

available as accession numbers AE011190 and AE011191. Shotgun electropherogram traces from the chromosome and plasmids were deposited in the National Center for Biotechnology Information Trace Archive (trace IDs 107051265 to 107116771).

Genome-based analysis will provide a powerful new tool for investigation of unexpected outbreaks of infectious disease, whether these represent biological warfare attacks, emerging agents, or more familiar pathogens. Because even highly monomorphic species such as *B. anthracis* are known to exhibit phenotypic variation in infection potential, genomic comparisons can be used to investigate the genetic basis of pathogenesis. To lay the groundwork for future investigations, databases derived from genome-based surveys of natural variation in all major pathogens, not just potential biological warfare agents, should be constructed.

References and Notes

1. Centers for Disease Control, *Morb. Mortal. Wkly. Rep.* **50**, 877 (2001).
2. L. B. Price, M. Hugh-Jones, P. J. Jackson, P. Keim, *J. Bacteriol.* **181**, 2358 (1999).
3. P. Keim et al., *J. Bacteriol.* **182**, 2928 (2000).
4. Centers for Disease Control, *Morb. Mortal. Wkly. Rep.* **50**, 909 (2001).
5. The index strain was sent from Fort Detrick to Porton Down, UK, in or after 1982 and was cured of the pXO1 and pXO2 virulence plasmids (Fig. 1). Two preparations of genomic DNA were sent to TIGR. The first was prepared in a laboratory at the University of California, Berkeley in 1998 ("Porton1"). The second was prepared at Porton Down in 2001 from a bacterium grown from the original frozen culture ("Porton2"). The reported history of storage of the bacteria used to create the cultures provides no obvious reasons for the seven intrastrain chromosomal SNPs observed (Table 1).
6. Materials and methods are available as supporting material on Science Online.
7. A. L. Delcher et al., *Nucleic Acids Res.* **27**, 2369 (1999).
8. R. T. Okinaka et al., *J. Bacteriol.* **181**, 6509 (1999).
9. R. T. Okinaka et al., *J. Appl. Microbiol.* **87**, 261 (1999).
10. B. Ewing, L. Hillier, M. C. Wendt, P. Green, *Genome Res.* **8**, 175 (1998).
11. B. Ewing, P. Green, *Genome Res.* **8**, 186 (1998).
12. For a full discussion of quality value determinations, see supporting material on Science Online.
13. R. D. Fleischmann, *Emerg. Infect. Dis.* **7**, 487 (2001).
14. If a small portion of the genome has low coverage, this will have a disproportionate effect on the overall accuracy for the genome. For example, suppose that 10% of the genome has low coverage and an error rate of 1/10,000, whereas 90% of the genome has an error rate of 1/100,000,000. At a hypothetical average error rate per sequence of 1% (corresponding to a quality value of 20), 2-fold coverage will yield an error rate of 1/10,000, whereas 4-fold coverage will have an error rate of 1/100,000,000. This distribution would produce an overall sequencing accuracy of 1/100,000, despite the fact that the accuracy for 90% of the genome was essentially perfect.
15. J. Pannucci, R. T. Okinaka, R. Sabin, C. R. Kuske, *J. Bacteriol.* **184**, 134 (2002).
16. X. X. Peterson et al., in preparation.
17. This deletion occurs in a complex repeat composed of four distinct units (labeled A to D). The order of these units in the Pasteur pXO2 plasmid is ACDCBACB, followed by another copy of D about 4.3-kbp downstream. The Florida isolate plasmid lacks the 1.4-kbp region between the two extreme copies of the C unit, leading to the ACB genotype.
18. D. L. Robertson, S. H. Leppla, *Gene* **44**, 71 (1986).
19. C. B. Thorne, in *Bacillus subtilis and Other Gram-*

Positive Bacteria (American Society for Microbiology, Washington, DC, 1993), pp. 113–124.

20. A. B. Kolstø et al., *J. Bacteriol.* **184**, 621 (2002).
21. Regions surrounding selected SNPs were amplified by PCR, and the products were sequenced. VNTRs were typed by sequencing and/or sizing on acrylamide gels (3). In cases where the SNPs or VNTRs differed from the Florida strain sequence (asterisk in Table 4) two additional sequence reactions were performed in order to obtain better overall quality values and at least 3-fold coverage for each SNP tested.
22. M. C. Enright, B. G. Spratt, *Trends Microbiol.* **12**, 482 (1999).
23. The sequencing of the Florida isolate was supported by National Science Foundation Small Grant for Exploratory Research (MCB-202304 to C.M.F. and

T.D.R.). The Porton genome sequence was supported by a grant (N00149610604) from the Office of Naval Research, which was supplemented by the NIH, Department of Energy (grant DEFCO295ER6196), and funds from the United Kingdom Defence Science Technology Laboratory. S.L.S and M.P. were supported in part by NSF (grant KDI-9980088) and NIH (grant R01-LM06845).

Supporting Online Material
www.sciencemag.org/cgi/content/full/1071837/DC1
 Materials and Methods
 Table S1

13 March 2002; accepted 25 April 2002
 Published online 9 May 2002;
 10.1126/science.1071837
 Include this information when citing this paper.

AID Enzyme-Induced Hypermutation in an Actively Transcribed Gene in Fibroblasts

Kiyotsugu Yoshikawa,^{1, 2} Il-mi Okazaki,¹ Tomonori Eto,¹ Kazuo Kinoshita,¹ Masamichi Muramatsu,¹ Hitoshi Nagaoka,¹ Tasuku Honjo^{1*}

Activation-induced cytidine deaminase (AID), a putative RNA-editing enzyme, is indispensable for somatic hypermutation (SHM), class switch recombination, and gene conversion of immunoglobulin genes, which indicates a common molecular mechanism for these phenomena. Here we show that ectopic expression of AID alone can induce hypermutation in an artificial *GFP* substrate in NIH 3T3 murine fibroblast cells. The frequency of mutations was closely correlated with the level of transcription of the target gene, and the distribution of mutations in NIH 3T3 cells was similar to those of SHM in B lymphocytes. These results indicate that AID is sufficient for the generation of SHM in an actively transcribed gene in fibroblasts, as well as B cells, and that any of the required cofactors must be present in these fibroblasts.

In order to protect against a huge number of pathogens, the vertebrate immune system increases the limited antigen receptor repertoire encoded in the genome by taking advantage of somatic DNA alterations. First, V(D)J recombination assembles two or three pieces of distant germ line segments to form a variable (V) exon of antigen receptor genes during the development of T and B lymphocytes. Subsequently, immunoglobulin genes of peripheral mature B lymphocytes are further modified by three types of genetic alterations: SHM (1) and gene conversion (GC) (2) in the V gene, and class switch recombination (CSR) in the heavy-chain constant region (C_H) gene (3). The immunoglobulins' specificity and affinity for antigen are augmented by either untemplated SHM or pseudogene-templated GC in the V region gene when coupled with selection by the antigen. Dis-

tinct from antigen specificity of the receptor, the C_μ gene is replaced with one of other CH genes by CSR, thereby changing immunoglobulin isotype and effector functions. Although molecular mechanisms for the three types of DNA alterations remain to be elucidated, recent isolation of a B cell-specific gene for AID and characterization of its function have shown that all three reactions are dependent on AID (4–6). We have shown that ectopic expression of AID induces CSR in an artificial construct introduced into NIH 3T3 murine fibroblasts (7); in this study, we ask whether AID expression induces SHM in nonlymphoid cells.

In order to examine hypermutation in nonlymphoid cells, we generated an NIH 3T3 cell line transfected with the tetracycline responsive (tet-off) transactivator and the pI plasmid (8), which carries a mutant *GFP* sequence driven by the inducible tetracycline (tet) promoter (Fig. 1A). The mutant *GFP* has a premature stop codon (TAG) in an RGYW [R, purine (A/G); Y, pyrimidine (C/T); W, A/T] SHM hotspot (9). Furthermore, the pI plasmid was shown to be useful for assaying SHM in a mouse pre-B cell line (8) and a

¹Department of Medical Chemistry and Molecular Biology, Graduate School of Medicine; and ²Radioisotope Research Center, Kyoto University, Yoshida Ko-noe-cho, Sakyo-ku, Kyoto, 606-8501, Japan.

*To whom correspondence should be addressed. E-mail: honjo@mfour.med.kyoto-u.ac.jp

REPORTS

CH12F3-2 B lymphoma line (10). AID was introduced by retrovirus infection into 19 clones of NIH 3T3-pI cells (11), each containing one to four copies of pl as determined by Southern blot analysis (10). Significant numbers of GFP⁺ cells (1 to ~1.8%) were detected in AID-infected NIH 3T3-pI cells cultured for 10 days in the absence of tetracycline. However, mock infection or a loss-of-function mutant of AID (AIDm-1), which lacks most of the cytidine deaminase motif, did not generate

GFP⁺ cells at all (Fig. 1B). We confirmed that the premature stop codon (TAG) in all 27 GFP sequences invariably reverted to a tyrosine codon (TAC) in sorted GFP⁺ cells derived from NIH 3T3-pI clone 19 (11), which carry one copy of pl plasmid (10).

The frequency of GFP⁺ cells increased in a dose-dependent manner with AID-virus transfection (Fig. 1C) but not AIDm-1-virus. The frequency of GFP⁺ cells also increased with reduction of the tetracycline concentration in

culture medium (Fig. 1D). In parallel with the frequency of GFP⁺ cells, the mutations in the GFP substrate (11) increased in a dose-dependent manner with the amount of AID-virus and also with the level of transcription induction (Fig. 1, C and D, gray bars). GFP⁺ cells began to appear on day 3 after AID expression and reached the maximal level (0.92 ± 0.14%) on day 10 (Fig. 1E). Although the frequency of GFP⁺ cells gradually decreased from day 10 to 20, the mutation frequency in the GFP sequence on day 20 (6.5 × 10⁻³ mutations per base pair) was slightly higher than that on day 10 (4.5 × 10⁻³ mutations/bp), indicating that additional mutations continued to accumulate in the GFP substrate.

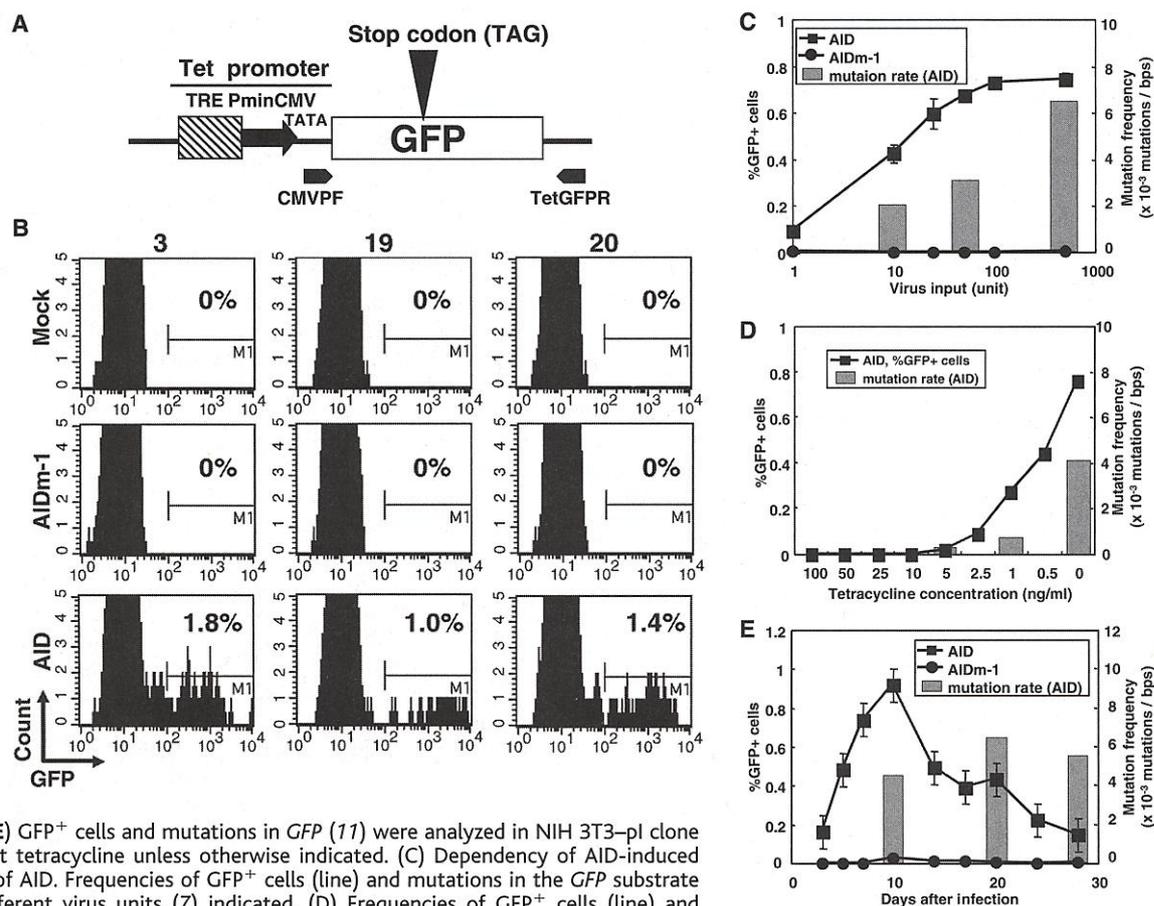
To assess the overall mutation frequency, we determined GFP sequences of 24 randomly picked clones of polymerase chain reaction (PCR) products that were amplified from bulk DNA of AID-transfected NIH 3T3-pI cells cultured for 10 days without tetracycline. Massive numbers of mutations (4.5 × 10⁻³ mutations/bp), including small numbers of deletions and duplications, were observed in 22 out of 24 GFP sequences (Table 1). Considering one division per day, we estimated the mutation frequency to be 4.5 ×

Table 1. AID-induced hypermutation in the GFP substrate in NIH 3T3-pI clone 19 containing one copy of pl. DNA was extracted from NIH 3T3-pI clone 19 cultured 10 days after infection with AID- or AIDm-1-expressing virus with or without tetracycline. The fragment (0.9 kb) containing the GFP coding sequence along with the 5' and 3' multiple cloning sites was amplified, subcloned, and then sequenced. Numbers indicate mutated bases per total bases sequenced in each category. Numbers in parentheses indicate mutated clones among total clones examined. For percentage of GFP⁺ cells, the result of triplicate experiments is shown as mean value ± SEM.

Retrovirus	Tetracycline	Percent GFP ⁺ cells	Mutated bases in GFP sequence
AID	+	0.00 ± 0.00	3/21,643 (2/23)*
	-	0.92 ± 0.14	102/22,584 (22/24)†
AIDm-1	+	0.00 ± 0.00	0/14,115 (0/15)
	-	0.03 ± 0.02	0/23,525 (0/25)

*All, three mutations were point mutations. †Of 102 mutations, 95 were point mutations, 5 were deletions (1 to 48 bp), and 2 were duplications (1 bp and 22 bp), as shown in Fig. 2A.

Fig. 1. Induction of hypermutation in the artificial GFP substrate in NIH 3T3 cells by expression of AID. (A) GFP substrate (pl) under tetracycline-inducible promoter (8). TRE, tetracycline-responsive element; PminCMV, minimal CMV promoter; TATA, TATA box; CMVFP and TetGFPR, primers for PCR amplification and sequencing. (B) Three representative NIH 3T3-pI clones 3, 19, and 20 were infected by retrovirus expressing mock, AIDm-1, or AID in the absence of tetracycline. After 10 days, the cells were analyzed for expression of GFP by fluorescence-activated cell sorting (FACS). Copy numbers of pl in clones 3, 19, and 20 are 4, 1, and 2, respectively, as determined by Southern blot analysis (10). (C to E) GFP⁺ cells and mutations in GFP (11) were analyzed in NIH 3T3-pI clone 19 cells cultured without tetracycline unless otherwise indicated. (C) Dependency of AID-induced hypermutation on dose of AID. Frequencies of GFP⁺ cells (line) and mutations in the GFP substrate (bar) on day 7 with different virus units (7) indicated. (D) Frequencies of GFP⁺ cells (line) and mutations (bar) on day 10 at different concentrations of tetracycline with 500 units of AID-expressing virus. (E) Time-course analysis of hypermutation. Frequencies of GFP⁺ cells (line) and mutations (bar) at different time points with 500 units of AID-expressing virus.



REPORTS

10^{-4} mutations/bp per generation, which is in agreement with that of immunoglobulin SHM (1×10^{-4} to $\sim 1 \times 10^{-3}$ mutations/bp per generation) (1). By contrast, only three mutations were found in 23 *GFP* sequences of AID-infected cells cultured with tetracycline, and no mutations were detected in AIDm-1-infected cells cultured with or without tetracycline. Not all transcribed genes appeared to accumulate such large numbers of mutations. For example, the *c-myc* gene, which accumulates 1/100th the mutations that V does in some diffuse large-cell B lymphomas (12), did not accumulate AID-induced mutations in NIH 3T3-p1 cells (10).

Specific features of SHM of immunoglobulin include the predominance of point mutations with occasional deletions or duplications, a preference for transition over transversion, and a targeting to the RGYW/WRCY motif (1, 9). We analyzed a pool of 247 mutations in 53

clones of the *GFP* sequence for mutation distributions (Fig. 2 and Table 2). The point mutations were statistically biased to the RGYW/WRCY motif (Fig. 2A and Table 2), and transitional mutations were predominant over transversion (Fig. 2B). Previously, we have shown that the CSR junction is distributed preferentially around computer-predicted stem-loop structures of single-stranded DNA (13, 14). Such secondary structures of DNA are also recognized around the complementarity-determining regions (CDR) in the V gene, which are preferred targets of SHM (15, 16). In the present study, AID-induced mutations were also biased to the DNA secondary structures in the *GFP* sequence (Fig. 2A and Table 2). Interestingly, AID-induced substitutions were biased to G/C base pairs: only 3 out of 247 mutations occurred at A/T base pairs (Fig. 2). Similar bias was reported in SHM of the V gene in AID-transfected hybridomas

Table 2. Mutation bias to RGYW/WRCY and secondary structure. Point mutations in the *GFP* substrate were analyzed for the association with sequence and structural motifs. Mutations associated with or not associated with RGYW/WRCY motif were counted as (+) and (-), respectively. Secondary structure was predicted by a computer program developed by Zuker (11) as described and shown in Fig. 2A (14). Both strands were analyzed for association with secondary structure and the results were combined. Statistical significance was determined by the chi-square test.

Sequences	RGYW/WRCY		Secondary structure	
	+	-	+	-
Mutated	186	61	377	117
Total	247	694	1094	788
P value	<0.001		<0.001	

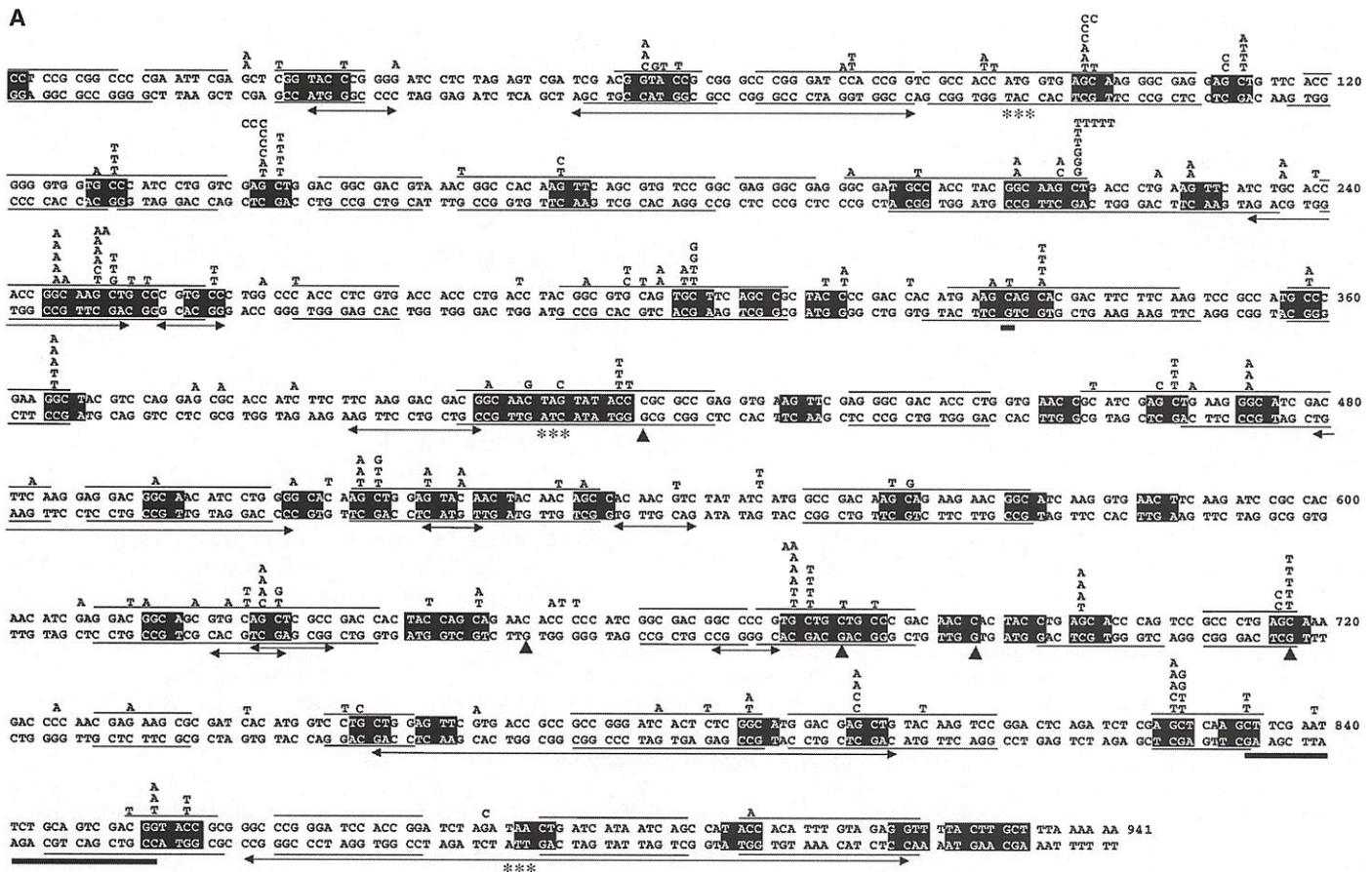
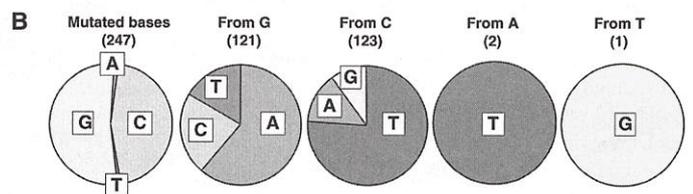


Fig. 2. Distribution of mutations on the *GFP* substrate and their properties. (A) A pool of 247 point mutations, 18 deletions (bidirectional arrow and arrowhead), and two duplications (bold underline) in 53 sequenced *GFP* clones (obtained in Fig. 1, C to E) are mapped on the sequence of the *GFP* substrate. Codons for initiation, stop mutation, and original stop are indicated by asterisks. RGYW/WRCY motifs are shown by white letters. The lines above or below the sequences indicate that these sequences form computer-predicted secondary structures in the sense or antisense strand, respectively (11). (B) Bias to G/C base pair and preference for transitional substitution. The relative mutation frequency of each base after correction for base composition in the *GFP* sense strand (A, 23%; C, 33%; G, 28%; T, 16%) are shown in the leftmost pie graph. In the four pie graphs on the right, the proportion of each substitution from four bases is shown. The number of mutations is shown in parentheses. The base substitutions are statistically biased to transition, relative to transversion (chi-square test, $P < 0.001$).



(17), a mouse pre-B cell line (18), human Burkitt lymphoma lines (19, 20), B cells of mismatch repair-deficient mice (21, 20), and an XRCC2/3-deficient chicken B-cell lymphoma line (23), although mutated bases in SHM of V genes in human and mouse B cells in vivo did not show such a strong bias to G/C base pairs (24, 25). This difference in nucleotide preference could be attributed to relative abundance of error-prone DNA polymerases (26) or specific DNA repair proteins. In addition to point mutations, 18 deletion and two duplication events were observed, consistent with the features of SHM of immunoglobulin (I) (Fig. 2A).

In summary, AID-induced hypermutation in NIH 3T3 cells has common properties with SHM of immunoglobulin genes; it shows strict dependence on AID, dependency on transcription of the target gene, and mutations biased to specific motifs. In addition, AID induced occasional deletions and duplications, along with a high mutation frequency. These properties are also shared by hypermutation in the S μ region in B cells stimulated with lipopolysaccharide (27). These results suggest that hypermutation in NIH 3T3 cells may use the same molecular machinery as that used in V and S μ sequences, and they indicate that AID is the only B cell-specific factor required to generate hypermutation experimentally in actively transcribed genes in nonlymphoid cells. Taken together with our previous observation that CSR could be induced by AID in the same cell line, it is likely that all trans-acting factors required for AID to accomplish CSR and hypermutation, are constitutively expressed in nonlymphoid cells, as well as B cells.

The AID enzyme is likely to be involved in the cleavage step of CSR, because AID deficiency abolishes accumulation of γ -H2AX and NBS1, which are involved in DNA repair and recruited to the site of DNA double-strand breakage, at the immunoglobulin C $_H$ locus in CSR-induced spleen B cells (28). Because hypermutation can be introduced into the germ-line S μ region when CSR takes place in the other C $_H$ allele (27, 28), AID appears to play a similar role in CSR and hypermutation, most likely at the DNA cleavage step itself (3, 16). However, we cannot completely exclude the possibility that AID is involved in the stage after cleavage in SHM (29). In that case, the SHM target genes in NIH 3T3 cells would be expected to receive continuous generation of cleavages, even in the absence of AID, and these cleavages would be repaired without inducing mutations. Detection of such cleavages in NIH 3T3 cells is a critical test for the latter model. In summary, widespread expression of all the components other than AID that are required for CSR and SHM provides an important clue for analyzing the molecular mechanisms for the two genetic events; the target and cofactors of AID are also widespread.

References and Notes

- M. S. Neuberger, C. Milstein, *Curr. Opin. Immunol.* **7**, 248 (1995).
- C. A. Reynaud, V. Anquez, H. Grimal, J. C. Weill, *Cell* **48**, 379 (1987).
- T. Honjo, K. Kinoshita, M. Muramatsu, *Annu. Rev. Immunol.* **20**, 165 (2002).
- M. Muramatsu et al., *Cell* **102**, 553 (2000).
- P. Revy et al., *Cell* **102**, 565 (2000).
- H. Arakawa, J. Hauschild, J.-M. Buerstedde, *Science* **295**, 1301 (2002).
- I. Okazaki, K. Kinoshita, M. Muramatsu, K. Yoshikawa, T. Honjo, *Nature* **416**, 340 (2002).
- J. Bachl, C. Carlson, V. Gray-Schopfer, M. Dessing, C. Olsson, *J. Immunol.* **166**, 5051 (2001).
- I. B. Rogozin, N. A. Kolchanov, *Biochim. Biophys. Acta* **1171**, 11 (1992).
- K. Yoshikawa et al, unpublished observations.
- Supplemental details of experimental procedures and analyses are available on Science Online.
- L. Pasqualucci et al., *Nature* **412**, 341 (2001).
- J. Tashiro, K. Kinoshita, T. Honjo, *Int. Immunol.* **13**, 495 (2001).
- X. Chen, K. Kinoshita, T. Honjo, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13860 (2001).
- N. A. Kolchanov, V. V. Solovyov, I. B. Rogozin, *FEBS Lett.* **214**, 87 (1987).
- K. Kinoshita, T. Honjo, *Nature Rev. Mol. Cell. Biol.* **2**, 493 (2001).
- A. Martin et al., *Nature* **415**, 802 (2002).
- J. Bachl, M. Wabl, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 851 (1996).
- S. Denepoux et al., *Immunity* **6**, 35 (1997).
- J. E. Sale, M. S. Neuberger, *Immunity* **9**, 859 (1998).
- C. Rada, M. R. Ehrenstein, M. S. Neuberger, C. Milstein, *Immunity* **9**, 135 (1998).
- M. Wiesendanger, B. Kneitz, W. Edelmann, M. D. Scharff, *J. Exp. Med.* **191**, 579 (2000).
- J. E. Sale, D. M. Calandrini, M. Takata, S. Takeda, M. S. Neuberger, *Nature* **412**, 921 (2001).
- D. S. Smith et al., *J. Immunol.* **156**, 2642 (1996).
- J. Spencer, M. Dunn, D. K. Dunn-Walters, *J. Immunol.* **162**, 6596 (1999).
- P. J. Gearhart, R. D. Wood, *Nature Rev. Immunol.* **1**, 187 (2001).
- H. Nagaoka, M. Muramatsu, N. Yamamura, K. Kinoshita, T. Honjo, *J. Exp. Med.* **195**, 529 (2002).
- S. Petersen et al., *Nature* **414**, 660 (2001).
- S. D. Fuggman, D. G. Schatz, *Science* **295**, 1244 (2002).
- We thank J. Bachl and C. Olsson for the plasmid; S. Sakakibara and S. Fagarasan for critical reading of the manuscript; T. Muto, T. Doi, V. Ta, and other members of the Honjo lab for sharing information and discussion; Y. Tabuchi, T. Toyoshima, and K. Yurimoto for their excellent technical assistance; E. Inoue for cell sorting; and T. Nishikawa for secretary help. Supported by grants from the Ministry of Education, Science, Sports and Culture of Japan (C.O.E.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/296/5575/2033/DC1
 Materials and Methods

6 March 2002; accepted 3 May 2002

Induction of T Helper Type 2 Immunity by a Point Mutation in the LAT Adaptor

Enrique Aguado,¹ Sylvie Richelme,¹ Selene Nuñez-Cruz,¹ Arkadiusz Miazek,¹ Anne-Marie Mura,¹ Mireille Richelme,¹ Xiao-Jun Guo,^{1*} Danielle Sainty,² Hai-Tao He,¹ Bernard Malissen,^{1†} Marie Malissen¹

The transmembrane protein LAT (linker for activation of T cells) couples the T cell receptor (TCR) to downstream signaling effectors. Mice homozygous for a mutation of a single LAT tyrosine residue showed impaired T cell development. However, later they accumulated polyclonal helper T (T $_H$) cells that chronically produced type 2 cytokines in large amounts. This exaggerated T $_H$ 2 differentiation caused tissue eosinophilia and massive maturation of plasma cells secreting to immunoglobulins of the E and G1 isotypes. This paradoxical phenotype establishes an unanticipated inhibitory function for LAT that is critical for the differentiation and homeostasis of T $_H$ cells.

The TCR recognizes peptides bound to major histocompatibility complex (MHC) molecules and relays this information to the T cell through adaptor proteins. The adaptor LAT

coordinates the assembly of signaling complexes through multiple tyrosine residues within its intracytoplasmic segment (1). Upon TCR-induced phosphorylation, each of these tyrosines manifests some specialization in the signaling proteins it recruits. For instance, mutation of tyrosine 136 (Y136) selectively eliminates binding of phospholipase C- γ 1 (PLC- γ 1), whereas the simultaneous mutation of Y175 and Y195 results in loss of binding of the Gads adaptor (2-4). Mice deficient in LAT (LAT $^{-/-}$) or having a mutation of the four COOH-terminal tyrosine residues revealed that LAT is essential for the

¹Centre d'Immunologie de Marseille-Luminy, INSERM- and CNRS-Université de la Méditerranée Parc Scientifique de Luminy, 13288 Marseille Cedex 9, France. ²Department of Biology, Institut Paoli-Calmettes, 232 Boulevard Sainte Marguerite, 13009 Marseille, France.

*On leave from Institut Méditerranéen de Recherche en Nutrition, UMR-INRA, Marseille, France.

†To whom correspondence should be addressed. E-mail: bernardm@ciml.univ-mrs.fr