1. Introduction

Microbial infection presents a major threat to the survival of higher organisms. In response, animals have evolved a variety of strategies to combat infections. The innate immune system employs a diversity of proteins and receptors that recognize structures shared by many micro-organisms (e.g. lipopolysaccharides). However, this strategy invites evasion through mutations. It is presumably this limitation of innate recognition that has driven the evolution of adaptive immunity. The humoral arm of the adaptive immune system is mediated by antibodies, which are produced by B cells. The adaptive immune system generates a specific antibody against any possible antigen and thus provides antibodies of great diversity.

The diversity of antibodies is generated in two steps: first, before antigen contact by random and imprecise joining of gene segments, encoding the antigen-binding part of the antibody. Second, upon encounter with an antigen, B cells can undergo three kinds of genetic alterations in their immunoglobulin (Ig) loci: somatic hypermutation (SHM) and gene conversion, respectively, in the antigen binding part; and class switch recombination (CSR) in the non-antigen binding part of the antibody. All three phenomena with obviously different molecular mechanisms require a common enzyme, activation-induced cytidine deaminase (AID), although its detailed role in these processes largely remains to be worked out. In the following paper, the characteristics of antibodies and the mechanisms underlying antibody diversity are described in more detail.

Structure of Antibodies

Antibodies are expressed on the surface of B cells or are secreted by plasma cells. All antibodies are Igs and consist of two identical heavy (H) chains and two identical light (L) chains (Fig. 1).

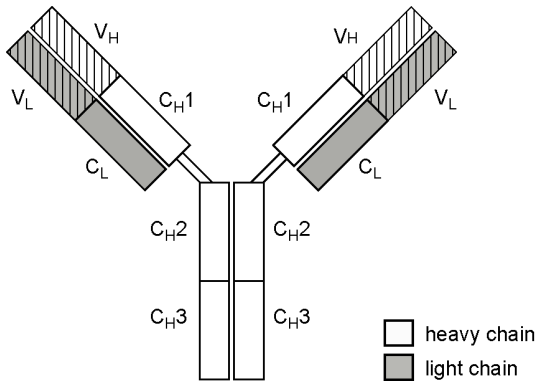

Fig. 1. Scheme of a secreted antibody (IgG isotype) with H (white) and L (grey) chains. The variable domains (V) are hatched. V domain of the H chain are labeled V_H; of the L chain V_L. C domains of the H chain are labeled C_H; of the L chain, C_L. The C_H1 is connected to the C_H2 via a hinge region.

Igs have two functions that are executed by two separate regions: the variable (V) regions of the L and H chains bind antigens, and the constant (C) region of the H chain binds effector molecules (Fig. 1).

Diversity of antibodies

The antibody repertoire of an individual human or mouse contains millions of different antibodies. Because there are no more than 30,000 genes in the human (Lander et al., 2001; Venter et al., 2001) or mouse (Waterston et al., 2002) genome, the question arises, of how the immune system can generate such a vast diversity of molecules. The answer includes a series of genetic alterations during lymphocyte differentiation. Part of antibody diversity is generated before interaction with an antigen (primary repertoire), the other part after antigen contact (secondary repertoire).

Primary repertoire

The primary repertoire is generated through gene rearrangement at the Ig loci during early lymphoid development in the bone marrow (Tonegawa, 1983). The gene rearrangement is catalyzed by the recombination-activating gene (RAG) recombinase (Gellert, 1992; Schatz et al., 1989). There are two gene segments encoding the V region of the L chain: a variable (V_L) and a joining (J_L) gene segment. Recombined, they form a complete V_L exon. Three gene segments, however, encode the V_H region: the V_H, the diversity (D_H), and the J_H segments, which together form a complete V_H exon (Fig. 2). RAG mediates the joining of the immunoglobulin V_H, D_H, and J_H gene segments, or V_L, J_L gene segments, to yield functional exons; they encode the V regions of the H and L chains of an antibody, respectively (Fig. 2).

Because the V, D, and J segments exist in multiple copies, and because the choice of segments is essentially random, a large diversity is possible already in the primary repertoire (Hozumi and Tonegawa, 1976). Moreover, the joining of V, D, and J gene segments is somewhat imprecise, and nucleotides may be inserted at the junction in a non-templated manner, generating even more diversity. This also contributes to the structural diversity of the third complementarity-determining region of the antibody—a region that plays a critical role in antigen recognition. In Fig. 2, VJ- and VDJ rearrangement at the κ L chain locus and μ H chain locus, respectively, are shown. But in mice there is also a λ L chain isotype, and there are eight H chain isotypes.

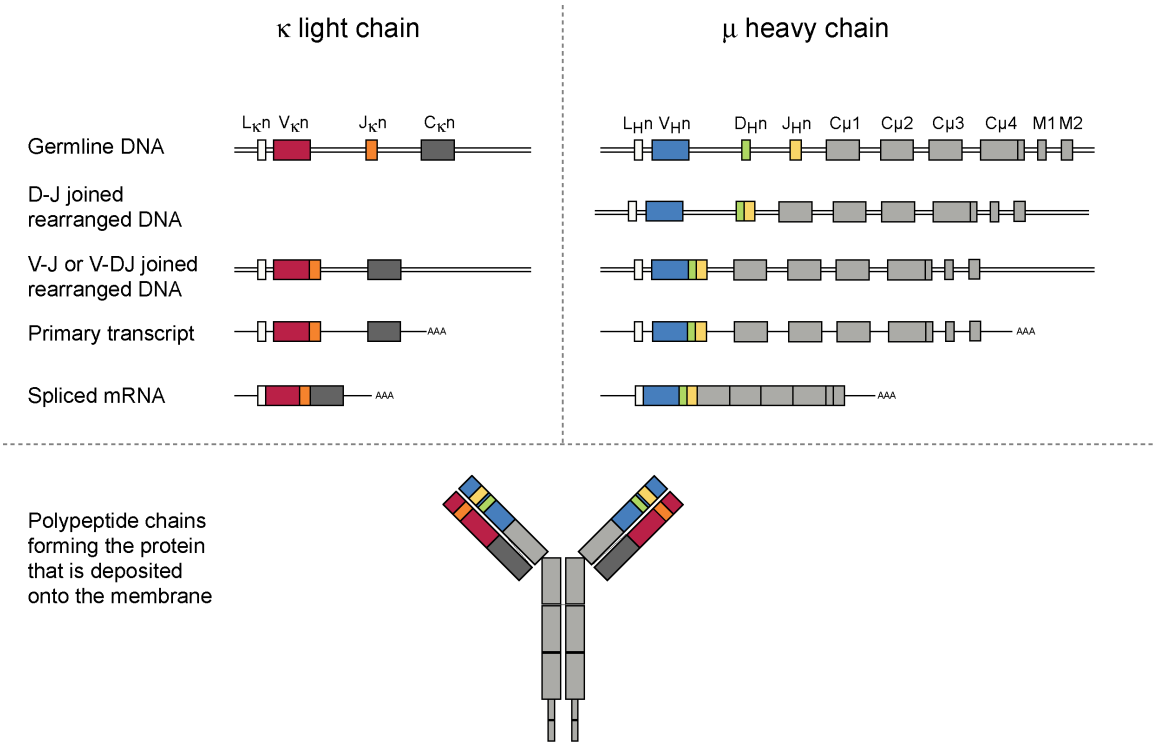

Fig. 2. Schematic presentation of the gene segment rearrangement exemplified for κ L chain (left) and μ H chain (right). Figure is not to scale. L, exon encoding the leader peptide; V, variable region; J, joining region; D, diversity region; C, constant region, n, multiple copies. For the κ L chain rearrangement, V and J segments are joined at the DNA level. The exons encoding the κ L chain C region (C_L), as well as the leader sequence, are spliced to the rearranged segments at the RNA level. Rearrangement of the H chain initiates with joining of the D-J segments. Then V is joined to the DJ rearrangement. The C and leader regions are spliced to the rearranged VDJ exon at the RNA level. The leader peptide directs the protein into the cell's secretory pathways and is cleaved off during movement into the lumen of the endoplasmic reticulum.

IgM antibodies of the primary repertoire bind with low affinity to the antigen. However, the low affinity is in part offset by the higher avidity of the IgM antibodies, as the secreted IgM is combined into pentamers or hexamers. Thus the ten or twelve identical antigen combining sites counterbalance low affinity to some extent.

Secondary repertoire

A secondary repertoire provides high affinity antibodies and is created after contact with an antigen. In mice and humans, point mutations are introduced into the V region by SHM—a non-templated mutational process in which single nucleotide substitutions are introduced in and around the rearranged V gene segments (Neuberger and Milstein, 1995).

The V region diversification process may be accompanied by a switch in the isotype class from IgM to IgG, IgA and IgE. The class switch serves to distribute a particular V region to a different C region, with each C region mediating a specialized effector

function. Secreted Ig classes other than IgM have only two (IgG, IgA, IgE) or four (IgA) instead of ten or twelve combining sites of the IgM. It is this high number of combining sites that confer high avidity upon the IgM molecule, which usually has low affinity combining sites. The loss of high avidity in molecules with a low number of combining sites is counterbalanced for the most part by affinity maturation via SHM. But CSR and SHM are two independent events, and neither of them is a prerequisite of the other (Rudikoff et al., 1984; Siekevitz et al., 1987).

In chickens and rabbits, gene conversion—a templated mutational process—plays a major role in V gene diversification. This process involves a family of V pseudogenes that is located upstream of the productively rearranged V gene. These V pseudogenes serve as templates for the gene conversion process. (Weill and Reynaud, 1992).

B cell development

B cells are derived from bone marrow, where they undergo three developmental stages: pro-B cell, pre-B cell, and immature B cell stage. In the early and late pro-B cell stage, the cell joins a D to a J segment, and then V to the DJ segment (Fig. 2). Together with the μ constant region segment the VDJ exon forms the μ gene. The μ chain in combination with the surrogate light chain forms the pre-B cell receptor, which is expressed on the cell surface. The next developmental stage is the immature B cell, during which the cell rearranges its L chain gene segments and expresses complete surface IgM. B cells that have acquired surface IgM and IgD are able to respond to antigens and are designated as mature B cells. As predicted by the clonal selection hypothesis, each B cell produces an antibody of a single specificity.

B cell activation

In order to be activated, the mature B cell must encounter the cognate antigen and, in some cases, a T helper (Th) cell with the appropriate antigen specificity. The chance of a random encounter between a B and a Th cell that recognize the same antigen is low. To nevertheless achieve such an encounter, mature B cells and naïve Th cells travel continuously via the blood stream to peripheral lymphoid organs (spleen and lymph nodes) and leave via the efferent lymph. Those cells that bind to antigens will be trapped at the T cell/B cell border in the peripheral lymphoid organs. Upon recognition of the same antigen from B cell and Th cell, the B cell receives signals from the Th cell, and the cells enter the cell cycle and migrate into a primary

lymphoid follicle. Here they continue to proliferate and ultimately form a germinal center, which is mainly composed of proliferating B cells, but antigen specific Th cells make up about 10% (Camacho et al., 1998). In the germinal centers, B cells undergo SHM, which alters the V region of the Ig, leading to affinity maturation and isotype switching, which allows selected B cells to express a variety of effector functions (Berek et al., 1991; Jacob et al., 1991; Liu et al., 1996; MacLennan, 1994; Pascual et al., 1994). Some of the B cells then differentiate into plasma cells, whereas others become memory cells.

Immunoglobulin Class Switch Recombination (CSR)

The class of the antibody is defined by the C region of the H chain (isotype). In mice, there are eight different Ig classes: IgM, IgD, IgG3, IgG1, IgG2b, IgG2a, IgE, and IgA. The corresponding H chains are designated by lower-case Greek letters: μ, δ, γ1, γ3, γ2a, γ2b, ε, and α (Fig. 3). The gene cluster encoding the different H chains spans 200 kb and is located 3' of the rearranged VDJ exon, with gene segments encoding μ chain in closest proximity to the VDJ (Fig. 3).

B cells start out with expression of IgM type immunoglobulin. In mature B cells, transcription is initiated at the V promoter and extends through both Cμ and Cδ exons. This long primary transcript is then processed by splicing and polyadenylation; in this way mature B cells express both IgM and IgD (Blattner and Tucker, 1984; Goding et al., 1977) (Fig. 3) .

Upon activation, the B cell can switch from IgM to other Ig classes, which recruit different effector functions. For example, pathogens coated with IgG3 or IgG1 are recognized by macrophages and neutrophils, which can thereby bind and engulf pathogens. IgE on the other hand is recognized by mast cells, baso- and eosinophils, which release inflammatory mediators. The Ig classes have different structures: IgG has three constant domains and a hinge region (Fig. 1), whereas IgM has no hinge but an extra constant domain (Fig. 2). IgM and IgA can form multimers and thereby increase the number of antigen binding sites (Hendrickson et al., 1995; Niles et al., 1995).

During the class switch the V_H region of an antibody is joined to a C_H region different from Cμ or Cδ (Honjo and Kataoka, 1978). This is mediated at the DNA level by CSR. The gene segments that encode the various C_H regions are consecutively encoded in a linkage group (Shimizu et al., 1982). Each C_H is flanked upstream by a switch (S) region, composed of tandem repetitive sequences with many palindromes

(Kinoshita and Honjo, 2001). CSR takes place between two S regions via a loop-out and deletion of the intervening DNA segments as circular DNA (Iwasato et al., 1990; Matsuoka et al., 1990; von Schwedler et al., 1990) (Fig. 3).

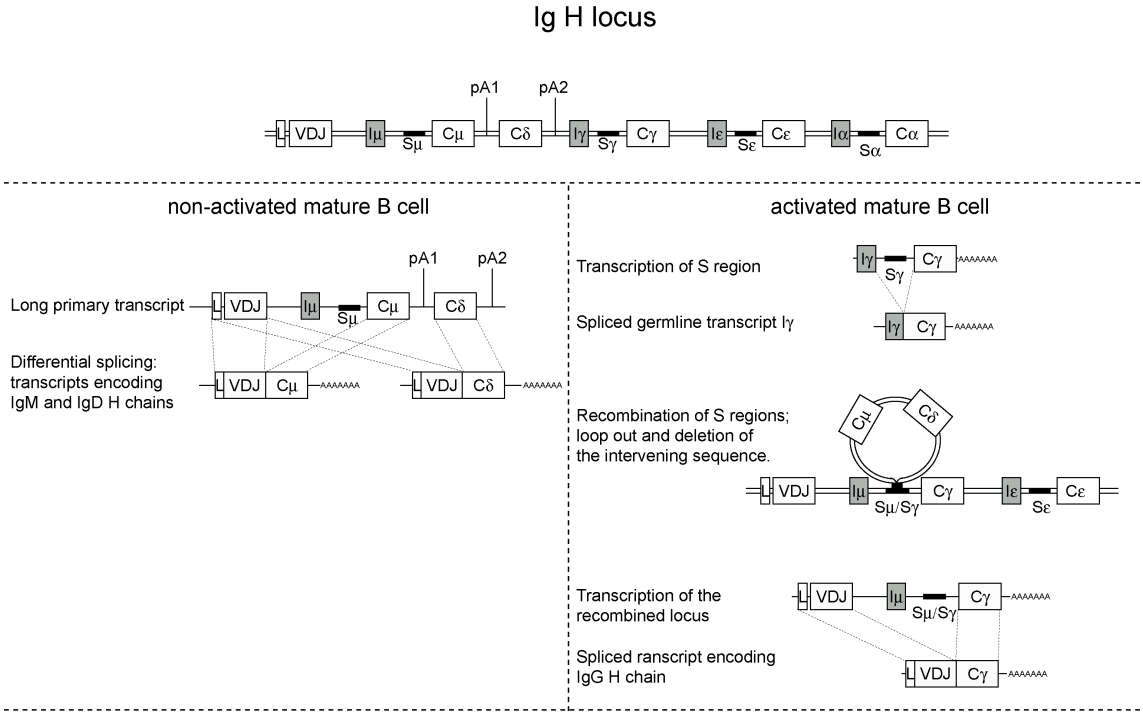

Fig. 3. Gene locus encoding C regions of the H chains (Cx), not to scale. For simplicity, not all individual C-region exons are shown. The locus for the exons encoding the four Cγ regions was simplified and named Cγ. Respective S regions (black) and I exons (grey); L, leader sequence. Left: Co-expression of IgM and IgD in mature B cells. Transcription initiated at the V_H promotor extends through both Cμ and Cδ exons. This long primary transcript is processed by cleavage, polyadenylation and splicing. pA1 and pA2 indicate polyadenylation signal for Cμ and Cδ, respectively. Right: Upon activation, B cells may undergo CSR. Shown is the switch from IgM to IgG. CSR is preceded by transcription of the S region and splicing of the transcript. For the recombination of two S regions (in this case Sμ und Sγ) DNA is looped out and deleted. The recombined locus encoding IgG H chain is transcribed and spliced.

CSR is preceded by transcription of the two S regions involved in recombination, starting from the I promoter located 5' to each S region. This non-coding transcript contains the I exon, the S region containing intron and the C region, and is spliced so that the I exon is joined to the C region (Gerondakis, 1990; Lutzker and Alt, 1988; Stavnezer-Nordgren and Sirlin, 1986). Such germline transcripts are harbingers of CSR (Stavnezer et al., 1988), and their splicing is necessary in order for CSR to occur (Jung et al., 1993; Lorenz et al., 1995). Before splicing, the newly synthesized RNA is thought to hybridize with the S region DNA template strand and to form R-loops (Reaban and Griffin, 1990; Reaban et al., 1994; Shinkura et al., 2003) or after splicing, transcription bubbles (reviewed in (Wang and Wabl, 2004)). Both formations

result in exposure of single strand DNA that can exceed 1 kb in length (Yu et al., 2003), and thus increase accessibility of the locus for factors mediating CSR.

Cytokines provided by Th cells and accessory cells in germinal centers regulate isotype specificity of CSR (Durandy et al., 2003; Snapper et al., 1997). Cytokine-driven isotype switching directly correlates with the ability of the given cytokine to selectively induce germline transcripts from a specific H chain locus (Fayette et al., 1997) and thus CSR to this Ig isotype (reviewed in (Kenter, 2003; Manis et al., 2002b)). Many groups dedicated their work to study how different lymphokines engage the switch recombinase system (reviewed in (Stavnezer, 2000)). Different antigen-presenting cells handle different types of pathogens. This is why, in the end, the pathogen determines the isotype of the antibody. However, the question remains as to how the lymphokines accomplish this. Obviously, as they bind to the appropriate receptor on the surface of the cell that is about to switch its Ig isotype, a signaling cascade is initiated that ultimately leads to transcription of the I region promoter (reviewed in (Stavnezer, 2000)). It is not clear whether this, in combination with the AID activity (see below), is all that is needed to set off the switch process. In other words, S region-specific factors may be needed (Kinoshita et al., 1998). S regions differ (Nikaido et al., 1982), and they differ more from one S region to another than between the same S region of different species (Mills et al., 1990; Sun and Butler, 1997). This may indicate other requirements (Ma et al., 2002; Shanmugam et al., 2000), in addition to transcription and subsequent splicing. Various factors that bind to S regions have been reported; however, none of these factors are expressed specifically in switching B cells. Some of these factors are well characterized in processes other than switch recombination (reviewed in (Wang and Wabl, 2004)).

Somatic hypermutation (SHM)

SHM contributes to antibody diversity through introduction of point mutations into the variable (V) region of immunoglobulin sequences (Weigert et al., 1970). In mice and humans, SHM occurs at rates of 10^{-5} to 10^{-3} mutations per base pair per generation, which is ~1 million–fold higher than the spontaneous rate of mutation in most other genes (Rajewsky et al., 1987; Wabl et al., 1999). The rate has been estimated based on the sequencing of large numbers of somatically mutated Ig genes that were compared to their germline gene segment counterparts (Gearhart and Bogenhagen, 1983; Gearhart et al., 1981; McKean et al., 1984; Weigert et al., 1970); and by fluctuation tests in hypermutating cell lines (Green et al., 1995; Wabl et al., 1985;

Wabl et al., 1987). The mutations are mainly single base substitutions, with occasional insertions and deletions. These mutations are targeted to the V region, beginning ~150 to 200 bp downstream of the promoter, and extend ~1.5 kb further downstream, ending before the intronic enhancer Eμ and mostly sparing the C region (reviewed in (Rada and Milstein, 2001). Although mutations occur throughout the rearranged V regions and its immediate flanking sequences, there is preferential targeting to C embedded in WRCY motifs (W = A or T, R = A or G, C, Y = T or C). Some potential hot spots are targets and others are not, suggesting that other local sequences or higher-order structures may also influence the targeting of mutations (Kinoshita and Honjo, 2001; Rogozin and Kolchanov, 1992). In addition, transition mutations (R→R or Y→Y) arise more frequently than transversions (R→Y or Y→R). The targeting of somatic hypermutation to the Ig loci remains a subject of debate, since genomic sites other than the Ig loci have been reported to support hypermutation to various degrees (Gordon et al., 2003; Shen et al., 1998; Wang et al., 2004a; Woo et al., 2003).

Gene conversion

In humans and mice, functional V, D and J gene segments are present in large clusters in the germline, and VDJ recombination generates combinatorial diversity by random assortment of the individual segments. Further modifications of the antibody occur after antigen stimulation via CSR and SHM. However, in species other than mouse and human, VDJ recombination may lead to the assembly and expression of a single functional gene. This was first demonstrated in chickens, where only single functional V, (D) and J segments are present in the L and H chain gene loci. In this case, diversity is introduced into the rearranged VDJ segments by gene conversion, using pseudo V genes as donors. Besides birds, rabbits and perhaps cattle, swine, and horses use gene conversion for B cell repertoire formation (Butler, 1998).

In chickens, a cluster of pseudo V genes is located upstream of the functional V gene segment to both the L and H chain loci. The L chain locus has 25 pseudo V genes (Reynaud et al., 1987), whereas the H chain locus has around 80 pseudogenes, which are similar to V-D joints (Reynaud et al., 1989). All pseudo V genes lack promoter, leader exon, and VDJ recombination signal sequences; some contain stop codons. During B cell development in the bursa, segments of pseudogene sequences appear in the rearranged V gene segments, but the pseudogene donor sequence does not change (Carlson et al., 1990; Reynaud et al., 1987). Conversion tracts range from 8 to 200 bp (McCormack et al., 1989). The 5' ends of gene

conversion always begin in regions of homology between pseudogene donor and recipient V segment, whereas the 3' end can occur in regions of non-homology and often has nucleotide insertions or deletions.

Activation-induced cytidine deaminase (AID)

AID is an enzyme expressed exclusively in germinal center B cells (Muramatsu et al., 1999); it is essential for somatic SHM and CSR in mice (Muramatsu et al., 2000) and human (Revy et al., 2000). Furthermore, disruption of AID in the chicken cell line DT40 completely blocks Ig gene conversion (Arakawa et al., 2002). These findings indicate that AID is a master enzyme for antibody diversification in vertebrate germinal center B cells. Although in mice two mRNAs encoding AID were described in the original publication (Muramatsu et al., 1999), only one canonical protein of 24 kD molecular mass is known. In humans, three splice forms were reported (Albesiano et al., 2003; McCarthy et al., 2003; Noguchi et al., 2001).

In fibroblasts, exogenous expression of AID is necessary and sufficient for both SHM (Yoshikawa et al., 2002) and CSR (Okazaki et al., 2002) of exogenous substrates. Presumably because the processes are strictly transcription dependent (Bachl et al., 2001; Stavnezer, 2000; Storb et al., 1998), there is no hypermutation, or switching at the endogenous loci in fibroblasts. However, introduction of AID alone into hybridomas may result in hypermutation of the active allele in hybridoma cells, which normally do not hypermutate their Ig genes (Martin et al., 2002). In a cell-free system, AID needs to be phosphorylated in order to retain specificity to the target sequence (Chaudhuri et al., 2004).

AID has some homology to apolipoprotein (apo) B mRNA editing catalytic polypeptide 1 (APOBEC-1), and therefore may also be an RNA editing enzyme (reviewed in (Honjo et al., 2004)). APOBEC-1 deaminates mRNA that encodes the cholesterol carrier protein in low-density lipoprotein (LDL) at a deoxycytidine (C) in position 6666 to deoxyuracil (U). By this means APOBEC-1 creates a premature stop codon (Powell et al., 1987; Teng et al., 1993), and thus generates apoB48, a carrier of triglyceride (Navaratnam et al., 1995). Both APOBEC-1 and AID form dimers (Lau et al., 1994; Ta et al., 2003) and are nucleocytoplasmic shuttling proteins (Chester et al., 2003; Ito et al., 2004) containing an N-terminal nuclear localization signal (NLS) and a C-terminal nuclear export signal (NES). GFP-tagged AID is predominantly located in the cytoplasm, which raises the question of how cytoplasmic AID reaches its substrate in the nucleus (Ito et al., 2004; Rada et al., 2002a).

While no RNA editing function has been documented for AID, it clearly can work directly on DNA (Bransteitter et al., 2003; Chaudhuri et al., 2003; Petersen-Mahrt et al., 2002). Specifically, AID converts C to U in DNA by cytidine deamination preferentially when the C is embedded in a WRCY motif (W = A or T, R = A or G, C, Y = T or C). If U is neither edited nor repaired, it will pair with A during DNA replication, which in turn will pair with T. Thus the result is a nucleotide substitution of C by T. In an alternative outcome, the non-canonical DNA base U may be removed by the uracil N-glycosylase (UNG) (Di Noia and Neuberger, 2002; Rada et al., 2002b), and translesion DNA synthesis of error-prone (mismatch) repair introduces a mutation (reviewed in (Neuberger et al., 2003) (Fig. 4).

Although AID mediates SHM, CSR and gene conversion, the three functions can be dissociated in the two separate functional domains by appropriate AID mutants. Mutants with changes in the C-terminal region of AID retain SHM and gene conversion but lose class switch activity (Barreto et al., 2003; Shinkura et al., 2004; Ta et al., 2003). Similarly, in hyper-IgM syndrome patients with mutations in the C-terminal region, class switch recombination is impaired but hypermutation seems to be unaffected (Ta et al., 2003) (Imai et al., 2005). Because in the B cells of a subgroup of these patients the (presumably switch-recombination-induced) double-strand DNA breaks in the Sμ regions were still present, it is thought that the C-terminal region may interact with the DNA repair machinery (Imai et al., 2005).

AID in SHM

SHM is triggered by and dependent on the expression of AID in activated B cells (Minegishi et al., 2000; Muramatsu et al., 2000; Muramatsu et al., 1999; Revy et al., 2000) after they have bound to a cognate antigen and formed a germinal center. Ectopic expression of AID in hybridomas (Martin et al., 2002), non-B cells (Martin and Scharff, 2002b; Yoshikawa et al., 2002), and even E. coli (Petersen-Mahrt et al., 2002) results in a mutator phenotype. In conjunction with RPA (Chaudhuri et al., 2004), AID acts on ssDNA during transcription and deaminates C into U (Chaudhuri et al., 2003; Dickerson et al., 2003; Pham et al., 2003). If the mismatch is not repaired, it results in a C to T transition, which is the predominant mutation during ectopic AID expression in cell lines and E. coli. After introduction of the G·U mismatch, various pathways fix the mutation in the genome (reviewed in (Li et al., 2004)). For example, the uracil N-glycosylase (UNG) removes the U generated by AID (Di Noia and Neuberger, 2002). If the resulting abasic site would be converted by an apyrimidinic endonuclease (AP-endonuclease) into a single-strand nick and

subsequently be repaired via the base excision pathway, no mutation would be generated. But if the abasic site is not being removed, any base could be introduced during replication (Neuberger et al., 2003) (Fig. 4). Alternatively, mismatch repair (MMR) may sense G·U mispairing and set off a sequence of events that introduces further mutations rather than restore the original base pair sequence (Cascalho et al., 1998) (reviewed in (Lee et al., 2004)).

AID in CSR

In CSR, AID is required to initiate DNA cleavage (Petersen et al., 2001). When the non-canonical DNA base U is removed by UNG (Di Noia and Neuberger, 2002; Rada et al., 2002b), thereby creating an abasic site on one DNA strand, this site may then be targeted by an AP-endonuclease, which creates single-strand DNA breaks. Because S regions have many repeats of AGCT—a palindromic RGYW (or WRCY on the complementary strand) motif—there will be many mutations (Reina-San-Martin et al., 2003), and, as a consequence, many nicks will be introduced into the S region DNA. Moreover, the G·U mismatch may be dealt with by mismatch repair, which removes longer stretches of single-strand DNA around the mismatch by an exonuclease (Bardwell et al., 2004); if this stretch happens to have a single nick on the other strand, a double-strand break would also result. In fact, components of the mismatch repair system are involved in the switch process (Ehrenstein and Neuberger, 1999; Ehrenstein et al., 2001; Martin et al., 2003; Schrader et al., 2002). Without them, there is less switching, and the recombination breakpoints are more likely to occur in consensus motifs.

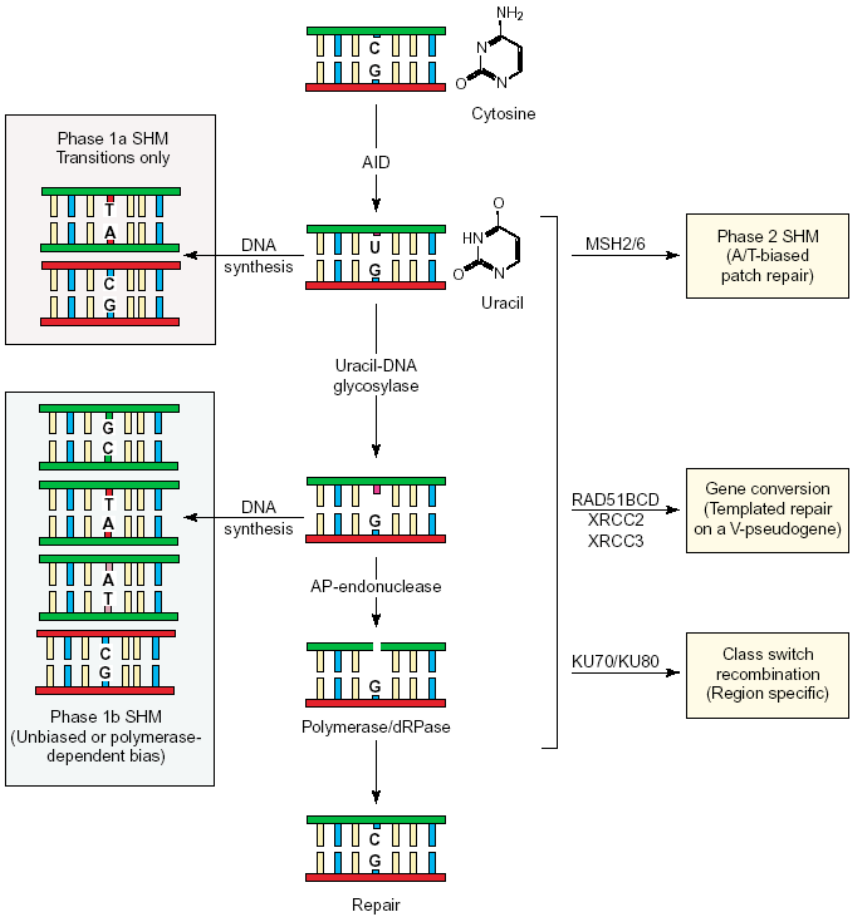

Fig. 4. DNA deamination model for antibody diversification (Neuberger et al., 2003). The figure illustrates how different mechanisms for resolving the initiating U·G lesion can lead to various patterns of antibody diversification. Phase 1A: SHM with AID triggering C to U deamination within the IgV domain. If DNA synthesis occurs opposite this U residue, a C to T (and a G to A) mutation will result. Phase 1B: if this U residue is subjected to base excision by a uracil-DNA glycosylase before DNA synthesis, then C to A, G or T (and G to A, C or T) mutations can all result from DNA synthesis opposite the apyrimidinic site. Phase 2: alternatively, recognition of the U·G mismatch by components of the mismatch repair pathway (probably involving MSH2/MSH6, possibly together with other factors) is proposed to trigger a short-patch mutagenic DNA repair synthesis leading to mutations at A and T that also occur during SHM. Thus, with regard to C to U deamination within the IgV domain in mouse or human, the initiating U·G lesion can be processed by one of a variety of distinct pathways (yielding Phase 1A, 1B or Phase 2 SHM or, alternatively, repair). In the case of chicken or rabbit, IgV gene conversion results when the initiating U·G lesion (or an intermediate in its repair) is resolved by a recombinational process templated on one of the proximal IgV pseudogenes. This recombination is probably dependent on the RAD51 paralogs XRCC2, XRCC3 and RAD51B. CSR is thought to be triggered when AID-mediated deamination is targeted to C residues in the μ switch region, with the resolution involving a partner S region from the downstream Ig isotype. The major pathway of switch recombinations might occur by a form of non-homologous end-joining involving Ku70 or Ku80. Abbreviations: AP-endonuclease, apyrimidinic endonuclease, dRPase, 5'-deoxyribophosphodiesterase.

Phenotypes of AID deficient organisms

AID deficient organisms are immune-deficient, because their B cells fail to switch the Ig isotype and to hypermutate the V region of the Ig. Concomitant with the lack of AID, for unknown reasons, immature B cells accumulate in abnormal germinal centers and cause enlargements of lymph nodes and spleen.

It is not surprising that AID deficient mice are relatively healthy under specified pathogen free (SPF) conditions, but it is not clear why they are resistant to infections with a virulent strain of influenza virus (Harada et al., 2003). However, AID deficient mice are more susceptible to secondary infections with higher doses. Thus, apparently non-mutated IgM protects against low-dose virus, but affinity-matured Igs are important for protection from infection with higher-dose virus.

Similar to AID deficient mice, Hyper-IgM-syndrome II (HIGM2) patients suffer from recurrent infections, which may cause hypertrophy of lymph nodes and enlarged germinal centers (Revy et al., 2000). AID deficient mice also have enlarged germinal centers (Muramatsu et al., 2000) and accumulation of activated IgM^+ B cells and IgM plasma cells in all lymphoid tissues (Fagarasan et al., 2002). Accumulation of IgM^+ B cells and plasma cells is explained by blockade of in situ CSR and sustained activation of the immune system.

Other factors required for CSR and SHM

For both SHM and CSR, repair enzymes are important. Pms2 (Cascalho et al., 1998), MSH2/6 (Phung et al., 1998; Rada et al., 1998), Mlh1 (Kim et al., 1999), and UNG (Di Noia and Neuberger, 2002) modify the incipient mutations introduced by AID, thereby extending the spectrum of the mutations generated; in CSR they may increase the chance of introducing a double-stranded break into the DNA (Ehrenstein and Neuberger, 1999; Schrader et al., 2002)—a prerequisite for the recombination to occur. A deficiency in UNG in humans (Imai et al., 2003) and mice (Rada et al., 2002b) is associated with impaired H chain class switch and perturbed hypermutation. For CSR, both Ku70 (Manis et al., 1998) and Ku80 (Casellas et al., 1998) are required. It is envisaged that one or more Ku70/Ku80 proteins bind to Sμ and to another S region and bring them into close proximity, either with (Rolink et al., 1996) or without (Bosma et al., 2002; Manis et al., 2002a) the help of DNA-PKcs, a DNA dependent phosphokinase that is recruited by Ku70/80 in conventional (non-homologous) double strand repair (Smith and Jackson, 1999). Furthermore, B cells

lacking H2AX show impaired CSR (Celeste et al., 2002; Petersen et al., 2001); H2AX, however, is not needed for SHM (Reina-San-Martin et al., 2003).

Supt6h – a potential candidate mediating CSR and SHM

Suppressor of Ty (SPT) genes were originally identified through a genetic screen for mutations in the yeast, which restore gene expression that was previously disrupted by the insertion of the transposon Ty (Winston, 1992; Winston and Carlson, 1992). Two out of 16 members of the SPT family, SPT5 and SPT6, have a highly negatively charged N-terminal region, indicating that they comprise part of a nuclear protein complex that is important for establishment and maintenance of chromatin structure (Swanson et al., 1990; Swanson and Winston, 1992; Winston, 1992). The homologues of SPT6 in humans and mice are denoted as SUPT6H and Supt6h, respectively (Chiang et al., 1996). The murine and human proteins are encoded by a 7-kb transcript; the human transcript encodes a 1603 amino acid ORF, and the murine transcript encodes a 1726 amino acid ORF (Chiang et al., 1996). Within the defined ORF, the sequence homology at the DNA level is 90% between human and murine Spt6 homologues (Chiang et al., 1996). Both have a calculated molecular mass of ~200 kD. Also, both have features that are consistent with a nuclear protein that regulates transcription: an S1-motif that binds oligonucleotides—most likely single strand RNA or single strand DNA—and an SH2 domain, which functions as a regulatory module of intracellular signaling cascades.

Supt6h interacts with high affinity with phospho-tyrosine-containing target peptides, whereby the SH2 domain recognizes 3 to 6 amino acid residues C-terminal to the phosphorylated tyrosine. The protein may regulate multiple steps of mRNA synthesis through its interaction with histones, elongating RNA polymerase II and possibly other components of the transcription machinery (Endoh et al., 2004).

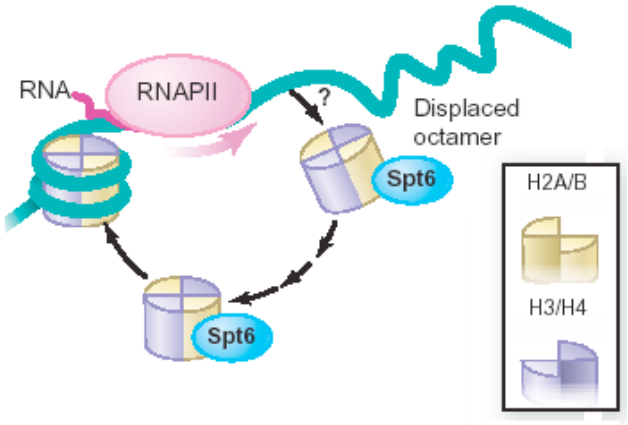

Fig. 5. Model of Supt6h function from (Svejstrup, 2003).

In a current model, with its histone chaperone activity, Supt6h displaces histone proteins in front of RNA polymerase II and reassembles them with DNA immediately behind it (Fig. 5). Because RNA polymerase II activity greatly increases AID activity, and Supt6h facilitates RNA polymerase II transcription in eukaryotes, there may be a functional interaction of AID and Supt6h. We have found previously that Supt6h binds AID in yeast two-hybrid screens, and verified the binding by co-immunoprecipitation.

TID as a candidate

TID is a member of the DnaJ family. Members of this family contain a conserved J domain, which serves as a binding domain for Hsp70 chaperones and orchestrates the interaction with specific substrates (Cyr et al., 1992; Kelley, 1999; Laufen et al., 1999). The DnaJ/Hsp70 complexes are involved in protein folding, assembling of multimeric protein complexes, and translocation of proteins across membranes (Hartl, 1996; Meacham et al., 1999; Medema et al., 1992). In mice three alternative splice variants of Tid-1 were found (Trentin et al., 2001), two of which correspond to the human TID-1_L and TID-1_S forms: the 43 kDa TID-1_L and 40 kDa TID-1_S (Syken et al., 1999). TID-1_L expression increases apoptosis, whereas a mutant J domain suppressed apoptosis (Syken et al., 1999). By contrast, TID-1_S expression suppresses apoptosis, whereas a mutant J domain increases apoptosis (Syken et al., 1999). Mouse Tid-1 binds to a GTPase-activating protein (GAP), and the GAP•Tid-1•HSP70 complex might play a role in the integration of mitogenic signaling pathways at the plasma membrane, which in turn may control apoptotic signal transduction at mitochondrial membranes (Trentin et al., 2001).

During this thesis work, we isolated a mutator mutant cell line that apparently gained the ability to hypermutate due to retroviral integration. The integrated virus either switched on a gene necessary for hypermutation, or disrupted a gene encoding an inhibitor of the process. Surprisingly, not only was the viral integration close to the gene encoding Tid-1, Tid-1 was also found as a potential interaction partner of AID in a yeast-two-hybrid screen (R. Harper, Wabl lab, unpublished data). These two findings prompted us to further investigate Tid-1 as a candidate for function in CSR and SHM in mouse B cells.

The WEHI-231 cell line

WEHI-231 is a mouse B lymphoma cell line of BALB/c x NZB F_1 origin, induced by mineral oil injection (Gutman et al., 1981). Lanier and Warner deposited this line into the ATCC in 1982. Since then, it has been studied and kept in culture in many laboratories, and thus one can expect a variety of cell lines due to the accumulation of variants. We obtained the line, designated FS, from Anthony DeFranco (UCSF). The line was subcloned in 1990, and one subclone with low surface IgM expression was designated as HM.

Although WEHI-231 has been studied in great detail and has revealed a wealth of information, the Ig loci of the cell line, and their potential to hypermutate or switch the Ig class, have not yet been characterized. Cells of this line display IgM on their surface, and in vitro growth seems to be dependent on continuous signaling through the antigen receptor (Jakway et al., 1986; Page and DeFranco, 1988). Thus, even after chemical mutagenesis, mutants without surface receptors could not be isolated. Because of its propensity to apoptose upon crosslinking, and because of the absence of IgD on its surface, WEHI-231 has been classified as a tumor of immature B lymphocytes (Boyd and Schrader, 1981). Nevertheless, in the course of our studies we found that AID is highly expressed in some sublines (and subclones) of WEHI-231. Based on this finding, in combination with the fact that the Ig alleles can be distinguished, we decided to use WEHI-231 as a model to study AID mediated mechanisms, i.e., SHM, CSR, and recombination between homologues.