Somatic hypermutation in the variable regions of immunoglobulin genes is required to produce high affinity antibody molecules. Somatic hypermutation results by processing G-U mismatches generated when activation-induced cytidine deaminase (AID) deaminates C to U. Mutations at C/G sites are targeted mainly at deamination sites, whereas mutations at A/T sites entail error-prone DNA gap repair. We used B-cell lysates to analyze salient features of somatic hypermutation with in vitro mutational assays. Tonsil and hypermutating Ramos B-cells convert C→U in accord with AID motif specificities, whereas HeLa cells do not. Using tonsil cell lysates to repair a G-U mismatch, A/T and G/C targeted mutations occur about equally, whereas Ramos cell lysates make fewer mutations at A/T sites (~24%) compared with G/C sites (~76%). In contrast, mutations in HeLa cell lysates occur almost exclusively at G/C sites (>95%). By recapitulating two basic features of B-cell-specific somatic hypermutation, G/C mutations targeted to AID hot spot motifs and elevated A/T mutations dependent on error-prone processing of G-U mispairs, these cell free assays provide a practical method to reconstitute error-prone mismatch repair using purified B-cell proteins.

Somatic hypermutation (SHM) and class switch recombination act together in mammalian germinal center B-cells to produce isotype switched high affinity antibody (Ab) molecules. SHM is characterized by exceptionally high mutation rates (~10^{-3}–10^{-4} per base) in the variable regions of immunoglobulin genes (1). CSR replaces the Ig heavy chain μ region by one of the downstream regions, converting an IgM Ab isotype to IgG, IgE, or IgA, acting in different locations throughout the immune system (2, 3).

SHM and CSR are initiated by B-cell-specific activation-induced cytidine deaminase (AID) (4, 5) and require that active transcription occurs in variable (V) and switch (S) regions (6, 7). AID belongs to the APOBEC family of nucleic acid-dependent cytidine deaminases that deaminate C using DNA or RNA as substrates (8, 9). Biochemical studies show that AID converts C→U on ssDNA (10–13) and on the nontranscribed strand of transcriptionally active dsDNA (13–17). AID preferentially targets WRC (W = A or T; R = purine) hot spot motifs (14, 15, 18, 19). It has been reported that Ig V- and S-regions are transcribed bidirectionally in B-cells, and it therefore seems likely that AID can produce U, i.e. U-G mismatches, on both DNA strands (20).

SHM is about equally divided between nontranscribed and transcribed strands with mutations occurring to a similar extent at C/G sites and A/T sites (21, 22). Mutations at C/G sites occur by copying U or by copying an abasic site caused by the removal of U by uracil DNA N-glycosylase (UNG). Mutations at A/T sites occur when processing of G-U mispairs by mismatch repair (MMR), and base excision repair (BER) proteins provide gapped DNA substrates for error-prone repair (23, 24). Mutations at A/T sites appear to favor 5’WA motifs (25), reflecting the specificity of errors caused by pol η (26). A majority (80%) of mutations at A/T sites are abolished in humans and mice lacking pol η (27–29).

Based on genetic data, proteins required in MMR and BER are also required in SHM and CSR. However, unlike the normal high fidelity repair pathways designed to avoid mutations, low fidelity gap-filling reactions produce mutations, primarily at A/T sites. Mice lacking MMR proteins MutSα (MSH2/MSH6) (30–34) or EXO1 (35) show about a 4-fold reduction in mutations targeted at A and T. The loss of BER in mice lacking uracil glycosylase (UNG) reduces mutations by about 10–20%. The loss of MMR and BER (msh2−/− ung−/− or msh6−/− ung−/−) eliminates mutations at A and T sites (22, 36), whereas mutations at G and C sites are retained. Mutations at A/T sites are also abolished in msh2−/− polη−/− double mutant mice (37). The genetic data in mice and cultured cells show that pol η is a key participant inmutagenic DNA repair pathways, although additional error-prone polymerases, e.g. pol ζ, θ, and Rev1, can substitute in the absence of pol η (38–42). Monoubiquitination of PCNA (PCNAub116) is necessary for pol η to function during gap filling synthesis (43, 44).

A biochemical approach to investigate error-prone DNA repair pathways, focused on SHM at A/T sites, is currently lacking. High fidelity MMR and BER repair pathways have been extensively characterized biochemically (45–48). Because
many of the proteins are shared between the high and low fidelity pathways, this is an apt time to develop strategies to reconstitute error-prone repair in vitro. An essential initial issue to address is whether or not elevated mutations at A/T sites can be demonstrated in cell-free DNA repair assays using B-cell lysates, which are absent in non-B-cell lysates. We report here a biochemical analysis in which MMR processing of a G-U mismatch by tonsil cell lysates results in significantly elevated levels of mutations at A/T sites, whereas a parallel assay with non-B HeLa cell lysates exhibits mutations solely at G/C sites. By simulating a benchmark aspect of MMR-dependent SHM, the assay offers a convenient approach to reconstitute SHM in vitro with purified B-cell proteins.

EXPERIMENTAL PROCEDURES

Materials—The M13mp2 gapped construct and Escherichia coli strains CSH50, MC1061, and NR9404 (ung−) were described previously (15). E. coli deficient in MMR and BER (ung−/mutS−) was constructed as a derivative of NR9404 by P1 transduction. Uracil glycosylase inhibitor (UGI), 1000 units/μl, was a generous gift from Drs. Dale Mosbaugh and Samuel Bennett (Oregon State University). M13mp2 mutant phage containing a mutation A→G at position −11 or G→A at position +89 of the lacZα gene were constructed by site-directed mutagenesis. Nicked heteroduplexes with a single G-U mismatch were constructed by primer-template DNA extension using a 5′-end phosphorylated oligonucleotide primer containing U and an M13 phage ssDNA template. A 33-mer oligonucleotide (5′-CAACACCAUTCCCGAGCACTAA3′) was annealed to circular ssDNA of A→G (−11) mutant phage, and a 31-mer oligonucleotide 5′-ACCGCTTTTCTCTAACACAUCAGGCGGAGCATAA3′ was annealed to ssDNA of G→A (+89) mutant phage, followed by extension with T7 DNA polymerase (unmodified form; New England Biolabs) in the presence of the four dNTP substrates (200 μM for each substrate). The nicked dsDNA constructs containing a G-U mismatch were separated using agarose gel electrophoresis (0.7%) and purified by electroelution. A control heteroduplex with a G-T mismatch was constructed as described previously (49).

Preparation and Stimulation of Human B-cells—Human tonsils were collected from patients who underwent a tonsillectomy, having obtained appropriate Institutional Review Board approval. Tonsillar B-cells were purified from tonsil mononuclear cells as described (50). Purified B-cells (2×10⁶/ml) were stimulated with 5 ng/ml of IL-4 (R & D Systems, Minneapolis, MN) and 1 μg/ml of anti-CD40 monoclonal Ab G28-5 (produced from a hybridoma cell obtained from the ATCC, Manassas, VA), for 48 h. The human B-cell line Ramos 2G6 (1×10⁶/ml) was stimulated with anti-CD40 mAb G28-5 (1 μg/ml) for 48 h. The stimulated cells were harvested and washed with phosphate-buffered saline buffer twice, and cell pellets were stored at −80°C prior to preparation of cytoplasmic and nuclear lysates.

Preparation of Cytoplasmic and Nuclear Lysates—Cytoplasmic and nuclear lysates were prepared as described previously (49). Cells (~1×10⁸) were thawed on ice and washed at 0°C in an isotonic buffer A (20 mM HEPES, pH 7.9, 5 mM KCl, 1.5 mM MgCl₂, 1 mM dithiothreitol, 250 mM sucrose) followed by a 4°C wash with hypotonic buffer B (20 mM HEPES, pH 7.9, 5 mM KCl, 1.5 mM MgCl₂, 1 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride). Cells were resuspended in two packed cell volumes of hypotonic buffer B, incubated on ice for 10 min, and then lysed using a 2-ml glass Dounce homogenizer. The nuclei were pelleted in a microcentrifuge for 20 min, and cytoplasmic lysates (100% fraction) were collected. The pelleted nuclei were resuspended in one packed cell volume of buffer C (20 mM HEPES, pH 7.9, 420 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride) and placed on a rocking platform for 30–45 min at 4°C. Nuclear lysate fractions were collected as the supernatant after centrifugation for 30 min. Cytoplasmic and nuclear fractions were dialyzed overnight in buffer D (20 mM HEPES, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol, 20% glycerol), and aliquots of the lysates were frozen in liquid nitrogen and stored at −70°C.

Measurement of Deamination Activity and Specificity of Cell Lysates—Enzyme activities and C-deamination specificities of cell lysates were measured using the following reaction conditions: a reaction volume (30 μl) contained HEPES (50 mM, pH 7.5), dithiothreitol (1 mM), MgCl₂ (10 mM), gapped DNA (500 ng), UGI (500 units), and cell lysate (5 μg). Following incubations for 30 min at 37°C, the reactions were quenched by a double extraction with phenol:chloroform:isoamyl alcohol (25:24:1). Conversions of C→U on the DNA substrate were detected as white or light blue plaques indicating C→T mutations in a lacZα target gene, after transfection into uracil glycosylase-deficient (ung−) E. coli, as described previously (15).

G-U Mismatch Repair Reactions—G-U mismatch repairs in cell lysates in vitro were performed with M13 phage heteroduplex circular dsDNA substrates containing a G-U mispair. A single strand nick required to initiate MMR (49, 51, 52) is located 15 nt at the 5′-side of U. The efficiency of G-U mismatch repair in cytoplasmic cell lysates was measured using a standard protocol for measuring MMR of normal mismatched base pairs (49). Reaction mixtures (25-μl total volume) containing 5 ng of substrate (G-U mismatch heteroduplex located at the position −11 of the lacZα gene) and cell lysates (75 μg) were incubated for 30 min at 37°C in the presence of HEPES (30 mM, pH 7.8), MgCl₂ (7 mM), ATP (4 mM), 4 dNTPs (100 μM each), creatine phosphate (40 mM), creatine phosphokinase (100 μg/ml), sodium phosphate (15 mM, pH 7.5), and UGI (1000 units). The reactions were quenched by addition of 125 μl of quenching buffer (Tris-HCl (10 mM, pH 7.5), EDTA (5 mM), SDS (0.1%), and proteinase K (200 μg/ml)), and incubation was carried out for an additional 30 min at 37°C. The DNA was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1), desalted in water, and concentrated to 10–15 μl using a Microcon YM30 centrifugal filter concentrator (Millipore). The repaired DNA (1 μl) was transfected into ung−mutS− double mutant E. coli competent cells by electroporation and plated on CSH50 host cells as described (15, 49). Plaques were scored as pure white, pure blue, or mixed (blue/white). The G-U repair efficiency (%) was calculated as 100 × (1 − the ratio of the percentage of mixed bursts obtained from lysate-treated and -untreated samples) (49).
The “error-prone” MMR repair specificity was measured using a G-U heteroduplex with TUA located opposite a TGA codon (position +89 of lacZα). The assay conditions used to measure error-prone MMR are the same as those used for the G-U heteroduplex located at position −11. The repair products were analyzed by transfection into an ung+ and mutS+ E. coli (strain MC1061). Reversion frequencies were calculated as the fraction of blue plaques to total number of plaques. To obtain the MMR reversion spectra for cell lysates, at least 50 revertants for each reaction were randomly picked, and the lacZα DNA region was sequenced.

**RESULTS**

Our main objective was to develop a model biochemical system to analyze a basic component of SHM, namely the error-prone MMR processing of G-U mismatches resulting in mutations at A/T sites. AID initiates SHM by introducing G-U mismatches in IgV DNA. Taking a first step toward reconstituting the error-prone component of MMR during SHM at A/T sites, we investigated the repair of a preformed G-U mispair by human tonsil and hypermutating Ramos B-cell lysates alongside a HeLa non-B-cell lysate. U is protected from excision by suppressing UNG activity with uracil glycosylase inhibitor (UGI). In a separate analysis, we investigated the specificity of C-deamination in tonsil, Ramos, and HeLa cell lysates.

**C→U Deamination Specificity in Human B-cell and Non-B-cell Lysates** — The specificity of C deaminations was determined in lysates from tonsil and Ramos B-cells compared with HeLa cells. Closed circular M13 phage DNA containing a lacZα reporter gene located in a single-stranded (ss) gap (Fig. 1) was incubated with cell lysates in the presence of a large excess of UGI, to protect against removal of U by the UNG-containing lysates (53, 54) (see “Experimental Procedures”). C→U deaminations on gapped ssDNA are detected as C→T mutations in mutant M13 phage (white and light blue plaques) following transfection into uracil glycosylase-deficient ung−E. coli (15). Following incubation (30 min) of the DNA with each lysate, there was an increase in the lacZ mutant frequency of 20–100-fold (−1–6 × 10−3) above background mutation levels (−7 × 10−5), i.e. in the absence of lysate (Table 1). Omitting UGI reduces mutations to background (8.2 ± 3.2 × 10−4).

Deamination spectra were determined by sequencing individual DNA clones isolated from mutant (clear and light blue) M13 plaques (Fig. 2). AID is expressed in tonsil and Ramos B-cells but not in HeLa cells. With tonsil B-cell nuclear lysates, more than a third (36%) of the C→T mutations occur in WRC hot spot motifs, which is indicative of deamination by AID (Table 1 and Fig. 2a). This number is about 13% in Ramos cell lysates but is virtually absent (1%) with HeLa cell lysates (Table 1 and Fig. 2b). AID is much more abundant in the cytoplasm of B-cells compared with the nucleus, and yet the occurrence of C deamination in WRC motifs with tonsil cytoplasmic lysates is 4-fold less (−8%) compared with nuclear lysates (Table 1). This reduction in AID-like activity may stem from the strong inhibition of cytoplasmic AID by bound RNA (10). There are at least 10 human APOBEC nucleic acid deaminases (8, 55, 56), any of which would give rise to C→T mutations in the lacZ assay. Perhaps the reduced fraction of WRC deaminations in Ramos cells (13%) compared with tonsils (36%) might reflect a reduced ratio of AID to other APOBEC proteins in Ramos compared with tonsil cells. However, because AID is the only APOBEC deaminase currently known to favor WRC motifs (8, 19), and because AID is expressed solely in B-cells, the tonsil and Ramos B-cell DNA-dependent deaminations in WRC motifs are almost certainly attributable to AID. The correlation in AID expression and observed deamination at WRC motifs suggests that an analysis of hot spot motifs can be used as an indication of AID activity.

**TABLE 1**

| C deamination motif profile of human cell lysates |
|-----------------|-----------------|-----------------|-----------------|
|                 | Tonsil | Ramos, | HeLa |
| Mutation frequency (×10−3) | NE† | S100 | NE† | S100 |
| 1.7 ± 0.7 | 3.7 ± 0.9 | 6.4 ± 1.2 | 1.3 ± 0.5 |
| No. of sequenced clones | 44 | 48 | 93 | 29 |
| No. of clones with WRC mutations | 16 (36%) | 4 (8%) | 13 (14%) | 1 (1%) |
| No. of clones with CCC43 mutation | 18 (41%) | 41 (85%) | 68 (73%) | 29 (30%) |

*NE indicates nuclear lysate; S100 indicates cytoplasmic lysate.
†The background mutant frequency for the gap substrate is 6.5 ± 2 × 10−4.
a functional signature for detection of AID expression, even in crude lysates.

A large fraction of deaminations common to tonsil (nuclear and cytoplasmic fractions), Ramos, and HeLa also take place in SYC motifs (S = G or C; Y = pyrimidine) (Fig. 2), which are disfavored by AID (8, 15). CCC motifs are deamination cold spots for AID but are hot spots for APOBEC3G (A3G) (57–59). A high proportion of deaminations occur at a single CCC motif (lacZ CCC_{108}), where the fraction of C→T mutations for tonsil nuclei is 41%, tonsil cytoplasm 85%, Ramos cells 73%, and HeLa cells 30% (Table 1 and Fig. 2). We suggest that A3G is a likely source for these mutations because we find that more than 98% of the mutated clones are deaminated at CCC_{108} by purified A3G.  

Efficiency of "Normal" G-U Mismatch Repair in B- and Non-B Human Cell Lysates—A circular heteroduplex M13mp2 phage dsDNA containing a G-U mismatch (Fig. 3, schematic) is used to measure MMR efficiencies for lysates prepared from tonsil and HeLa cells. The minus phage strand contains a nick located 15 nt at the 5'-side of U. Following incubation of DNA with cell lysates in the presence of UGI and ATP, lysate-specific MMR was detected by transfecting DNA into MMR- and BER-deficient E. coli (mutS– ung–), and plating on an α-complementation host (CSH50 strain) in the presence of 5-bromo-4-chloro-3-indolyl-D-galactopyranoside (X-gal) (see “Experimental Procedures”). Plus strand repair yields “wild type” dark blue plaques (Fig. 3a, B plaques); minus strand repair yields “mutant” white plaques (Fig. 3a, W plaques); no repair yields “mixed” plaques (Fig. 3a, M plaques). The MMR efficiency is determined from the ratio of mixed plaques/total plaques (49) (see “Experimental Procedures”). The control plaque distribution in the absence of lysate is about 45% white, 27% blue, and 28% mixed (Table 2 and Fig. 3a, No cell lysate). In the presence of tonsil-derived activated B-cell lysate, and requiring ATP, there is a marked reduction in mixed plaques to <2% (Table 2 and Fig. 3b, Tonsil B-cell lysate), demonstrating efficient repair (~95%) of G-U mismatches.

A control experiment using a G-T pair shows that the lysates are proficient for normal mismatch repair (Table 2). We have measured a similar efficiency of G-U repair (~96%) in HeLa lysates (Table 2). As reported previously (49), repair occurs primarily on the nicked strand, i.e. B plaques are essentially absent (Fig. 3b, Tonsil B-cell lysate).

A control experiment using a G-T pair shows that the lysates are proficient for normal mismatch repair (Table 2).
Error-prone Mismatch Repair in Somatic Hypermutation

Similar to repair of G·T (49, 51, 52), the repair of the single G·U mispair is strongly biased toward the nicked strand (Table 2), and repair requires the presence of ATP in reaction. The repair efficiency appears to be somewhat higher for the G·U mispair (96%) compared with the normal G·T mispair (58%), which might result from repair involving G·T mismatch-specific thymine-DNA glycosylase (TDG) or single strand selective monofunctional uracil-DNA glycosylase (SMUG) (60, 61). However, because the addition of a neutralizing antibody to SMUG along with UGI has no discernible effect on the repair efficiency (data not shown), it would appear that SMUG has at most a minor role in processing G·U mispair in the cell lysates used here.

![Figure 3](image-url)  
**FIGURE 3.** An M13 plaque assay designed to measure the efficiency of MMR of a single G·U mismatch with B-cell lysates. a, M13 plaques in the absence of cell lysates. b, M13 plaques reflecting MMR of a G·U mismatch in the presence of tonsil B-cell lysates. G·U heteroduplex repair is carried out by B-cell lysates in the presence of UGI, followed by transfection into MMR and BER (ung mutS) competent cells. The G·U mismatch repair efficiency is measured as the reduction in the fraction of mixed plaques (M) relative to the total number of plaques. Arrows with W, B, or M indicate examples of pure white, pure blue, or mixed plaques, respectively.

**TABLE 2**

G·U mismatch repair in tonsil B-cell and HeLa cell lysates

| M13 phage plaques | White | Blue | Mixed | Repair efficiency % |
|--------------------|-------|------|-------|---------------------|
| No lysate          | 45 ± 5| 27 ± 4| 28 ± 2| 0                   |
| G·U B-cell lysate  | 96 ± 9| 25 ± 0| 15 ± 2| 95                  |
| G·U B-cell lysate (−ATP) | 57 ± 8| 17 ± 5| 28 ± 5| 7                   |
| G·U HeLa cell lysate | 98 ± 8| 0 ± 0| 1.1 ± 0.3| 96               |
| G·U HeLa cell lysate (−ATP) | 70 ± 5| 12 ± 3| 18 ± 3| 38                  |
| T·G no lysate      | 14 ± 3| 29 ± 4| 58 ± 4| 0                   |
| T·G HeLa cell lysate | 10 ± 1| 66 ± 5| 24 ± 3| 58                  |

**FIGURE 4.** Schematic representation of a reversion assay designed to measure error-prone MMR accuracy and specificity with cell lysates. Schematic at top shows a DNA molecule containing a G·U heteroduplex with a nonsense codon (TGA) in the plus strand of the lacZα gene and an A·T in the minus strand located opposite TGA. Repair of G·U is carried out by incubation of the heteroduplex with cell lysates in the presence of UGI. Accurate MMR of either the plus strand or the minus strand results in a phage carrying a nonsense codon in the lacZα gene that results in a white plaque phenotype. Error-prone MMR of either strand converts the nonsense codons into missense codons, resulting in a blue plaque phenotype.

**TABLE 3**

Reversion of G·U mismatched heteroduplex DNA caused by MMR in cell lysates

| No. of experiments | Mutant frequency (× 10−3) |
|--------------------|--------------------------|
| No cell lysate     | 6 | 132 | 563,388 | 2.3 ± 0.2 |
| Tonsil B-cell lysate | 6 | 139 | 342,808 | 4.1 ± 0.4 |
| Ramos cell lysate  | 3 | 38 | 87,480 | 4.3 ± 0.3 |
| HeLa cell lysate   | 10 | 282 | 582,293 | 4.8 ± 0.8 |

**Error-prone G·U Mismatch Repair in B-cell Lysates Generates Mutations at A·T Sites**—We have developed a sensitive reversion assay to measure the specificity of repairing G·U mismatches. We compare B-cell lysates, in which MMR is expected to be error-prone, to non-B-cell HeLa lysates, in which MMR is known to be accurate (49, 52). The substrate is an M13 phage heteroduplex carrying a nonsense codon (TGA) in the plus strand of the lacZα gene, and the minus strand contains a U in a TUA opposite TGA (Fig. 4). The nick is on the minus strand, 15 nt on the 5’ side of the U, similar to the substrate used to measure normal MMR (Fig. 3, schematic). The assay conditions used to measure error-prone MMR are the same as those for normal MMR, in which G·U mismatch repair is carried out by incubation of the heteroduplex with cell lysates in the presence of UGI, transfection into E. coli, and plating on CSH50 host cells (see “Experimental Procedures”). If the repair of either the plus strand or the minus strand is accurate, then the resulting phage will carry a nonsense codon in the lacZα gene, resulting in a white “W” plaque, see, e.g. Fig. 3). Inaccurate repair of either strand will convert the nonsense codon into a missense codon, resulting in a blue “B” plaque (see, e.g. Fig. 3). This assay detects mutations at a single G/C site and two adjacent A/T sites.

We measured the reversion specificity of G·U in activated human tonsil B-cells, Ramos cells, and HeLa cells. The data show that error-prone MMR is occurring in each of the cell lysates about 2–3-fold above background (−4 − 5 × 10−4 reversion frequency compared with a background level of ~2 × 10−4, see Table 3) in the absence of lysate, demonstrating the presence of inaccurate G·U MMR in B-cell and non-B-cell lysates. Although the increase in mutations is similar for B-cell and non-B-cell lysates, the mutational specificities differ sub-
The increase in mutations at A and T reflects the favor G/C over A/T mutations by about a 3:1 ratio (Table 5). Levels of mutations at A/T sites (62–64), and Ramos cells show far fewer mutations at A/T and G/C sites, with the same frequency of mutations at A/T and G/C sites (Tables 4 and 5). Tonsil cell lysates generate about equal numbers of A/T and G/C mutations, with the specificity of SHM, the frequency of mutations at A/T and G/C sites (Table 5), which concur with in vivo data (62–64). Activated Ramos B-cell lysates generate about 25% A/T and 75% G/C mutations, also in accord with in vivo cell culture studies (65–67). Yet it is puzzling why in hypermutating Burkitt lymphoma cell lines, Ramos and BL2, less than 20% of the mutations are at A/T sites (65–67), despite the presence of functional MMR and BER, including normally expressed pol η and UNG (67).

Mutation at A/T sites is uniquely achieved in activated B-cells (Tables 4 and 5). HeLa, a non-B-cell line, makes mutations almost exclusively at C/G sites (Tables 4 and 5). Our biochemical data conform with a recent in vivo study reporting mutation spectra of germinal center B-cells, naive B-cells, and non-B somatic cells of mice with transgenic lacZ genes (68). The lacZ transgene mutations in germinal center B-cells contain about equal numbers of A/T and G/C mutations, with the A/T mutations being dependent on the presence of pol η (68). Naïve B-cells and non-B-cells generate significantly less spontaneous A/T mutations (68). Although the transgene data agree with the specificity of SHM, the frequency of lacZ mutations are ~5 orders of magnitude less than SHM frequencies. Whereas AID generates substantial numbers of G/U mismatches in actively transcribed Ig V regions, it appears inactive on a lacZ transgene expressed in B-cells (68), perhaps because of much lower transcriptional activity. Although the lacZ transgene mutations are far fewer compared with V-gene mutations, the balance of A/T and G/C mutations are similar, probably because the same error-prone MMR is operating on mismatches arising spontaneously in lacZ, i.e. which are not dependent on AID. The biochemical (Tables 4 and 5) and in vivo model systems reinforce the idea that error-prone DNA repair is responsible for generating A/T mutations in activated B-cells undergoing SHM.

We used the same three cell lysates, tonsil, Ramos and HeLa, to investigate the in vitro lacZ mutational signature associated with C deamination (Fig. 2 and Table 1). Although AID is expressed solely in B-cells, there are at least 10 APOBEC nucleic acid-dependent deaminases present in eukaryotic cells (8, 9, 56). Among these, AID has a unique mutational signature substantially. The B-cell lysates, tonsil and Ramos, have significantly elevated levels of mutations at A/T sites, whereas HeLa cell mutations are observed almost exclusively at G/C sites (Tables 4 and 5).

To compare in vitro and in vivo mutational specificities, we have determined the mutation frequency per site after subtracting the background mutagenesis frequency (no lysate, Table 4). Since there are two A/T sites in the TGA codon, the frequency per site for A/T mutations was obtained by dividing by 2.

**DISCUSSION**

The formation of high affinity Ab molecules takes place in germlinal B-cells by SHM, a “hypermutation” process in which the IgV regions accumulate and transversion mutations at C/G and A/T sites in roughly equal numbers on both DNA strands (21, 22). The B-cell mutations originate during active transcribing of Ig molecules (6, 7) and are initiated by AID-catalyzed deamination of C→U giving rise to G-U mismatches base pairs throughout the V-regions (23, 24). Mutations at C/G sites occur by copying U or by copying an abasic site caused by the removal of U by UNG (23, 24). Mutations at A/T sites occur when processing of G-U mismatches by MMR and BER proteins provide gapped DNA substrates suitable for error-prone repair in involving pol η (23, 24).

The biochemical mechanisms of normal MMR and BER are well understood in bacteria and higher organisms (45–48). The biochemical steps entailed in making mutations at A/T sites in hypermutating B-cells are not understood. An important step in discerning the biochemical mechanisms of error-prone MMR, and thus to differentiate between low and high fidelity repair, would be to identify elevated mutations at A/T in B-cell lysates. In this study, we have established biochemical assays using cell-free lysates to measure the capacity of hypermutating B-cells to generate A/T mutations during MMR of G-U mismatches. As expected, MMR is present to eliminate G-U heteroduplex DNA in human tonsil activated B-cells and in a Ramos B-cell line (Fig. 3 and Table 2). Tonsil cell lysates generate about equal numbers of mutations at A/T and C/G sites (Table 5), which concur with in vivo data (62–64). Activated Ramos B-cell lysates generate about 25% A/T and 75% G/C mutations, also in accord with in vivo cell culture studies (65–67). Yet it is puzzling why in hypermutating Burkitt lymphoma cell lines, Ramos and BL2, less than 20% of the mutations are at A/T sites (65–67), despite the presence of functional MMR and BER, including normally expressed pol η and UNG (67).

Mutation at A/T sites is uniquely achieved in activated B-cells (Tables 4 and 5). HeLa, a non-B-cell line, makes mutations almost exclusively at C/G sites (Tables 4 and 5). Our biochemical data conform with a recent in vivo study reporting mutation spectra of germinal center B-cells, naive B-cells, and non-B somatic cells of mice with transgenic lacZ genes (68). The lacZ transgene mutations in germinal center B-cells contain about equal numbers of A/T and G/C mutations, with the A/T mutations being dependent on the presence of pol η (68). Naïve B-cells and non-B-cells generate significantly less spontaneous A/T mutations (68). Although the transgene data agree with the specificity of SHM, the frequency of lacZ mutations are ~5 orders of magnitude less than SHM frequencies. Whereas AID generates substantial numbers of G/U mismatches in actively transcribed Ig V regions, it appears inactive on a lacZ transgene expressed in B-cells (68), perhaps because of much lower transcriptional activity. Although the lacZ transgene mutations are far fewer compared with V-gene mutations, the balance of A/T and G/C mutations are similar, probably because the same error-prone MMR is operating on mismatches arising spontaneously in lacZ, i.e. which are not dependent on AID. The biochemical (Tables 4 and 5) and in vivo model systems reinforce the idea that error-prone DNA repair is responsible for generating A/T mutations in activated B-cells undergoing SHM.

We used the same three cell lysates, tonsil, Ramos and HeLa, to investigate the in vitro lacZ mutational signature associated with C deamination (Fig. 2 and Table 1). Although AID is expressed solely in B-cells, there are at least 10 APOBEC nucleic acid-dependent deaminases present in eukaryotic cells (8, 9, 56). Among these, AID has a unique mutational signature

**TABLE 4**

Mutations resulting from G-U error-prone MMR in cell lysates

|          | No lysate | Tonsil B-cells | Ramos | HeLa |
|----------|----------|----------------|-------|------|
| T→A      | ~4.0 (0) | 15.3 (3)       | ~4.4 (0) |
| T→G      | ~4.0 (0) | 15.3 (3)       | ~4.4 (0) |
| T→C      | ~4.0 (0) | 15.3 (3)       | ~4.4 (0) |
| A→G      | ~4.0 (0) | 15.3 (3)       | ~4.4 (0) |
| G→A      | ~4.0 (0) | 15.3 (3)       | ~4.4 (0) |

* Numbers in parentheses represent number of revertants.

**TABLE 5**

Compilation of mutations at A/T and G/C sites caused by error-prone MMR

|          | Estimated mutation frequency per site × 10^-6 |
|----------|---------------------------------------------|
|          | Tonsil B-cells | Ramos | HeLa |
| A/T mutations | 57 | 40   | 1    |
| G/C mutations | 65 | 119  | 247  |

* Mutation frequency per site for A/T and G/C mutations was calculated from Table 4 by subtracting the background mutation frequency (no lysate, Table 4). Since there are two A/T sites in the TGA codon, the frequency per site for A/T mutations was obtained by dividing by 2.
favoring WRC hot spot motifs (8, 15), whereas in others, such as A3G, 3F, 3C, Apobec1 tends to avoid deaminating WRC motifs (8, 19). Therefore, the correlation in AID expression and observed deamination at WRC motifs suggest that deamination at WRC hot spots can be used as a functional signature for detection of AID expression. The three cell types have robust deamination activities, about a thousand-fold above spontaneous C deamination background levels, but their deamination profiles in lacZα are distinct. Activated tonsil and Ramos B-cells deaminate WRC motifs to a significant extent, whereas HeLa cells fail to deaminate even a single such motif (Fig. 2 and Table 1). Instead, HeLa cells favor deamination in YC motifs (Y = pyrimidine), which is a hot spot for A3G and other APOBEC proteins (8, 19, 57–59) but a cold spot for AID (14, 15).

An examination of individual DNA clones reveals a processive pattern of deaminations, based on the presence of clones with multiple deaminations (from 2 to 8 deaminations per clone). Deaminations were measured under conditions strongly favoring single enzyme-ssDNA substrate encounters (i.e., > 95% of the clones have 0 deaminations (69)). The processivity of AID, and possibly A3G, is less in the lysates compared with purified enzymes (14, 15, 18). We speculate that single strand DNA-binding proteins present in the lysates may reduce deaminase processivity (18), whose action involves sliding and jumping along the ssDNA backbone (18, 59, 70).

Although a large majority of B-cell-expressed AID is found in the cytoplasm (71–73), the fraction of WRC mutations is significantly less in the cytoplasmic S100 fraction (8%) compared with the nuclear fraction (36%) (Table 1 and Fig. 2). The low levels of cytoplasmic AID activity may account for its inability to protect the cell against retroviral infection, even when over-expressed (74). Cytoplasmic expression of several of the other APOBEC family members results in retroviral restriction, notably A3G and A3F acting in T cells to inactivate HIV-1 via deamination of viral cDNA (75, 76). It has been suggested that post-translational modification of AID, e.g., possibly phosphorylation (77–79), may play an essential role in nuclear transport and perhaps interaction with nuclear proteins, including specific transcription factors and other DNA-binding proteins (77–79). It has been reported that phosphorylated AID is highly enriched in the chromatin fraction of the nucleus (79). We have not observed a significant effect of phosphorylation per se on the biochemical properties of AID (18). However, we have observed that amino acid substitutions at three Ser residues that can undergo phosphorylation (S38A, S41A, and S43A) exhibit altered deamination specificities compared with wild type AID (18). Furthermore, an active AID mutant S43P is found in a patient with hyper-IgM-2 syndrome (80), characterized by the accumulation of IgM caused by the inability to perform isotype switching (CSR) (5).

This is an apt time to develop an in vitro approach to investigate the biochemical basis of SHM. A considerable amount of recent effort has been devoted to looking at the biochemical properties of AID (24, 81) and pol η (23, 24, 82). A seemingly good entry point is to address error-prone DNA repair, a process in which AID deaminates C to create G-U mispairs on actively transcribed DNA, presumably on the nontranscribed strand, leading to C/G mutations targeted to the deamination site and to mutations at A/T. There are numerous questions relating to the biochemical mechanisms of error-prone DNA repair, acting downstream from one or perhaps many G-U mismatches formed by AID.

A sampling of a few of the most important general questions are as follows. (i) What are the interactions that ensure that mutations occur in actively transcribed IgV but not IgC regions (83)? (ii) A closely related question concerns the regulatory process that generates mutations within the V-gene, namely beginning ~200 bp from the transcription initiation site, reaching a maximum within the V(D)J coding exon, and decaying exponentially ~1.5 kb downstream from the end of the promoter. (iii) What determines the distribution of mutations, especially mutations at A/T sites, on each strand (21)? This could involve differential G-U repair efficiencies (84) or pol η targeting the nontranscribed strand.

A sampling of several more specific questions include the following. (i) Which endonuclease is used to nick the DNA backbone to provide an entry point for EXO1 digestion during MMR? (ii) How is pol η recruited in the presence of monoubiquinated PCNA to the repair gap? (iii) What interactions determine whether MMR (MSH2/MSH6) or BER (UNG) gains access to a U-G mispair to initiate a long or short repair gap?

Studies with mutant mice suggest that mutations at A/T sites occur in the vicinity of G-U sites (22, 36). A recent study further suggests that error-prone MMR of G-U intermediates is asymmetric and that A/T mutations may be concentrated within 30 bp of G-U sites (84). Whereas BER-generated ssDNA gaps are from 1 to 17 nt long (47, 48), typical MMR-generated gaps can often exceed several hundred nucleotides in length (45, 46). How error-prone MMR might be confined to work proximally to an AID-catalyzed C deamination is not understood.

The use of cell-free assays to investigate error-prone DNA repair is an important step toward defining the biochemical reactions involved in generating Ab diversity. We have chosen to begin looking at error-prone MMR, taking a “bottom-up” approach using cell lysates to process a preformed U-G mispair. By recapitulating two basic features of B-cell-specific SHM, G/C mutations targeted to AID hot spot motifs and elevated A/T mutations dependent on error-prone processing G-U mispairs, these cell-free assays provide a viable approach to start to reconstitute error-prone MMR with purified B-cell proteins.

Acknowledgments—We thank Drs. Dale Mosbaugh and Samuel Bennett for their generous gift of UGI protein, and we also thank Dr. Matthew D. Scharff for reading and commenting on the manuscript.

REFERENCES
1. Rajewsky, K., Forster, I., and Cumano, A. (1987) Science 238, 1088–1094
2. Stavnezer, J., Guikema, J. E., and Schrader, C. E. (2008) Annu. Rev. Immunol. 26, 261–292
3. Chaudhuri, J., and Alt, F. W. (2004) Nat. Rev. Immunol. 4, 541–552
4. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000) Cell 102, 553–563
5. Rey, P., Muto, T., Levy, Y., Geissmann, F., Pleban, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersøy, F., Kayserili, H., Uagazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A., and Durandy, A. (2000) Cell 102, 565–575
Error-prone Mismatch Repair in Somatic Hypermutation

75. Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., Neuberger, M. S., and Malim, M. H. (2003) *Cell* **113**, 803–809
76. Bishop, K. N., Holmes, R. K., Sheehy, A. M., and Malim, M. H. (2004) *Science* **305**, 645
77. Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J. P., Manis, J. P., and Alt, F. W. (2005) *Nature* **438**, 508–511
78. Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 395–400
79. McBride, K. M., Gazumyan, A., Woo, E. M., Barreto, V. M., Robbiani, D. F., Chait, B. T., and Nussenzweig, M. C. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8798–8803
80. Zhu, Y., Nonoyama, S., Morio, T., Muramatsu, M., Honjo, T., and Mizzuti, S. (2003) *J. Med. Dent. Sci.* **50**, 41–46
81. Goodman, M. F., Scharff, M. D., and Romesberg, F. E. (2007) *Adv. Immunol.* **94**, 127–155
82. Seki, M., Gearhart, P. J., and Wood, R. D. (2005) *EMBO Rep.* **6**, 1143–1148
83. Li, Z., Luo, Z., Ronai, D., Kuang, F. L., Peled, J. U., Iglesias-Ussel, M. D., and Scharff, M. D. (2007) *Adv. Exp. Med. Biol.* **596**, 93–109
84. Unniraman, S., and Schatz, D. G. (2007) *Science* **317**, 1227–1230