MutSα undergoes an ATP-dependent conformational change, which converts it into a sliding clamp that causes it to leave the mismatch and diffuse along the DNA contour in search of a nick[187, 230, 380]. We argued that a MutSα sliding clamp loaded at one U/G might interfere with uracil processing by BER at a distal U/G site. In order to test this hypothesis, we constructed a supercoiled substrate

containing a U/G mismatch situated 54 base pairs from a G/T mispair in the *Acl*I restriction site. [A G/T rather than a U/G mispair in the *Acl*I site was used in these experiments because it is refractory to BER (Fig. 1b, lane 5), yet its processing by MMR is comparable to U/G (cf. lanes 4 and 6).] In the control reaction, the substrate contained, in addition to the two mispairs, also a nick 350 nucleotides 5' from the G/T in the uracil-containing strand. As shown in Fig. 2a, the G/T mismatches in the nicked U/G-G/T substrate (lane 3) and in the nicked G/T control substrate (lane 5) were corrected with high efficiency to A/T. The former reaction was slightly inhibited by the addition of Ugi (lane 4), which might be explained by an interference of MMR and BER at the U/G mispair. Importantly, the G/T mismatch in the supercoiled U/G-G/T substrate was also corrected with appreciable efficiency (lane 1), and this repair could be almost totally inhibited by the addition of Ugi (lane 2). This shows that uracil processing in these extracts is not concerted and that BER intermediates are accessible to other enzymes, possibly because polymerase-β levels (and thus presumably also in BER efficiency) are reduced in B-cells[381]. In this case, MutSα loaded at the downstream G/T mismatch apparently activates EXO1 to load at the cleaved AP-site and initiate a strand displacement reaction that results in the conversion of the G/T mismatch to an A/T base pair.

Having demonstrated that a uracil residue in a U/G mispair can act as a cryptic strand discontinuity and thus provide an entry point for MMR, we wanted to test whether uracils incorporated opposite adenines behave in a similar way. To this end, we constructed a substrate where the uracil residue was placed opposite an adenine 54 nucleotides 5' from the G/T mismatch in the *Acl*I site. As shown (Fig. 2b, lane 1), a U/A base pair was used at least as efficiently as a U/G mispair (Fig. 2a, lane 1) for MMR initiation.

Should AID act in a processive manner, it would be anticipated to generate several uracil residues, and thus several U/G mispairs, within the transcribed region. In this scenario, processing of the uracils by BER, or of the U/G mispairs by MMR, could theoretically begin at any one of the deaminated sites. Due to the short repair patch generated during BER, no interference between the processing of sites as close as ~10 nucleotides would be expected. In contrast, if both BER and MMR were involved, the outcome of the repair would depend on where the first strand break was introduced. In the above experiments, we used a combination of U/G and G/T

mispairs to show that uracil processing provides an entry point for mismatch-activated strand displacement. In order to demonstrate the relevance of this finding to SHM, we constructed a substrate containing two nearby U/G mismatches (Fig. 2c). Incubation of this supercoiled substrate with B-cell extracts resulted in an efficient repair of the U/G within the *Acl*I site to a C/G (lane 1), which indicated that the MMR-catalysed strand displacement initiated at the upstream uracil. Addition of Ugi inhibited the reaction (lane 2), which showed that the strand displacement process was absolutely dependent on UNG2 activity. When the U/G-U/G substrate contained also a nick 350 nucleotides upstream, the repair by MMR was more efficient (lane 3) and was only slightly inhibited by the addition of Ugi (lane 4), which showed that MMR is the predominant pathway of U/G processing in B-cell extracts.

Mismatch repair is a bidirectional process, where the strand degradation reaction can take place in both 5'→3' and 3'→5' directions[187, 311]. If mismatch- and ATP-activated MutSα encounters a nick 5' from the mispair, it can load EXO1, which then catalyses the degradation of the nicked strand in a 5'→3' direction until the mispair had been removed. Importantly, while this reaction is stimulated by the MLH1/PMS2 heterodimer MutLα, it is not absolutely dependent on it. In contrast, when the strand discrimination signal lies 3' from the mismatch, MutSα recruits MutLα, which then introduces several nicks into the discontinuous strand, both upstream and downstream from the mispair[221, 275]. These nicks are then used as excision initiation sites and the strand degradation proceeds in the 5'→ 3' direction, as dictated by the polarity of EXO1. When the strand degradation begins at nicks situated 5' from the mispair, the repair polymerase filling-in the EXO1-generated gap will replace the mispaired nucleotide with a correct one. In order to test whether this factor requirement applies also to the uracil-directed MMR, we incubated the U/G-G/T substrate with extracts lacking either MutSα or MutLα. Extracts of 293T-Lα[+] cells express all four MMR proteins[382] and are MMR-proficient, as judged also by the >70% efficiency of G/T → A/T repair in a 5' nicked heteroduplex substrate (Fig. 3a, lane 5). In these extracts, the G/T → A/T repair in the supercoiled U/G-G/T substrate was ~40% (lane 1). Downregulation of MLH1 expression in the 293T-Lα cells by doxycycline gives rise to a MMR-defect[382]. In these 293T-Lα[-] extracts, the efficiency of G/T repair in the supercoiled U/G-G/T substrate was reduced to ~20% (lane 2). In contrast, the same substrate incubated with extracts of MutSα-deficient LoVo cells

was fully refractory to *Acl*I cleavage (lane 3). Repair efficiency could be restored by the addition of purified recombinant MutSα (lane 4), which shows that solely the lack of the latter factor was responsible for the MMR defect in these extracts.

The above results show that a MutSα defect abolishes 5'→3' repair, whereas MutLα deficiency only decreases its efficiency. This agrees with the finding from the Modrich laboratory, which demonstrated that MutLα increases the efficiency of 5'→3' MMR, but that it is essential for the 3'→5'reaction[221, 275]. In order to test whether this applies also to the uracil-directed MMR, we constructed the G/T-U/G substrate, which contained the uracil residue 3' from the G. This substrate was refractory to MMR in both 293T-Lα⁻ (Fig. 3b, lane 2) and LoVo (lane 3) extracts. In contrast, the 3' G/T-U/G and the 3'-nicked control G/T substrate were repaired with similar efficiencies in the MMR-proficient 293T-La+ extracts (lanes 1,5) and in LoVo extracts supplemented with MutSα (lane 4). The absolute dependence of the MMR-dependent U/G repair on MutSα but not MutLα corroborate the results of genetic studies, which implicated the *MSH2* and *MSH6*, but not the *MLH1* and *PMS2* loci in SHM[21] and imply that the MutSα-dependent 5'→3' process predominates during this phase of diversification of immunoglobulin genes.

Based on the experiments presented in this study, and taking into account the large body of genetic evidence from other laboratories, we propose that SHM results from an interference between MMR and BER during the processing of multiple U/G mispairs, such as those arising in the immunoglobulin locus of activated B-cells through the processive action of AID. As outlined in Fig. 4, these data help explain why AID-induced deamination of cytidines is error-prone. If a mismatch- and ATP-activated MutSα sliding clamp were to interfere with the processing of a U/G mispair immediately after the removal of the uracil by UNG2, but before cleavage of the sugar-phosphate backbone, the AP-site might persist until S-phase, where its by-pass might give rise to class I mutations. Should MMR interfere with U/G processing after the AP-site had been cleaved, the EXO1-generated gap might be filled-in by an error-prone polymerase such as pol-η. Because this polymerase has a propensity to introduce non-complementary nucleotides opposite Ts[93], this would give rise to class II mutations. Biochemical studies aimed at verifying the latter hypothesis are currently in progress in our laboratory.

METHODS

Substrates, nuclear extracts and *in vitro* MMR assays. The detailed procedure has been described previously[383]. Briefly, heteroduplex DNA substrates containing a G/T or a U/G mismatch within an *Acl*I restriction site in the 46-bp polylinker of a pGEM13Zf(+) derivative were constructed by primer extension, using the mismatch-containing 50-mer oligonucleotides (G/T: 5'-GGC CGC GAT CTG ATC AGA TCC AGA CGT CTG TCA AUG TTG GGA AGC TTG AG-3'; U/G: 5'-GGC CGC GAT CTG ATC AGA TCC AGA CGT CTG TCG ACG TTG GGA AGC TTG AG-3') as primers and the single-stranded phagemid DNA as template. To introduce a second U/G mismatch or U/A base pair 5' or 3' from the *Acl*I restriction site, 89(85)-mer oligonucleotides containing an additional uracil at the indicated site were used (U/G-U/G: 5'-CCA GTG AAT TGT AAT AUG AAC ACT ATA GGG CGA ATT GGC GGC CGC GAT CTG ATC AGA TCC AGA CGT CTG TCA AUG TTG GGA AGC TTG AG-3'; U/G-G/T: 5'-CCA GTG AAT TGT AAT AUG AAC ACT ATA GGG CGA ATT GGC GGC CGC GAT CTG ATC AGA TCC AGA CGT CTG TCG ACG TTG GGA AGC TTG AG-3'; U/A-G/T: 5'- CCA GTG AA T TGT AAU ACG AAC ACT ATA GGG CGA ATT GGC GGC CGC GAT CTG ATC AGA TCC AGA CGT CTG TCG ACG TTG GGA AGC TTG AG-3'; G/T-U/G: 5'-CCA GAC GTC TGT CG A CGT TGG GAA GCT TGA GTA TTC TAT AGT GTC ACC TAA ATA GCT TGG CGT AAT UAT GGT CAT AGC TGT TTC C-3'). Isolation of the desired supercoiled heteroduplex substrates and the MMR assays were carried out as described, using 100 ng (47.5 fmol) heteroduplex DNA substrate and 150 μg of nuclear extracts from BL2, 293T-Lα+ (MLH1[+]), 293T-Lα- (MLH1[-]) or Lovo cells in a total volume of 30 μl.

UDG inhibition and TDG-immunodepletion of nuclear extracts. Protein A Dynabeads were washed twice with 30 mM Hepes-KOH, pH 7.5, 7 mM MgCl₂. Anti-TDG antibody (1:10,000) was added and the beads were incubated for 2h at 4°C. They were then washed three times with the above buffer and stored at 4°. The extracts were immunodepleted of TDG by incubating with 6.3 μl antibody-preadsorbed Dynabeads for 30 min at 4°C and subsequently used for *in vitro* MMR assays. Where indicated, UNG2 was inhibited by the addition of 3.6 μl of UGI (7.2 units) per 150 μg nuclear extracts and incubation for 10 min at 37°C.

Antibodies and reagents. The anti-TDG antibody (rabbit polyclonal) was a generous gift of Primo Schar. The UDG inhibitor UGI was obtained from New England Biolabs and the Protein A Dynabeads were obtained from Dynal Biotech, Invitrogen.

REFERENCES

1. Di Noia, J. M. & Neuberger, M. S. Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem 76, 1-22 (2007).
2. Vallur, A. C., Yabuki, M., Larson, E. D. & Maizels, N. AID in antibody perfection. Cell Mol Life Sci 64, 555-65 (2007).
3. Muramatsu, M. *et al.* Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem 274, 18470-6 (1999).
4. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102, 553-63 (2000).
5. Neuberger, M. S. & Rada, C. Somatic hypermutation: activation-induced deaminase for C/G followed by polymerase eta for A/T. J Exp Med 204, 7-10 (2007).
6. Arakawa, H. *et al.* A role for PCNA ubiquitination in immunoglobulin hypermutation. PLoS Biol 4, e366 (2006).
7. Kannouche, P. L., Wing, J. & Lehmann, A. R. Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. Mol Cell 14, 491-500 (2004).
8. Odegard, V. H. & Schatz, D. G. Targeting of somatic hypermutation. Nat Rev Immunol 6, 573-83 (2006).
9. Duquette, M. L., Pham, P., Goodman, M. F. & Maizels, N. AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. Oncogene 24, 5791-8 (2005).
10. Nambu, Y. *et al.* Transcription-coupled events associating with immunoglobulin switch region chromatin. Science 302, 2137-40 (2003).

11.   Li, Z., Woo, C. J., Iglesias-Ussel, M. D., Ronai, D. & Scharff, M. D. The generation of antibody diversity through somatic hypermutation and class switch recombination. Genes Dev 18, 1-11 (2004).

12.   Sharma, R. A. & Dianov, G. L. Targeting base excision repair to improve cancer therapies. Mol Aspects Med 28, 345-74 (2007).

13.   Jiricny, J. The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 7, 335-46 (2006).

14.   Fang, W. H. & Modrich, P. Human strand-specific mismatch repair occurs by a bidirectional mechanism similar to that of the bacterial reaction. J Biol Chem 268, 11838-44 (1993).

15.   Rada, C., Di Noia, J. M. & Neuberger, M. S. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. Mol Cell 16, 163-71 (2004).

16.   Marra, G. *et al.* Expression of human MutS homolog 2 (hMSH2) protein in resting and proliferating cells. Oncogene 13, 2189-96 (1996).

17.   Bransteitter, R., Pham, P., Calabrese, P. & Goodman, M. F. Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. J Biol Chem 279, 51612-21 (2004).

18.   Pham, P., Bransteitter, R., Petruska, J. & Goodman, M. F. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature 424, 103-7 (2003).

19.   Goodman, M. F., Scharff, M. D. & Romesberg, F. E. AID-initiated purposeful mutations in immunoglobulin genes. Adv Immunol 94, 127-55 (2007).

20.   Jiricny, J. MutLalpha: at the cutting edge of mismatch repair. Cell 126, 239-41 (2006).

21.   Iaccarino, I., Marra, G., Dufner, P. & Jiricny, J. Mutation in the magnesium binding site of hMSH6 disables the hMutSalpha sliding clamp from translocating along DNA. J Biol Chem 275, 2080-6 (2000).

22.   Poltoratsky, V., Prasad, R., Horton, J. K. & Wilson, S. H. Down-regulation of DNA polymerase beta accompanies somatic hypermutation in human BL2 cell lines. DNA Repair (Amst) 6, 244-53 (2007).

23.   Iyer, R. R., Pluciennik, A., Burdett, V. & Modrich, P. L. DNA mismatch repair: functions and mechanisms. Chem Rev 106, 302-23 (2006).

24.   Modrich, P. Mechanisms in eukaryotic mismatch repair. J Biol Chem 281, 30305-9 (2006).

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

26.   Cejka, P. *et al.* Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. Embo J 22, 2245-54 (2003).

27.   Lehmann, A. R. *et al.* Translesion synthesis: Y-family polymerases and the polymerase switch. DNA Repair (Amst) 6, 891-9 (2007).

28.   Baerenfaller, K., Fischer, F. & Jiricny, J. Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro. Methods Enzymol 408, 285-303 (2006).

29.   Neddermann, P. & Jiricny, J. Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. Proc Natl Acad Sci U S A 91, 1642-6 (1994).

30.   Gallinari, P. & Jiricny, J. A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. Nature 383, 735-8 (1996).

LEGENDS TO FIGURES

Figure 1│ BER- and MMR-catalysed repair of U/G and G/T mispairs in extracts of BL2 cells. a, Schematic representation of the circular heteroduplex substrates carrying either a single U/G or G/T mispair in the recognition site of *Acl*I endonuclease at nucleotide 44 or 46, respectively. These heteroduplex substrates were constructed by primer extension as described previously[383]. The positions of the other *Acl*I cleavage sites and of the N.BstNBI site where a nick can be introduced selectively into the outer strand are indicated. The numbering relates to the inner (viral) strand of the heteroduplex. The substrates were incubated with the cell extracts in the presence or absence of the UNG inhibitor Ugi, either covalently-closed or nicked as shown above the respective figures. In the absence of repair, digestion of the phagemid DNA gave rise to two fragments of 2823 and 373 bp. Repair of the U/G mispair to C/G, or of the G/T to A/T regenerated the third *Acl*I restriction site, such that the phagemid DNA was cut into three fragments of 1516, 1307 and 373 nucleotide. The efficiency of the repair reaction was estimated from PhosphoImager scans of the ethidium bromide stained agarose gels. *b,* Repair of the U/G or G/T mismatches in covalently-closed (lanes 1,2,5) or nicked (lanes 3,4,6) heteroduplex substrates in extracts of BL2 cells. (NB: In all repair experiments in this study, the extracts were immunodepleted of TDG, because this enzyme could be shown to process U/G mispairs with high efficiency in human cell extracts[168, 170] (data not shown). MBD4 and SMUG1, the other known uracil glycosylases, are inactive under our reaction conditions, such that UNG2 is the only remaining uracil-processing enzyme in the TDG-depleted extracts.) The U/G mispair in the unnicked substrate was efficiently processed by BER (lane 1) and this reaction could be inhibited by the addition of Ugi (lane 2). In a nicked substrate, the U/G mispair was addressed also by MMR (lanes 3). Correspondingly, only the BER and not the MMR reaction could be inhibited by Ugi (lane 4). The control experiments show that the unnicked substrate G/T is refractory to MMR (lane 5), but is efficiently processed when a nick is introduced into the G strand by N.BstNBI (lane 6). M, molecular size marker.

Figure 2│ Intermediates of uracil processing can be used as MMR initiation sites in substrates containing a uracil residue upstream from a mispair. *a*, A covalently-closed circular substrate carrying a U/G mispair 54 nucleotides 5' from a
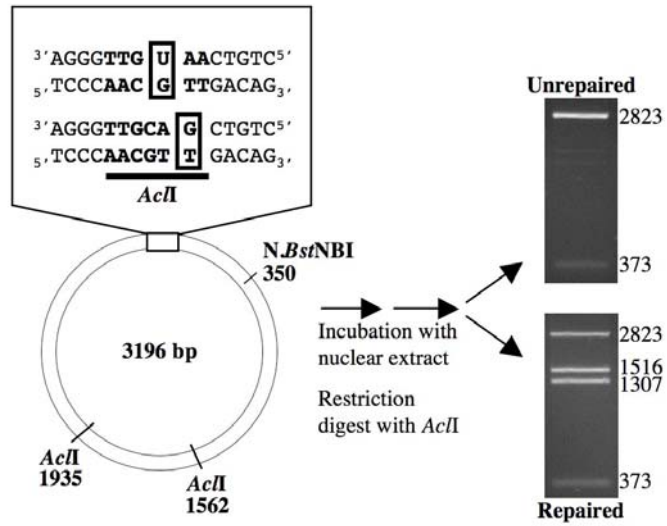
G/T mispair in the *Acl I* site was repaired with high efficiency (lane 1), but this reaction could be inhibited by the addition of Ugi (lane 2). This shows that uracil processing by UNG2 was indispensable for this MMR-catalysed process. Ugi inhibition could be overcome by the introduction of a nick into the substrate (lanes 3,4). The nicked G/T substrate was used as a control (lane 5). M, molecular size marker. *b,* A uracil residue opposite adenine can function similarly to that in a U/G mispair. In this experiment, the G/T substrate contained a U/A base pair 54 nucleotides 5' from the G. Lanes as above. *c,* In a covalently-closed substrate containing two U/G mispair, such as would arise through the processive action of AID, the mispair in the *Acl*I site (lane 1) was repaired to C/G with an efficiency similar to that observed in *a* above. The higher background in the Ugi-inhibited reaction (lane 2) is due to the fact that *Acl*I partially cleaves DNA containing a U/G mispair in its recognition sequence, as shown in the digest of the untreated supercoiled substrate (lane 6). Control lanes as above.

Figure 3│ Dependence of the uracil-initiated MMR reaction on MutSα and MutLα. *a,* The unnicked U/G-G/T substrate carrying a uracil residue 54 nucleotides 5' from the G was processed with high efficiency in MMR-proficient extracts of 293T-Lα$^+$ cells (lane 1). The G/T repair was less efficient in extracts of 293T-Lα$^-$ cells that do not express MutLα^{382} (lane 2). In contrast, extract of LoVo cells, which lack MutSα, were deficient in G/T repair (lane 3), unless supplemented with purified recombinant MutSα (lane 4). Lane 5 contains the nicked G/T control. *b,* The unnicked G/T-U/G substrate carrying a uracil residue 54 nucleotides 3' from the G was processed with high efficiency in MMR-proficient extracts of 293T-Lα$^+$ cells (lane 1). The G/T repair was detected neither in extracts of 293T-Lα$^-$ (lane 2) nor of LoVo (lane 3) cells. When the LoVo extracts were supplemented with purified recombinant MutSα, the substrate was processed efficiently (lane 4). Lane 5 contains the nicked G/T control. M, molecular size marker.
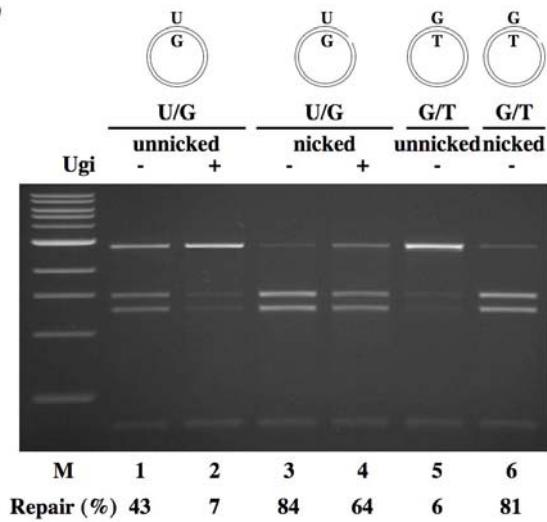
Figure 4│ Putative mechanism of somatic hypermutation. AID-induced deamination of cytosines in single-stranded DNA of a transcription bubble gives rise to a U/G mispair in duplex DNA once the bubble has moved on. This mispair may be detected by MutSα, but cannot be processed by MMR because it has no nicks in the

vicinity. It will therefore be addressed by BER. If BER were to be interrupted immediately after the action of UNG2, the uncleaved AP-site may persist until DNA replication, where its by-pass by REV1 would give rise to class I mutations at C/G sites. This situation changes once a second cytosine deamination takes place in the moving transcription bubble. Should a partially-processed deamination site lie within ~1kb (the maximum distance between a mismatch and a nick), the MutSα sliding clamp activated by the second U/G may interrupt the BER process at the upstream uracil. If MutSα were to encounter an AP-site cleaved either by APE1 or MRE11, it might load EXO1 and the subsequent 5'→3' strand degradation would give rise to a single-stranded region spanning the distance between the first deamination site and ~150 nucleotides past the second one. This gap would normally be filled-in by the high-fidelity replicative polymerase-δ. However, because rearrangements of the immunoglobulin locus in activated B-cells require ubiquitylation of the polymerase processivity factor PCNA[6], and because ubiquitylated PCNA preferentially associates with polymerase-η[360], it is possible that a subset of the MMR repair patches is filled-in by the latter enzyme. As polymerase-η is known to introduce non-complementary nucleotides (N) opposite Ts[93], such a process would give rise to the observed class II mutations at A/T base pairs.

a

M	1	2	3	4	5
Repair (%)	42	3	78	62	81

b

M	1	2	3	4	5
Repair (%)	72	6	75	64	79

c

M	1	2	3	4	5	6
Repair (%)	36	9	68	50	77	10

97

a
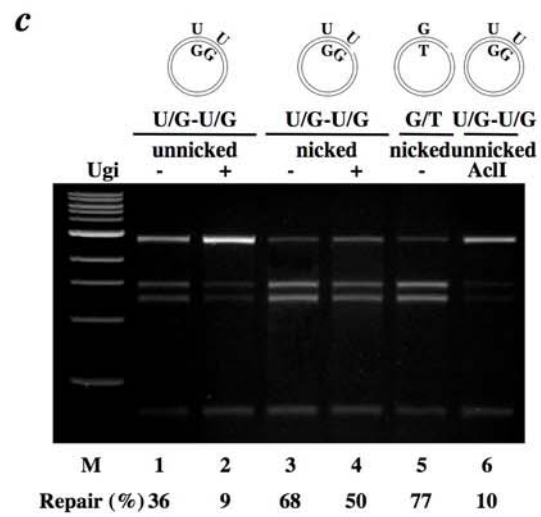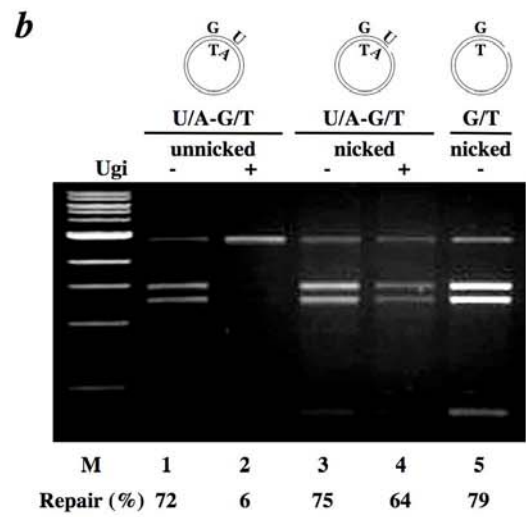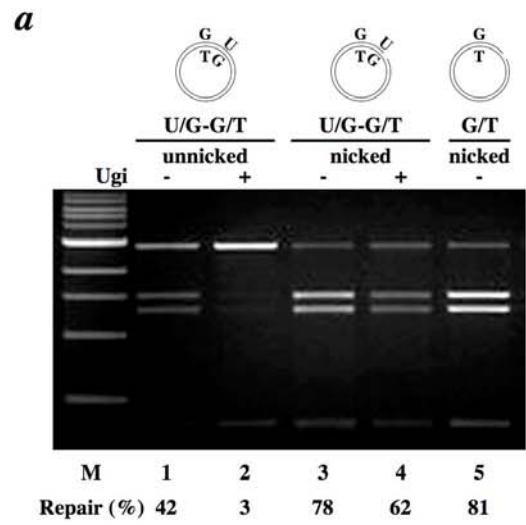
4.2.2   Characterization of cell extracts used in MMR assays

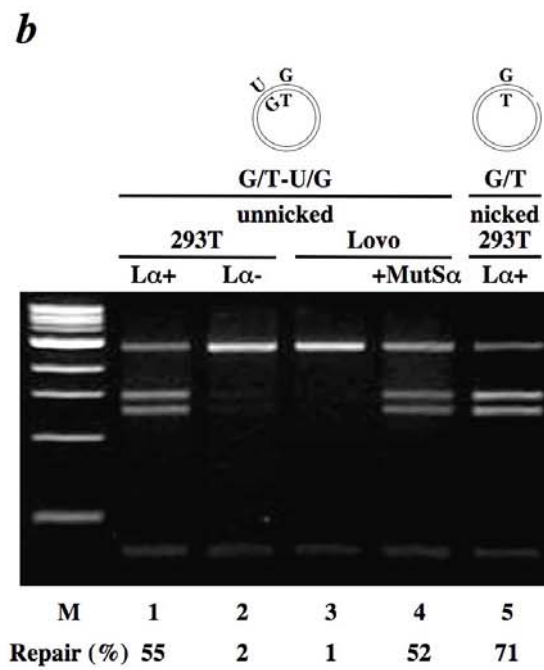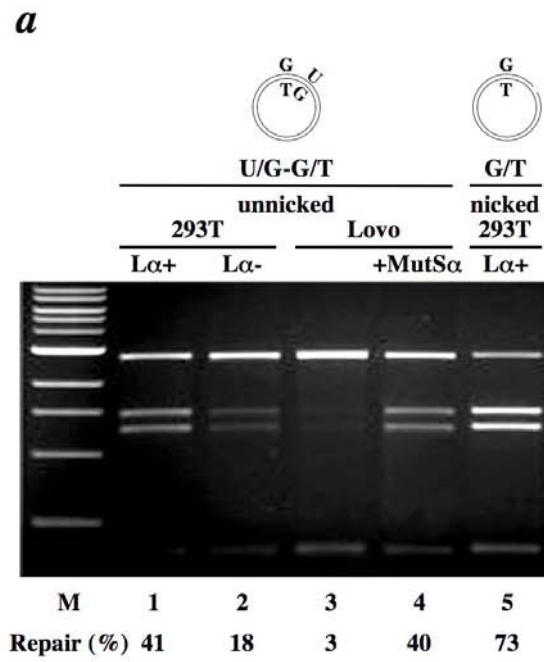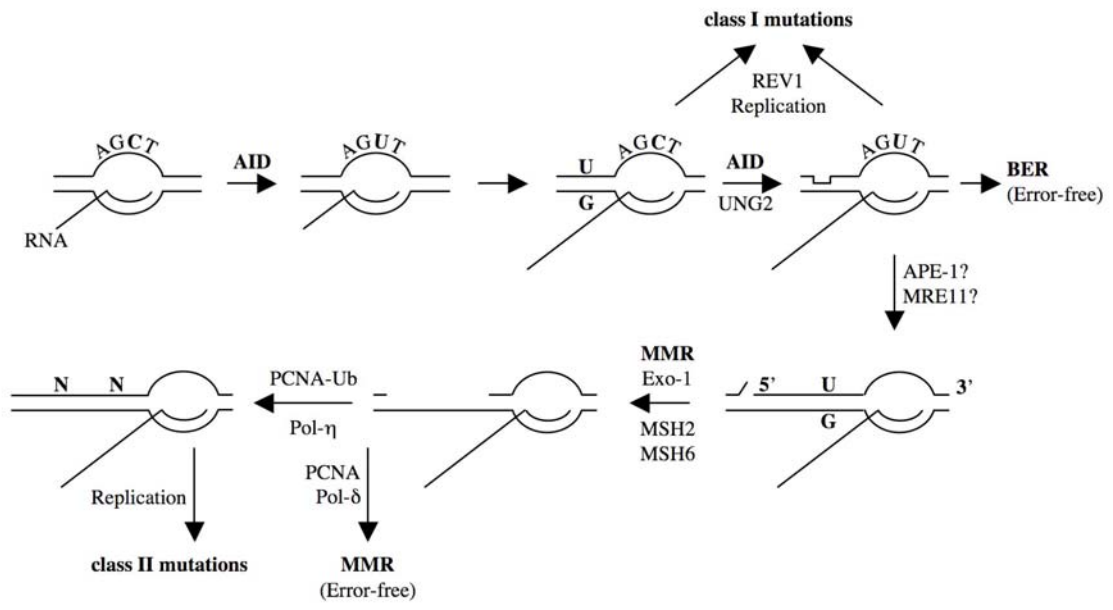We chose the Burkitt's lymphoma cell line BL2 as a model system for our *in vitro* mismatch repair assays. As described earlier, BL2 cells represent the transformed counterpart of germinal center B cells, where antibody diversification takes place.

Our studies were largely focused on the interplay between MMR and BER in the processing of U/G mispairs arising through cytosine deamination (see 4.2.1). We asked the question whether nicks appearing as intermediates during uracil processing could serve as entry point for MMR.

In order to investigate the role of UDG, the major uracil glycosylase in SHM, we ensured that no other uracil-processing activity was present in the cell extracts. Previous studies in our lab have shown that SMUG1 and MBD4 are inactive under the reaction conditions used for the MMR assay[384]. However, TDG is known to efficiently remove uracil residues[168-170] and is highly active on U/G DNA substrates (data not shown), implicating a role for TDG in the repair of deaminated cytosines[174]. To avoid a processing of our substrates by TDG, we depleted the BL2 cell extracts of this glycosylase by means of TDG-specific antibodies. As shown in Fig 4.9, depletion of both the high molecular weight and the low molecular weight forms of TDG was highly efficient and specific, as indicated by the unaffected levels of MMR proteins.
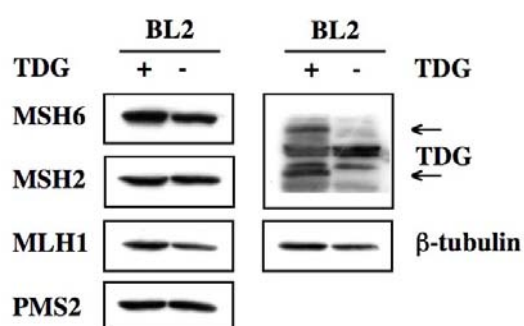
Fig. 4.9 Western blots showing MMR proteins and TDG in BL2 cell extracts before (+) and after (-) TDG depletion. TDG was efficiently removed, as judged by the disappearance of both the higher and lower molecular weight TDG variants (indicated by arrows). The level of MMR proteins before and after depletion remained unaffected. β-tubulin was used as a loading control.

In order to investigate the requirement for MutLα in uracil-directed MMR on heteroduplex DNA substrates, we opted to exploit the 293TLα cell system, established previously in our lab[382]. The expression of MLH1 can be down-regulated in these cells by doxycycline, resulting in a MMR defect. The lack of MLH1 simultaneously leads to degradation of its interaction partner PMS2. Thus, addition of

doxycycline to the cell culture medium gives rise to a complete loss of MutLα in these cells, referred to as 293TLα- (Fig. 4.10). In the absence of doxycycline, 293TLα cells express all relevant MMR proteins (Fig. 4.10, 293TLα+). The LoVo cell line is a MSH2-deficient colon cancer cell line that was used to test the MutSα requirement for repair of U/G mismatches on heteroduplex DNA substrates. MSH2 stabilizes MSH6 *in vivo*. Accordingly, the lack of MSH2 gives rise to a loss of MSH6 (Fig. 4.10).

293TLα+/- and LoVo cell extracts used in MMR assays were immunodepleted of TDG. The depletion did not affect the level of MMR proteins, as shown in Fig. 4.10.

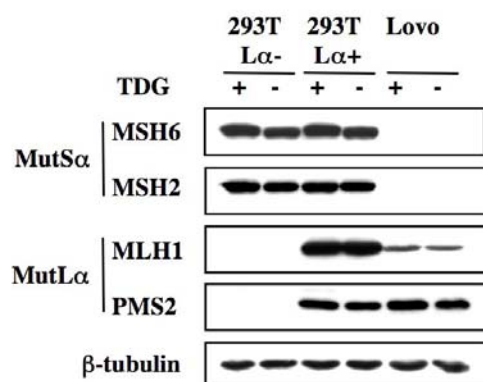**Fig. 4.10** Western blots depicting MMR proteins in 293TLα+/- and LoVo cell extracts before (+) and after (-) TDG depletion. Upon addition of doxycycline, 293TLα(-) cells lack MLH1, resulting in a complete loss of MutLα. In the absence of doxycycline, however, 293TLα(+) cells express MSH2/6 and MLH1/PMS2. In LoVo cells, MSH6 is degraded due to the absence of MSH2. β-tubulin indicates equal loading.

5   DISCUSSION

5.1   Mass spectrometry based search for AID interaction partners

Activation-induced cytidine deaminase (AID) has recently been the focus of intensive research efforts. Since the discovery of AID in 1999[17], numerous scientists have tried to shed light on the processes of antibody diversification triggered by this enzyme. However, many questions still remain to be solved, such as how AID is specifically targeted to immunoglobulin loci. Transcription seems to be a prerequisite for SHM and CSR[40, 59], yet, most transcribed genes are not mutated, while immunoglobulin genes are frequently mutated. To date, AID is the only B cell specific factor described to be necessary for antibody diversification. But how can a single enzyme trigger two processes that are as different as SHM and CSR? It is speculated that AID recruits help from different factors to initiate either SHM or CSR. Studies with AID deletion mutants, describing the N-terminus as being important for SHM and the C-terminus as crucial for CSR, support this hypothesis. Genetic evidence specifically implied the involvement of MMR proteins in both processes[21]. We were thus particularly interested in investigating the role of mismatch repair proteins in somatic hypermutation (SHM) and class switch recombination (CSR) and attempted to reveal a protein network possibly linking AID with MMR proteins.

5.1.1   The tandem affinity purification (TAP) approach

In order to identify proteins that are involved in SHM and CSR, we searched for interacting partners of AID, using the TAP dual-affinity strategy that has been shown to be invaluable in the characterization of protein complexes. This technique was first established in *S. cerevisae* [361] but has recently also been successfully applied in other organisms, including human cells. When TAP was compared with single-step purification strategies or immunoprecipitation experiments, it was shown to be significantly more specific, resulting in a dramatic reduction of false positive identifications[363]. In *S. cerevisae*, where large datasets are available, TAP was shown to exhibit an error-rate of only about 15%, while single-epitope purification methods

revealed an error-rate of up to 50%[385]. Another advantage is that TAP uses mild washing conditions, allowing for the recovery of native complexes. Taken together, these facts made TAP the method of choice as a first approach to search for interaction partners of AID.

Applying TAP, the level of the tagged protein is an important determinant for the success of the experiment. It is recommended to work with protein amounts that are comparable to the endogenous protein levels, in order to avoid artifacts induced by protein overexpression. We set out to establish a B cell line that stably expresses TAP-AID, since this allows selection of clones expressing the recombinant protein at levels comparable to wild type. In contrast, recombinant proteins tend to be overexpressed in transiently transfected cells, increasing the probability of identifying unspecific interactors binding to the overexpressed or misfolded protein. This not only falsifies the results, but also hampers the identification of low abundant binding partners. Another issue is the competition for binding partners between the endogenous protein and the TAP-tagged protein. To avoid this problem, we used an AID-deficient BL2 Burkitt lymphoma cell line. Faili and colleagues created these cells by targeted inactivation of the AID gene through homologous recombination in the parental BL2 cell line[66].

The TAP has a size of 21 kDa, roughly equal to the size of the full-length AID protein with 26 kDa. Thus, the tag-to-protein ratio was very high, possibly disturbing *in vivo* protein interactions. To minimize the risk of missing interacting proteins, we created two different expression constructs, with the TAP-tag located either at the N- or C-terminus of AID.

The stable transfection of the BL2 AID$^{-/-}$ cell line turned out to be the first obstacle. B cells in general are known to be difficult to transfect. We tried various transfection methods, including conventional calcium-phosphate- or lipid-based techniques and eventually succeeded with electroporation. However, the selected puromycin-resistant single cell clones did not show any AID expression when analyzed on western blots (data not shown). To see whether AID could be successfully expressed in non-B cells, where the transfection efficiency is higher, we transiently transfected 293T cells with the AID-N/C-TAP constructs. Indeed, AID was expressed in 293T cells. Subsequent RT-PCR analysis revealed that the AID gene was also integrated into our BL2 clones, transfected with TAP-tagged AID (Fig. 4.3). However, a comparison between 293T and BL2 cells showed that the level of AID transcript in the transiently-transfected

non-B cells was substantially higher than that in BL2 cells (Fig. 4.4). Notably, our AID-transfected BL2 cell clones (AID-N-TAP BL2, AID-C-TAP BL2) showed levels of AID transcripts comparable to the positive control, AID-expressing BL2 cells (Fig. 4.4). We thus concluded that this is the maximum level of AID transcript that can be reached after stably transfecting the gene into BL2 cells. Interestingly, BL2 transfectants that revealed higher levels of AID transcripts than the control cells died after a few days in culture. This suggests that strong AID expression leads to a high mutation rate, eventually resulting in cell death. Thus, cell clones that express AID at a high level are counter-selected.

We performed TAP with BL2 cells that were positive for the AID transcript in RT-PCR. Since each step of the procedure is inevitably accompanied by loss of protein, only relatively small quantities of proteins, albeit highly pure, can be recovered. Unfortunately, we were not able to isolate TAP-AID from BL2 cells and thus could not perform a MS-based search for novel interaction partners. We attributed this to a low overall expression level of the transgene. As our method of transfection favors integration of a single copy of our construct, expression of AID in BL2 cells might be insufficient for the two-step purification. In contrast, we were able to enrich and purify AID-TAP from transiently transfected 293T cells. During transient transfection, multiple copies of our construct are taken up by the cell, thus AID was overexpressed. The feasibility of tandem-affinity purification of recombinant proteins from 293T cells was also previously shown in studies in our lab and resulted in the identification of novel interacting partners of the MMR proteins MLH1 and PMS2[386]. Beside the difference in copy number, there are various other possibilities to explain the absence or low expression of AID in BL2 cells. Although integrated, as judged by RT-PCR, the transcription of the transgene might be down-regulated e.g. by promotor hypermethylation. Furthermore, translation can be inhibited or blocked, or, if successfully translated, the protein could precipitate and aggregate. Another likely explanation is, as described earlier, that AID *per se* is a mutator and thus the cells try to minimize expression for the sake of the genomic integrity and survival.

5.1.2   The GST-AID pull-down approach

After our initial attempt to express AID in mammalian cells and its purification *via* the two-step TAP approach, we decided to modify the underlying experimental set-up. In our first approach we encountered one major problem: the expression of AID in high quantity, which is a well-known issue in the AID-field[371]. AID and other nucleic acid deaminases have been notoriously difficult to purify in an active form and in quantities necessary for structural studies. Thus, despite extensive research since the discovery of AID in 1999, the AID crystal structure has not been solved yet. However, a hypothetical AID model was set up based on a structural alignment with the yeast cytidine deaminase CDD1 (Fig. 4.6b).

In an attempt to overexpress and purify AID from *E. coli* we established a purification protocol exploiting the GST-gene fusion system. The basic procedure was recently published by Goodman and colleagues[30, 32], and we modified it for our purposes. Glutathione-S-transferase (GST) is commonly used as an N-terminal fusion partner when expressing proteins in *E. coli*. The 220 amino acid GST-tag sequence has been reported to enhance the solubility of its fusion partners[387]. In contrast to TAP, purification *via* the GST-glutathione-sepharose system is a single-step procedure, thus minimizing the loss of protein during the purification.

We succeeded in expressing GST-AID in significant amounts in *E. coli* and purified it to high homogeneity (Fig. 4.8a). Together with GST-AID migrating at ~ 44 kDa, a second protein, migrating at ~ 60 kDa was repeatedly co-purified and eventually identified to *E. coli* Hsp60. AID has the tendency to precipitate and misfold upon overexpression. The chaperone Hsp60 may thus be inseparably bound to AID, preventing structural abnormalities and aggregation of the deaminase.

The recombinant GST-hAID from *E. coli* was incubated with extracts from the human B cell lymphoblastoid cell line, Ramos, another Burkitt's lymphoma cell line. In contrast to BL2 cells that need to be induced to hypermutate[347, 365], Ramos cells were reported to undergo SHM constitutively[388]. Furthermore, they were shown to be proficient in BER and MSH2/MSH6-mediated MMR[389] and express polymerase η. We thus considered Ramos cells ideal for our purposes to reveal a protein network possibly linking SHM and CSR to repair mechansims.

After incubation with Ramos cell extracts, GST-AID was pulled-down with its interation partners *via* glutathione sepharose columns. The protein eluate was subsequently analyzed by MS and the peptide data obtained on the LTQ FT[TM] mass spectrometer were used in a Mascot search. By this means, a high number of proteins in complex with GST-AID were identified (Table 4.1).

We first validated our approach by demonstrating that the bait protein AID was among the most abundant proteins in the analyzed complex. AID was detected by MS with a very high score and was clearly visible on a silver-stained SDS gel (Fig. 4.5a). In addition, DNA-PK, a known interaction-partner of AID[50] was pulled-down specifically with GST-AID (Table 4.1), further validating our pull-down approach as a valuable method for the search for new interaction partners of AID.

Our analysis revealed several potentially interesting AID-interacting proteins. Some are directly linked to DNA metabolism, such as PARP, DNA Topoisomerase I, replication factor C (RFC) and chromodomain helicase-DNA-binding protein 4 (Table 4.1). A high number of PARP-1 derived peptides was detected, as illustrated by a high Mascot score. PARP-1 and PARP-2 are implicated in the BER pathway[390, 391], suggesting an involvement in U/G mismatch processing during immunoglobulin diversification. DNA topoisomerase was also previously implicated in the processes of immunoglobulin diversification. It was observed that during transcription, which is required for SHM and CSR, DNA upstream of the transcribing RNA polymerase is negatively supercoiled. This transient change in DNA topology, involving DNA topoisomerase, may allow AID to access both DNA strands[392]. Overall, there are a number of proteins that, in our view, are of particular interest and will be the subject of future studies.

We focused our attention on nucleolin that was also specifically pulled-down with the GST-AID bait (Fig. 4.5b). In particular, the very high Mascot score associated with this protein induced us to investigate the possibility that this molecule may play a role in AID-mediated antibody diversification. This assumption was strengthened by recent studies, establishing a link between nucleolin and CSR and possibly SHM. Nucleolin is a component of the B cell-specific transcription factor LR1, that regulates transcription in activated B cells[367]. LR1 binds G4 DNA structures in Ig heavy chain switch region sequences[393] and its binding activity was increased in B cells transfected with a nucleolin cDNA expression construct. Homology between histone H1 and nucleolin suggests that nucleolin may alter DNA organization in

response to cell cycle controls; nucleolin may thus function to organize switch regions before, during, or after switch recombination.

In an attempt to investigate the nucleolin-AID interaction *in vivo*, we performed immunoprecipitation (IP) experiments in 293T cells, using either HA-AID and WT-nucleolin or HA-nucleolin and WT-AID expression constructs. Unfortunately, we were not able to confirm the existence of a direct interaction between the two proteins in neither of the IP approaches. This might be due to the fact that the initial identification was based on a GST-pulldown of AID expressed in bacteria. The use of different methods in different organisms might strongly influence protein recovery. In addition, the strength of the interaction and the location of the GST- or HA-tag at one extremity of the bait protein might have a significant influence on the binding of certain molecules. Furthermore, post-translational modifications impact on the binding to potential interaction partners. The mammalian protein AID, or its interactors, may have to be modified in order to bind to each other. The interaction with an unmodified, *E. coli* expressed AID may thus not take place.

Yet, since the MS data obtained from the initial GST-AID pull-down needs to be validated with independent techniques and organisms or cell lines, and we failed to do so, we do not consider the AID-nucleolin interaction biologically relevant in mammalian cells.

5.1.3   The AID-(phospho)-peptide pulldown approach

In order to avoid the problem associated with expressing the full-length AID protein, we chose a peptide pull-down strategy as a third approach to search for AID interacting proteins. In this method, chemically synthesized peptides are biotinylated and coupled to streptavidin-coated magnetic beads. Using phospho-specific peptides and their unphosphorylated counterparts, this method allows for differentiation between phospho-dependent or –independent interactions upon incubation of the bead-coupled peptides with the cell extracts of interest.

The feasibility of this approach was recently demonstrated by a study of Jackson and colleagues where they demonstrated that MDC1 directly binds phosphorylated but not unphosphorylated histone H2AX to regulate cellular responses to DNA double strand breaks[370]. They could clearly distinguish between proteins that bound to the

phosphorylated and those that bound to the unphosphorylated peptides, indicating phospho-dependent interactions. Based on their protocol, we established a phospho-peptide pull-down assay. Designing the AID-specific peptides, we chose sequence stretches that comprised either serine 38 (S38) or tyrosine 184 (Y184), sites that were described to be phosphorylated in activated B cells[51]. We were particularly interested in the S38 peptide, since this residue lies within a cAMP-dependent PKA consensus phosphorylation site, conserved in AID across all species able to perform CSR. Although unphosphorylated AID is enzymatically active, this posttranslational modification is considered to be important for the biological activity of the enzyme. Furthermore, the phosphorylation was observed to facilitate its association with RPA[54], which in turn significantly increases the efficiency of AID-catalyzed deamination. This is likely due to RPA-mediated stabilization of ssDNA, allowing AID access to its substrate. However, the role of AID phosphorylation is still a matter of debate and is widely considered to have regulatory rather than an activating function. We were thus interested in interaction partners that were pulled down with the phosphorylated- and the unphosphorylated AID-specific peptides.

We succeeded in identifying a number of proteins in the S38 peptide pull-down fraction (Fig. 4.7). Notably, DNA-PK was among the most abundant proteins, verifying a known interaction partner of AID[50]. In addition, this identification confirmed our results obtained in the GST-AID pull-down experiment, thus simultaneously validating both approaches as suitable tools in the search for AID interacting partners.

The pull-down of the DNA repair protein RAD50 attracted our attention. The MS identification of RAD50 was subsequently confirmed by western blotting. Also MRE11 and NBS1 were immunodetected in the protein mixture. In the cell, these proteins form the trimeric MRN complex. The fact that no NBS1- or MRE11-specific peptides were identified in the MS screen was striking, since the same eluate was analyzed by MS and western blotting. However, in electrospray ionization mass spectrometry (ESI-MS), which we used for the analysis, there is a poor correlation between the amounts of the peptide present and the measured signal intensity. The peptides produce specific signal intensities depending on their chemical composition, on the matrix in which they are present and on a number of other variables like unsteady ionization, variable ion transmission and ion suppression. Accordingly,

MRE11 and NBS1 might well be present in the mixture, but their peptide ion signals might be below the threshold we set for significant hits.

The MRN complex has previously been linked to immunoglobulin diversification and thus we opted to follow up on the suggested AID-RAD50 interaction. MRN is a ubiquitous and conserved nuclease complex critical for DNA break repair. The two components MRE11 and RAD50 form the core of the MRN complex, with DNA binding activity vested in RAD50 and both endo- and exonucleolytic activities in MRE11[394-396]. The MRN complex was reported to be essential for CSR[89]. Furthermore, Maizels and colleagues showed that overexpression of NBS1, the regulatory subunit of the MRN complex, accelerated hypermutation in the human B cell line Ramos and increased the gene conversion efficiency in the chicken B cell line DT40. In both cases, accelerated diversification depended on MRN complex formation, suggesting that MRN promotes mutagenic repair of lesions initiated by AID. In addition, MRN was observed to undergo AID-dependent localization to nuclear foci at the Ig heavy chain locus in B cells activated for switch recombination[67]. In addition, MRE11/RAD50 was reported to cleave DNA at abasic sites, the predicted products of successive processing by AID and UNG, and MRE11 was shown to associate with the rearranged Ig heavy chain variable region (V_H) in Ramos cells[88].

In order to confirm the AID-MRN interaction *in vivo*, we used human 293T cells. However, our attempts to immunoprecipitate tagged- or untagged AID in complex with either tagged- or untagged NBS1 or RAD50 and MRE11 failed. This might have several reasons. As for the GST-AID pull-down, one possibility is the influence of protein tags on the binding of the bait protein to its potential interaction partners. Epitopes needed for specific protein interactions could be masked by the tag and thus binding could be sterically hindered. Furthermore, the peptide-pull-down was performed with whole cell extracts from Ramos B cells. The IP experiments, however, were carried out using human epithelial kidney cells (293T). Different cell lines in combination with different pull-down approaches might explain different protein recovery. Proteins expressed in different cell lines might be differently modified and thus the interaction pattern might differ significantly when comparing different cell lines.

We chose the AID-specific peptides based on the findings that they comprise potential phosphorylation sites. Post-translational modifications such as

phosphorylation generally occur on the surface of a protein. However, we cannot exclude that AID overexpressed in 293T cells is folded in such a way that the peptide sequence used for the pull-down is hidden in the core of the protein and thus is inaccessible for potential interactors. In contrast, AID-specific bead-coupled peptides are fully unfolded and exposed to interact with any protein present in the cell extracts.

A number of additional proteins, the peptides of which were identified in our samples, are in our view of particular interest and thus became subject of our studies. Among these, a complex of two proteins related to the bacterial ATP-dependent helicase RuvB, RuvB-like1 (TIP49a) and RuvB-like2 (TIP49b), was mainly identified in association with the unphosphorylated and, to a lesser extent, the phosphorylated S38 peptide (Fig. 4.7). RuvB-like1 and RuvB-like2 are highly conserved in evolution and are essential in yeast. Although the precise role of these ATPase-helicases is not known, they were implicated in several cellular processes. They were reported to modulate apoptosis[397], oncogenic transformation[398, 399] and were shown to be associated with transcription factors[400, 401], chromatin remodeling complexes in yeast[402, 403] and with the histone acteyltransferase TIP60 in human cells[404, 405]. Besides, an interaction with RuvB-like helicases and RNA PolII was reported[373], emphasizing their role in transcription and demonstrating a possible link to AID-mediated processes. [399]

We were able to confirm the interaction between AID and RuvB-like helicases, as suggested in the AID-(phospho)-peptide pull-down *in vivo*, by successfully immunoprecipitating AID in complex with either RuvB-like1 or RuvB-like2 from human 293T cells. We also performed the reciprocal reaction and succeeded in pulling-down RuvB-like1 in complex with transiently expressed AID. Based on these observations, further characterization of the RuvB-like1/2 proteins by various cell-based approaches eventually revealed a novel function of these helicases in cell division (Castor *et al.*, manuscricpt in preparation). Finally, in order to biochemically confirm the AID-RuvB-like1 interaction, we carried out enzyme-linked immunosorbent assay (ELISA)-based protein interaction experiments (Fig. 4.8). ELISA allows for detection of specific binding of an immobilized bait protein with increasing amounts of a prey protein. Incubation with an antibody against the prey protein and subsequent addition of a secondary, enzyme-linked antibody eventually leads to emission of a visible signal, indicating specific binding. Interaction between

the bait and prey proteins was only observed when RuvB-like1 was used as the prey and GST-AID as the bait, but not for control baits. Thus, this experiment confirmed specific binding of RuvB-like1 to GST-AID. To map the exact interaction domain, we performed further ELISA studies using full length-GST-AID and various RuvB-like1 deletion mutants. We hereby took advantage of the structural information from the recently solved crystal-structure of RuvB-like1[374]. Domain I, comprising amino acids (aa) 1-120 and 295-366 of RuvB-like1, and, to a lesser extent, domain II (121-295 aa) seem to mediate the interaction with AID, whereas domain III (366-456 aa) seems to be dispensable.

In our peptide pull-down screen we used four different peptides. However, the AID-derived peptide comprising Y184 pulled down very few interaction partners, both in the phosphorylated and in the non-phosphorylated form. We concluded that either the synthesis of the peptides was not carried out correctly, or the chosen peptide sequence does not comprise a protein interaction domain. In contrast, the peptide containing the S38 residue efficiently pulled-down a number of potential AID-interacting proteins. This residue lies within a cAMP-dependent PKA consensus phosphorylation site conserved in AID across all species, thus partners interacting with this site might indeed play a role in SHM or CSR. Strikingly, we pulled-down more proteins with the unphosphorylated than with the phosphorylated S38 (S38P) peptide (Fig. 4.7). However, the correct identity of the S38P remains to be confirmed, since it was very difficult to dissolve prior to use in the pull-down experiment. We expected phosphorylation of the S38 residue to have a regulatory function, in that it recruited additional interaction partners that help AID to perform SHM and CSR. However, phosphorylation might also regulate these processes by blocking binding of inhibitory molecules, thus promoting the enzymatic activity of AID through the release of these inhibitory factors. In this sense, MRN and RuvB-like1/2, the proteins we identified as potential interaction partners of AID, might indeed play an important role in the processes of SHM and CSR. The biological function of these interactions will thus be addressed in future studies in our lab.

5.2 Mechanistic insights into the interplay between MMR and BER in immunoglobulin diversification

Three metabolic processes mediate antibody diversification in vertebrates: V(D)J recombination, somatic hypermutation (SHM) and class switch recombination (CSR). The latter two are triggered by the enzyme activation-induced cytidine deaminase (AID) that deaminates cytosines at certain hot spots in the immunoglobulin locus to uracils. While uracil repair is generally error-free, this repair appears to be error-prone in the context of immunoglobulin diversification. Accordingly, mutations are introduced rather than repaired, leading to a highly diverse repertoire of antigen-specific antibodies. Despite extensive research in this field, this mechanism of error-prone repair pathway remains to be elucidated. However, many recent studies provided data for an implication of MMR and BER proteins in these processes. Genetic studies with knock-out mice revealed the involvement of MutSα, UDG, EXO1 and many more DNA repair proteins[21, 91, 327, 328].

In the first part of my studies we were searching for new interaction partners of AID, in order to elucidate or identify a protein network that would reveal a link between immunoglobulin (Ig) diversification and the DNA repair pathways MMR and BER. In the second part, we investigated the role of MMR and BER in the processing of U/G mispairs at the Ig locus, exploiting an *in vitro* MMR assay system. This system was previously established in our lab[383, 384] and allows the determination of the mismatch repair efficiency of nuclear extracts. Using heteroduplex DNA substrates with specific mismatches within an *Acl*I restriction site, repair can be monitored *via* the restoration of the restriction site. In case of faithful repair, the *Acl*I restriction site is restored and the plasmid will be cut into three fragments. However, in MMR deficient cell extracts, the mismatch will persist and the plasmid will be subsequently cut into only two fragments (see manuscript 4.2.1, Fig. 1a).

5.2.1 BER- and MMR-catalyzed repair of U/G and G/T mispairs in B cell extracts

SHM and CSR take place in germinal center B cells in secondary lymphoid organs. In order to mimic this physiological situation, we investigated the interplay between MMR and BER in U/G processing at the Ig locus in extracts of the Burkitt's lymphoma cell line BL2. These cells represent the transformed counterpart of germinal center cells and express all relevant MMR proteins (Fig. 4.2; Fig. 4.9). We showed in preliminary experiments that BL2 cells are MMR proficient (data not shown).

Recruitment of AID to its sites of action is thought to be dependent on transcription[40, 59]. A single deamination event in a transcription bubble generates a uracil in one of the strands and gives rise to a U/G mismatch once the transcription bubble has moved on. Under normal circumstances, the uracil is addressed by the BER pathway. Uracil DNA glycosylase removes the uracil, creating an AP-site[21]. Cleavage of the sugar-phosphate backbone by an AP-endonuclease provides a nick, serving as the entry point for polymerase β that inserts a dCMP residue.

To see whether this canonical BER pathway also functions in our B cell extracts we performed an *in vitro* MMR assay with DNA heteroduplex substrates carrying a single U/G or a G/T mismatch within the *Acl*I restriction site. The mispairs render the DNA refractory to cleavage. Faithful repair restores the cleavage site and the plasmid will be cut in three fragments instead of two fragments (see manuscript 4.2.1, Fig. 1a). The circular closed U/G substrate, but not the G/T substrate, was successfully repaired in the BL2 nuclear cell extracts, and when the glycosylases TDG and UDG were inhibited (by immunodepletion and by the uracil glycosylase inhibitor Ugi, respectively), also repair of the U/G substrate was blocked (see manuscript 4.2.1, Fig. 1b). This indicated that BER was responsible for the repair. The closed circular G/T substrate was repaired neither by BER, nor by MMR, since a nick or a gap, essential prerequisites for entry of MMR proteins, were missing. Prenicked U/G and G/T substrates, however, were repaired irrespectively of the presence of UDG. Here, MutSα recognizes the mismatch, EXO1 is loaded at the preexisting nick and the error-containing strand is subsequently degraded (see manuscript 4.2.1, Fig. 1b).

With these experiments we showed that a single U/G mispair on a closed circular DNA substrate is not processed by MMR, even though it is efficiently recognized by the mismatch recognition factor MutSα (MSH2/MSH6). In the context of SHM and CSR, multiple U/G mismatches are likely to occur due to the processivity of AID[32]. We argued that a MutSα sliding clamp loaded at one U/G might slide along the DNA contour in search of a nick and thereby interfere with uracil processing by BER at distal U/G sites.

5.2.2 Use of intermediates of uracil processing as MMR initiation sites

We asked how MMR proteins could be involved in U/G mismatch processing in the absence of nicks or other potential entry sites for EXO1. To address this point, we first tested a substrate that carried a U/G mispair situated 54 base pairs from a G/T, rather than a second U/G mismatch, in the *Acl*I restriction site. G/T is refractory to BER, therefore U/G is the only mispair addressed by BER, while both mispairs are recognized by MutSα. Notably, the G/T mismatch in our supercoiled U/G-G/T substrate was corrected with appreciable efficiency when incubated with the BL2 cell extracts (see manuscript 4.2.1, Fig. 2a). The repair was blocked when Ugi was added, indicating that G/T correction was fully dependent on the activity of UDG. Since no nick was present in the substrate, we concluded that BER intermediates were accessible to other enzymes. A likely explanation is that upon excision of uracil from DNA and cleavage by APE, the nick is temporarily accessible due to low levels of Pol β in BL2 cells[381]. During this window of accessibility, MutSα loaded at the downstream G/T mismatch triggers EXO1 loading at the cleaved AP-site and initiates a strand replacement reaction that results in the conversion of the G/T mismatch to an A/T base pair.

We hereby demonstrated that a uracil residue in a U/G mispair can act as a cryptic strand discontinuity and thus provide an entry point for MMR. Using U/A-G/T substrates, we furthermore showed that the same scenario might apply when a uracil is incorporated opposite an adenine (see manuscript 4.2.1, Fig. 2b). In order to demonstrate the relevance of this finding to SHM, we tested a DNA heteroduplex carrying two nearby U/G mispairs, mimicking the processive action of AID in the

immunoglobulin locus. In this setting, processing of the uracils by BER, or of the U/G mispairs by MMR, could theoretically begin at any of the deaminated sites. Due to the short repair patch generated during BER, no interference between the processing of sites as close as ~10 nucleotides would be expected. In contrast, if both BER and MMR were involved, the outcome of the repair would depend on where the first strand break was introduced.

Incubation of the U/G-U/G substrate with B cell extracts indeed led to an efficient repair of the U/G within the *Acl*I site to C/G (see manuscript 4.2.1, Fig. 2c), indicating that the repair reaction including the strand displacement was initiated at the upstream uracil. The repair could be specifically inhibited by the addition of Ugi, again indicating that the strand displacement reaction is fully dependent on UDG activity. The observation that the repair of a U/G-U/G substrate that contained also a nick 350 nucleotides upstream was more efficient and was only slightly inhibited upon addition of Ugi, showed that U/G processing in B cell extracts is mainly mediated by the MMR pathway.

5.2.3 Dependence of the uracil-initiated MMR reaction on MutSα and MutLα

Mismatch repair is a bidirectional process, where the strand degradation can take place in both 5'-3' and 3'-5' directions[187, 378]. If the nick, serving as a strand discriminating signal, lies 5' to the mismatch, ATP-activated MutSα loads EXO1, which then catalyzes the strand degradation in a 5'-3' direction, according to its polarity. In this case, MutLα, the MLH1/PMS2 heterodimer, stimulates the reaction, but the repair process is not dependent on this factor. In contrast, 3'-5' repair fully depends on the presence of MutLα that is recruited by MutSα. MutLα induces several nicks into the error-containing strand, both upstream and downstream of the mismatch. EXO1 then initiates excision at these nicks in a 5'-3' direction. Thus, a 3' substrate is converted into several 5' substrates, which explains the fact that no exonuclease with a 3'-5'-polarity is needed, and that EXO1 with its 5'-3'-excision activity is sufficient also for MMR in the 3'-5'-direction. To summarize, MMR necessitates MutSα for repair in both 5'-3'- and 3'-5'-direction, whereas MutLα is only required for 3'-5'-repair but is dispensable for 5'-3'-repair.

To address the question whether this factor requirement also applies to uracil-directed MMR, we constructed various DNA heteroduplex substrates that carried a U/G mispair either 5' or 3' from a G/T mismatch within the *Acl*I restriction site. To determine the role of MutSα and MutLα, we took advantage of cell lines deficient for these heterodimers. 293T Lα cells were previously established in our lab[382] and allow the doxcycline-regulated expression of MLH1. In the absence of doxcycline, these cells express MLH1 and PMS2 and are referred to as 293T Lα+ cells. Upon addition of doxcycyline, however, MLH1 expression is suppressed, PMS2 is subsequently degraded and the cells named 293T Lα- (Fig. 4.10). The colon cancer cell line LoVo is deficient for MSH2, inevitably leading to degradation of its partner MSH6 and thus to a loss of the complete MutSα heterodimer (Fig. 4.10).

When we incubated our DNA substrates with 293T Lα- cell extracts, we observed that the U/G-G/T substrate was still repaired with appreciable efficiency, while the G/T-U/G substrate was not repaired at all. This implied that MLH1 is indeed dispensable for uracil-directed 5'-repair, whereas its presence is crucial for uracil-directed 3'-repair. This finding is in line with what was published for nick-directed MMR, as described earlier[221, 275]. Accordingly, we found a consistency with regard to the role for MutSα. Both uracil-directed 5'- and 3'-repair are absolutely dependent on MutSα, as indicated by the inhibition of repair of both substrates, U/G-G/T and G/T-U/G, in the absence of MSH2 in LoVo extracts (see manuscript 4.2.1, Fig. 3). The repair could be restored completely upon the addition of recombinant MutSα, further emphasizing the crucial role of this factor in the repair reaction.

Remarkably, these observations also correlate with results obtained in genetic studies that revealed a major role for the *MSH2* and *MSH6*, but not for *MLH1* and *PMS2* loci in SHM[21]. The predominance of MutSα over MutLα further implies that the MutSα-dependent 5'-3'-process is the major pathway followed during this phase of diversification of immunoglobulin genes.

5.2.4 Putative mechanism of SHM and MMR implication on CSR

Based on the observations we made in the experiments presented in this work and taking into account the large amount of genetic evidence obtained in other laboratories, we propose a model of how MMR and BER might interfere during the processing of multiple U/G mispairs in SHM (see manuscript 4.2.1, Fig. 4).

AID-mediated deamination of cytosines in single-stranded DNA of a transcription bubble gives rise to a U/G mismatch, once the transcription bubble has moved on. Due to the absence of a nick, U/G cannot be processed by MMR, but instead is addressed by BER. If BER were to be interrupted immediately after removal of the uracil by UNG2, but before the sugar-phosphate backbone is cleaved, the AP-site might persist until S-phase, where its by-pass by translesion synthesis polymerases such as REV1 might give rise to class I mutations at C/G sites. If BER were not to be interrupted, the AP-site would be cleaved and the gap filled by insertion of dCMP by polymerase β. In this scenario, the initial U/G mismatch is faithfully repaired and no mutation arises. The situation becomes more complex when a second cytosine deamination occurs in the moving transcription bubble. Should a partially-processed deamination site lie within ~1kb (maximum distance between a mismatch and a nick), the MutSα sliding clamp activated by the second U/G may interrupt the BER process at the upstream uracil. It might encounter an AP-site cleaved either by APE1 or MRE11[88] and use it as an entry point for Exo1 loading. The exonuclease then starts to degrade the strand in 5'-3'-direction, resulting in a stretch of single-stranded DNA, spanning the distance between the first deamination site until ~ 150 nucleotides past the second U/G mismatch. Under normal circumstances, the high-fidelity polymerase Pol δ would fill in the gap, thus faithfully repairing the lesion. However, in the case of SHM, mutations are fixed rather than repaired. This might be explained by the recruitment of translesion synthesis polymerases, such as Pol η. Ubiquitylation of PCNA was shown to be crucial for SHM[358] and ubiquitylated PCNA was described to physically interact with Pol η[360]. Hence, a likely scenario would be the recruitment of Pol η to the Ig locus by ubiquitinated PCNA. Thus, it is possible that a subset of MMR repair patches is filled in by the error-prone Pol η that is known to preferentially introduce non-complementary nucleotides opposite Ts[93], which would eventually lead to the observed class II mutations at A/T base pairs.

It would be interesting to monitor the expression pattern of high- and low-fidelity polymerases in activated- versus non-activated B cells. In order to do so, ongoing studies in the lab aim to compare the expression of Pol δ, Pol β and Pol η in primary B cells isolated from mouse spleen, either stimulated with lipopolysaccharide (LPS) or unstimulated. LPS is generally used to mimic the activation of B cells that, *in vivo*, occurs in germinal centers by interaction with antigen and T helper cells.

We speculate that the expression of Pol η activity might be upregulated following B cell stimulation; Pol η might thus compete with high-fidelity polymerases for resynthesis of the EXO1 degraded strand. This could be supported by a possible down-regulation of Pol δ or post-translational modifications, leading to a decreased activity of Pol δ. Besides, cellular relocation of the polymerases of interest might influence the outcome of the repair. Thus, Pol δ could be partially excluded from entry into the nucleus, whereas Pol η might be specifically recruited. However, one has to take into account that, in most cases, repair will still be carried out faithfully and only in a small subset of cases will the mutations be fixed rather than repaired. As for Pol β, we imagine a down-regulation in activated primary B cells, possibly in combination with an increased expression or activity of UDG. Pol β was previously reported to be down-regulated in the B cell line BL2[381], thus a down-regulation in primary B cells may also seem likely. Hence, canonical BER might be overwhelmed in activated B cells, leaving BER intermediates that might be alternatively processed and eventually give rise to class I and class II mutations in SHM.

Our data also imply that MMR and BER interference in the processing of uracil residues on opposite strands could give rise to double strand breaks (DSB), which are postulated to trigger CSR. This nicely fits a model for MMR protein involvement in CSR that was recently postulated by Stavnezer and colleagues[113, 114]: If single strand breaks (SSB) are near each other on opposite strands, DSBs can form spontaneously, due to thermal lability. However, if SSBs are distal from each other, as is most likely the case within large switch regions, conversion to DSBs is necessary. A likely scenario is the following: MutSα recognizes and binds to AID-mediated U/G mismatches, MutLα is recruited and EXO1 excises from the nearest 5' SSB created by AID-UDG-APE activity toward the mispaired U/G. EXO1 is then thought to continue the excision past the mismatch until a SSB is reached on the opposing strand, thus forming a DSB, which would trigger CSR[105].

5.3   Conclusions

The goal of this thesis was to shed light on the involvement of mismatch repair proteins in the processes of antibody diversification, including somatic hypermutation (SHM) and class switch recombination (CSR).

Both processes are triggered by a single enzyme, activation-induced cytidine deaminase (AID)[17]. However, it is assumed that additional proteins are necessary to correctly perform two mechanisms as different as SHM and CSR and recent studies suggested an involvement of MMR and BER proteins[21, 111]. In an attempt to gain new insights into these processes, we applied various protein purification and interaction approaches to search for binding partners of AID. In doing so, we encountered several problems and pitfalls. AID is a mutator and therefore toxic to the cells when expressed at high levels. We therefore struggled with obtaining AID protein and its binding partners in quantities necessary for reliable MS based analysis. However, while we could not observe any MMR proteins directly interacting with AID, we were eventually able to identify several other proteins that demonstrate potential interaction partners of AID. Among these, the helicases RuvB-like1/2 were identified and confirmed to physically interact with AID, using various cell lines and interaction approaches. The identification of DNA-PK as a known interaction partner of AID[50] in the GST-AID pull-down, as well as in the phospho-peptide pull-down, validated these approaches as suitable tools to search for new interaction partners of AID.

In order to investigate the interference of MMR and BER in the processing of AID-generated U/G mispairs, we exploited an *in vitro* MMR assay system. Using heteroduplex DNA substrates containing two U/G mismatches, such as might arise through the processive deamination of cytosines by AID, we could show that uracil residues can act as initiation sites for MMR-dependent repair of neighboring mispairs. We furthermore showed that this uracil-directed repair is dependent on the MSH2/MSH6 heterodimer, MutSα, while the MLH1/PMS2 heterodimer, MutLα, seems to be of minor importance. These findings are in line with observations made with regard to nick-directed MMR[221, 275] and help explain results obtained in genetic studies that revealed a dominant role for MutSα in SHM[21].

6 MATERIALS AND METHODS

6.1 Cell lines

The Burkitt's lymphoma cell line BL2 was kindly provided by Jean-Claude Weill. Ramos B cells were a kind gift of Silvio Hemmi. Both suspension B-cell lines were cultured in RPMI 1640 medium (GIBCO), supplemented with 10% fetal bovine serum (FBS, GIBCO), 100 U/ml penicillin and 100 μg/ml streptomycin. HeLa and LoVo cell lines used in this study were obtained from the cell line repository of the Cancer Network Zurich. HEK293T cells were derived from the parental cell line HEK293 by transfection with SV40 large T antigen and immortalization with adenovirus 5 DNA[406]. Promoter hypermethylation leads to the epigenetic silencing of the *hMLH1* gene in these cells[407]. The 293T Lα cell line was generated in our laboratory[382]. In this cell line, the hMLH1 c-DNA was stably integrated under the control of the tetracycline response promoter using the Tet-Off system (Clontech, Palo Alto, CA). Under normal culture conditions, the cells express MLH1 and are MMR-proficient. Upon addition of doxycycline the MLH1 expression is shut off and the cells become MMR-deficient.

All aforementioned non-B cell lines were cultivated in Dulbecco's modified Eagle's medium (DMEM, Gibco/Invitrogen, San Diego, CA) with supplements as above. All cell lines were kept in culture at 37°C in a 5% CO_2-humified atmosphere.

6.2 SDS-PAGE analysis and western blotting

Cells were harvested and washed twice in phosphate buffered saline (PBS, GIBCO) before the pellet was dissolved in an appropriate volume of lysis buffer (50 mM Tris-HCl pH 8.0, 125 mM NaCl, 1% NP40, 2 mM EDTA, 1 mM phenylmethylsulfonyl fluoride (PMSF), complete protease inhibitor, 500 μM NaVan, 20 mM NaF). Cell lysis was performed during 30-60 min on ice and the extracts clarified at 12'000 x g, 2°C, 3min. The protein-containing supernatant was then aliquoted, snap-frozen and stored at -80°C.

Generally, 50 μg of total or nuclear protein were denatured, reduced and separated on 7.5 - 12% SDS gels. In preparation for western blot analysis, the proteins were transferred to Hybond-P PVDF membranes (Amersham Pharmacia Biotech, Little Chalfont, UK) according to standard protocols. The membranes were blocked by 30 min incubation in 5% non-fat dry milk in TBS-T (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1 % Tween-20). Primary antibodies were then added for 3 h at room temperature or overnight, the membranes washed three times in TBS-T and incubated with a horseradish peroxidase (HRP)-coupled secondary antibody. Following another three wash steps, immunoreactive proteins were detected using enhanced chemiluminescence (ECL, Amersham Pharmacia Biotech, Little Chalfont, UK).

6.3　Tandem affinity purification

BL2 cells stably transfected with plasmids expressing the N-terminally or C-terminally TAP-tagged human AID were grown in T175 flasks. 293T cells transiently transfected with the same expression vectors, were plated in 15 cm dishes. Both cell lines were cultured to 80% confluency, washed twice in cold PBS and lysed on ice in 50 mM Tris-HCl pH 8.0, 125 mM NaCl, 1% NP40, 2 mM EDTA, 1 mM PMSF, 1x complete inhibitory cocktail (Roche Molecular Biology), 0.5 mM sodium orthovanadate, 20 mM sodium fluoride and 5 nM okadaic acid. The lysates were cleared by centrifugation at 12'000 x g for 3 min and the soluble material was collected. Protein concentration was determined using the Bradford assay (BIORAD). Batch tandem affinity purification was performed according to the original protocol[408] with minor changes. All purification steps were performed at 4°C or on ice. For each experiment, 60 mg of whole cell extract was incubated for 4 h with gentle agitation with 100 μl of IgG Sepharose beads (Amersham Biosciences), equilibrated with lysis buffer. Beads were then washed 3x with 1 ml lysis buffer and 3x with 1 ml TEV buffer (10 mM Hepes-KOH pH 8.0, 150 mM NaCl, 0.1% NP40, 0.5 mM EDTA, 1 mM DTT, 1 mM PMSF and 1x complete inhibitory cocktail). Bound TAP-tagged proteins were released by overnight incubation in TEV buffer containing 16 U of acTEV protease (Invitrogen) in tubes mounted on a rotating platform. The supernatant from the TEV reaction was collected and transferred to a new tube. 1 volume of

calmodulin binding buffer (CBB: 10 mM β-mercaptoethanol, 10 mM Hepes-KOH pH 8.0, 150 mM NaCl, 1 mM MgOAc, 1 mM imidazole, 0.1% NP40, 2 mM $CaCl_2$, 1 mM PMSF and 1x complete inhibitory cocktail) was added to the collected supernatant and centrifuged at 1'500 rpm for 3 min. The supernatant was then transferred to a new tube and the procedure described above was repeated two more times. 1/250 volume of 1 M $CaCl_2$ was then added and the supernatant was batch-purified by binding to 100 μl of calmodulin affinity resin (Stratagene) equilibrated in CBB, for 4 h on a rotating platform. Beads were washed 3x with 1.2 ml of CBB and 2x with 1.2 ml of calmodulin rinsing buffer (CRB: 50 mM ammonium bicarbonate pH 8.0, 75 mM NaCl, 1 mM MgOAc, 1 mM imidazole and 2 mM $CaCl_2$) and eluted with 100 μl of calmodulin elution buffer (CEB: 50 mM ammonium bicarbonate pH 8.0 and 35 mM EGTA). One third of the eluate was separated by SDS-PAGE and visualized by silver staining. As negative control, the purification was performed with extracts prepared from parental cells not expressing the tagged protein.

6.4   Quantitative RT-PCR

Total RNA was isolated from 5 x 10^6 AID-TAP expressing BL2, 293T or control cells, using an affinity resin column (RNeasy, Qiagen). RNA was then converted to cDNA using a cDNA sythesis kit (Invitrogen). RT-PCR was performed with the Roche LightCycler System using the Light Cycle-RNA Master Sybr Green I Kit (Roche) according to the manufacturer's instructions, 0.3 μM of each oligonucleotide primer (AID or GAPDH, Microsynth) and 300 ng of total RNA in 20 μl reaction volume. The generation of PCR products can be detected by measurement of green fluorescence. SYBR green (Roche), a dye that is added to the RT-PCR reaction mix, will emit a fluorescence signal only when bound to double-stranded DNA. Thus, during PCR, the increase of measured fluorescence is directly proportional to the amount of double-strand DNA generated. All of the experiments were performed in duplicate, and the specificity of each amplification product was verified by agarose gel electrophoresis.

6.5 GST-AID expression and pull-down studies

To perform GST-AID pull-down experiments from bacterial cell extracts, the *AID* gene was cloned into the inducible pGEX-2TK expression vector (Amersham Pharmacia Biotech, Little Chalfont, UK) *via* the *BamH*I and *EcoR*I restriction sites. This confers an N-terminal glutathione-S-transferase (GST) tag to the protein of interest. GST-fusion proteins can then be purified from the expression culture through interaction with glutathione Sepharose (GSH).

The pGEX-AID construct was transformed into *E.coli* BL21 bacteria and grown on LB-Amp plates. A fresh colony was used to inoculate a starting culture of typically 200 ml LB medium containing 100 μg/ml ampicilin. After growing the bacteria for 8 h at 30°C, protein expression was induced through addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, SIGMA). The culture was incubated overnight (12 h) at 16°C and then spun down (4'000 rpm, 10 min, 4°C). The resulting pellet was resuspended in 10 ml lysis buffer (PBS, 1 mM PMSF, 1x complete protease inhibitor) and the cells lysed by sonification (Bandelin, Sonoplus; 5 min, 100 % amplitude, 50 % pulse). Cell debris were pelleted for 5 min at 10'000 x g in an ultra-centrifuge (SORVALL ULTRA, Pro80) and the protein-containing supernatant used for binding to 500 μl of GSH-Sepharose beads, pre-equilibrated in lysis buffer. Binding was carried out using chromatography columns (BIORAD) on a roller at 4°C over night. The beads were washed three times with 10 ml lysis buffer and the GST-AID protein was eluted by incubation with elution buffer (20 mM glutathione, 50 mM Tris-HCl pH 8.0) for 1 h at 4°C. Proteins were eluted drop-wise and stored in aliquots at -80°C.

If GST-AID was used for subsequent pull-down studies with human cell extracts, the protein was not eluted, but left bound on the beads. In this case, 1 mg of the respective total cell extract (Ramos or 293T cells) was added to the GSH-beads and further incubated for 2 h on the rolling wheel at 4°C. After this secondary incubation step, the beads where washed three times as above to get rid of unspecifically-bound protein. GST-AID plus specifically-interacting proteins were eluted through a 5 min boiling procedure at 95°C in 1x SDS-loading buffer. The protein samples were then separated and analyzed on 7.5 - 12% SDS gels.

6.6   AID-(phospho)peptide pull-down

To perform peptide pull-down studies, four 18 amino acid containing peptides derived from the AID wild type sequence were designed. The peptides comprised known phosphorylation sites on AID, namely serine 38 and tyrosine 184. For both sites a phosphorylated peptide and its unphosphorylated derivative were created (pS38: biotin-VKRRDpS(38)ATSFSLDFGYLR, S38: biotin-VKRRDS(38) ATSFSLDFGY LR, pY184: biotin-LRRILLPLpY(184)EVDDLRDAF, Y184: biotin-LRRILLPLY (184)EVDDLRDAF) and obtained from Franz Fischer (Friedrich-Miescher-Institut, Basel, Switzerland).

In an initial step, the biotinylated peptides were coupled to streptavidin-coated magnetic beads (Dynabeads M-280, Dynal Biotech, Invitrogen). In order to do so, the beads were washed thrice in TBS, 0.1% Tween and 100 μl peptide (100 mM in PBS, 10% acetic acid) were added. The mixture was incubated and rotated for 30 min at 4°C. Following another three washing steps as above, 1 mg of the whole cell extracts (Ramos, 293T, made up to 1 ml with TBS-T) was incubated with the coated beads for 1 h on a rotating wheel at 4°C. Beads were washed extensively with TBS, 0.1% Tween and bound proteins eluted with SDS sample buffer and subjected to SDS-PAGE. The protein composition was investigated using silver-stained gels, mass spectrometry analysis and results were confirmed by immunoblotting.

6.7   Enzyme-linked immunosorbent assay (ELISA)

Enzyme-linked immunosorbent assays (ELISA) can be used to study protein interactions. This approach necessitates highly pure protein. To investigate binding between AID and RuvB-like1, we thus expressed recombinant GST-AID and His-tagged RuvB-like1 protein in *E. coli* and purified them by glutathione Sepharose- and Ni-NTA-columns, respectively. The ELISA plate was coated with constant amounts of GST-AID (250 ng) diluted in carbonate buffer (0.016 M Na_2CO_3, 0.034 M $NaHCO_3$, pH 9.6) and incubated overnight at 4°. As controls, the plate was additionally coated with equal amounts of GST-only, BSA and with carbonate buffer only. The ELISA plate was then washed in washing buffer (PBS + 0.1% Tween

(PBS-T) and the plate surface was blocked with blocking buffer (PBS-T + 3% BSA) for 2 h. After three additional washing steps in PBS-T, the samples were incubated with increasing amounts of His-RuvB-like1 protein (50, 100, 250, 500 ng) for 2 h at RT. The plate was then washed and again blocked for 2 h in blocking buffer. Subsequent incubation of the samples with an anti-RuvB-like1 antibody for 2 h was followed by washing and further blocking of the samples in washing and blocking buffer, respectively. The samples were incubated on a shaker for 1 h with a secondary, HRP-conjugated antibody at RT, followed by washing of the plate. Addition of a specific substrate (o-phenylenediamine dihydrochloride, SIGMA) in citrate buffer (17.25 mM citrate, 37.5 mM Na_2HPO_4, 0.03% H_2O_2) and incubation for 5-30 min produces a soluble orange-brown colored end product, if a reaction with the HRP conjugates occurs. This indicates specific binding of the prey to the bait and the color intensity can be measured spectrophotometrically at 450 nm. If no binding occurs, no signal will be emitted.

6.8   Immunoprecipitation

Whole cell extracts (Ramos, 293T cells, 1 mg) were incubated in a total volume of 500 μl NP40 lysis buffer (50 mM Tris-HCl, pH 8.0, 125 mM NaCl, 1% NP40, 2 mM EDTA, 1 mM PMSF, 1x complete protease inhibitory cocktail (Roche Molecular Biochemicals, Basel, Switzerland)) with the respective antibody (anti-HA, mouse monoclonal, SIGMA; anti-AID, mouse monoclonal, Cell Signaling; anti-RuvB-like1, goat polyclonal, Santa Cruz; 6 μg each were used for the IP) for 3 h at 4°C. The immunoprecipitates were captured by incubation of the antibody containing lysates with 50 μl of 50 % slurry of Protein A/G PLUS agarose (Santa Cruz Biotechnology). The incubation was carried out for only 30 min (4°C, on rolling wheel) to reduce background binding. The agarose beads were then washed three times with cold NP40 lysis buffer and the bound proteins eluted through boiling in SDS buffer. The samples were subsequently subjected to SDS-PAGE and analyzed by western blot. Control experiments were done either in the absence of antibody or in the presence of a control antibody.

6.9 Silver staining of SDS gels and peptide preparation for mass spectrometry

In order to detect also non-abundant proteins, highly sensitive silver staining of proteins following SDS-PAGE was performed.

The gels were fixed for 1 h in 40% methanol, 10% acetic acid and washed three times for 20 min in 30% ethanol. The proteins were then reduced for 1 min in 0.02% sodium thiosulfate and washed in water prior to incubation with the silver nitrate reagent (0.2% silver nitrate, 0.02% formaldehyde, 37%) for 20 min. After washing, gel developer (3% sodium carbonate, 0.05% formaldehyde (37%), 0.0005% sodium thiosulfate) was added until protein bands became visible (5-20 min). Developing was stopped shortly before the optimal staining occurred through addition of stop reagent (0.5% glycine).

The gel was cut into several pieces containing the proteins of interest and subjected to in-gel tryptic digestion. For silver destaining, the gel slices were repeatedly overlaid with H_2O_2 solution (1% v/v) and then washed and dehydrated in 50% acetonitrile. Protein disulphide bridges were reduced through addition of sufficient 10 mM DTT (in 25 mM ammonium bicarbonate pH 8.0) and incubation for 45 min at 56°. Cysteines were then alkylated via 50 mM iodoacetamide (in 25 mM ammonium bicarbonate pH 8.0) that was left on the gel pieces for 1 h in the dark. After washing in 50% acetonitrile, the slices were dried under vacuum and subjected to digestion with 100 ng of sequencing modified trypsin (Promega) at 37°C overnight. To extract the peptides, 5% trifluoroacetic acid (TFA) in 50% acetonitrile was repeatedly added, the supernatant collected and dried under vacuum.

The tryptic peptides were analyzed on a LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany). Prior to MS analysis, the peptides were separated online on a nano-HPLC (Eksigent, Dublin, CA) following a C_{18} reversed phase column using an acetonitrile/water system at a flow rate of 200 nl/min. Tandem mass spectra were then obtained in a data dependent manner. Generally, 6 MS/MS were acquired after each high accuracy spectral acquisition range. To evaluate the protein data, the human portion of the UniProt database (http://www.uniprot.org, taxonomy ID: 9606) was interrogated using the Mascot search algorithm[409]. Mascot is a search engine that uses mass spectrometry peptide data to identify proteins from primary sequence databases.

6.10   Preparation of nuclear extracts

The cells (10-20 T175 flasks) were grown and harvested in log-phase (2-5 x 10^8) with complete medium, counted and centrifuged at 500 x g for 5-10 min. The pellet was then resuspended in 50 ml cold isotonic buffer (20 mM Hepes, 5 mM KCl, 1.5 mM MgCl2, 0.2 mM PMSF, complete EDTA-free tablet, 0.7 µg/ml pepstatin, 0.5 µg/ml leupeptin, 250 mM sucrose, 1 mM DTT) and recentrifuged as above. Following two washing steps in hypotonic buffer (isotonic buffer without sucrose), the cells were brought to a density of 1.8 - 2 x 10^8/ml, transferred to a tissue grinder and lysed. The nuclei were then spun down at 2'000-3'000 x g for 8 min and resuspended in cold extraction buffer (500 µl buffer/1.5 ml packed nuclei; extraction buffer: 25 mM HEPES/KOH, pH 7.5, 10% sucrose, 1 mM PMSF, 0.5 mM DTT, 1 µg/ml leupeptine, 1 tablet/5 ml of protease inhibitor cocktail "complete mini" (Roche)). The pellet volume was measured and 0.031 volumes of 5 M NaCl (final concentration: 0.155 M) were slowly added dropwise to avoid an uneven increase in salt concentration. After allowing the proteins to leave the nucleus through rotating the mixture for 1h at 4°C, the nuclear debris were pelleted for 20 min at 2°C and 14'500 rpm. The supernatant was dialyzed twice for 1 h against 2 liters of dialysis buffer (25 mM HEPES/KOH pH 7.5, 50 mM KCl, 0.1 mM EDTA, 10% sucrose, 1 mM PMSF, 2 mM DTT, 1 µg/ml leupeptine). Subsequently, the extract was clarified by centrifugation in an Eppendorf centrifuge for 15 min at 20'000 x g at 2°C. The nuclear proteins were then aliquoted, snap-frozen in liquid nitrogen and stored at -80°C. The salt concentration was measured by a conductivity meter and the protein concentration determined with a Bradford assay.

6.11   UDG inhibition and TDG-immunodepletion of nuclear extracts

Protein A Dynabeads (Dynal Biotech, Invitrogen) were washed twice with 30 mM Hepes-KOH, pH 7.5, 7 mM $MgCl_2$ and anti-TDG antibody (rabbit polyclonal, a kind gift of Primo Schaer, 1:10000) was added and incubated for 2h at 4°C on the roller. The beads were washed three times and stored at 4° in the above buffer.

In order to inhibit UDG, 3.6 µl of UGI (uracil glycosylase inhibitor, New England Biolabs) per 150 µg nuclear extract were added and incubated for 10 min at 37°C. To immunodeplete the extracts of TDG, 150 µg of extract were incubated with 6.3 µl pretreated Dynabeads for 30 min at 4°C on the roller.

The nuclear extracts were then used for *in vitro* MMR assays and western blots.

6.12 *In vitro* mismatch repair (MMR) assay

To determine the mismatch repair efficiency of nuclear extracts, restriction enzymes are required in the below described *in vitro* repair assay. The specific enzyme and restriction sites being used for interpretation are shown in Fig. 1a of the manuscript (see 4.2). Substrates with G/T or U/G within the *Acl*I restriction site are refractory to cleavage. After correction of the mismatch from G/T to A/T or U/G to C/G, respectively, the restriction site is restored and the plasmid will be cut three times instead of twice.

To start the mismatch repair reaction, 100 ng (47.5 fmol) of the respective DNA substrate, water, 1 M KCl, 10x MMR buffer (200 mM Tris-HCl, pH 7.6, 400 mM KCl, 50 mM MgCl$_2$, 10 mM glutathione, 500 µg/ml BSA, 1 mM dNTPs) and 15 mM ATP were mixed in 30 µl reaction volumes to a final concentration of 20 mM Tris-HCl, pH 7.6, 110 mM KCl, 5 mM MgCl$_2$, 1 mM glutathione, 50 µg/ml BSA, 0.1 mM dNTPS each, 1.5 mM ATP and 100 ng DNA substrate. If the assay was performed with [α-^{32}P]dATP, dATP was omitted from the dNTP mix.

The repair reactions were incubated for 45 min at 37°C and then stopped by the addition of 45 µl of stop solution (1.12% SDS, 41.67 mM EDTA, 83.33 µg/ml proteinase K) for 30 min at 37°C. Subsequently, the MiniElute^{TM} Reaction Cleanup Kit Protocol (QIAGEN) was used to purify the reactions, resulting in elution of the DNA in 10 µl EB buffer (10 mM Tris-HCl, pH 8.5). The DNA was then submitted to a 3 h *Acl*I digest in a 50 µl reaction at 37°C. Following the digest, 2 µl of RNAse A (2 mg/ml) was added to the samples for 15 min at 37°C to remove the RNA. To free the DNA from residual restriction enzymes, 0.2% SDS and 3.6 µg proteinase K were added and further incubated for another 15 min at 37°C.

To subsequently precipitate the DNA, 0.3 M NaOAc pH 5.5, 40 µl of TE pH 8.0 and 2.5 volumes of cold 100 % ethanol were added and incubated at -20°C overnight. The samples were then centrifuged for 30 min at 4°C and 14'000 rpm in an Eppendorf centrifuge, the pellet washed with 70% ethanol and recentrifuged for 15 min. After drying of the pellet, the DNA was dissolved in 10.5 µl of 10 mM Tris-HCl pH 8.5 and 6x bromophenol blue loading buffer (15%[w/v] Ficoll 400, 0.25% [w/v] bromophenol blue) was added to a final concentration of 1.5x. The samples were loaded on a 1% TAE agarose gel and separated at constant 90V for 45 min. DNA was visualized on a UV transilluminator. In the case of radioactive assays, the gels were first vacuum-dried at 80°C for 75-90 min and then autoradiographed on a PhosphoScreen (Molecular Dynamics).

7 REFERENCES

1. Murphy, K. T., P.; Walport, M. Janeway's immunobiology (Garland Science, Taylor and Francis Group, LLC, New York, 2008).
2. Alberts, B. J., A; Lewis, J; Raff, M; Roberts, K; Walter, P. Molecular biology of the cell (Garland Science, Taylor and Francis Group, New York, 2002).
3. Stefan Karlsson, L. U. (Stefan Karlsson, Lund, 2006).
4. Janeway, C. T., P; Walport, M; Shlomchik, M. Immunologie (Spektrum Akademischer Verlag Heidelberg, Berlin, 2002).
5. Billingham, R. E., Brent, L. & Medawar, P. B. The antigenic stimulus in transplantation immunity. Nature 178, 514-9 (1956).
6. Jaeck, H. M.
7. Burnet, F. The clonal selection theory of acquired immunity. (Cambridge University Press, London, 1959).
8. Rajewsky, K. Clonal selection and learning in the antibody system. Nature 381, 751-8 (1996).
9. Health, N. I. o. (National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases, 2003).
10. Harris, L. J. et al. The three-dimensional structure of an intact monoclonal antibody for canine lymphoma. Nature 360, 369-72 (1992).
11. Tonegawa, S. Somatic generation of antibody diversity. Nature 302, 575-81 (1983).
12. Soulas-Sprauel, P. et al. V(D)J and immunoglobulin class switch recombinations: a paradigm to study the regulation of DNA end-joining. Oncogene 26, 7780-91 (2007).
13. Di Noia, J. M. et al. Dependence of antibody gene diversification on uracil excision. J Exp Med 204, 3209-19 (2007).
14. Buerstedde, J. M. & Arakawa, H. Immunoglobulin gene conversion or hypermutation: that's the question. Subcell Biochem 40, 11-24 (2006).
15. Neuberger, M. S., Harris, R. S., Di Noia, J. & Petersen-Mahrt, S. K. Immunity through DNA deamination. Trends Biochem Sci 28, 305-12 (2003).
16. Muramatsu, M. et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102, 553-63 (2000).
17. Muramatsu, M. et al. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem 274, 18470-6 (1999).
18. Arakawa, H., Hauschild, J. & Buerstedde, J. M. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science 295, 1301-6 (2002).
19. Muramatsu, M., Nagaoka, H., Shinkura, R., Begum, N. A. & Honjo, T. Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. Adv Immunol 94, 1-36 (2007).
20. Maizels, N. Immunoglobulin gene diversification. Annu Rev Genet 39, 23-46 (2005).
21. Di Noia, J. M. & Neuberger, M. S. Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem 76, 1-22 (2007).

22. Li, Z., Woo, C. J., Iglesias-Ussel, M. D., Ronai, D. & Scharff, M. D. The generation of antibody diversity through somatic hypermutation and class switch recombination. Genes Dev 18, 1-11 (2004).

23. Longerich, S., Basu, U., Alt, F. & Storb, U. AID in somatic hypermutation and class switch recombination. Curr Opin Immunol 18, 164-74 (2006).

24. Teng, G. & Papavasiliou, F. N. Immunoglobulin somatic hypermutation. Annu Rev Genet 41, 107-20 (2007).

25. Casali, P., Pal, Z., Xu, Z. & Zan, H. DNA repair in antibody somatic hypermutation. Trends Immunol 27, 313-21 (2006).

26. Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature 418, 99-103 (2002).

27. Revy, P. et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell 102, 565-75 (2000).

28. Durandy, A., Peron, S., Taubenheim, N. & Fischer, A. Activation-induced cytidine deaminase: structure-function relationship as based on the study of mutants. Hum Mutat 27, 1185-91 (2006).

29. Dedeoglu, F., Horwitz, B., Chaudhuri, J., Alt, F. W. & Geha, R. S. Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. Int Immunol 16, 395-404 (2004).

30. Bransteitter, R., Pham, P., Scharff, M. D. & Goodman, M. F. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci U S A 100, 4102-7 (2003).

31. Bransteitter, R., Pham, P., Calabrese, P. & Goodman, M. F. Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. J Biol Chem 279, 51612-21 (2004).

32. Pham, P., Bransteitter, R., Petruska, J. & Goodman, M. F. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature 424, 103-7 (2003).

33. Chaudhuri, J. et al. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature 422, 726-30 (2003).

34. Dickerson, S. K., Market, E., Besmer, E. & Papavasiliou, F. N. AID mediates hypermutation by deaminating single stranded DNA. J Exp Med 197, 1291-6 (2003).

35. Sohail, A., Klapacz, J., Samaranayake, M., Ullah, A. & Bhagwat, A. S. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. Nucleic Acids Res 31, 2990-4 (2003).

36. Besmer, E., Market, E. & Papavasiliou, F. N. The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. Mol Cell Biol 26, 4378-85 (2006).

37. Larijani, M. et al. Methylation protects cytidines from AID-mediated deamination. Mol Immunol 42, 599-604 (2005).

38. Ramiro, A. R., Stavropoulos, P., Jankovic, M. & Nussenzweig, M. C. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. Nat Immunol 4, 452-6 (2003).

39. Goodman, M. F., Scharff, M. D. & Romesberg, F. E. AID-initiated purposeful mutations in immunoglobulin genes. Adv Immunol 94, 127-55 (2007).

40. Nambu, Y. et al. Transcription-coupled events associating with immunoglobulin switch region chromatin. Science 302, 2137-40 (2003).

41. Peled, J. U. et al. The biochemistry of somatic hypermutation. Annu Rev Immunol 26, 481-511 (2008).

42. Ito, S. et al. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. Proc Natl Acad Sci U S A 101, 1975-80 (2004).

43. McBride, K. M., Barreto, V., Ramiro, A. R., Stavropoulos, P. & Nussenzweig, M. C. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. J Exp Med 199, 1235-44 (2004).

44. Brar, S. S., Watson, M. & Diaz, M. Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. J Biol Chem 279, 26395-401 (2004).

45. Barreto, V., Reina-San-Martin, B., Ramiro, A. R., McBride, K. M. & Nussenzweig, M. C. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. Mol Cell 12, 501-8 (2003).

46. Shinkura, R. et al. Separate domains of AID are required for somatic hypermutation and class-switch recombination. Nat Immunol 5, 707-12 (2004).

47. Ta, V. T. et al. AID mutant analyses indicate requirement for class-switch-specific cofactors. Nat Immunol 4, 843-8 (2003).

48. MacDuff, D. A., Neuberger, M. S. & Harris, R. S. MDM2 can interact with the C-terminus of AID but it is inessential for antibody diversification in DT40 B cells. Mol Immunol 43, 1099-108 (2006).

49. Wu, X. et al. Immunoglobulin somatic hypermutation: double-strand DNA breaks, AID and error-prone DNA repair. J Clin Immunol 23, 235-46 (2003).

50. Wu, X., Geraldes, P., Platt, J. L. & Cascalho, M. The double-edged sword of activation-induced cytidine deaminase. J Immunol 174, 934-41 (2005).

51. Basu, U. et al. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. Nature 438, 508-11 (2005).

52. McBride, K. M. et al. Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. Proc Natl Acad Sci U S A 103, 8798-803 (2006).

53. Pasqualucci, L., Kitaura, Y., Gu, H. & Dalla-Favera, R. PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. Proc Natl Acad Sci U S A 103, 395-400 (2006).

54. Chaudhuri, J., Khuong, C. & Alt, F. W. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature 430, 992-8 (2004).

55. Fraenkel, S. et al. Allelic 'choice' governs somatic hypermutation in vivo at the immunoglobulin kappa-chain locus. Nat Immunol 8, 715-22 (2007).

56. Li, Z., Luo, Z. & Scharff, M. D. Differential regulation of histone acetylation and generation of mutations in switch regions is associated with Ig class switching. Proc Natl Acad Sci U S A 101, 15428-33 (2004).

57. Wang, L., Whang, N., Wuerffel, R. & Kenter, A. L. AID-dependent histone acetylation is detected in immunoglobulin S regions. J Exp Med 203, 215-26 (2006).

58. Yang, S. Y. & Schatz, D. G. Targeting of AID-mediated sequence diversification by cis-acting determinants. Adv Immunol 94, 109-25 (2007).

59. Odegard, V. H. & Schatz, D. G. Targeting of somatic hypermutation. Nat Rev Immunol 6, 573-83 (2006).

60. Perlot, T., Alt, F. W., Bassing, C. H., Suh, H. & Pinaud, E. Elucidation of IgH intronic enhancer functions via germ-line deletion. Proc Natl Acad Sci U S A 102, 14362-7 (2005).

61. Wang, C. L., Harper, R. A. & Wabl, M. Genome-wide somatic hypermutation. Proc Natl Acad Sci U S A 101, 7352-6 (2004).

62. Duquette, M. L., Pham, P., Goodman, M. F. & Maizels, N. AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. Oncogene 24, 5791-8 (2005).

63. Yu, K. & Lieber, M. R. Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. DNA Repair (Amst) 2, 1163-74 (2003).

64. Michael, N. et al. The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. Immunity 19, 235-42 (2003).

65. Kodama, M. et al. The PU.1 and NF-EM5 binding motifs in the Igkappa 3' enhancer are responsible for directing somatic hypermutations to the intrinsic hotspots in the transgenic Vkappa gene. Int Immunol 13, 1415-22 (2001).

66. Faili, A. et al. AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. Nat Immunol 3, 815-21 (2002).

67. Petersen, S. et al. AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature 414, 660-5 (2001).

68. Reina-San-Martin, B. et al. H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. J Exp Med 197, 1767-78 (2003).

69. Kosak, S. T. et al. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science 296, 158-62 (2002).

70. Greeve, J. et al. Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas. Blood 101, 3574-80 (2003).

71. Oppezzo, P. et al. Different isoforms of BSAP regulate expression of AID in normal and chronic lymphocytic leukemia B cells. Blood 105, 2495-503 (2005).

72. Oppezzo, P. et al. Chronic lymphocytic leukemia B cells expressing AID display dissociation between class switch recombination and somatic hypermutation. Blood 101, 4029-32 (2003).

73. McCarthy, H. et al. High expression of activation-induced cytidine deaminase (AID) and splice variants is a distinctive feature of poor-prognosis chronic lymphocytic leukemia. Blood 101, 4903-8 (2003).

74. Albesiano, E. et al. Activation-induced cytidine deaminase in chronic lymphocytic leukemia B cells: expression as multiple forms in a dynamic, variably sized fraction of the clone (2003).

75. Noguchi, E. et al. Association between a new polymorphism in the activation-induced cytidine deaminase gene and atopic asthma and the regulation of total serum IgE levels. J Allergy Clin Immunol 108, 382-6 (2001).

76. Gordon, M. S., Kanegai, C. M., Doerr, J. R. & Wall, R. Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). Proc Natl Acad Sci U S A 100, 4126-31 (2003).

77. Shen, H. M., Peters, A., Baron, B., Zhu, X. & Storb, U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science 280, 1750-2 (1998).

78. Muschen, M. et al. Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. J Exp Med 192, 1833-40 (2000).

79. Kotani, A. et al. A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. Proc Natl Acad Sci U S A 102, 4506-11 (2005).

80. Li, Z. et al. A role for Mlh3 in somatic hypermutation. DNA Repair (Amst) 5, 675-82 (2006).

81. Kim, N., Bozek, G., Lo, J. C. & Storb, U. Different mismatch repair deficiencies all have the same effects on somatic hypermutation: intact primary mechanism accompanied by secondary modifications. J Exp Med 190, 21-30 (1999).

82. de Yebenes, V. G. & Ramiro, A. R. Activation-induced deaminase: light and dark sides. Trends Mol Med 12, 432-9 (2006).

83. Rada, C., Ehrenstein, M. R., Neuberger, M. S. & Milstein, C. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. Immunity 9, 135-41 (1998).

84. Phung, Q. H. et al. Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. J Exp Med 187, 1745-51 (1998).

85. Frey, S. et al. Mismatch repair deficiency interferes with the accumulation of mutations in chronically stimulated B cells and not with the hypermutation process. Immunity 9, 127-34 (1998).

86. Bachl, J. & Wabl, M. An immunoglobulin mutator that targets G.C base pairs (1996).

87. Petersen-Mahrt, S. DNA deamination in immunity. Immunol Rev 203, 80-97 (2005).

88. Larson, E. D., Cummings, W. J., Bednarski, D. W. & Maizels, N. MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. Mol Cell 20, 367-75 (2005).

89. Yabuki, M., Fujii, M. M. & Maizels, N. The MRE11-RAD50-NBS1 complex accelerates somatic hypermutation and gene conversion of immunoglobulin variable regions. Nat Immunol 6, 730-6 (2005).

90. Wilson, T. M. et al. MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. J Exp Med 201, 637-45 (2005).

91. Bardwell, P. D. et al. Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. Nat Immunol 5, 224-9 (2004).

92. Tippin, B. & Goodman, M. F. A new class of errant DNA polymerases provides candidates for somatic hypermutation. Philos Trans R Soc Lond B Biol Sci 356, 47-51 (2001).

93. Lehmann, A. R. et al. Translesion synthesis: Y-family polymerases and the polymerase switch. DNA Repair (Amst) 6, 891-9 (2007).

94. Martomo, S. A., Yang, W. W. & Gearhart, P. J. A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. J Exp Med 200, 61-8 (2004).

95. Wiesendanger, M., Kneitz, B., Edelmann, W. & Scharff, M. D. Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-

deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. J Exp Med 191, 579-84 (2000).

96. Rada, C., Di Noia, J. M. & Neuberger, M. S. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. Mol Cell 16, 163-71 (2004).

97. Neuberger, M. S. & Milstein, C. Somatic hypermutation. Curr Opin Immunol 7, 248-54 (1995).

98. Fukita, Y., Jacobs, H. & Rajewsky, K. Somatic hypermutation in the heavy chain locus correlates with transcription. Immunity 9, 105-14 (1998).

99. Longerich, S., Tanaka, A., Bozek, G., Nicolae, D. & Storb, U. The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. J Exp Med 202, 1443-54 (2005).

100. Peters, A. & Storb, U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. Immunity 4, 57-65 (1996).

101. Beale, R. C. et al. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J Mol Biol 337, 585-96 (2004).

102. Manis, J. P., Tian, M. & Alt, F. W. Mechanism and control of class-switch recombination. Trends Immunol 23, 31-9 (2002).

103. Dunnick, W., Hertz, G. Z., Scappino, L. & Gritzmacher, C. DNA sequences at immunoglobulin switch region recombination sites. Nucleic Acids Res 21, 365-72 (1993).

104. Min, I. M., Rothlein, L. R., Schrader, C. E., Stavnezer, J. & Selsing, E. Shifts in targeting of class switch recombination sites in mice that lack mu switch region tandem repeats or Msh2. J Exp Med 201, 1885-90 (2005).

105. Schrader, C. E., Linehan, E. K., Mochegova, S. N., Woodland, R. T. & Stavnezer, J. Inducible DNA breaks in Ig S regions are dependent on AID and UNG. J Exp Med 202, 561-8 (2005).

106. Chaudhuri, J. & Alt, F. W. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol 4, 541-52 (2004).

107. Jung, S., Rajewsky, K. & Radbruch, A. Shutdown of class switch recombination by deletion of a switch region control element. Science 259, 984-7 (1993).

108. Stavnezer, J. Immunoglobulin class switching. Curr Opin Immunol 8, 199-205 (1996).

109. Huang, F. T. et al. Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. Mol Cell Biol 27, 5921-32 (2007).

110. Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E. & Lieber, M. R. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol 4, 442-51 (2003).

111. Stavnezer, J., Guikema, J. E. & Schrader, C. E. Mechanism and regulation of class switch recombination. Annu Rev Immunol 26, 261-92 (2008).

112. Christmann, M., Tomicic, M. T., Roos, W. P. & Kaina, B. Mechanisms of human DNA repair: an update. Toxicology 193, 3-34 (2003).

113. Schrader, C. E., Guikema, J. E., Linehan, E. K., Selsing, E. & Stavnezer, J. 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, 6064-71 (2007).

114. Stavnezer, J. & Schrader, C. E. Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. Trends Genet 22, 23-8 (2006).

115. Rooney, S., Chaudhuri, J. & Alt, F. W. The role of the non-homologous end-joining pathway in lymphocyte development. Immunol Rev 200, 115-31 (2004).

116. Sonoda, E., Hochegger, H., Saberi, A., Taniguchi, Y. & Takeda, S. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. DNA Repair (Amst) 5, 1021-9 (2006).

117. Manis, J. P. et al. Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. J Exp Med 187, 2081-9 (1998).

118. Casellas, R. et al. Ku80 is required for immunoglobulin isotype switching. Embo J 17, 2404-11 (1998).

119. Pan-Hammarstrom, Q. et al. Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells. J Exp Med 201, 189-94 (2005).

120. Soulas-Sprauel, P. et al. Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination. J Exp Med 204, 1717-27 (2007).

121. Chen, L., Trujillo, K., Sung, P. & Tomkinson, A. E. Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. J Biol Chem 275, 26196-205 (2000).

122. Nick McElhinny, S. A., Snowden, C. M., McCarville, J. & Ramsden, D. A. Ku recruits the XRCC4-ligase IV complex to DNA ends. Mol Cell Biol 20, 2996-3003 (2000).

123. Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P. A. & Vindigni, A. Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. DNA Repair (Amst) 6, 712-22 (2007).

124. Walker, J. R., Corpina, R. A. & Goldberg, J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature 412, 607-14 (2001).

125. Spagnolo, L., Rivera-Calzada, A., Pearl, L. H. & Llorca, O. Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. Mol Cell 22, 511-9 (2006).

126. Moreno-Herrero, F. et al. Mesoscale conformational changes in the DNA-repair complex Rad50/Mre11/Nbs1 upon binding DNA. Nature 437, 440-3 (2005).

127. Lee, J. H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science 308, 551-4 (2005).

128. Difilippantonio, S. et al. Role of Nbs1 in the activation of the Atm kinase revealed in humanized mouse models. Nat Cell Biol 7, 675-85 (2005).

129. Falck, J., Coates, J. & Jackson, S. P. Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. Nature 434, 605-11 (2005).

130. Cerosaletti, K., Wright, J. & Concannon, P. Active role for nibrin in the kinetics of atm activation. Mol Cell Biol 26, 1691-9 (2006).

131. Hoeijmakers, J. H. Genome maintenance mechanisms for preventing cancer. Nature 411, 366-74 (2001).

132. Lindahl, T. The Croonian Lecture, 1996: endogenous damage to DNA. Philos Trans R Soc Lond B Biol Sci 351, 1529-38 (1996).

133. Friedberg, E. C. et al. DNA Repair and Mutagenesis (ed. Friedberg, E. C.) (ASM Press, Washington, DC, 2006).

134. Friedberg, E. C. DNA damage and repair. Nature 421, 436-40 (2003).

135. Sancar, A. Structure and function of DNA photolyase. Biochemistry 33, 2-9 (1994).

136. Todo, T. Functional diversity of the DNA photolyase/blue light receptor family. Mutat Res 434, 89-97 (1999).

137. Todo, T. et al. A new photoreactivating enzyme that specifically repairs ultraviolet light-induced (6-4)photoproducts. Nature 361, 371-4 (1993).

138. Koike, G., Maki, H., Takeya, H., Hayakawa, H. & Sekiguchi, M. Purification, structure, and biochemical properties of human O6-methylguanine-DNA methyltransferase. J Biol Chem 265, 14754-62 (1990).

139. Roos, W. P. & Kaina, B. DNA damage-induced cell death by apoptosis. Trends Mol Med 12, 440-50 (2006).

140. Jiricny, J. DNA repair: bioinformatics helps reverse methylation damage. Curr Biol 12, R846-8 (2002).

141. Wood, R. D. DNA damage recognition during nucleotide excision repair in mammalian cells. Biochimie 81, 39-44 (1999).

142. Tornaletti, S. & Hanawalt, P. C. Effect of DNA lesions on transcription elongation. Biochimie 81, 139-46 (1999).

143. Hanawalt, P. C. Subpathways of nucleotide excision repair and their regulation. Oncogene 21, 8949-56 (2002).

144. Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K. & Linn, S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu Rev Biochem 73, 39-85 (2004).

145. Dianov, G. L. et al. Base excision repair in nuclear and mitochondrial DNA. Prog Nucleic Acid Res Mol Biol 68, 285-97 (2001).

146. Fortini, P. & Dogliotti, E. Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. DNA Repair (Amst) 6, 398-409 (2007).

147. Dianov, G. & Lindahl, T. Reconstitution of the DNA base excision-repair pathway. Curr Biol 4, 1069-76 (1994).

148. Kubota, Y. et al. Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. Embo J 15, 6662-70 (1996).

149. Scharer, O. D. & Jiricny, J. Recent progress in the biology, chemistry and structural biology of DNA glycosylases. Bioessays 23, 270-81 (2001).

150. Lindahl, T. An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. Proc Natl Acad Sci U S A 71, 3649-53 (1974).

151. Crosby, B., Prakash, L., Davis, H. & Hinkle, D. C. Purification and characterization of a uracil-DNA glycosylase from the yeast. Saccharomyces cerevisiae. Nucleic Acids Res 9, 5797-809 (1981).

152. Sekiguchi, M., Hayakawa, H., Makino, F., Tanaka, K. & Okada, Y. A human enzyme that liberates uracil from DNA. Biochem Biophys Res Commun 73, 293-9 (1976).

153. Wist, E., Unhjem, O. & Krokan, H. Accumulation of small fragments of DNA in isolated HeLa cell nuclei due to transient incorporation of dUMP. Biochim Biophys Acta 520, 253-70 (1978).

154. Krokan, H. & Wittwer, C. U. Uracil DNa-glycosylase from HeLa cells: general properties, substrate specificity and effect of uracil analogs. Nucleic Acids Res 9, 2599-613 (1981).

155. Krokan, H. E., Standal, R. & Slupphaug, G. DNA glycosylases in the base excision repair of DNA. Biochem J 325 (Pt 1), 1-16 (1997).

156. Mol, C. D. et al. Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. Cell 80, 869-78 (1995).

157. Slupphaug, G. et al. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. Nature 384, 87-92 (1996).

158. Parikh, S. S. et al. Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. Proc Natl Acad Sci U S A 97, 5083-8 (2000).

159. Kavli, B. et al. Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. Embo J 15, 3442-7 (1996).

160. Slupphaug, G. et al. Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. Biochemistry 34, 128-38 (1995).

161. Dizdaroglu, M., Karakaya, A., Jaruga, P., Slupphaug, G. & Krokan, H. E. Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage. Nucleic Acids Res 24, 418-22 (1996).

162. Otterlei, M. et al. Post-replicative base excision repair in replication foci. Embo J 18, 3834-44 (1999).

163. Nilsen, H. et al. Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. Mol Cell 5, 1059-65 (2000).

164. Wiebauer, K. & Jiricny, J. In vitro correction of G.T mispairs to G.C pairs in nuclear extracts from human cells. Nature 339, 234-6 (1989).

165. Wiebauer, K. & Jiricny, J. Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. Proc Natl Acad Sci U S A 87, 5842-5 (1990).

166. Neddermann, P. & Jiricny, J. The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. J Biol Chem 268, 21218-24 (1993).

167. Neddermann, P. et al. Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. J Biol Chem 271, 12767-74 (1996).

168. Gallinari, P. & Jiricny, J. A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. Nature 383, 735-8 (1996).

169. Hardeland, U., Bentele, M., Jiricny, J. & Schar, P. The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs. Nucleic Acids Res 31, 2261-71 (2003).

170. Neddermann, P. & Jiricny, J. Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. Proc Natl Acad Sci U S A 91, 1642-6 (1994).

171. Hardeland, U., Bentele, M., Jiricny, J. & Schar, P. Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. J Biol Chem 275, 33449-56 (2000).

172. Waters, T. R., Gallinari, P., Jiricny, J. & Swann, P. F. Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. J Biol Chem 274, 67-74 (1999).

173. Steinacher, R. & Schar, P. Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. Curr Biol 15, 616-23 (2005).

174. An, Q., Robins, P., Lindahl, T. & Barnes, D. E. C --> T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. Embo J 24, 2205-13 (2005).

175. Chen, D. S., Herman, T. & Demple, B. Two distinct human DNA diesterases that hydrolyze 3'-blocking deoxyribose fragments from oxidized DNA. Nucleic Acids Res 19, 5907-14 (1991).

176. Klungland, A. & Lindahl, T. Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1). Embo J 16, 3341-8 (1997).

177. Matsumoto, Y. & Kim, K. Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. Science 269, 699-702 (1995).

178. Podlutsky, A. J., Dianova, II, Podust, V. N., Bohr, V. A. & Dianov, G. L. Human DNA polymerase beta initiates DNA synthesis during long-patch repair of reduced AP sites in DNA. Embo J 20, 1477-82 (2001).

179. Kunkel, T. A. & Bebenek, K. DNA replication fidelity. Annu Rev Biochem 69, 497-529 (2000).

180. Hubscher, U., Maga, G. & Spadari, S. Eukaryotic DNA polymerases. Annu Rev Biochem 71, 133-63 (2002).

181. Kunkel, T. A. DNA replication fidelity. J Biol Chem 279, 16895-8 (2004).

182. Peltomaki, P. et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science 260, 810-2 (1993).

183. Fishel, R. et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 75, 1027-38 (1993).

184. Leach, F. S. et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 75, 1215-25 (1993).

185. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. Nature 396, 643-9 (1998).

186. Grady, W. M. Genomic instability and colon cancer. Cancer Metastasis Rev 23, 11-27 (2004).

187. Jiricny, J. The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 7, 335-46 (2006).

188. Karran, P. & Marinus, M. G. Mismatch correction at O6-methylguanine residues in E. coli DNA. Nature 296, 868-9 (1982).

189. Duckett, D. R. et al. Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct. Proc Natl Acad Sci U S A 93, 6443-7 (1996).

190. Mazurek, A., Berardini, M. & Fishel, R. Activation of human MutS homologs by 8-oxo-guanine DNA damage. J Biol Chem 277, 8260-6 (2002).

191. Mello, J. A., Acharya, S., Fishel, R. & Essigmann, J. M. The mismatch-repair protein hMSH2 binds selectively to DNA adducts of the anticancer drug cisplatin. Chem Biol 3, 579-89 (1996).

192. Kolodner, R. Biochemistry and genetics of eukaryotic mismatch repair. Genes Dev 10, 1433-42 (1996).

193. Modrich, P. & Lahue, R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem 65, 101-33 (1996).

194. Jiricny, J. Replication errors: cha(lle)nging the genome. Embo J 17, 6427-36 (1998).

195. Kolodner, R. D. & Marsischky, G. T. Eukaryotic DNA mismatch repair. Curr Opin Genet Dev 9, 89-96 (1999).

196. Li, G. M. The role of mismatch repair in DNA damage-induced apoptosis. Oncol Res 11, 393-400 (1999).

197. Karran, P. Mechanisms of tolerance to DNA damaging therapeutic drugs. Carcinogenesis 22, 1931-7 (2001).

198. Stojic, L., Brun, R. & Jiricny, J. Mismatch repair and DNA damage signalling. DNA Repair (Amst) 3, 1091-101 (2004).

199. de la Chapelle, A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 4, 769-80 (2004).

200. Rowley, P. T. Inherited susceptibility to colorectal cancer. Annu Rev Med 56, 539-54 (2005).

201. Eshleman, J. R. & Markowitz, S. D. Microsatellite instability in inherited and sporadic neoplasms. Curr Opin Oncol 7, 83-9 (1995).

202. Peltomaki, P. DNA mismatch repair and cancer. Mutat Res 488, 77-85 (2001).

203. Lahue, R. S., Au, K. G. & Modrich, P. DNA mismatch correction in a defined system. Science 245, 160-4 (1989).

204. Wildenberg, J. & Meselson, M. Mismatch repair in heteroduplex DNA. Proc Natl Acad Sci U S A 72, 2202-6 (1975).

205. Wagner, R., Jr. & Meselson, M. Repair tracts in mismatched DNA heteroduplexes. Proc Natl Acad Sci U S A 73, 4135-9 (1976).

206. Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. T. & Modrich, P. In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. Proc Natl Acad Sci U S A 98, 6765-70 (2001).

207. Cooper, D. L., Lahue, R. S. & Modrich, P. Methyl-directed mismatch repair is bidirectional. J Biol Chem 268, 11823-9 (1993).

208. Viswanathan, M., Burdett, V., Baitinger, C., Modrich, P. & Lovett, S. T. Redundant exonuclease involvement in Escherichia coli methyl-directed mismatch repair. J Biol Chem 276, 31053-8 (2001).

209. Su, S. S. & Modrich, P. Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. Proc Natl Acad Sci U S A 83, 5057-61 (1986).

210. Grilley, M., Griffith, J. & Modrich, P. Bidirectional excision in methyl-directed mismatch repair. J Biol Chem 268, 11830-7 (1993).

211. Au, K. G., Welsh, K. & Modrich, P. Initiation of methyl-directed mismatch repair. J Biol Chem 267, 12142-8 (1992).

212. Grilley, M., Welsh, K. M., Su, S. S. & Modrich, P. Isolation and characterization of the Escherichia coli mutL gene product. J Biol Chem 264, 1000-4 (1989).

213. Welsh, K. M., Lu, A. L., Clark, S. & Modrich, P. Isolation and characterization of the Escherichia coli mutH gene product. J Biol Chem 262, 15624-9 (1987).

214. Allen, D. J. et al. MutS mediates heteroduplex loop formation by a translocation mechanism. Embo J 16, 4467-76 (1997).

215. Blackwell, L. J., Martik, D., Bjornson, K. P., Bjornson, E. S. & Modrich, P. Nucleotide-promoted release of hMutSalpha from heteroduplex DNA is

consistent with an ATP-dependent translocation mechanism. J Biol Chem 273, 32055-62 (1998).

216. Acharya, S., Foster, P. L., Brooks, P. & Fishel, R. The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. Mol Cell 12, 233-46 (2003).

217. Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P. & Yang, W. Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. Mol Cell 7, 1-12 (2001).

218. Dao, V. & Modrich, P. Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. J Biol Chem 273, 9202-7 (1998).

219. Holmes, J., Jr., Clark, S. & Modrich, P. Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines. Proc Natl Acad Sci U S A 87, 5837-41 (1990).

220. Thomas, D. C., Roberts, J. D. & Kunkel, T. A. Heteroduplex repair in extracts of human HeLa cells. J Biol Chem 266, 3744-51 (1991).

221. Modrich, P. Mechanisms in eukaryotic mismatch repair. J Biol Chem 281, 30305-9 (2006).

222. Bjornson, K. P. et al. Assembly and molecular activities of the MutS tetramer. J Biol Chem 278, 34667-73 (2003).

223. Jiricny, J., Su, S. S., Wood, S. G. & Modrich, P. Mismatch-containing oligonucleotide duplexes bound by the E. coli mutS-encoded protein. Nucleic Acids Res 16, 7843-53 (1988).

224. Jiricny, J. Mismatch repair: the praying hands of fidelity. Curr Biol 10, R788-90 (2000).

225. Lamers, M. H. et al. The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. Nature 407, 711-7 (2000).

226. Dufner, P., Marra, G., Raschle, M. & Jiricny, J. Mismatch recognition and DNA-dependent stimulation of the ATPase activity of hMutSalpha is abolished by a single mutation in the hMSH6 subunit. J Biol Chem 275, 36550-5 (2000).

227. Malkov, V. A., Biswas, I., Camerini-Otero, R. D. & Hsieh, P. Photocross-linking of the NH2-terminal region of Taq MutS protein to the major groove of a heteroduplex DNA. J Biol Chem 272, 23811-7 (1997).

228. Gorbalenya, A. E. & Koonin, E. V. Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination/repair systems. J Mol Biol 213, 583-91 (1990).

229. Obmolova, G., Ban, C., Hsieh, P. & Yang, W. Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature 407, 703-10 (2000).

230. Iaccarino, I., Marra, G., Dufner, P. & Jiricny, J. Mutation in the magnesium binding site of hMSH6 disables the hMutSalpha sliding clamp from translocating along DNA. J Biol Chem 275, 2080-6 (2000).

231. Iaccarino, I., Marra, G., Palombo, F. & Jiricny, J. hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSalpha. Embo J 17, 2677-86 (1998).

232. Modrich, P. Mechanisms and biological effects of mismatch repair. Annu Rev Genet 25, 229-53 (1991).

233. Drotschmann, K., Aronshtam, A., Fritz, H. J. & Marinus, M. G. The Escherichia coli MutL protein stimulates binding of Vsr and MutS to heteroduplex DNA. Nucleic Acids Res 26, 948-53 (1998).

234. Dutta, R. & Inouye, M. GHKL, an emergent ATPase/kinase superfamily. Trends Biochem Sci 25, 24-8 (2000).

235. Spampinato, C. & Modrich, P. The MutL ATPase is required for mismatch repair. J Biol Chem 275, 9863-9 (2000).

236. Ban, C., Junop, M. & Yang, W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. Cell 97, 85-97 (1999).

237. Bende, S. M. & Grafstrom, R. H. The DNA binding properties of the MutL protein isolated from Escherichia coli. Nucleic Acids Res 19, 1549-55 (1991).

238. Kosinski, J., Steindorf, I., Bujnicki, J. M., Giron-Monzon, L. & Friedhoff, P. Analysis of the quaternary structure of the MutL C-terminal domain. J Mol Biol 351, 895-909 (2005).

239. Robertson, A., Pattishall, S. R. & Matson, S. W. The DNA binding activity of MutL is required for methyl-directed mismatch repair in Escherichia coli. J Biol Chem 281, 8399-408 (2006).

240. Guarne, A. et al. Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. Embo J 23, 4134-45 (2004).

241. Hall, M. C., Jordan, J. R. & Matson, S. W. Evidence for a physical interaction between the Escherichia coli methyl-directed mismatch repair proteins MutL and UvrD. Embo J 17, 1535-41 (1998).

242. Ban, C. & Yang, W. Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. Cell 95, 541-52 (1998).

243. Matson, S. W. & Robertson, A. B. The UvrD helicase and its modulation by the mismatch repair protein MutL. Nucleic Acids Res 34, 4089-97 (2006).

244. Fishel, R. et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 77, 167 (1994).

245. Kolodner, R. D. et al. Structure of the human MLH1 locus and analysis of a large hereditary nonpolyposis colorectal carcinoma kindred for mlh1 mutations. Cancer Res 55, 242-8 (1995).

246. Papadopoulos, N. et al. Mutation of a mutL homolog in hereditary colon cancer. Science 263, 1625-9 (1994).

247. Nicolaides, N. C. et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature 371, 75-80 (1994).

248. Alani, E., Sokolsky, T., Studamire, B., Miret, J. J. & Lahue, R. S. Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. Mol Cell Biol 17, 2436-47 (1997).

249. Habraken, Y., Sung, P., Prakash, L. & Prakash, S. ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2-MSH6 and MLH1-PMS1 protein complexes. J Biol Chem 273, 9837-41 (1998).

250. Johnson, R. E., Kovvali, G. K., Prakash, L. & Prakash, S. Requirement of the yeast MSH3 and MSH6 genes for MSH2-dependent genomic stability. J Biol Chem 271, 7285-8 (1996).

251. Marsischky, G. T., Filosi, N., Kane, M. F. & Kolodner, R. Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev 10, 407-20 (1996).

252. Reenan, R. A. & Kolodner, R. D. Characterization of insertion mutations in the Saccharomyces cerevisiae MSH1 and MSH2 genes: evidence for separate mitochondrial and nuclear functions. Genetics 132, 975-85 (1992).

253. Culligan, K. M., Meyer-Gauen, G., Lyons-Weiler, J. & Hays, J. B. Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. Nucleic Acids Res 28, 463-71 (2000).

254. Hollingsworth, N. M., Ponte, L. & Halsey, C. MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in Saccharomyces cerevisiae but not mismatch repair. Genes Dev 9, 1728-39 (1995).

255. Ross-Macdonald, P. & Roeder, G. S. Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. Cell 79, 1069-80 (1994).

256. Alani, E. The Saccharomyces cerevisiae Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs. Mol Cell Biol 16, 5604-15 (1996).

257. Habraken, Y., Sung, P., Prakash, L. & Prakash, S. Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. Curr Biol 6, 1185-7 (1996).

258. Iaccarino, I. et al. MSH6, a Saccharomyces cerevisiae protein that binds to mismatches as a heterodimer with MSH2. Curr Biol 6, 484-6 (1996).

259. Marsischky, G. T. & Kolodner, R. D. Biochemical characterization of the interaction between the Saccharomyces cerevisiae MSH2-MSH6 complex and mispaired bases in DNA. J Biol Chem 274, 26668-82 (1999).

260. Flores-Rozas, H. & Kolodner, R. D. The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations. Proc Natl Acad Sci U S A 95, 12404-9 (1998).

261. Prolla, T. A., Christie, D. M. & Liskay, R. M. Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene. Mol Cell Biol 14, 407-15 (1994).

262. Kramer, W., Kramer, B., Williamson, M. S. & Fogel, S. Cloning and nucleotide sequence of DNA mismatch repair gene PMS1 from Saccharomyces cerevisiae: homology of PMS1 to procaryotic MutL and HexB. J Bacteriol 171, 5339-46 (1989).

263. Harfe, B. D., Minesinger, B. K. & Jinks-Robertson, S. Discrete in vivo roles for the MutL homologs Mlh2p and Mlh3p in the removal of frameshift intermediates in budding yeast. Curr Biol 10, 145-8 (2000).

264. Genschel, J., Littman, S. J., Drummond, J. T. & Modrich, P. Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. J Biol Chem 273, 19895-901 (1998).

265. Drummond, J. T., Li, G. M., Longley, M. J. & Modrich, P. Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. Science 268, 1909-12 (1995).

266. Palombo, F. et al. hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. Curr Biol 6, 1181-4 (1996).

267. Peltomaki, P. Role of DNA mismatch repair defects in the pathogenesis of human cancer. J Clin Oncol 21, 1174-9 (2003).

268. Peltomaki, P. Lynch syndrome genes. Fam Cancer 4, 227-32 (2005).

269. Edelmann, W. et al. The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. Cancer Res 60, 803-7 (2000).

270. Bocker, T., Ruschoff, J. & Fishel, R. Molecular diagnostics of cancer predisposition: hereditary non-polyposis colorectal carcinoma and mismatch repair defects. Biochim Biophys Acta 1423, O1-O10 (1999).

271. Snowden, T., Acharya, S., Butz, C., Berardini, M. & Fishel, R. hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. Mol Cell 15, 437-51 (2004).

272. Li, G. M. & Modrich, P. Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. Proc Natl Acad Sci U S A 92, 1950-4 (1995).

273. Raschle, M., Marra, G., Nystrom-Lahti, M., Schar, P. & Jiricny, J. Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1. J Biol Chem 274, 32368-75 (1999).

274. Cannavo, E. et al. Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. Cancer Res 65, 10759-66 (2005).

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

276. Constantin, N., Dzantiev, L., Kadyrov, F. A. & Modrich, P. Human mismatch repair: reconstitution of a nick-directed bidirectional reaction. J Biol Chem 280, 39752-61 (2005).

277. Zhang, Y. et al. Reconstitution of 5'-directed human mismatch repair in a purified system. Cell 122, 693-705 (2005).

278. Genschel, J., Bazemore, L. R. & Modrich, P. Human exonuclease I is required for 5' and 3' mismatch repair. J Biol Chem 277, 13302-11 (2002).

279. Genschel, J. & Modrich, P. Mechanism of 5'-directed excision in human mismatch repair. Mol Cell 12, 1077-86 (2003).

280. Yuan, F., Gu, L., Guo, S., Wang, C. & Li, G. M. Evidence for involvement of HMGB1 protein in human DNA mismatch repair. J Biol Chem 279, 20935-40 (2004).

281. Wei, K. et al. Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. Genes Dev 17, 603-14 (2003).

282. Jiricny, J., Hughes, M., Corman, N. & Rudkin, B. B. A human 200-kDa protein binds selectively to DNA fragments containing G.T mismatches. Proc Natl Acad Sci U S A 85, 8860-4 (1988).

283. Hughes, M. J. & Jiricny, J. The purification of a human mismatch-binding protein and identification of its associated ATPase and helicase activities. J Biol Chem 267, 23876-82 (1992).

284. Palombo, F., Hughes, M., Jiricny, J., Truong, O. & Hsuan, J. Mismatch repair and cancer. Nature 367, 417 (1994).

285. Aravind, L., Walker, D. R. & Koonin, E. V. Conserved domains in DNA repair proteins and evolution of repair systems. Nucleic Acids Res 27, 1223-42 (1999).

286. Palombo, F. et al. GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. Science 268, 1912-4 (1995).

287. Acharya, S. et al. hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. Proc Natl Acad Sci U S A 93, 13629-34 (1996).

288. Warren, J. J. et al. Structure of the human MutSalpha DNA lesion recognition complex. Mol Cell 26, 579-92 (2007).

289. Martik, D., Baitinger, C. & Modrich, P. Differential specificities and simultaneous occupancy of human MutSalpha nucleotide binding sites. J Biol Chem 279, 28402-10 (2004).

290. Drummond, J. T., Genschel, J., Wolf, E. & Modrich, P. DHFR/MSH3 amplification in methotrexate-resistant cells alters the hMutSalpha/hMutSbeta ratio and reduces the efficiency of base-base mismatch repair. Proc Natl Acad Sci U S A 94, 10144-9 (1997).

291. Marra, G. et al. Mismatch repair deficiency associated with overexpression of the MSH3 gene. Proc Natl Acad Sci U S A 95, 8568-73 (1998).

292. Raschle, M., Dufner, P., Marra, G. & Jiricny, J. Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. J Biol Chem 277, 21810-20 (2002).

293. Prolla, T. A. et al. Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. Nat Genet 18, 276-9 (1998).

294. Pang, Q., Prolla, T. A. & Liskay, R. M. Functional domains of the Saccharomyces cerevisiae Mlh1p and Pms1p DNA mismatch repair proteins and their relevance to human hereditary nonpolyposis colorectal cancer-associated mutations. Mol Cell Biol 17, 4465-73 (1997).

295. Hoffmann, E. R. & Borts, R. H. Meiotic recombination intermediates and mismatch repair proteins. Cytogenet Genome Res 107, 232-48 (2004).

296. Santucci-Darmanin, S. et al. The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination. Hum Mol Genet 11, 1697-706 (2002).

297. Chen, P. C. et al. Contributions by MutL homologues Mlh3 and Pms2 to DNA mismatch repair and tumor suppression in the mouse. Cancer Res 65, 8662-70 (2005).

298. Harfe, B. D. & Jinks-Robertson, S. DNA mismatch repair and genetic instability. Annu Rev Genet 34, 359-399 (2000).

299. Surtees, J. A., Argueso, J. L. & Alani, E. Mismatch repair proteins: key regulators of genetic recombination. Cytogenet Genome Res 107, 146-59 (2004).

300. Matic, I., Radman, M. & Rayssiguier, C. Structure of recombinants from conjugational crosses between Escherichia coli donor and mismatch-repair deficient Salmonella typhimurium recipients. Genetics 136, 17-26 (1994).

301. Petit, M. A., Dimpfl, J., Radman, M. & Echols, H. Control of large chromosomal duplications in Escherichia coli by the mismatch repair system. Genetics 129, 327-32 (1991).

302. Rayssiguier, C., Thaler, D. S. & Radman, M. The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants. Nature 342, 396-401 (1989).

303. Chen, S., Bigner, S. H. & Modrich, P. High rate of CAD gene amplification in human cells deficient in MLH1 or MSH6. Proc Natl Acad Sci U S A 98, 13802-7 (2001).

304. Worth, L., Jr., Clark, S., Radman, M. & Modrich, P. Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. Proc Natl Acad Sci U S A 91, 3238-41 (1994).

305. Bocker, T. et al. hMSH5: a human MutS homologue that forms a novel heterodimer with hMSH4 and is expressed during spermatogenesis. Cancer Res 59, 816-22 (1999).

306. Kunz, C. & Schar, P. Meiotic recombination: sealing the partnership at the junction. Curr Biol 14, R962-4 (2004).

307. Lipkin, S. M. et al. Meiotic arrest and aneuploidy in MLH3-deficient mice. Nat Genet 31, 385-90 (2002).

308. Nakagawa, M., Iwasa, T., Kikkawa, S., Tsuda, M. & Ebrey, T. G. How vertebrate and invertebrate visual pigments differ in their mechanism of photoactivation. Proc Natl Acad Sci U S A 96, 6189-92 (1999).

309. Baker, S. M. et al. Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. Cell 82, 309-19 (1995).

310. Kunkel, T. A. & Erie, D. A. DNA mismatch repair. Annu Rev Biochem 74, 681-710 (2005).

311. Iyer, R. R., Pluciennik, A., Burdett, V. & Modrich, P. L. DNA mismatch repair: functions and mechanisms. Chem Rev 106, 302-23 (2006).

312. Bowater, R. P. & Wells, R. D. The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. Prog Nucleic Acid Res Mol Biol 66, 159-202 (2001).

313. van den Broek, W. J. et al. Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. Hum Mol Genet 11, 191-8 (2002).

314. Wei, K., Kucherlapati, R. & Edelmann, W. Mouse models for human DNA mismatch-repair gene defects. Trends Mol Med 8, 346-53 (2002).

315. Bertocci, B. et al. Probing immunoglobulin gene hypermutation with microsatellites suggests a nonreplicative short patch DNA synthesis process. Immunity 9, 257-65 (1998).

316. Li, Z. et al. Examination of Msh6- and Msh3-deficient mice in class switching reveals overlapping and distinct roles of MutS homologues in antibody diversification. J Exp Med 200, 47-59 (2004).

317. Li, Z. et al. The mismatch repair protein Msh6 influences the in vivo AID targeting to the Ig locus. Immunity 24, 393-403 (2006).

318. Phung, Q. H., Winter, D. B., Alrefai, R. & Gearhart, P. J. Hypermutation in Ig V genes from mice deficient in the MLH1 mismatch repair protein. J Immunol 162, 3121-4 (1999).

319. Ehrenstein, M. R., Rada, C., Jones, A. M., Milstein, C. & Neuberger, M. S. Switch junction sequences in PMS2-deficient mice reveal a microhomology-mediated mechanism of Ig class switch recombination. Proc Natl Acad Sci U S A 98, 14553-8 (2001).

320. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F. & Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev 18, 1618-29 (2004).

321. Woo, C. J., Martin, A. & Scharff, M. D. Induction of somatic hypermutation is associated with modifications in immunoglobulin variable region chromatin. Immunity 19, 479-89 (2003).

322. Larson, E. D., Duquette, M. L., Cummings, W. J., Streiff, R. J. & Maizels, N. MutSalpha binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions. Curr Biol 15, 470-4 (2005).

323. Ehrenstein, M. R. & Neuberger, M. S. Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. Embo J 18, 3484-90 (1999).

324. Schrader, C. E., Edelmann, W., Kucherlapati, R. & Stavnezer, J. Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. J Exp Med 190, 323-30 (1999).

325. Martin, A. et al. Msh2 ATPase activity is essential for somatic hypermutation at a-T basepairs and for efficient class switch recombination. J Exp Med 198, 1171-8 (2003).

326. Krokan, H. E., Drablos, F. & Slupphaug, G. Uracil in DNA--occurrence, consequences and repair. Oncogene 21, 8935-48 (2002).

327. Di Noia, J. & Neuberger, M. S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature 419, 43-8 (2002).

328. Rada, C. et al. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr Biol 12, 1748-55 (2002).

329. Di Noia, J. M. & Neuberger, M. S. Immunoglobulin gene conversion in chicken DT40 cells largely proceeds through an abasic site intermediate generated by excision of the uracil produced by AID-mediated deoxycytidine deamination. Eur J Immunol 34, 504-8 (2004).

330. Saribasak, H. et al. Uracil DNA glycosylase disruption blocks Ig gene conversion and induces transition mutations. J Immunol 176, 365-71 (2006).

331. Imai, K. et al. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. Nat Immunol 4, 1023-8 (2003).

332. Bardwell, P. D. et al. Cutting edge: the G-U mismatch glycosylase methyl-CpG binding domain 4 is dispensable for somatic hypermutation and class switch recombination. J Immunol 170, 1620-4 (2003).

333. Di Noia, J. M., Rada, C. & Neuberger, M. S. SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. Embo J 25, 585-95 (2006).

334. Xue, K., Rada, C. & Neuberger, M. S. The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/-ung-/- mice. J Exp Med 203, 2085-94 (2006).

335. Shen, H. M., Tanaka, A., Bozek, G., Nicolae, D. & Storb, U. Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice. J Immunol 177, 5386-92 (2006).

336. Zeng, X. et al. DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. Nat Immunol 2, 537-41 (2001).

337. Faili, A. et al. DNA polymerase eta is involved in hypermutation occurring during immunoglobulin class switch recombination. J Exp Med 199, 265-70 (2004).

338. Delbos, F. et al. Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. J Exp Med 201, 1191-6 (2005).

339. Mayorov, V. I., Rogozin, I. B., Adkison, L. R. & Gearhart, P. J. DNA polymerase eta contributes to strand bias of mutations of A versus T in immunoglobulin genes. J Immunol 174, 7781-6 (2005).

340. Martomo, S. A. et al. Different mutation signatures in DNA polymerase eta- and MSH6-deficient mice suggest separate roles in antibody diversification. Proc Natl Acad Sci U S A 102, 8656-61 (2005).

341. Rogozin, I. B., Pavlov, Y. I., Bebenek, K., Matsuda, T. & Kunkel, T. A. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. Nat Immunol 2, 530-6 (2001).

342. Delbos, F., Aoufouchi, S., Faili, A., Weill, J. C. & Reynaud, C. A. DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. J Exp Med 204, 17-23 (2007).

343. Masutani, C. et al. The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. Nature 399, 700-4 (1999).

344. Lehmann, A. R. et al. Xeroderma pigmentosum cells with normal levels of excision repair have a defect in DNA synthesis after UV-irradiation. Proc Natl Acad Sci U S A 72, 219-23 (1975).

345. Kannouche, P. et al. Localization of DNA polymerases eta and iota to the replication machinery is tightly co-ordinated in human cells. Embo J 21, 6246-56 (2002).

346. Faili, A. et al. Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase iota. Nature 419, 944-7 (2002).

347. Poltoratsky, V. et al. Expression of error-prone polymerases in BL2 cells activated for Ig somatic hypermutation. Proc Natl Acad Sci U S A 98, 7976-81 (2001).

348. McDonald, J. P. et al. 129-derived strains of mice are deficient in DNA polymerase iota and have normal immunoglobulin hypermutation. J Exp Med 198, 635-43 (2003).

349. Diaz, M., Verkoczy, L. K., Flajnik, M. F. & Klinman, N. R. Decreased frequency of somatic hypermutation and impaired affinity maturation but intact germinal center formation in mice expressing antisense RNA to DNA polymerase zeta. J Immunol 167, 327-35 (2001).

350. Zan, H. et al. The translesion DNA polymerase zeta plays a major role in Ig and bcl-6 somatic hypermutation. Immunity 14, 643-53 (2001).

351. Zan, H. et al. The translesion DNA polymerase theta plays a dominant role in immunoglobulin gene somatic hypermutation. Embo J 24, 3757-69 (2005).

352. Nelson, J. R., Lawrence, C. W. & Hinkle, D. C. Deoxycytidyl transferase activity of yeast REV1 protein. Nature 382, 729-31 (1996).

353. Prakash, S., Johnson, R. E. & Prakash, L. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. Annu Rev Biochem 74, 317-53 (2005).

354. Ross, A. L. & Sale, J. E. The catalytic activity of REV1 is employed during immunoglobulin gene diversification in DT40. Mol Immunol 43, 1587-94 (2006).

355. Jansen, J. G. et al. Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J Exp Med 203, 319-23 (2006).

356. Bachl, J., Ertongur, I. & Jungnickel, B. Involvement of Rad18 in somatic hypermutation. Proc Natl Acad Sci U S A 103, 12081-6 (2006).

357. Simpson, L. J. et al. RAD18-independent ubiquitination of proliferating-cell nuclear antigen in the avian cell line DT40. EMBO Rep 7, 927-32 (2006).

358. Arakawa, H. et al. A role for PCNA ubiquitination in immunoglobulin hypermutation. PLoS Biol 4, e366 (2006).

359. Langerak, P., Nygren, A. O., Krijger, P. H., van den Berk, P. C. & Jacobs, H. A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. J Exp Med 204, 1989-98 (2007).

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

361. Rigaut, G. S., A; Rutz, B; Wilm, M; Mann, M; Seraphin, B. A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotechnol. 17, 1030-1032 (1999).

362. Puig, O. C., F; Rigaut, G; Rutz, B; Bouveret, E; Bragado-Nilsson, E; Wilm, M; Seraphin, B. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods 24, 218-229 (2001).

363. Gingras, A. C., Aebersold, R. & Raught, B. Advances in protein complex analysis using mass spectrometry. J Physiol 563, 11-21 (2005).

364. Sale, J. E. & Neuberger, M. S. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. Immunity 9, 859-69 (1998).

365. Denepoux, S. et al. Induction of somatic mutation in a human B cell line in vitro. Immunity 6, 35-46 (1997).

366. Zan, H. et al. Induction of Ig somatic hypermutation and class switching in a human monoclonal IgM+ IgD+ B cell line in vitro: definition of the requirements and modalities of hypermutation. J Immunol 162, 3437-47 (1999).

367. Hanakahi, L. A., Dempsey, L. A., Li, M. J. & Maizels, N. Nucleolin is one component of the B cell-specific transcription factor and switch region binding protein, LR1. Proc Natl Acad Sci U S A 94, 3605-10 (1997).

368. Hanakahi, L. A. & Maizels, N. Transcriptional activation by LR1 at the Emu enhancer and switch region sites. Nucleic Acids Res 28, 2651-7 (2000).

369. Williams, M. & Maizels, N. LR1, a lipopolysaccharide-responsive factor with binding sites in the immunoglobulin switch regions and heavy-chain enhancer. Genes Dev 5, 2353-61 (1991).

370. Stucki, M. et al. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell 123, 1213-26 (2005).

371. Pham, P., Bransteitter, R. & Goodman, M. F. Reward versus risk: DNA cytidine deaminases triggering immunity and disease. Biochemistry 44, 2703-15 (2005).

372. Gallant, P. Control of transcription by Pontin and Reptin. Trends Cell Biol 17, 187-92 (2007).

373. Qiu, X. B. et al. An eukaryotic RuvB-like protein (RUVBL1) essential for growth. J Biol Chem 273, 27786-93 (1998).

374. Matias, P. M., Gorynia, S., Donner, P. & Carrondo, M. A. Crystal structure of the human AAA+ protein RuvBL1. J Biol Chem 281, 38918-29 (2006).

375. Vallur, A. C., Yabuki, M., Larson, E. D. & Maizels, N. AID in antibody perfection. Cell Mol Life Sci 64, 555-65 (2007).

376. Neuberger, M. S. & Rada, C. Somatic hypermutation: activation-induced deaminase for C/G followed by polymerase eta for A/T. J Exp Med 204, 7-10 (2007).

377. Sharma, R. A. & Dianov, G. L. Targeting base excision repair to improve cancer therapies. Mol Aspects Med 28, 345-74 (2007).

378. Fang, W. H. & Modrich, P. Human strand-specific mismatch repair occurs by a bidirectional mechanism similar to that of the bacterial reaction. J Biol Chem 268, 11838-44 (1993).

379. Marra, G. et al. Expression of human MutS homolog 2 (hMSH2) protein in resting and proliferating cells. Oncogene 13, 2189-96 (1996).

380. Jiricny, J. MutLalpha: at the cutting edge of mismatch repair. Cell 126, 239-41 (2006).

381. Poltoratsky, V., Prasad, R., Horton, J. K. & Wilson, S. H. Down-regulation of DNA polymerase beta accompanies somatic hypermutation in human BL2 cell lines. DNA Repair (Amst) 6, 244-53 (2007).

382. Cejka, P. et al. Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. Embo J 22, 2245-54 (2003).

383. Baerenfaller, K., Fischer, F. & Jiricny, J. Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro. Methods Enzymol 408, 285-303 (2006).

384. Fischer, F., Baerenfaller, K. & Jiricny, J. 5-Fluorouracil is efficiently removed from DNA by the base excision and mismatch repair systems. Gastroenterology 133, 1858-68 (2007).

385. Dziembowski, A. S., B. Recent developments in the analysis of protein complexes. FEBS Lett. 556, 1-6 (2004).

386. Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R. & Jiricny, J. Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2. J Biol Chem 282, 2976-86 (2007).

387. Khattar, S. K. et al. Enhanced soluble production of biologically active recombinant human p38 mitogen-activated-protein kinase (MAPK) in Escherichia coli. Protein Pept Lett 14, 756-60 (2007).

388. Harris, R. S., Croom-Carter, D. S., Rickinson, A. B. & Neuberger, M. S. Epstein-Barr virus and the somatic hypermutation of immunoglobulin genes in Burkitt's lymphoma cells. J Virol 75, 10488-92 (2001).

389. Xiao, Z. et al. Known components of the immunoglobulin A:T mutational machinery are intact in Burkitt lymphoma cell lines with G:C bias. Mol Immunol 44, 2659-66 (2007).

390. Menissier de Murcia, J. et al. Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse. Embo J 22, 2255-63 (2003).

391. Hochegger, H. et al. Parp-1 protects homologous recombination from interference by Ku and Ligase IV in vertebrate cells. Embo J 25, 1305-14 (2006).

392. Shen, H. M. & Storb, U. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. Proc Natl Acad Sci U S A 101, 12997-3002 (2004).

393. Dempsey, L. A., Sun, H., Hanakahi, L. A. & Maizels, N. G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. J Biol Chem 274, 1066-71 (1999).

394. Hopfner, K. P. et al. Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. Cell 105, 473-85 (2001).

395. Paull, T. T. & Gellert, M. The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. Mol Cell 1, 969-79 (1998).

396. Trujillo, K. M. & Sung, P. DNA structure-specific nuclease activities in the Saccharomyces cerevisiae Rad50*Mre11 complex. J Biol Chem 276, 35458-64 (2001).

397. Dugan, K. A., Wood, M. A. & Cole, M. D. TIP49, but not TRRAP, modulates c-Myc and E2F1 dependent apoptosis. Oncogene 21, 5835-43 (2002).

398. Wood, M. A., McMahon, S. B. & Cole, M. D. An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. Mol Cell 5, 321-30 (2000).

399. Feng, Y., Lee, N. & Fearon, E. R. TIP49 regulates beta-catenin-mediated neoplastic transformation and T-cell factor target gene induction via effects on chromatin remodeling. Cancer Res 63, 8726-34 (2003).

400. Bauer, A. et al. Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. Embo J 19, 6121-30 (2000).

401. Cho, S. G. et al. TIP49b, a regulator of activating transcription factor 2 response to stress and DNA damage. Mol Cell Biol 21, 8398-413 (2001).

402. Shen, X., Mizuguchi, G., Hamiche, A. & Wu, C. A chromatin remodelling complex involved in transcription and DNA processing. Nature 406, 541-4 (2000).

403. Jonsson, Z. O., Jha, S., Wohlschlegel, J. A. & Dutta, A. Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. Mol Cell 16, 465-77 (2004).

404. Ikura, T. et al. Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell 102, 463-73 (2000).

405. Frank, S. R. et al. MYC recruits the TIP60 histone acetyltransferase complex to chromatin. EMBO Rep 4, 575-80 (2003).

406. DuBridge, R. B. et al. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol Cell Biol 7, 379-87 (1987).

407. Trojan, J. et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. Gastroenterology 122, 211-9 (2002).

408. Puig, O. et al. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods 24, 218-29 (2001).

409. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551-67 (1999).

8   ACKNOWLEDGEMENTS

I would like to especially thank Prof. Joe Jiricny for convincing me to do my PhD thesis in his group. I want to thank him for his scientific supervision and great support, stimulating discussions and the right ideas at the right moment. I highly appreciate his constant effort to maintain the IMCR a wonderful place to work.

I would like to express my gratitude to all those people who made this thesis possible. Giancarlo who was always there when I needed his help with the cells. Kai for his great help and guidance especially at the end of my thesis. Dennis for fruitful discussions and his collaboration on the AID project. Katja, Elda and Bertran for their help with the MS analysis. Jacob for his support with the RT-PCR. Franziska who introduced me into the world of *in vitro* mismatch repair and Myri for excellent technical assistance.

Furthermore, I would like to thank my other former and present labmates for creating a fantastic atmosphere that made my life at the IMCR so enjoyable. It was not just a place to work but also a place to make friends. Special thanks to Anne for many joyful moments, Daniela, Esther and Elisa with whom I spent great lunch and coffee breaks and we will continue to do so. And to Nina for her valuable input regarding my future career.

Very special thanks to my family and my friends. To my wonderful parents who continuously support me in every phase of my life and who made the realization of my studies and this thesis possible. To my great sisters Stefanie and Melanie, although they are sometimes far away, they are always here when I need them.
Last but not least, thank you very much to Florence, Bettina, Cathy and Freddy who make me feel at home in Zurich.

9   CURRICULUM VITAE

Silvia Schanz
Kurvenstrasse 38
CH-8006 Zürich
Tel.:   +41 43 810 08 29
Mobil: +41 76 321 34 78
Email: schanz@imcr.uzh.ch

Nationality: German
Marital status: Single
Date of birth: 03.04.1978
Place of birth: Reutlingen, Germany

EDUCATION

01/05 – 08/08   **Federal Institute of Technology Zurich (ETHZ), Switzerland**
PhD thesis at the Institute of Molecular Cancer Research, University of Zurich
• Title: Involvement of MMR in AID-mediated antibody diversification

10/06 – 02/07   **University of Zurich, Switzerland,** Swiss Banking Institute
Certificate Program in Finance: „eCF Basic Corporate Finance"

10/01 – 10/04   **Albert-Ludwigs-University Freiburg i. Br., Germany**
Major: Biochemistry; Minors: Genetics, Immunology, Economics
Diploma thesis at the Institute of Molecular Medicine and Cell Research
• Title: Investigation of the acetylation of TCF4 by p300
Diploma, grade: 1,0

01/01 – 07/01   **Ecole Normale Supérieure (ENS) Lyon, France**
External studies; Major: Neurobiology

10/98 – 09/00   **Albert-Ludwigs-University Freiburg i. Br., Germany**
Undergraduate Courses

09/88 – 07/97   **Gymnasium Gammertingen, Germany**
Matura, grade: 1,1

WORK EXPERIENCE

05/05 – 10/07   **Krebsliga Zurich, Switzerland**
• Lecture assistant; writer of medical summaries for publication on the Krebsliga Zurich homepage

05/03 – 06/03   **Institute for Molecular Medicine, Freiburg i. Br., Germany**
• Academic assistant in the field of Wnt-signaling

12/02 – 03/03   **Max-Planck-Institute for Immunobiology, Freiburg i. Br.**
• Academic assistant in the field of immunological signal transduction

07/02 – 10/02   **Genentech, Inc., South San Francisco, CA, USA**
• Internship in the "Department of Protein Engineering"

07/01 – 09/01	**Genentech, Inc., South San Francisco, CA, USA**
	• Internship in the "Department of Assay and Automation Technology"
03/01 – 05/01	**Institut National de la Santé et de la Recherche Médicale (INSERM U433), Lyon, France**
	• Internship in the department of "Neurobiologie expérimentale et physiopathologique", Faculty of Medicine Laennec
10/00 – 12/00	**Roche, Basel, Switzerland**
	• Internship in the department „Pharma Research Basel Neuroscience (PRBN)"
10/97 – 12/97	**University Hospital, Freiburg i. Br., Germany**
	• Internship in the „Division of cardiology"

PUBLICATIONS AND CONFERENCES

Publications
- Schanz S, Castor D, Fischer F, Jiricny J: „Interference of mismatch repair in the processing of U/G mispairs arising through cytosine deamination." Manuscript submitted, 2008.
- Stephan JP, Schanz S, Wong A, Schow P, Wong WL: "Development of a frozen cell array as high-throughput approach for cell-based analysis." American Journal of Pathology 2002, 161:787-797.

Conferences
- "Maintenance of genome stability." Puerto Vallarta (Mexico), March 2008; Poster presentation: "The role of mismatch repair in SHM and CSR."
- "2nd EU-US DNA repair meeting on endogenous stress, base excision repair and related processes." Erice, Sicily (Italy), November 2005; Poster presentation: "Identifying new interaction partners of AID."

SKILLS AND HOBBIES

Languages
- German (mother tongue)
- English (very good)
- French (basic knowledge)

IT
- Operating systems: MS Windows, Mac OS
- MS Office: Word, Excel, Power Point
- Adobe: Acrobat, Photo Shop

Sports
- Sailing, Skiing, Biking

Hobbies
- Traveling, Piano