4. Discussion

AID – key to the secondary antibody repertoire

AID expression in WEHI-231

AID is the key enzyme for processes that give rise to a secondary antibody repertoire, namely, CSR, SHM, and gene conversion. We have found that the WEHI-231 successor line FS and its subclone HM express canonical AID (Spillmann and Wabl, 2004). Anthony DeFranco (UCSF) provided us with the WEHI-231 line designated as FS that has been in culture for an undefined time. AID expression in WEHI-231 seems contradictory, since AID is a marker for activated B cells (Muramatsu et al., 1999) and WEHI-231 is classified as an immature B-cell line: it has IgM but no IgD (Tisch et al., 1988) on its surface, and its cells apoptose upon reaction with antibody to IgM (Gottschalk and Quintans, 1995). Immature B cells are thought not to express AID or to hypermutate or switch their Ig genes. However, IgD has been reported on WEHI-231 (Haggerty et al., 1993). Indeed, we found IgD expression in the original WEHI-231 lines. The expression of IgD, and the expression of an 18-kD protein specific for non-activated B cells (see below) would indicate that the original WEHI-231 line is a mature B-cell line. However, we did not find IgD-positive FS or HM cells. In addition, FS and HM express canonical AID, which may identify them as an activated B cell line.

By quantitative PCR we detected 3.5-fold more AID transcripts in the FS line than in the switching and hypermutating 18-81 line or in activated spleen cells. This finding concurs with the results obtained from AID Western blots (Fig. 7). If AID is normalized to the actin loading control, it is obvious that there is a higher expression of AID protein in FS and HM than in 18-81. However, although there is 250-fold less AID transcript in LPS activated spleen cells compared to FS (Fig. 9), the amount of AID protein in the activated spleen cells is much higher than in FS (Fig. 8B, 10). This result suggests that AID expression may be controlled on translational level.

We concluded that AID is functional in FS and HM, because full-length AID mRNA did not reveal any mutations. Expression of non-mutated AID is also important for the localization of AID, as AID is a nucleocytoplasmic shuttling protein with NLS and NES in its N and C termini, respectively (Ito et al., 2004). Although we have not examined localization of AID in WEHI-231, because currently available reagents cannot visualize endogenous AID protein by immunostaining, the fact that AID is not mutated is a good indication that nucleocytoplasmic shuttling of AID potentially works in FS.

The 18-kD band

Since FS and HM had been in culture for a long time and may have differentiated, we wished to compare their AID expression to that of the original WEHI-231 line. We obtained lines from ATCC, Chris Paige (CP) and Paul Kincade (PK) that had been in culture for only 2 to 3 generations and therefore are closer to the original WEHI-231 line. Comparing all WEHI-231 lines, we found that the lines from PK, CP and ATCC express none or very little of the canonical 24-kD AID protein. Thus, FS and HM clones must have gained AID expression during culture. The WEHI-231 precursor lines, however, seem to express an 18-kD AID protein; this difference in proteins may be reflected in normal resting and activated B cells. In fact, in AID Western blots, the 18-kD band was expressed in non-stimulated spleen cells. We suspect that the 18-kD protein expressed in the precursor lines and in non-stimulated spleen cells is an immature form of AID. The 18-kD band may be encoded by an AID splice variant (see below) or by a transcript containing a premature stop codon. The latter may be edited in such a way that the stop codon is reverted and the canonical 24-kD AID is expressed. We further hypothesize that the 24-kD AID itself edits the immature transcript. However, the fact that transcripts with premature stop codons are targeted by nonsense-mediated mRNA degradation (Jack et al., 1989) contradicts this theory, because we did not detect diminished expression levels of the 18-kD band in comparison to the 24-kD band.

In the most recent AID Western blot, we found that the 18-kD band is expressed in non-stimulated B cells (B220+) from wild-type (WT) and AID knockout (KO) mice. Yet, expression of the 18-kD band in non-stimulated KO spleen cells is not confirmed, because of the lack of AID KO mice and therefore AID KO cells. The presence of the 18-kD band in non-stimulated AID KO cells would identify this band as non-AID but specific for non-stimulated B cells, because after stimulation with LPS, WT and KO cells do not express the 18-kD band. Three days after LPS treatment, B cells from WT mice express the canonical 24-kD AID band, whereas B cells from AID KO mice express neither the 18- nor the 24-kD band.

Another reason for the expression of the 18-kD band in AID KO mice could be leakage of the knockout. The AID KO mice were generated by replacing AID exon 2 and part of exon 3 (containing the cytidine deaminase motif) with a neo cassette and the flanking vector sequences (Muramatsu et al., 2000). When we performed RT-PCR with primers in the neo cassette and exon 5, we found four different neo-AID splice forms. Still, if there were leakage of the KO, we should see the 18-kD band in LPS-stimulated KO mice as well. Whether the 18-kD band is encoded by an AID

splice variant or by a transcript specific for non-stimulated B cells, the goal of our present work is to identify this band.

AID splice variants

Six AID splice variants but only one AID protein of 24 kD have been reported, including all splice variants for humans: retention of intron 4 (Albesiano et al., 2003; McCarthy et al., 2003), omission of exon 4 (Albesiano et al., 2003; McCarthy et al., 2003; Noguchi et al., 2001), a 10-amino-acid (30 bp) deletion in exon 4 (Albesiano et al., 2003; Noguchi et al., 2001; Oppezzo et al., 2003), a 51-amino-acid deletion that includes exons 4 and 5 (Oppezzo et al., 2003), and a variant containing introns 4 and 5 (Albesiano et al., 2003). We performed PCR with primers just 5' and 3' of the coding region of AID in the five WEHI-231 lines and found the canonical AID mRNA in all of them. This is no surprise, because, as shown by conventional and quantitative PCR, the expression level of AID in the PK, CP and ATCC lines is minor. In addition to the canonical AID, we found two of the human splice variants: a 30-bp deletion at the 5' end of exon 4 in CP and 18-81 and a complete deletion of exon 4 in 18-81.

In order to find other splice variants, we carried out RT-PCR on AID with 5' primers in AID exons 1-4 and a constant 3' primer in exon 5 of the non-coding region. With a primer pair in AID exons 3 and 5, we found only a small amount of amplicon in the 18-81 line, although the primer pair 1-5 amplifies the canonical fragment perfectly well. Because we know that exon 3 and the sequence covered by the 3' primer are present, this is quite puzzling, especially because the cytidine deaminase motif is located in exon 3 (Muramatsu et al., 2000; Muramatsu et al., 1999). We know from amplification in PK that the exon 3-5 primer pair is prone to artifacts, because sequencing confirmed this amplification product (Fig. 12) to be due to non-specific priming of the '3-5' primer pair. However, the primers amplify AID sequences perfectly well in FS and HM, and we repeatedly get little amplification product for 18-81. Most recently, we have tried to discover AID splice forms by RACE.

Other factors required for CSR and SHM

Besides AID, we verified the expression of other factors known to be important for CSR and SHM in FS and HM. We focused on UNG, Ku70, Ku80 and transcription of the CSR targeted loci. UNG is the enzyme that excises U from DNA after deamination from C to U by AID. In UNG-deficient humans (Imai et al., 2003) and mice (Rada et al., 2002b), CSR is impaired and SHM is perturbed. When we

analyzed expression levels and the sequence from UNG in FS and HM of the nuclear UNG2, we found normal expression of UNG.

Ku70 and Ku80 are both required for switch recombination (Casellas et al., 1998; Manis et al., 1998), because they form heterodimers and bind to single- and double-stranded DNA ends. One hypothesis is that Ku70/80 bind to the switched S regions in order to bring them into close proximity (reviewed in (Wang and Wabl, 2004)). However, we have examined the expression and sequences of both factors, and they did not reveal any abnormalities.

For CSR and SHM to occur, the targeted loci have to be transcribed. One way to measure transcription of the loci is to verify the expression of germline transcripts. Indeed, we found germline transcripts expressed in FS and HM. From sequencing UNG, Ku70, Ku80 and germline transcripts, we concluded that CSR and SHM should be possible in FS and HM.

Of course, other factors are known to be involved in SHM and CSR. These factors are not as potent as the ones we tested for and therefore not as essential. However, it might be worthwhile to check on these factors in future experiments. For instance, deficiencies in mismatch repair enzymes such as MSH2, MSH6, Mlh1, and Pms2 result not only in reduced SHM but also in a shift of mutation spectra. In humans and mice, the MSH2–MSH6 heterodimer recognizes single or very short base-pair mismatches (Buermeyer et al., 1999). In mice or cells deficient in MSH2 or MSH6, the frequency of SHM is decreased (reviewed in (Martin and Scharff, 2002a)) and the majority of mutations are G to A or C to T transitions in WRCY motifs. In MSH2-deficient mice, a 2- to 10-fold reduction in CSR was reported (Ehrenstein and Neuberger, 1999; Schrader et al., 1999). The phenotype of MLH1- and PMS2-deficient mice is less dramatic and appears to give rise to increased mutations at G·C pairs (Cascalho et al., 1998; Martin and Scharff, 2002a) and a 2- to 4-fold reduction in switching (Ehrenstein et al., 2001; Schrader et al., 1999).

Non-homologous end joining (NHEJ) plays an essential role during V(D)J recombination and in the repair of double-strand DNA breaks. NHEJ enzyme H2AX-deficient mice showed impaired CSR activity, whereas they are normal for SHM (Reina-San-Martin et al., 2003), suggesting that ligation of two double-strand-break ends is not a prominent process during SHM. For ligation of the double-strand break in CSR, the Ku70/80-recruited, conventional DNA-Pkcs may (Rolink et al., 1996) or may not (Bosma et al., 2002; Manis et al., 2002a) be required.

Early explanations of the generation of antibody diversity invoked error-prone polymerases (Brenner and Milstein, 1966), which were later suggested to be involved in SHM (Goodman and Tippin, 2000). This has been confirmed by the finding that patients with deficiencies in pol η have a decreased frequency of mutations at A and T (Zeng et al., 2001). Interference with pol ζ in mice and in a Burkitt's lymphoma cell line decreases the rate of SHM (Diaz et al., 2001; Zan et al., 2001). In addition, Burkitt's cells in which pol ι has been inactivated have a 75% reduction in SHM frequency (Faili et al., 2002). However, mice deficient in pol ι do not change their rate and spectrum of mutations (McDonald et al., 2003). We have completed a cDNA chip comparison (data not shown) with 18-81 and FS and found that pol ε is expressed at a higher level in 18-81 (M-value: 3.38); therefore, pol ε deficiency may give rise to an altered and diminished SHM pattern in FS.

In chickens, variability of Ig is mainly provided by gene conversion. In chicken B cell lines deficient in the RAD51 paralogues XCCR2 and XCCR3 (RAD51B), gene conversion is transformed into SHM (Sale et al., 2001). The observed SHM in the absence of XRCC2/XRCC3 could reflect either that a sister chromatid-dependent recombinational repair has become error prone or that it has been inhibited, thereby revealing an alternative, non-templated mechanism of break resolutions. Because in the WEHI-231 line FS recombination between the Ig alleles may in part replace the lack of SHM (see below), we consider to verify the expression of XRCC2/XRCC3 in FS.

CSR and SHM in WEHI-231

Although both the cDNA sequence and protein expression of AID appear normal in WEHI-231, and additional factors required for SHM and CSR are present and functional, the question arose whether WEHI-231 is able to perform SHM and CSR. We examined SHM and CSR in WEHI-231 in the endogenous Ig locus and in exogenous loci.

SHM in the exogenous locus

To measure mutation rates and frequency in WEHI-231 lines, we used a mutation reporter construct. The reporter construct encodes GFP with a premature termination codon TAG. No fluorescent protein is produced unless a mutation reverts the stop codon into a sense codon and cells carrying the mutated plasmid appear green. AID mediates C to U mutations, preferentially when the C is embedded in a WRCY motif. The equivalent mutation on the complementary strand is G to A in an RGYW motif. In

our reporter construct, the G of the termination codon is embedded in an RGYW motif, but deamination of C would yield another stop codon TAA (ATC to ATU on the complementary strand), and mutations due to AID activity are not detected. Intriguingly, in comparison to reporter plasmids that detect AID-induced mutations, the TAG reporter works the best (Wang and Wabl, 2005b).

The reporter construct encoding GFP is introduced by means of a retroviral vector. When the cells are infected with retrovirus, the viral genome is present as RNA. Before the viral genome is integrated into the mouse genomic DNA, the viral RNA is reverse transcribed into DNA. This is mediated by the reverse transcriptase of the virus, which has an error rate in the same order of magnitude as the hypermutation process (10^{-4} per copying event). As a result, the cells transduced with the reporter construct will express GFP on mutations caused by the reverse transcriptase. Mutated cells that are due to the reverse transcriptase appear to be lower in fluorescence intensity than the mutated cells that are due to SHM, probably because the number of mutations introduced by the viral reverse transcriptase is higher; therefore GFP loses fluorescence intensity because of accumulation of mutations. Also the mutation spectrum of the reverse transcriptase is different than that of AID, thus resulting in other amino-acids changes.

Because the viral vector is not replication competent, mutations introduced by the reverse transcriptase will occur only once. To score the mutations generated by SHM, the transduced cultures were selected 3 days after transduction against green cells. That way, reverse transcriptase-induced mutants were purged from the culture. For rate measurements this step is not strictly necessary, since the mutation frequencies are determined over time and the reverse transcriptase contributes to the frequency of mutated cells as a base level. Only the number of AID-generated mutated cells increases with time in culture, and the slope of the increase determines the rate. In general, it is important for SHM experiments to have a large number of transduced cells ($>10^5$) to avoid position effects of the integrated virus and to dilute mutated cells due to processes other than SHM.

While the 18-81 line mutates the transduced GFP construct at a rate similar to that at the endogenous locus (Wang et al., 2004a), we could not determine a rate in FS, since there was no increase in mutated cells over time. For the other WEHI-231 lines—HM, PK and CP—we also could not determine a mutation rate, because after purging GFP-positive cells due to the viral reverse transcriptase, no mutants were present in these transduced cultures.

In fact, there were almost no GFP-positive cells in PK and CP cultures to start with, and we did not expect any mutants in these cultures because they lack the canonical AID.

In FS and HM, which do synthesize canonical AID, two mutant populations are present in non-sorted cultures: one that is close to the right of the quadrant partition (predominant in FS) and another that is close to the left of the quadrant partition (predominant in HM). In the HM clone, there is another cell population defined by fluorescence intensity: in addition to the low-intensity cell population to the right of the quadrant partition, there is one with even lower fluorescence intensity to the left of the partition. This low-intensity GFP population is by far the largest low-intensity GFP cell population we have seen so far. We have sorted the intermediate population from transduced FS and HM and sequenced GFP from genomic DNA. In FS we found that TAG reverted to TTG (which encodes leucine), and to TGG (which encodes tryptophan) in HM. The amino acid changes could well be the reason for the shift in fluorescence intensity. Because both stop codon reversions have mutations untypical of AID, they are most likely due to the reverse transcriptase of the virus. In addition, these GFP-positive populations were absent when the cultures were sorted after transduction. It might be the combination of AID and the reverse transcriptase that introduces mutations, similar to APOBEC3G (Harris et al., 2002): APOBEC3G introduces C to T mutations at high frequency into the HIV first-strand cDNA in viral factor-deficient (vif-) cells (Harris et al., 2003a; Harris et al., 2003b).

From other experiments with the mutation reporter constructs, it is clear that FS does not have the hypermutation spectrum of 18-81 (Wang and Wabl, 2005b); i.e., it seems that the C to U transition cannot be modified by replacing the incipient U with another nucleotide. Because our reporter construct does not score unmodified C to U transitions, we cannot be certain whether the AID expressed in WEHI-231 performs this initial step in hypermutation. However, the G to A mutations at the endogenous locus indicate that this might be the case (Wang and Wabl, 2005b). If so, then it is interesting to note that, because UNG2 seems to be functional in FS, the downstream events are believed to be set off by the removal of deoxyuridine by UNG2 (Rada et al., 2002b) must differ between 18-81 and WEHI-231: in 18-81, the U is replaced by a G or an A; this is not the case in FS.

SHM at the endogenous locus

Since we did not find evidence for SHM with the exogenous reporter construct, we examined the endogenous locus for mutations. Because FS has been in culture for an undefined time and expresses AID at a very high level, we expected to find an accumulation of mutations in the V region. When compared to the Ensembl database, we found 13 nucleotide differences for the active allele and 2 for the silent allele (not counting the two differences at the very beginning of the sequence). We do not know to what extent these differences represent mutations or polymorphism. As mentioned above, the WEHI-231 line originated in a (BALB/c x NZB)F_1 mouse and therefore has one allele of the a allotype and one of the b allotype. The changes in two positions on the silent allele (non-productive VDJ rearrangement) are both from T to C. These transitions do not bear the signature activity of AID. One might argue that the transcription-dependent activity of AID ought to be similar at the rearranged active and silent alleles. Although, in general, the levels of mRNA encoding the silent allele are much lower because of nonsense-mediated mRNA degradation (Jack et al., 1989), transcription levels are similar to that of the active allele. But it is also possible that mRNA is involved in the hypermutation process, which may decrease the rate at the silent allele. Some of the 13 nucleotide differences between the active allele and the database may represent mutations; however, only 5 of these would be in a G·C base pair, and only 2 are embedded in the RGYW motif (Bachl et al., 1997; Rogozin and Kolchanov, 1992).

CSR endogenous locus

We confirmed the absence of CSR in FS by staining for all eight Ig classes. For all classes we found a few positive cells, which we selected by FACS and cultured. The flow cytometric analysis proved that these cells were false positive, because they appeared to be positive for IgM and negative for the Ig class for which they were originally selected. We also stained HM for Ig classes IgG2b, IgE and IgA, but none of the cells were positive for an Ig subclass other than IgM. Interestingly, when we stained with anti-IgM antibodies that distinguish between a and b allotypes, we discovered that FS was positive for IgM^b but HM and the WEHI precursor lines PK and CP were positive for IgM^a. The switching of allotypes is discussed on page 83.

Incubation with LPS makes B cells switch their Igs preferentially to IgG1 and IgE. Even after stimulation with LPS for 3 days, not a single FS cell had switched to either of these Ig classes. One reason for CSR not to occur is the absence of germline transcripts. Expression of germline transcripts precedes CSR, presumably to

increase the accessibility of the switched loci (Stavnezer-Nordgren and Sirlin, 1986). We have analyzed the expression of the germline transcripts Iγ2b, Iα and Iε, which are expressed before switching to IgG2b, IgA and IgE, respectively. All were present in FS and HM when analyzed by conventional RT-PCR. We quantified Iγ2b and Iε by Taqman, because we suspected diminished germline transcript expression. Indeed, we found 74 times fewer Iε transcripts in FS and 195 times fewer in HM in comparison to the positive control. There were 151 times fewer Iγ2b transcripts in FS and 370 times fewer in HM. Based on these results, we conclude that there may not be enough germline transcript expressed for the cells to switch. To enable the cells for CSR, we plan on overexpress germline transcripts in FS and HM. On the other hand, there have been no reports of the number of germline transcripts necessary to enable CSR. Another reason for the absence of CSR is mutated Ig loci such as large-scale deletions. We therefore sequenced the Ig loci for IgA, IgG2a and IgG2b, but the loci did not reveal any mutations.

CSR exogenous locus

The growth of all WEHI-231 lines, and thus FS, depends on the presence of signals from IgM (Jakway et al., 1986; Page and DeFranco, 1988). It is possible that cells with another isotype no longer provide this signal and therefore undergo apoptosis. To avoid the loss of IgM, we used an exogenous switch construct so that switched cells would retain IgM, and CSR would be detected by means other than the actual switched surface Ig. We have used the tetracycline-controlled CSR substrate SCI(μ,α) (Okazaki et al., 2002), in which the gene for the fusion protein hygromycin phosphotransferase-thymidine kinase is inserted between the two transcription units of the S region. SCI(μ.α) contains the Sμ and Sα regions directed by the elongation factor 1α promoter and a tetracycline-responsive promoter (pTET), respectively. Both S sequences are removed by splicing from each transcript. The coding sequences for the extracellular domain of CD8α and the sequence for the transmembrane domain of CD8α fused with GFP are separated into two transcription units. The extracellular domain of CD8α can be anchored on the cell surface only after its fusion with the transmembrane domain by recombination between the two S regions, allowing detection of switched cells by FACS.

We transfected the SCI(μ,α) in 18-18, FS and 70Z. We chose 18-81 as a positive control because it switches from endogenous μ to γ2b at a rate of 10^{-2} per cell generation. One could argue that 18-81 does not have factors that are specifically

required for a switch to IgA. However, since SCI(μ,α) was reported to work in fibroblasts expressing exogenous AID, we concluded that 18-81 would be an appropriate control. As a result, we found that the SCI(μ,α) construct does not work particularly well. Based on the finding of only a few CD8α-positive cells in 18-81 and none in FS or 70Z, we take this result as further proof of absent CSR in FS.

Allele switch

Because the WEHI-231 cell line was generated by mineral oil injection of a (BALB/c x NZB)F_1 mouse (Gutman et al., 1981), it has two alleles. From FS, which expresses IgM of the b allotype (IgM^b), a subclone (designated HM) was isolated that was positive for IgM of the a allotype (IgM^a). Further analysis revealed that the WEHI-231 precursor lines PK, CP and ATCC also express IgM^a. Thus, WEHI-231 originally expresses IgM^a allotype and FS is exceptional in its expression of IgM^b.

We found in all WEHI-231 lines—FS, HM and the precursor lines—that both H chain alleles are rearranged, generating one productive VDJ exon (V_P) and one non-productive VDJ exon (V_{NP}). However, FS has lost one $C\mu^a$ allele, whereas HM and the precursor WEHI-231 lines retain both alleles. We conclude this from sequencing the cDNA encoding for the productive V_P or the V_{NP} rearrangement linked to Cμ from FS, HM, PK and CP (Fig. 37), and from sequencing of the Cμ locus spanning from before Cμ1 to after the membrane exon M2 from FS and HM. The BALB/c and NZB alleles (a and b, respectively) differ in Cμ by four single nucleotides (SNPs) (Fig. 34). The allelic difference that translates into an amino acid residue difference is located at the 3' end of Cμ1 and is the basis for monoclonal antibodies that distinguish between the IgM^a and IgM^b. The next allelic difference consists of a missing G in the a allele, in intron 1 of Cμ. The SNPs further downstream are located in introns 2 and 3. FS has none of the a allele SNPs. Interestingly, in HM, CP and PK we found the V_P spliced to $C\mu^a$, whereas in FS the V_P is spliced to $C\mu^b$. Furthermore, we found the silent allele transcripts in HM and the precursor lines CP and PK, and we found that they are composed of V_{NP} spliced to IgM^b. For FS we could not find a cDNA encoding the silent allele. Based on these results we conclude that HM precursor line (Fig. 6) at least one cell recombined its productive VDJ rearrangement with Cμ of b by switching the alleles, and thus gave rise to the FS line. Because the HM precursor line, FS, and HM express AID, it is possible that such switching is AID mediated. Before it is worthwhile to consider whether this switch in *trans* is mediated by AID, we need to determine the frequency and rate of the phenomenon in HM, because this line expresses canonical AID and retains both alleles.

At very low frequency, any genetic event can occur in cells without necessarily being physiologically relevant. Switching between homologs was reported to occur in (non-AID-expressing) plasmacytomas (Radbruch et al., 1980) at a frequency of 10^{-7} after cells were expanded to large cell cultures. This would set the rate per cell generation even lower. Today the looping-out and deletion model is generally accepted to be the physiologically most relevant. In the AID-expressing 18-81 cell line, the rate of switching (in *cis*) from μ to γ2b is 10^{-2} per cell generation. Nevertheless, there have been other reports on switching in *trans*. For example, in a transgenic mouse, cells switched their transgenic V regions, linked to Cμ, to an endogenous Cγ (Durdik et al., 1989). In that case, however, the μ transgene did not contain a Cγ linked to the Cμ, and so the mouse selected for recombination events at an unknown intrinsic rate. But switching in *trans* was also reported in non-transgenic rabbits, in which the V and C region allotypes can easily be determined on the Ig molecules (Kingzette et al., 1998; Knight et al., 1995). Unlike rabbits, mice do not diversify their V region repertoire by hyperconversion, which is also mediated by AID. However, interallelic CSR was recently reported to contribute significantly to CSR in mouse B cells (Reynaud et al., 2005). Because AID seems to be functional in the FS line, we speculate that, in the absence of the missing (switch) factors, switching occurs in *trans* rather than in *cis*.

To measure the frequency of allele-switched cells, we used HM as a model, because this line still has both alleles and expresses the canonical AID. In a pilot experiment (data not shown), we stained HM (positive for IgM^a) for surface IgM^b. The IgM^b-positive HM cells were selected by FACS and cultured. In the flow cytometric analysis, the cells appeared to be positive for IgM^a; thus, at the time of selection, because of nonspecific binding of the antibody, they must have been IgM^b positive. We have observed a high rate of false positives in flow cytometry assays using antibodies against surface Igs. Especially for the allele switch experiments, it might be useful to have a GFP knock-in into the IgM^b locus of HM and screen for GFP-positive cells.

Another way to exclude false positives is to stain with two monoclonal antibodies to IgM^a and IgM^b at the same time. Those that are IgM^a negative and IgM^b positive will then be selected. If we find a reasonable frequency of allele-switched cells (10^{-5}), we will do rate measurements. To prove that the allele switching is AID mediated, we will reduce AID mRNA by RNAi based on a lentiviral system. The WEHI-231 lines CP, PK and ATCC are a perfect negative control for these experiments, since they do not express the canonical AID.

We already know that the switch regions in the major intron differ in FS and HM. Because of this and because of AID's function in class switching, it seems likely that the allele switch is Sμ mediated. However, the deletions in the Sμ regions might be due to independent events, and allele switching may not be mediated by AID.

Switch rearrangement is thought not to be a homologous recombination event, in which homologous sequences line up to initiate recombination. Nevertheless, such a mechanism may operate in this case, where there is near-identity over large sections of DNA: the μ alleles differ by only a few SNPs. In further experiments, we wish to test whether there is an increased frequency of homologous recombination in the WEHI-231 lines by introducing a reporter construct that scores such events. The reporter contains tandemly arranged unidirectional GFP genes with an insertion that creates an XhoI site that inactivates the upstream GFP (the downstream GFP gene lacks a promoter). Homologous recombination events are scored by flow cytometry as the number of GFP-positive cells (Limoli et al., 2005). A pilot experiment has already been executed in the chicken B cell line DT40. With a positive outcome of these experiments, we could entertain the possibility that AID might mediate such events.

IgM in HM

Reduced IgM

The FS subclone HM has diminished surface IgM. Since HM was selected and subcloned for low IgM expression, this feature is not surprising. Yet, because HM has retained both alleles, and because of the mutations we find in the Cμ region in FS, we believe that HM represents the original transformed B cells. But even compared with WEHI-231 precursor lines, HM expresses significantly lower IgM. Because surface IgM expression is essential for growth and survival of immature B cells (Boyd and Schrader, 1981; Page and DeFranco, 1988; Page and DeFranco, 1990; Page et al., 1991), we investigated the reason for diminished surface IgM. One possibility is differentiation of FS, and consequently its subclone HM, to a mature B cell, which would be in line with AID expression of these cell lines. Mature B cells are not dependent on the signaling of IgM, and therefore the amount of surface IgM would not matter. In fact, we found that the WEHI-231 precursor line CP expresses the mature B cell marker IgD, whereas the progeny line FS expresses no IgD. We concluded from this result and the expression of canonical AID and potential AID-precursors that the original WEHI-231 line is a mature B cell line, and its successor lines FS and HM are activated B cell lines. Because mature B cells are able to

undergo CSR, we stained HM for Ig subclass IgG2b, IgE and IgA. As in FS, we did not find a single switched cell in HM, although the respective germline transcripts are expressed. As in FS, the expression level of germline transcript in HM is lower than that in switching cells and may cause an inability to switch.

To exclude secretion as the reason for decreased surface IgM in HM, we ran an ELISA with media supernatant from FS and HM cells cultured with the same conditions. We did not find secreted IgM in HM cultures, but there was abundant IgM in the medium harvested from FS cultures, presumably due in part to shedding. By flow cytometry we found that decreased surface μ H chain expression in HM correlates with decreased surface κ L chain expression. By immunofluorescence and Western blotting, we found equal μ H chain expression levels in FS and HM, whereas κ L chain expression was significantly decreased in HM. Because μ chain complexes in L chain-deficient cells remain associated with the endoplasmic reticulum-resident chaperone BiP (GRP78) and are not secreted (Bornemann et al., 1995), we expected an accumulation of μ chain in HM. By immunofluorescence we detected equal amounts of μ chain in FS and HM but significantly less κ L chain in HM. Thus, μ chain may accumulate in the lumen of the endoplasmic reticulum in HM and does not reach the cell surface.

To find the reason for decreased κ L chain expression in HM, we sequenced the κ L chain promoter region from FS and HM, which did not reveal mutations. We then sequenced κ L chain from all WEHI-231 strains and discovered that HM exclusively carries three coding mutations in the Vκ region that encode two amino acid changes. Two mutations are located in the CDR1 at the second and third positions of one codon, and they change tyrosine to tryptophan. Both mutations may or may not result from AID activity, since they are A to G and T to G mutations. The other coding mutation is a C to T mutation in FR2 and changes leucine to phenylalanine. A fourth, non-coding mutation is located in Jκ1. Because the latter two mutations are C to T, with the C embedded in a WRCY motif, they may have AID signature. Using conventional RT-PCR we found that the κ mRNA is expressed at same levels in all WEHI-231 lines. Yet, we found the κ protein significantly diminished in HM. We suspect, that due to the mutations and consequently the amino acid changes in κ, this protein may be prone to degradation and, in addition, may be incapable to form functional Ig molecules with the μ chain. To prove that the mutated κ chain is the cause for diminished surface IgM, we plan to overexpress the mutated κ from HM, and non-mutated κ from one of the other WEHI-231 lines in a cell line deficient for L

chain (e.g. 18-81). We then would quantify Ig expression by flow cytometry. Taken together, this result may suggest that low IgM expression in HM is due to (hyper)mutation in the Vκ region.

IgM double band

In Western blots with lysates from FS and HM, we not only detected diminished κ L chain but also found that all WEHI-231 lines except HM have a double band for μ H chain. Because we considered the double band a reason for low IgM expression, we followed up on this result. We hypothesized that the two bands represent membrane and secreted forms of IgM and that membrane IgM is ablated in HM due to mutations in the membrane exons. Based on our sequencing of the Cμ membrane exons, we can exclude this possibility.

Igs are glycoproteins. Glycosylation of IgM may (Hickman and Kornfeld, 1978) or may not (Sibley and Wagner, 1981) be required for membrane localization or secretion of IgM in mouse cell lymphoma. Whatever the case, we executed a deglycosylation assay and found that, indeed, the lack of the upper IgM band in HM is caused by a lack of glycosylation, since after deglycosylation all WEHI-231 samples revealed the same band sizes. We further hypothesize that glycosylation in HM is inhibited because of to the disabled Ig formation due to mutated κ.

Search for unknown factors

Although the WEHI-231 successor lines FS and HM have all the prerequisites to undergo CSR and SHM, we found no evidence that these mechanisms actually take place. We believe that FS and HM either lack one or more factors required for SHM and CSR or express a CSR/SHM inhibitor. To find the cause of the failure of these lines to undergo SHM/CSR, we compared expression patterns of the switching and hypermutating line 18-81 with FS in a cDNA chip experiment. However, nothing in FS that could contribute to SHM and CSR was differentially expressed. Only two candidates appeared to be interesting—polymerase ε (M-value: 3.38) and a guanine deaminase (M-value: 2.75)—but both factors are expressed at a higher rate in 18-81.

We then decided to enable FS for CSR and SHM by retroviral insertional mutagenesis. With this method, cells are infected with retrovirus, which integrates into the cellular genome. Integration of the virus has three possible outcomes: (i) The virus integrates in an exon and thereby disrupts a gene. (ii) It integrates in an intron in close proximity (<200 kb) to a gene, thereby switching it on via the viral LTRs. (iii)

It integrates far away from a gene and has no effect. Assuming that the viral integrations are random and equally distributed throughout the genome, 3 x 10^5 cells, each infected with one integration, would cover the whole mouse genome. A virus commonly used for this purpose is AKV. Because we had very low transduction efficiency in FS with this virus, we chose to complete the retroviral insertional mutagenesis with the Moloney virus encoding the mutation reporter construct. That way we had an immanent screening system for mutated cells. We started out with 4.8 x 10^5 infected FS cells, which we selected and subcloned for GFP-positive cells by FACS. In all subclones we found GFP-negative populations. We hypothesized that the loss of GFP was due to an accumulation of mutations resulting from the virally activated mutator. To gain a closer look at the populations with diminished GFP intensity, we picked one FS clone (#11) and compared the high-intensity GFP population with the intermediate-intensity GFP population. We found equal expression levels of GFP using Taqman. When we sequenced GFP from genomic DNA obtained from the high- and intermediate-intensity populations, we found the reverted stop codon and a second mutation in the intermediate-intensity population that results in an amino acid change from proline to serine. We consider this second mutation the reason for diminished GFP intensity in FS. The question is whether the second mutation is due to hypermutation, because we strongly selected for these cells many times.

In the FS clone #11 we found the integrated virus close to the gene encoding Tid-1, a chaperone. Tid-1 had been found before in our lab in a yeast two-hybrid screen with AID as a bait. Therefore, we thought at first that the integrated virus was more proof that Tid-1 is involved in SHM. But because we found no elevated Tid-1 expression in clone 11 and a lack of hypermutation (see above), we suspect that the two Tid-1 occurrences were a coincidence. However, another way to find unknown factors for CSR and SHM is to transduce FS with a cDNA library obtained from activated B cells and then screen for mutated cells with the mutation reporter construct.

Supt6h

Although there is general agreement that AID works directly on DNA to mutate it, there is less agreement on how AID mediates switch recombination. Clearly, these two functions can be separated, in part, in the protein domains of AID (Imai et al., 2005; Shinkura et al., 2004), and it is thought that proteins binding to AID contribute to its activity. Indeed, AID is associated with chromatin at the transcribed switch regions (Nambu et al., 2003). Chromatin remodeling factors are also implied in the

transcription dependence of AID activity for SHM (Bachl et al., 2001) and CSR (Okazaki et al., 2002), and most researchers postulate factors that guide the hypermutation activity of AID to the Ig locus. But what targets AID to the transcribed region? Recently, it was shown in an in *vitro* bacterial assay that the frequency of deamination caused by AID increases proportionally, up to 80-fold, with the amount of (bacterial) RNA polymerase II. We speculate that it may be difficult to piggyback RNA polymerase with a DNA mutator and maintain function. It would be easier to load AID onto another protein that is closely associated with RNA polymerase II in the transcription process.

In the yeast two-hybrid screen using AID as bait, we isolated the chromatin regulator Supt6h from two different libraries 11 times. Supt6h 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). With its histone chaperone activity, Supt6h displaces histone proteins in front of RNA polymerase II and reassembles them with DNA immediately behind it. Because RNA polymerase II activity greatly increases AID activity, and because Supt6h facilitates RNA polymerase II transcription in eukaryotes, we hypothesize that AID binds to Supt6h. To confirm the binding of AID and Supt6h, we provided both binding partners with tags (Supt6h-Myc and AID-FLAG/HA), overexpressed them in HEK cells, and co-immunoprecipitated both proteins.

Both the Myc-tagged and HA-tagged constructs are well expressed in HEK cells, but endogenous Supt6h is not expressed at great levels in these cells. Supt6h is processed either in the HEK cells or in their lysates. The antibody to Supt6h was raised to an epitope at the C-terminal end, and the Myc tag is N-terminal. So it looks as if the protein is cleaved at, or trimmed from, the N-terminal end. In line with this are the other, smaller bands that stain with antibody to Supt6h. But the protein seems to be trimmed from the C-terminal end also, since there are lower-molecular-weight bands that strongly react with the antibody to the Myc tag. Therefore, we plan to develop our blots with antibody to the HA tag and to put the Myc and HA tags at the C-terminal end. It would also be useful to obtain a polyclonal antibody to the middle of Supt6h.

Since we wanted to get an indication of whether the exogenous AID and Supt6h products can bind to each other in HEK cells, we used antibody to FLAG to precipitate the lysates of HEK cells transduced with Supt6h-Myc, AID-FLAG, or both. The composition of the precipitates was then analyzed with antibodies to AID, the FLAG tag, Supt6h, or the Myc tag. In lysates of the non-transduced HEK cells,

antibody to Supt6h revealed endogenous Supt6h; anti-Myc tag stained endogenous Myc at 65 kD, in addition to another (non-specific) band between 100 and 150 kD; and anti-FLAG stained a non-specific band above 75 kD. As expected, the anti-AID and anti-FLAG antibodies detected the exogenous AID in both lysates and precipitates, and the anti-Supt6h and anti-Myc antibodies detected exogenous Supt6h in the lysates. But these antibodies also stained exogenous Supt6h in the anti-FLAG precipitates. The anti-Supt6h staining was barely visible, but the anti-Myc staining was quite strong.

There are various interpretations of the results of the co-immunoprecipitation experiments. A trivial explanation would be that the Myc tag mediates the binding to AID rather than to the Supt6h portion of the molecule. To exclude this possibility, the experiment has to be repeated with a random Myc-tagged protein. Alternatively, the tags could be switched: Supt6h HA tagged on either the N- or C-terminal end, in combination with FLAG-tagged AID. Changing both epitope tags should rule out the most likely artifacts caused by the tags themselves. Other interpretations are possible, but in the end it comes down to the general question of defining the functional importance of the interaction, i.e., does it target AID to (specific locations of) chromatin?

To address the functionality of Supt6h in CSR and SHM, we ran a pilot experiment in which we aimed to knock down Supt6h in the activated B-cell line 18-81 using a lentiviral, conditional shRNA construct designated pSico (Ventura et al., 2004). Upon expression of Cre, shRNA is expressed, which would silence the target gene. Based on the NCBI reference sequence NM_009297 from the Swiss Webster mouse strain, three different shRNAs were designed in the 3'UTR of Supt6h. The reason for using the 3'UTR as an shRNA target is to preserve the possibility of downstream rescuing experiments. Of all tested shRNAs, only shRNA #3 had a weak silencing effect on Supt6h. Since most of the sequence differences between mouse strains are located in the UTRs, and because we did not have access to Supt6h sequence from BALB/c (the background of 18-81), we suspect that the design of the reference sequence was the cause for the failed siRNA experiment. For future experiments, we plan to sequence BALB/c Supt6h or simply design shRNAs in the coding region.