Chem. Rev. **2006,** *106,* 700–719

Evaluation of Molecular Models for the Affinity Maturation of Antibodies: Roles of Cytosine Deamination by AID and DNA Repair

Mala Samaranayake,[†] Janusz M. Bujnicki,[‡] Michael Carpenter,[†] and Ashok S. Bhagwat*[,†]

Department of Chemistry, Wayne State University, Detroit, Michigan 48202, Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Trojdena 4, PL-02-109 Warsaw, Poland, and Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, PL-61-614 Poznan, Poland

Received August 9, 2005

Contents

1. Introduction

One of the most complex and dispersed organs in the human body is the immune system, which functions to identify and destroy invading infectious agents such as bacteria and viruses. It includes cells of discrete organs such as the spleen and thymus, but also components of other organs including bones (bone marrow) and the intestine (Peyer's patch). Additionally, it uses a network of blood and lymphatic vessels that circulate molecules and cells through much of the body. When an infection threatens the body, various cells and molecules of the immune system work together to destroy the infectious particles. This represents a formidable defensive wall in healthy individuals against foreign invaders and is rarely breached. Of interest here are a series of programmed DNA alterations initiated by an enzyme, activation-induced deaminase (AID), that are essential for an effective immune response. The molecular mechanism of AID action and the response of the cell in the form of DNA repair will be discussed in detail below.

A useful way to look at the immune system is to divide the immune response to an infection into two parts—the cellular response and the humoral response. The first of these refers to the action of cells such as the killer T cells and involves direct interactions of these cells with other cells of the immune system and infected cells in the body. The other part, the humoral response, acts instead through antibodies. These proteins can have such a variety of structures that they bind an apparently limitless number of different small molecules such as fragments of proteins and lipopolysaccharides (collectively called antigens) derived from infectious agents. Antibodies are made by B lymphocytes (B-cells) and form tight specific complexes with the antigens. This recognition of foreign antigens by antibodies helps other molecules and cells of the immune system to kill and destroy the infectious organism. The two types of immune responses are not completely separate and, in fact, work together. In

---

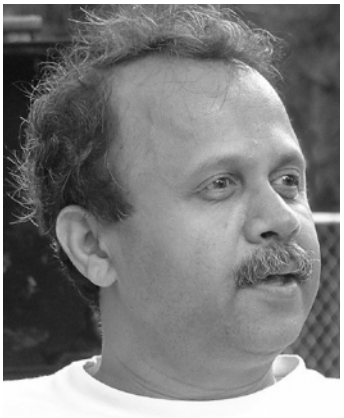
* To whom correspondence should be addressed. E-mail: axb@chem.wayne.edu. Telephone: (313) 577-2547. Fax: 313 577-8822.
[†] Wayne State University.
[‡] International Institute of Molecular and Cell Biology and Adam Mickiewicz University.

10.1021/cr040496t CCC: $59.00

Mala Samaranayake was born in Colombo, Sri Lanka, and earned B.Sc. (Chemistry) Honors in 1993 from University of Ruhuna. She started her career as an Assistant Lecturer in the Department of Chemistry, University of Ruhuna, and later joined the Department of Biochemistry, University of Sri Jayewardenepura, as a Teaching Assistant. There she completed an M.Phil. degree in Biochemistry under the supervision of Dr. S. M. D. N. Wickramasinghe. She then returned to University of Ruhuna and worked as a lecturer in Biochemistry for two years. Currently, she is pursuing a Ph.D. degree in Biological Chemistry at the Department of Chemistry, Wayne State University, under the direction of Ashok Bhagwat.

Janusz Bujnicki received his M.Sc. and Ph.D. degrees in Biology from the University of Warsaw (Poland) in 1998 and 2001, respectively, and his D.Habil. degree in Bioinformatics from the Polish Academy of Sciences in 2005. In 2002 he was selected as a Young Investigator by EMBO and the Howard Hughes Medical Institute. Since 2002 he has been a Professor and head of the Laboratory of Bioinformatics and Protein Engineering at the International Institute of Molecular and Cell Biology in Warsaw. In 2004, he also joined the Faculty of Biology at the Adam Mickiewicz Univeristy in Poznan. His research interests revolve around the development and application of methods for modeling and engineering of proteins and protein–nucleic acid interactions.

Michael Carpenter was born in Grand Rapids, Michigan, in 1979. He earned a B.S. degree in Biochemistry at Northern Michigan University in 2001 and is pursuing a Ph.D. degree in Biochemistry at Wayne State University in Ashok Bhagwat's lab. He is a member of the American Chemical Society, the Golden Key International Honor Society, and Phi Lambda Upsilon honor society.



Ashok S. Bhagwat is a Professor of Chemistry at Wayne State University. He obtained B.Sc. and M.Sc. degrees in Physics from University of Bombay and Indian Institute of Technology, Bombay, respectively. His interest in biological research was stimulated at Indian Institute of Science, Bangalore, and he eventually obtained a Ph.D. in Biophysics from Pennsylvania State University. Following postdoctoral work at Cold Spring Harbor Laboratory, he joined Wayne in 1988. His research interests include the biochemistry of DNA-acting enzymes, and he is a hopeless romantic about the importance of basic research in science.

particular, T cells play a crucial role in activating B-cells to undergo the genetic rearrangements described below. We will focus only on the humoral response in this review and cover the progress made in the field since 1999. However, we shall first describe some aspects of the immune response relevant to these alterations in an outline form, and the reader is referred to a standard immunology textbook (see ref 1, for example) for additional details.

2. Background

2.1. General Structure of Antibody Genes and Proteins

An antibody is a homodimer of a heterodimer consisting of a longer polypeptide chain (called the heavy chain) and a shorter (light) chain (Figure 1A). The homodimer as well as the heterodimer is partly held together by disulfide bridges, and the complete protein can bind two identical antigen molecules. The amino terminal parts of the heavy and light chains, which form the binding pocket, accomplish antigen binding. These protein segments are called variable domains because antibodies that bind different antigens have different primary sequences within these segments. Although the remaining part of each chain is referred to as the "constant" domain, there are five different types of constant domains—α, γ, δ, ϵ, and μ. The antibodies with these domains are respectively said to be of IgA, IgG, IgD, IgE, and IgM isotypes.[1]

The variable and the constant domains of the antibodies are coded by separate exons in the immunoglobulin (Ig) gene (Figure 1B). The multiple constant domains are encoded by separate exons, and the choice of which constant domain is combined with a particular variable domain is made through genetic recombination (see section 2.4 below). The transcription from promoters for the Ig genes occurs at high levels due to the presence of enhancers, which for the heay chain

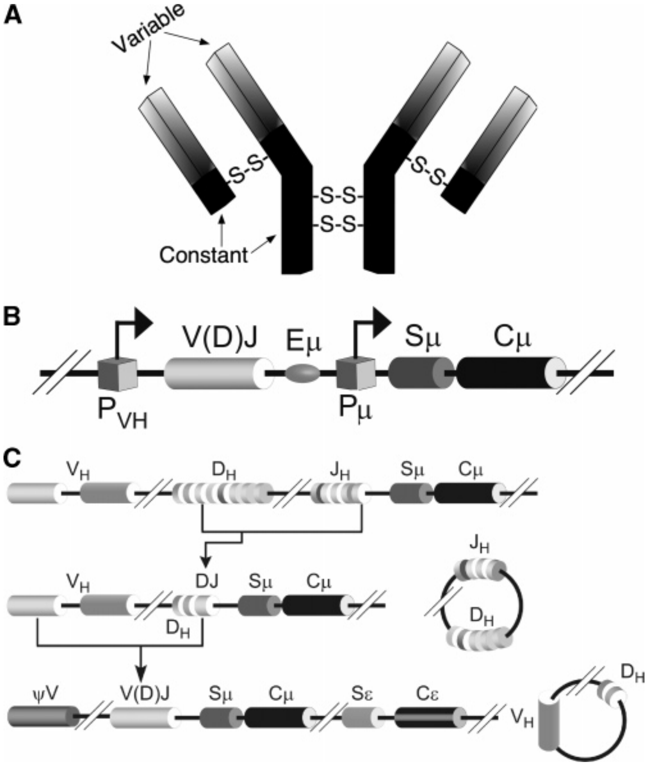


Figure 1. Antibody structure and V(D)J recombination. (A) Schematic representation of an antibody molecule. The longer and shorter chains are respectively called heavy (Ig_H) and light (Ig_L) chains. Disulfide links between the chains (−S−S−) are also shown. Each chain is divided into variable (lightly shaded) and constant (dark) domains. For convenience, only Ig_H genes are shown in parts B and C. (B) Schematic representation of an Ig_H gene. P_{VH} and P_μ are promoters, and V(D)J and C_μ are exons that code the variable and constant domains for an IgM isotype antibody. E_μ and S_μ are respectively an enhancer for the promoter P_{VH} and the switch sequence for C_μ. (C) V(D)J recombination. The human chromosome contains multiple tandem segments for V (variable), D (diversity), and J (junction) sequences. Recombination occurs in two steps: first involving a D and a J segment followed by recombination between a V segment and the already rearranged DJ segment. The recombined VDJ segment is the exon that codes for the variable domain. This is typically shown as V(D)J in recognition of the fact that the segment that codes the light chain variable domain does not contain a D segment.

lie downstream of the exon for the variable segment (Figure 1B). The level of transcription of the Ig genes is regulated in part by genetic rearrangements within B-cells that bring the downstream enhancers closer to the promoter.[2]

2.2. Generation of Antibody Diversity

A remarkable feature of the immune response is its ability to produce secreted antibodies and cell surface receptors that recognize a limitless number of foreign molecules, antigens, using only a limited number of genes. The antigen-binding pockets of antibody proteins are very malleable in their three-dimensional structure, and this diversity arises because the variable domain can acquire an almost limitless diversity of amino acid sequences. Consequently, the immune system is thought to be capable of producing antibodies that can collectively bind over 10^{11} different antigens. One of the early paradoxes regarding the immune system was that the antibody proteins can bind so many different antigens although the total number genes in the human genome is thought not to exceed 50 000. Some of the molecular mechanisms that create this amazing diversity within the antibodies are the subject of this review.

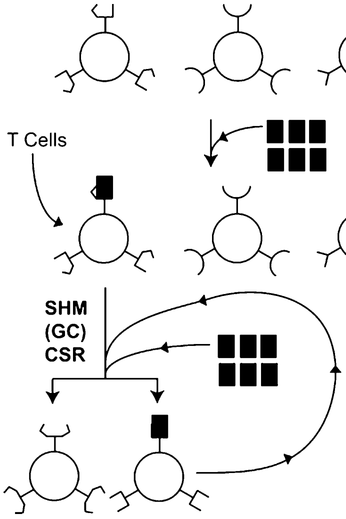


Figure 2. Clonal selection theory. V(D)J recombination creates clones of B-cells each coding for a different antibody. Three such B-cells with different cell surface antibodies are schematically shown. When a specific antigen (filled rectangle) appears, it has significant affinity toward only one of the clonal antibodies. T cells recognize this antigen•antibody complex and stimulate the B-cell to divide. The dividing cells also undergo genetic rearrangements abbreviated as SHM, GC, and CSR (see text for details). This changes the structure of the antibodies made, creating antibodies with worse (bottom left) or better (bottom right) affinity toward the antigen. The cell producing the antibodies that bind the antigen can undergo the same selection and amplification to further increase antibody affinity (semicircular arrow).

This molecular diversity is due, in part, to a series of recombination events that create the variable segment called V(D)J in Figure 1B. This is a combinatorial process that combines three types of protein-coding DNA units called V, D, and J segments. There are scores of different V segments and a few copies each of the D (only for the heavy chains) and the J segments in the genome (Figure 1C). During early development, each B-cell creates a variable segment from a unique combination of V, D, and J (for heavy chains) or V and J (light chains) segments (Figure 1C). This genetic rearrangement (V(D)J recombination) occurs *prior* to the exposure of B-cells to any antigen and creates millions of clones, each capable of making a distinct antibody. These antibodies are of IgM isotype and are displayed on the cell surface such that they can bind antigens. The molecular mechanisms underlying V(D)J recombination are widely covered in advanced biology textbooks and reviews (e.g., see refs 1, 3, and 4) and will not be discussed here.

2.3. Clonal Selection Theory

In higher vertebrates, B lymphocytes undergo additional genetic changes when the cells are exposed to an antigen. This helps many of these cells produce antibodies that bind antigens with higher affinity. This evolutionary process of making better antibodies is explained by the “clonal selection theory” of Burnet and Talmage,[5,6] and a modern version of this proposal is presented in Figure 2.

The current version of this model for antibody maturation starts with V(D)J recombination creating millions of clones of B-cells, with each clone expressing a unique antibody on its cell surface. When the organism is exposed to a foreign agent such as a virus, only a small fraction of these clones

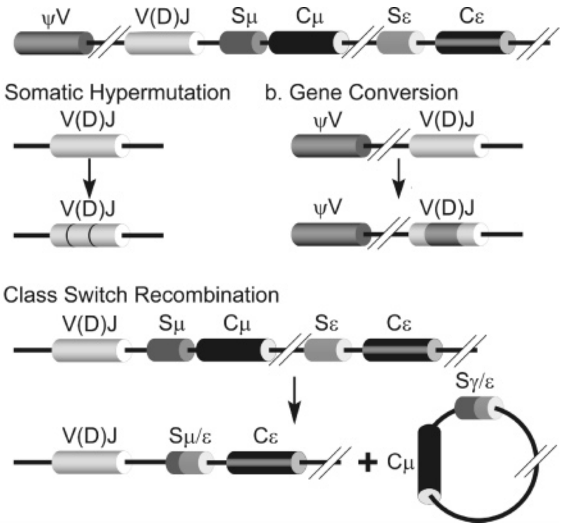


Figure 3. Genetic rearrangements during affinity maturation of antibody genes. An Ig$_H$ gene resulting from V(D)J recombination is shown at the top. Depending on the organism, the gene undergoes somatic hypermutations (SHMs; part a) or gene conversion (GC; part b). Point mutations introduced in the V(D)J segment during SHM are shown by darker lines. The part of V(D)J converted to the sequence of a pseudo-V segment (ψV) during GC is shown as a dark patch. Ig$_H$ genes also undergo class switch recombination (CSR; part c). In this case double-strand breaks within two different switch regions (S_μ and S_ϵ) and the rejoining of open ends create two products. One contains an Ig$_H$ gene that codes for IgE isotype antibody and a circular DNA product with DNA between the two double-strand breaks.

are capable of binding foreign antigens using the antibodies displayed on their surface (Figure 2). These antigen presenting B-cells interact with T cells, which then stimulate the B-cells to undergo division and further differentiation. This results in the amplification of only those B-cell clones that are capable of producing antibodies specific for the foreign antigen.[7] At the same time, the cells undergo additional genetic alterations that create antibodies of even higher affinity toward the antigen. These latter alterations in the Ig genes are a critical part of the "affinity maturation" of antibodies.

2.4. Genetic Alterations during Affinity Maturation

The vertebrate Ig genes in maturing B lymphocytes are known to undergo three genetic changes—somatic hypermutations (SHMs), class switch recombination (CSR), and gene conversion (GC; Figure 3; ref 8). Of these, SHM and GC are principally mutational processes that introduce (mostly) base substitutions within the V(D)J rearranged Ig genes at a rate of $\sim 10^{-3}$ per base pair per generation. This mutation frequency is $\sim 10^6$-fold higher than normal[9] and is restricted to the V(D)J segment of Ig genes. GC involves recombination between a rearranged V(D)J segment and a pseudo-V gene and is presumed to require homologous recombination events (Figure 3). It is found in some animals (rabbits and chickens), but not in humans, and will not be discussed here.

SHM introduces point mutations in the Ig gene starting at the promoter for the Ig gene and ending around the 5′ end of the intron between V(D)J and the Cμ segments. They do not extend into the constant domain segments.[10,11] These mutations are scattered over the variable segment and include transitions as well as transversions. The hypermutations occur about equally at C:G and A:T base pairs, creating approximately one amino acid change per cell per generation. Among the many interesting features of SHM are its ability to target a $\sim$1500 bp segment out of a genome of $\sim 3 \times 10^9$ bp and the presence of hypermutational "hot spots" within the V(D)J segment. Another curious feature of SHM is its strict requirement for transcription of the Ig gene.[12−14] These and other aspects of SHM are described below in some detail.

Some of these mutated B-cell clones produce antibodies that have higher affinity toward the foreign antigen and are further selected for cell division and amplification (Figure 2). This is an iterative process involving mutations in the Ig variable segment and selection of antigen-binding antibody-producing cells. This means that if the infection that triggered affinity maturation persists in the body, then the humoral response creates antibodies with higher and higher affinities for the antigens with the passage of time. For the same reason, repeated immunization of an animal with the same vaccine makes it better able to combat an infection. In contrast to cells that produce antibodies that can bind the antigen, cells that express mutated antibodies that do not bind the antigen are no longer stimulated for cell division and are eliminated from the B-cell population. The final stage of the development of B-cells producing antibodies against circulating antigens is their conversion to plasma cells that secrete the antibody molecules, which then diffuse into blood and lymphatic vessels.[1]

The introns separating the exons for the different constant segments contain two features that are relevant to the third genetic rearrangement, CSR. One feature is a sequence referred to as the "switch" (S) region, and the second is a promoter within the intron that transcribes each switch region prior to the genetic rearrangements within the constant domains. The S regions contain short repetitive sequences (GGGGT and GAGCT, for example) and typically have different base compositions in the two DNA strands. CSR is a region-specific recombination process that requires double-strand breaks in two different S regions and the joining of the open DNA ends eliminating intervening constant segments as a circle (Figure 3). In maturing B-cells this exchanges the μ constant segments of the immature Ig genes with one of the other constant segments (say ϵ) causing a switch from IgM type antibodies to a different isotype (IgE; Figure 3; refs 2 and 8). The strand breaks essential for CSR occur within the S regions and require transcription (but not translation) of these sequences. The molecular mechanism of CSR is poorly understood and will be discussed below mostly in the context of SHM.

2.5. Antibody Maturation and Immunodeficiency Syndrome

Defects in affinity maturation of antibodies lead to an immune deficiency referred to as hyper-IgM syndrome (HIGM). HIGM is a rare immunodeficiency characterized by normal or elevated serum IgM levels with absence of IgG, IgA, and IgE, resulting in a profound susceptibility to bacterial infections and an increased susceptibility to opportunistic infections. While the lack of antibody types other than IgM in these patients is due to defective CSR, many of these patients are also defective in SHM. It is the latter defect that reduces the ability of these patients to fight infections. HIGM is divided into five subgroups, HIGM1 through 5. While two of these subgroups (HIGM1 and HIGM3) have genetic defects that prevent activation of B lymphocytes for maturation by an antigen, two others (HIGM2 and HIGM5) have defects in the DNA processing that create more diverse antibodies. The latter two types of genetic defects will be discussed in sections 3, 4, and 6. The remaining subgroup (HIGM4) may also be defective in a DNA processing step required for CSR, but its molecular cause is unknown (see the OMIM database for additional details: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM).

3. Discovery and Biology of AID

In recent years, there have been two conceptual breakthroughs in our understanding of the molecular processes by which immunoglobulin genes are altered in response to the exposure of naive B-cell clones to antigen. The first of these was the discovery of a gene whose protein product, activation-induced deaminase (AID), is required for both SHM and CSR in murine lymphoid cells.[15] The second breakthrough in our understanding of the mechanics of antibody maturation came with the observation that AID is a mutator in *E. coli*[16] and the suggestion that AID may be a DNA−cytosine deaminase.

We discuss below the molecular mechanisms underlying antibody maturation with a greater emphasis on trying to understand the role of the mutator protein, AID, and of a number of DNA repair processes involved in SHM. The mechanisms of CSR and GC will be discussed only in the context of SHM.

3.1. AID Is Required for Antibody Maturation

Honjo and colleagues discovered AID while studying a cell line that required stimulation by cytokines to undergo CSR. They found that expression of AID from a tet-contolled promoter alleviated the requirement for cytokines, suggesting that cytokines may stimulate the expression of AID to promote CSR. Furthermore, an AID$^{-/-}$ mouse was defective in both CSR and SHM,[15] demonstrating the absolute requirement for AID in these processes. In other experiments, the expression of AID was studied in various murine and human tissues and the protein was detected in germinal center B-cells[17] and various lymphoid organs.[18] Another study investigated the importance of this protein in a clinical setting by sequencing AID gene from 18 patients with one form of the hyper-IgM syndrome (HIGM2). All the patients had mutations in their AID gene, and 10 different mutant alleles, which included missense mutations as well as frame-shift mutations that cause premature chain termination, were discovered (refs 19 and Table 1).

AID is required for two additional mutational processes associated with antibody maturation. As noted earlier, some animals use GC instead of SHM as the principal mechanism for generating sequence diversity in maturing B lymphocytes. Using targeted gene disruption in chicken cells, Arakawa *et al.*[20] and Harris *et al.*[21] showed that AID was required for gene conversion in Ig genes. The other mutational process occurs within the switch regions upstream of the constant domain segments in Ig genes. When B lymphocytes are stimulated to undergo CSR, the switch regions acquire point mutations and small addition/deletions regardless of whether they have undergone recombination.[22] Nagaoka *et al.*[23] found that in a murine AID$^{-/-}$ cell line the S_μ region upstream of nonswitched C_μ segments did not acquire mutations when B-cells were stimulated. Transfection of these cells with an AID expressing retrovirus restored the hypermutation phenotype in the switch region. Thus, AID is required for all known mutational and recombinational processes involved in antibody maturation and plays a critical role in producing a robust immune response against infections.

3.2. AID as an RNA-Editing Enzyme

A comparison of the AID sequence with available sequences suggested a function for the protein. The protein shares sequence similarity with bacterial cytidine (rC) and cytidylate (rCMP) deaminases.[17] This suggested that AID may also be a cytidine deaminase, and it was accordingly named activation-induced cytidine deaminase. This activity was apparently confirmed biochemically for a GST−AID hybrid protein purified from *E. coli*, and indirect results also suggested that the protein may contain catalytically important zinc ion(s).[17] More intriguingly, AID was most similar in sequence to a RNA−cytosine deaminase, APOBEC1.[17] This enzyme converts the cytosine at position 6666 in the mRNA for apolipoprotein B100 to uracil, changing a glutamine codon to a termination codon. The resulting shortened protein (apolipoprotein B48) has different physical properties and is processed differently by liver cells. The sequence conservation between AID and APOBEC1 led Muramatsu *et al.*[15] to suggest that AID may act on an mRNA encoding as yet unknown protein, changing its product into a CSR recombinase and hypermutator. In the latter case, double-strand breaks (DSBs) caused by this protein within the variable segments of Ig genes would be repaired and the errors in rejoining the broken DNA ends would result in hypermutations. They further suggested that, like APOBEC1,[24] AID may also require an accessory protein factor(s) to provide its substrate specificity.[15]

DSBs are a clear prerequisite for CSR, and hence, a model that invokes the synthesis of a new DNA endonuclease in response to AID induction is attractive. However, several key pieces in this hypothesis are missing, and there are serious questions about its validity. First, if the repair of DSBs in the variable region leads to SHM, then the predominant signature of this event should be addition/deletion mutations and not base substitutions. Typically, less than ∼10% of mutations in hypermutating cells are addition/deletion type.[25] Second, there may not be a strict requirement for DSBs for SHM although some reports do suggest a correlation between the two events.[26−28] Instead, single-strand breaks in Ig genes may be converted to DSBs during replication.[29] Third, SHM does not require a DNA-dependent protein kinase catalytic subunit,[30] or Rad54 and Rad54B,[31] suggesting that neither nonhomologous-end joining (NHEJ) nor homologous recombination machinery is required for SHM. Thus, the process by which the proposed DSBs in the variable segments would be repaired remains unclear. Fourth, although AID does bind nonspecific RNA pools,[32,33] as yet no specific mRNA has been identified as its target for cytosine deamination. Finally, if AID does require another protein to target it to a specific mRNA, the identity of this accessory protein is also currently unknown. It seems clear that much work remains to be done to validate the RNA editing model for AID action.

3.3. AID as a Mutator

Neuberger and colleagues used four different forward mutation assays to show that expression of AID in *E. coli* was moderately mutagenic. In wild-type (WT) cells, AID increased the frequency of mutations ∼3- to 6-fold and

Table 1. Phenotypes of AID and UDG Mutations[k]

amino acid change			nucleotide change	phenotype		additional comments[l]	ref
WT (aa)	AA position	mutation (aa)		SHM	CSR		
AID, Point Mutations							
S, N	3, 168	G, S	NR	low	low	artificial	48
K	10	R	NR	moderate	moderate	artificial	48
F, C	11, 117	V, X	T 31 G, 226 ins 1 bp	no	no	HIGM2	145
Y	13	H	NR	low	normal	murine, nuclear transport defect	57
V	18	R	NR	low	normal	murine, nuclear transport defect	57
V, R	18, 19	S, V	NR	low	normal	murine, nuclear transport defect	57
W	20	K	NR	low	normal	murine, nuclear transport defect	57
G	23	S	NR	low	normal	murine	57
R	24	W	C 70 T	no	no	HIGM2	19, 48, 145
S	43	P	T 127 C	low	no	HIGM2	145
H	56	R	NR	NR	NR	artificial, no deamination	44
H	56	Y	NR	no	no	HIGM2	48
H	56	X	C 166 T	NR	no	HIGM2	146
E	58	Q	NR	NR	NR	artificial, no deamination	44, 33, 35
L, W	59, 68	F, X	175 del 9 bp, G 203 A	no	no	HIGM2	19
W	68	X	G 203 A	no	no	HIGM2	48
W	80	R	T 238 C	no	no	HIGM2	19, 48
W	84	X	G 251 A	NR	no	HIGM2	147
C	87	A	NR	NR	NR	artificial, no deamination	33
C	87	R	T 259 C	NR	no	HIGM2	81, 146
C	90	A	NR	NR	NR	artificial, no deamination	33
L	98	R	T 292 G	NR	no	HIGM2	81
L	106	P	T 317 C	no	no	HIGM2	19, 48, 146
R	112, 208	C, X	C 334 T, 544 del 1 bp	moderate	no	HIGM2	145, 48
R	112	C	C 334 T	no	no/normal[j]	HIGM2	147, 48, 81
R	112	H	G 335 A	no/low[j]	no/low[j]	HIGM2	147, 48, 145
I	136	K	T 407 A	NR	no	HIGM2	145
M	139	V	A 415 G	low	no	HIGM2	19, 48
C	147	X	C 441 A	low	no	HIGM2	19, 48
F	151	S	T 452 C	NR	no	HIGM2	19, 146
R	174	S	A 522 C	NR	no	HIGM2	146
AID, addition/deletions							
R	190[a]	X	C 568 T	normal	no	HIGM2, dominant negative	148, 48
N[b]	7	X	21 del 19 bp	NR	no	HIGM2	146
F[b]	15	X	21 del 19 bp	no	no	HIGM2	19, 48
W	68	X	175 del 9 bp, G 203 A	no	no	HIGM2	19
		3 aa del	235 del 9 bp	no	NR	HIGM2	145
R	112, 208	C, X	C 334 T, 544 del 1 bp	moderate	no	HIGM2	145, 48
	183−208		544 del 1 bp[c]	normal	no	artificial	48
	182−215		182 ins 102 bp[d]	normal	no	HIGM2	48
	189−198		565 del 30 bp[e]	normal	no	artificial	56
			G +1 T (exon 2 ret)[f]	NR	low	HIGM2	149
			G +1 C (exon 4 del)[g]	NR	no	HIGM2	145
UNG							
	159	X	497 del TA	normal[h]	low	HIGM5	80, 81
F	151	S					
	141, 224	X, X	462 del C, 639 del TA	normal[h]	low	HIGM5	80
F	251	S	T 822 C	NR[i]	no	HIGM5	80, 79

[a] Seven patients with $\mathrm{AID}^{\mathrm{R190X}/+}$ genotype have been described. [b] Deletion of 19 bp starting from position 21 led to a premature stop codon at position 26. This is reported differently in refs 146 and 19. [c] Frame-shift replacement of C-terminus with 26 amino acids (CMRLMTYETHFVL-WDFDSNFQECHTR) at position 183. [d] Insertion of 34 amino acids (VTKPSTQFRRLSGPTDPQPRFEAIHSICFSLSLR) at position 182. [e] C-terminal deletion of 10 amino acids starting at position 189. [f] Splice donor site mutation (G > T) at position of +1 of intron 2 leading to retention of part of intron 2. [g] Splice donor site mutation (G > C) at position of +1 of intron 4 leading to deletion of exon 4. [h] Normal mutation frequency but biased toward transitions at G/C residues. [i] Mutant protein was fully active when purified from *E. coli*. [j] Different patients show different phenotypes. [k] Abbreviations: aa, amino acid; X, stop codon; ins, insertion; del, deletion; ret, retention; NR, not reported. [l] HIGM patients with different genetic defects are classified as follows: CD40 ligand (CD40L), HIGM1; AID, HIGM2; CD40, HIGM3; unknown defect, HIGM4; UDG, HIGM5.

shifted the spectrum of mutations in favor of transition mutations at C:G base pairs. Specifically, in the absence of AID, only 31% of mutations in the *rpoB* gene creating a rifampicin-resistant phenotype ($\mathrm{Rif}^{\mathrm{R}}$) had a C:G to T:A change (hereafter referred to as C to T mutation), but these mutations were 80% of the total in AID expressing cells. The mutations in the *gyrA* gene (phenotype: nalidixic acid-resistance) showed a similar picture. In this case, C:G to T:A transitions were 34% of all the mutations without AID and 70% with the enzyme.[16] Subsequently, the mutator effect of AID was confirmed in other genetic selection systems in *E. coli*[34−36] and yeast.[37,38]

Petersen-Mahrt *et al.*[16] suggested that AID acts directly on DNA, converting cytosines to uracils. As uracil in DNA pairs with adenine causing C:G to T:A transitions (Figure 4), this explains the increase in this class of mutations when AID is expressed in *E. coli*. Recently, Martomo *et al.*[39] used biochemical techniques to confirm that uracil accumulates

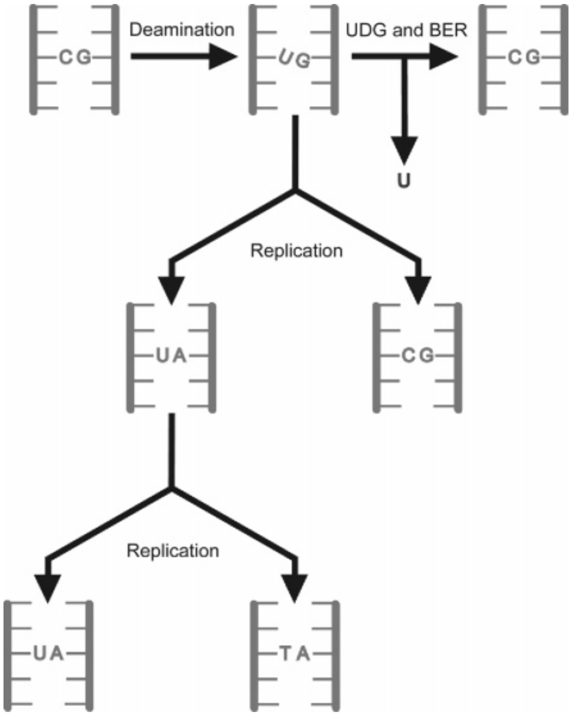


Figure 4. Cytosine deamination and C to T mutations. The possible consequences of the deamination of cytosine to uracil are shown. Repair of the lesion through the action of UDG and base excision repair (BER) restores the original C:G pair. Instead, if replication occurs prior to repair, half the daughter molecules contain C to T mutations.

in *E. coli* DNA upon expression of AID. Uracil is excised from DNA by the uracil−DNA glycosylase (UDG), which is present in all organisms (refs 40−43 and Figure 4). It hydrolyzes the *N*-glycosidic linkage between the uracil and deoxyribose sugar to create an abasic site, which is processed further by the base excision repair (BER) pathway to restore the cytosine base (Figure 4). The role of this enzyme in SHM and CSR is discussed more fully in section 5.1.

3.4. AID is a DNA–Cytosine Deaminase, Not a Cytidine Deaminase

Direct evidence for the ability of AID to catalyze deamination of cytosines in DNA was obtained by four different research groups. Bransteitter *et al.*[32] used a GST−AID fusion protein purified from insect cells to show that it converts cytosines in single-stranded (SS) DNA to uracil, but not in double-stranded (DS) DNA, in SS RNA, or in a DNA−RNA hybrid. Furthermore, they found that different pools of nonspecific RNAs from *E. coli* or mammalian cells were inhibitory toward the SS DNA deamination activity of AID.[32] Chaudhuri *et al.*[44] showed that partially purified B-cell extracts were capable of converting ^{3}H-cytosines in DNA to ^{3}H-uracils, which could then be released from DNA using UDG. This activity was inhibited by 20 μM tetrahydrouridine and was confirmed further by converting the abasic site created by UDG to strand breaks using alkali.[44] Sohail *et al.* and Dickerson *et al.*[33,36] used respectively GST- and Strep-tagged AID purified from *E. coli* and demonstrated its activity on DS DNA with a bubble and SS DNA. The GST−AID purified partially from *E. coli* specifically deaminated cytosines in a 5 nt SS bubble to uracils without affecting the cytosines in the DS portion of the same molecule, and the reaction was inhibited by 1,10-phenathroline but not by EDTA.[36] It is known that the Zn^{2+} ion within APOBEC1 can be extracted with 1,10-phenathroline but not EDTA,[45] and hence, these data suggest that AID also contains Zn^{2+} in its active site. Other investigators have also shown that AID can act on the SS portion of a DNA bubble substrate and that molecules with larger bubbles are better for it.[32] Dickerson *et al.*[33] found that Strep−AID bound tightly to SS RNA and DNA but deaminated cytosines in only the latter nucleic acid. These and other studies have established firmly that AID is a SS DNA-specific DNA−cytosine deaminase that has little effect on RNAs that have been tested.

Despite its original naming, AID is not a cytidine deaminase. It does not complement *E. coli cdd* defective in cytidine deaminase activity (M.C. and A.S.B., unpublished results). It is also not a cytidylate or deoxycytidylate deaminase. Dickerson *et al.*[33] reported that rC, rCTP, and dCTP were not detectably converted to their deamination products by Strep−AID. In contrast, an earlier study reported that GST-AID can deaminate cytidine.[17] A possible difference between these two studies is the level of purity of the protein used for the biochemical assays. While the latter group purified the protein hybrid on an affinity column for GST, the former group used two ion exchange columns to purify the protein. The Strep−AID protein was shown to be nearly homogeneous by silver staining, while the purity of GST−AID was not reported. It is possible that the GST−AID used by Muramatsu *et al.*[17] was contaminated with *E. coli* Cdd protein. Beale *et al.*[34] raised this very possibility in their study of AID, APOBEC1, and APOBEC3G (another enzyme in the AID−Apobec family). They found that the level of deoxycytidine (dC) deaminase activity in their proteins purified from *E. coli* varied from preparation to preparation and could be completely eliminated by the addition of tertrahydrouridine (THU), a known inhibitor of cytidine deaminase. Furthermore, a preparation of the catalytically inactive mutant of APOBEC1 (C93A) also had high level dC deaminase activity, which could also be inhibited with THU.[34] In contrast, the DNA−cytosine deaminase activity of APOBEC1 was unaffected by THU.[34,46] These data suggest that purified AID (as well as APOBEC1 and APOBEC3G) is not a nucleoside or mononucleotide deaminase and should be considered a DNA−cytosine deaminase. For this reason, we prefer to call it an activation-induced deaminase, rather than a cytidine deaminase.

4. Structure of AID

4.1. Gene for AID

The AID gene is located on chromosome 12 in *Homo sapiens* in a region of microsynteny (12p13) from mammals to pufferfish[47] and is close to the APOBEC1 gene. The human gene contains 5 exons over 10 677 bp and is transcribed into a 2791 nt mRNA. This message is translated into a small 198 amino acid protein (MW 23 954). Mutations that lie in the AID gene exons and in intron−exon boundaries have been discovered in the human population, and these individuals suffer from hyper-IgM type 2 (HIGM2) (ref 19 and Table 1).

4.2. Subunit Composition

Several lines of evidence suggest that AID dimerizes or forms higher order multimers, but the number of subunits within active AID remains unclear. One of the HIGM2 mutations (8 aa deletion from the C-terminus, Table 1) has a dominant negative phenotype[48] suggesting multimer formation. Additionally, when AID genes with two different tags were expressed in murine cells, they immunoprecipitated together when antibody against either tag was used.[48] The structure of yeast CDD1, which is an orthologue of APOBEC1, has been used to argue that AID may be a dimer.[49] However, the biochemical evidence regarding the AID composition is conflicting. Chaudhuri *et al.*[44] partially purified AID from mammalian cells and found that it sediments on a glycerol gradient as a 30 000−60 000 MW size range, and they have suggested that AID may exist as a dimer.[50] However, Dickerson *et al.*[33] reported that Strep−AID purified from *E. coli* was strongly resistant to dissociation and migrated on the gel as a tetramer. Consequently, the subunit composition of AID remains a matter of debate.

4.3. Subcellular Localization Signals

When AID is tagged at its N-terminus with GFP and expressed in Ramos cells, the protein is predominantly found in the cytoplasm.[51] This observation initially suggested that AID does not directly act on DNA. However, Ito *et al.*[52] constructed AID tagged at its C-terminus with GFP and found that the protein shuttles between the cytoplasm and the nucleus. Specifically, the fusion protein accumulated in the nucleus following the treatment of cells with an inhibitor of nuclear export. They also found that the C-terminal 16 amino acids in AID were essential for the export.[52] Similar results were also reported by two other groups.[53,54] Furthermore, Ito *et al.* reported the existence of a nuclear localization signal (NLS) in the N-terminus, and a similar motif has been found in the N-terminus of APOBEC1.[52] However, this sequence may not constitute a true NLS as its removal does not eliminate AID from the nuclei.[53] It is possible that AID is kept in the cytoplasm by specific chaperones until the stimulation of B-cells for maturation actively translocates it to the nucleus.[55] Additional work is needed to fully illuminate the mechanisms that regulate AID transport in and out of the nucleus.

4.4. Functional Domains

The carboxy terminus of AID has a second biochemical function; it is required for CSR but not for SHM. One HIGM2 patient had an AID allele with the terminal 8 amino acids deleted (Table 1), and this protein was shown to be defective in CSR.[48] However, this mutant has normal SHM activity in the Rif^R assay in *E. coli*. Additionally, a mutant with a 34 aa insertion after codon 181 and another frameshift mutant with changes starting after codon 182 also had a similar split phenotype.[48] Similarly, Baretto *et al.*[56] found that a deletion of 10 aa from the carboxy terminus of AID eliminates CSR but not SHM. Furthermore, these investigators found that hypermutations in the S_μ region were normal, showing that CSR was not required for switch region mutations.[56]

Shinkura *et al.*[57] reported several mutations near the N-terminus of AID that had reduced SHM activity, but had near normal CSR (Table 1). Based on these mutations, these investigators argue that SHM-specific factors bind near the N-terminus of AID that are not required for CSR. However, there are some concerns regarding such a conclusion. First, none of the mutants is completely defective in SHM. Depending on the assay used, some mutants have up to 50% of the WT SHM activity. Second, all the mutations lie within the putative NLS mentioned above. Consistent with their location, five out of the six mutants described are defective in transport into the nucleus. Thus, a possible simple explanation for their reduced SHM phenotype is a reduced accumulation in the nucleus.[57] Despite these reservations, there appear to be functionally distinct domains at the two ends of the protein. The N-terminal domain (1−23 aa) has a role in nuclear localization and may bind SHM-specific factors, while the C-terminal domain (∼180−198 aa) is required for export from the nucleus and is required only for CSR. All the known mutations in the central region affect both SHM and CSR and have little effect on protein localization (Table 1).

4.5. Catalytic Mechanism

Figure 5 shows a model for the active site of AID and a possible reaction mechanism. It is based on the structure and mechanism of *E. coli* cytidine deaminase[58] and properties of some of the AID mutants listed in Table 1. Briefly, a water molecule is activated and split by the combined action of glutamate 58 and the zinc(II) cation within the active site (step 1). A cytosine within SS DNA is inserted into the active site and stabilized by π interactions with Trp-80 (Figure 5A). The positioning of W80 within the active site is based on a suggestion regarding the APOBEC1 structure by Harris *et al.*[59] This allows the coordinated hydroxide, acting as a nucleophile, to attack at C4 of the cytosine. The π bond between C4 and N3 is lost, and the N3 deprotonates glutamate 58 (step 2; Figure 5B). The result is interrupted ring resonance, as C4 is now tetrahedral (step 3). Some rearrangement follows as glutamate 58 deprotonates the hydroxyl and protonates the amine, making it a good leaving group (steps 3 and 4). The reaction cycle completes as the negative charge on O4 forms a π bond with C4, kicking off the positively charged ammonium as ammonia and restoring the ring resonance (step 5; Figure 5B). No mechanism-based inhibitors of AID or other DNA-cytosine deaminases have been reported, and the ability of the product-mimic tetrahydrouridine (THU) to inhibit AID is controversial (see refs 17, 34, 44, and 50 and section 3.4). Consequently, much work remains to be done to validate the proposed mechanism.

4.6. Structural Model for AID

An X-ray crystal structure is currently unavailable for AID, APOBEC1, or other members of this family. A structure is available for the yeast RNA editing cytosine deaminase yCDD1.[49] Ta *et al.* have suggested dividing AID into four domains—helix, active site, linker, and pseudo-active site.[48] This division was based on a similar proposed division for APOBEC1[60] but is not found in the yCDD1 or modeling efforts based on yCDD1 done by Xie *et al.*[49] and by us.

An alignment of the AID sequence with its sequence homologues with known structures is shown in Figure 6A. The model of the human AID protein was constructed using the fold-recognition approach,[61] followed by recombination of fragments and the optimization of the sequence-structure fit of the "Frankenstein monster" approach,[62] combined with remodeling of uncertain regions with ROSETTA.[63] All fold-recognition servers generated reliable matches between the AID sequence and the structure of several different deaminases, with the yeast cytosine deaminase (yCD; ref 64)

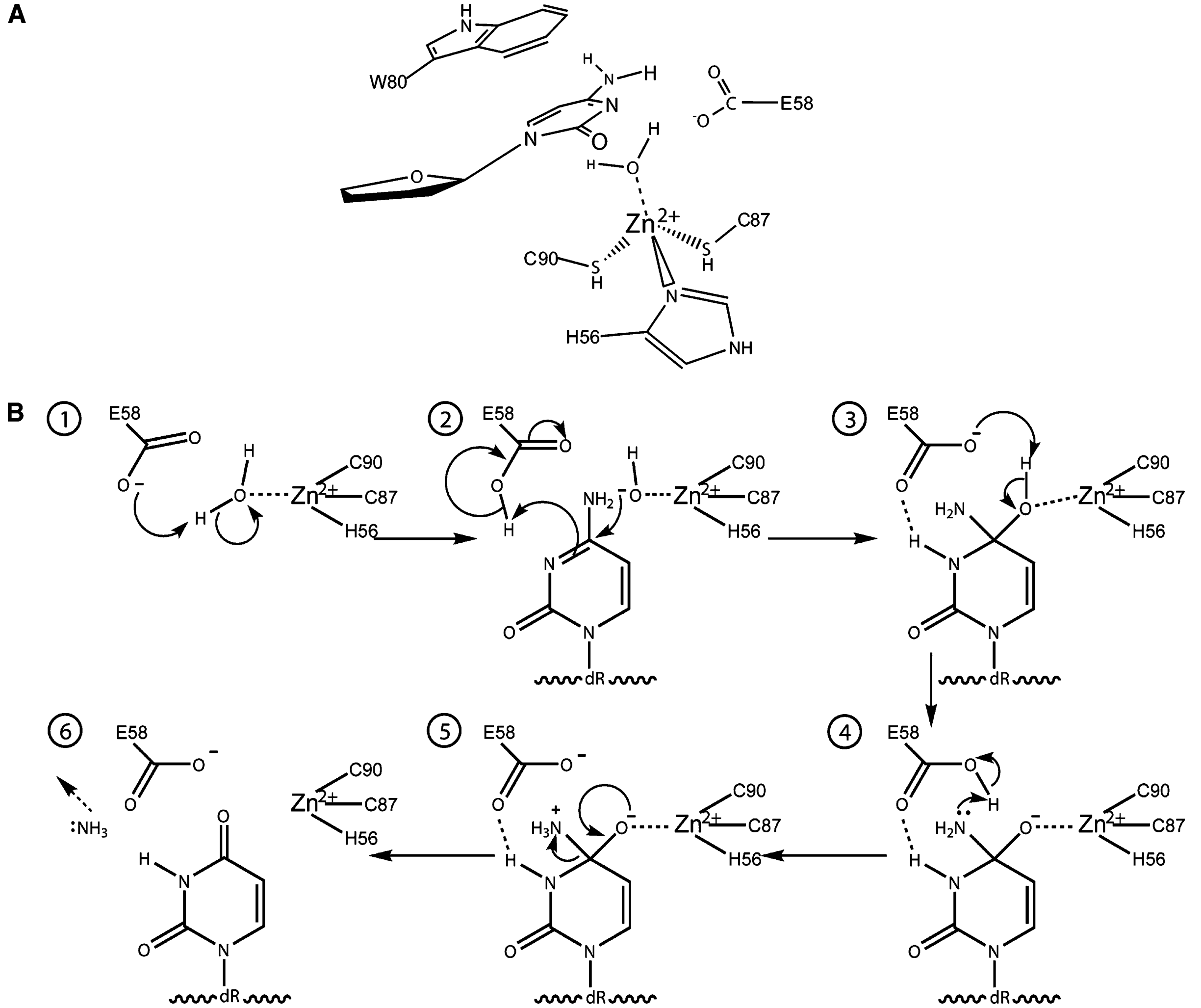


Figure 5. Active site structure and proposed reaction mechanism for AID. (A) Structure of the active site. The enzyme is presumed to contain a zinc atom, which is coordinated by two cysteines, a histidine, and a water molecule. These residues have been identified based on sequence alignments and mutational studies. It is expected that a tryptophan or some other aromatic residue in the protein will stabilize the cytosine. Harris *et al.*[59] have suggested that Trp-80 serves this function. The amino acid residues are numbered to correspond to the human AID sequence. (B) Reaction mechanism of AID. The proposed mechanism is based on a mechanism of *E. coli* cytidine deaminase.[58] E-58 alternately acts as a general base and a general acid, activating the water molecule bound to the zinc atom for an attack at C4 of cytosine and protonating N3. The same residue undergoes one more round of acid−base catalysis to protonate N4 and making it a better leaving group.

singled out as the unequivocally best template for modeling of hAID, in agreement with the earlier suggestion.[65] Importantly, no server produced a match that would agree with another prediction, that AID comprises two domains similar to the yeast yCDD1 enzyme.[49] Thus, we modeled AID based on the structure of the yCD dimer. The substrate SS DNA was docked manually based on the superposition of the target base with the ligand in the yCD structure.

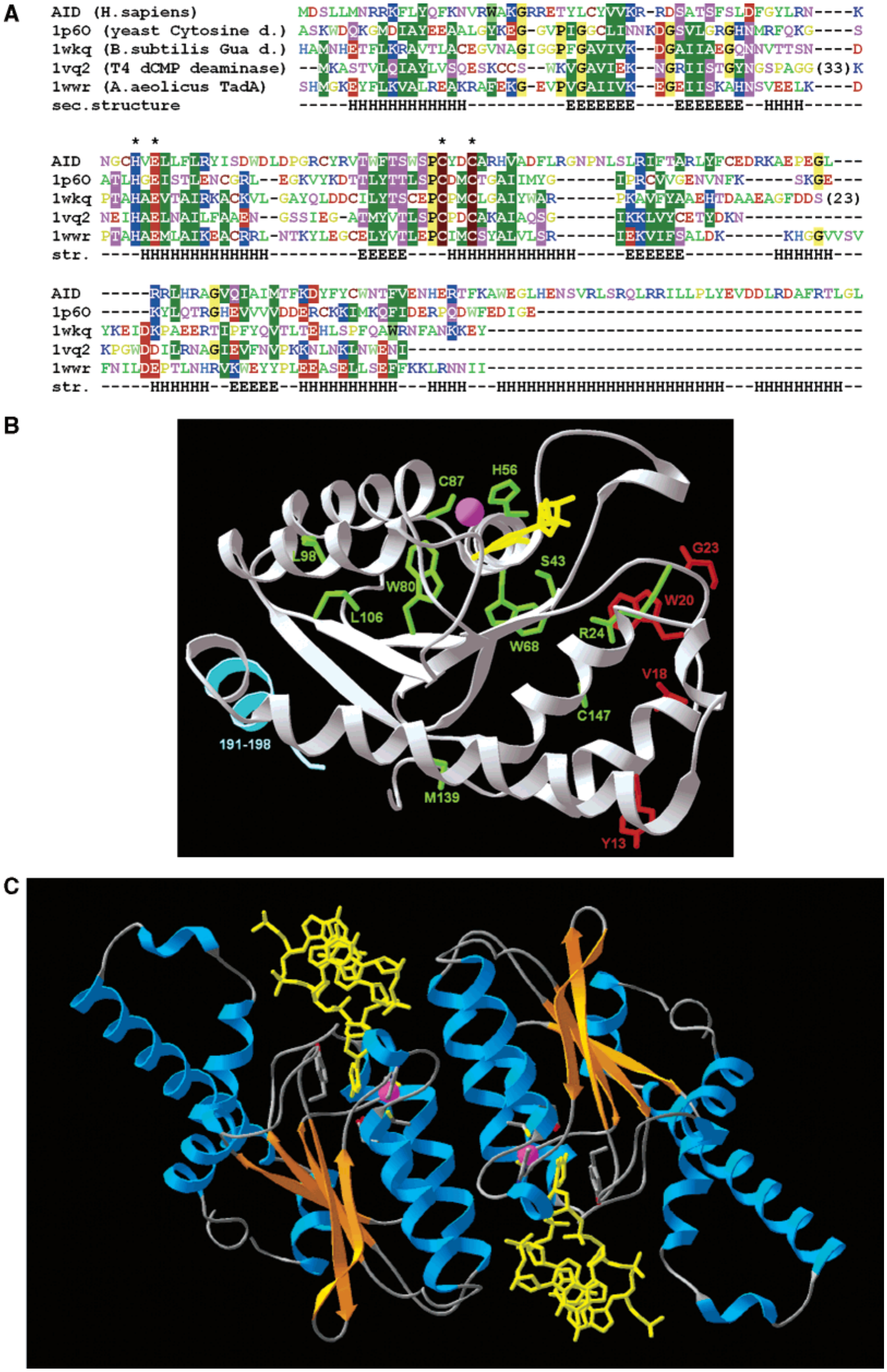
The monomer structure contains a five-stranded β sheet which is sandwiched between multiple α helices (Figure 6B). Significantly, the C-terminal residues of the protein required for CSR, but not SHM, fold partly into a helix (aa 190−198; light blue in Figure 6B) and are well separated from the residues thought to be required for SHM, but not CSR (shown in red). While some of the mutations that affect both SHM and CSR (shown in green) are within the proposed active site, some, such as M139, are quite far away. Presumably, these latter classes of mutations disrupt overall protein stability.

We modeled the protein as a dimer and docked two SS DNAs into it (Figure 6C). The protein dimerizes as a result of interactions between two central α helices which are also involved in catalysis. The two active sites may interact through the dimer interface and may be sensitive to each other's structural changes during catalysis. Consequently, it is possible to visualize a model for the enzyme in which binding of the substrate (or catalysis) by one active site affects the structure of the second active site. These structural models serve only as a basis for designing experiments and will have to be modified when additional biochemical or physical data become available.

5. Enzymatic Activity of AID

5.1. Sequence-Specificity of AID

One of the key features of SHMs is that a significant fraction of them appear within the consensus sequence WRCY (W is A or T, R is purine, and Y is pyrimidine[66]) or TW.[67] These will be referred to respectively as C:G and T:A hot spots. These hot spots could, in principle, have several different origins. They could represent susceptible DNA structures present in Ig genes undergoing SHM, Ig protein



```
AID (H.sapiens)            MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKR-RDSATSFSLDFGYLRN----K
1p60 (yeast Cytosine d.)  ASKWDQKGMDIAYEEAALGYKEG-GVPIGGCLINNKDGSVLGRGHNMRFQKG----S
1wkq (B.subtilis Gua d.)  HAMNHETFLKRAVTLACEGVNAGIGGPFGAVIVK--DGAIIAEGQNNVTTSN----D
1vq2 (T4 dCMP deaminase)  --MKASTVLQIAYLVSQESKCCS--WKVGAVIEK--NGRIISTGYNGSPAGG(33)K
1wwr (A.aeolicus TadA)    SHMGKEYFLKVALREAKRAFEKG-EVPVGAIIVK--EGEIISKAHNSVEELK----D
sec.structure              -----HHHHHHHHHHHH----------EEEEEEE----EEEEEEE--HHHH------

          * *                                   *  *
AID   NGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGL---
1p60  ATLHGEISTLENCGRL--EGKVYKDTTLYTTLSPCDMCTGAIIMYG------IPRCVVGENVNFK-----SKGE---
1wkq  PTAHAEVTAIRKACKVL-GAYQLDDCILYTSCEPCPMCLGAIYWAR------PKAVFYAAEHTDAAEAGFDDS(23)
1vq2  NEIHAELNAILFAAEN--GSSIEG-ATMYVTLSPCPDCAKAIAQSG------IKKLVYCETYDKN------------
1wwr  PTAHAEMLAIKEACRRL-NTKYLEGCELYVTLEPCIMCSYALVLSR------IEKVIFSALDK------KHGGVVSV
str.  ----HHHHHHHHHHHHH---------EEEEE----HHHHHHHHHHHHH-----EEEEEE---------HHHHHH---

AID   -----RRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL
1p60  -----KYLQTRGHEVVVVDDERCKKIMKQFIDERPQDWFEDIGE---------------------------------
1wkq  YKEIDKPAEERTIPFYQVTLTEHLSPFQAWRNFANKKEY--------------------------------------
1vq2  KPGWDDILRNAGIEVFNVPKKNLNKLNWENI----------------------------------------------
1wwr  FNILDEPTLNHRVKWEYYPLEEASELLSEFFKKLRNNII--------------------------------------
str.  -----HHHHHH--EEEEE--HHHHHHHHHH---HHHH---HHHHHHHHHHHHHHHHHHHHHHHHH---HHHHHHHHH--
```

Figure 6. Model for the structure of AID. (A) Structure-based alignment of AID with other deaminases. The deaminases other than AID are identified by their protein data bank, PDB, identification number (1p60, 1wkq, etc). Consensus alignment between the human AID and deaminases with known structure was constructed using the protein fold-recognition methods. The mutual alignment of deaminase structures was guided by their structural superposition. Residues identified as being important for the reaction mechanism figure are indicated with asterisks. Amino acids omitted for clarity are indicated in brackets. (B) Structure of the AID monomer. The residues thought to be involved in SHM, but not CSR, are shown in red. The residues in green, when mutated, are known to abolish both SHM and CSR. The C-terminal residues involved in CSR, but not SHM, are shown in light blue. The zinc is shown as a magenta sphere. The substrate deoxycytidine is shown in yellow. (C) Structure of the AID dimer. The protein backbone is shown in the ribbon representation, with helices in cyan and strands in orange. SS DNA is shown in yellow and was docked with the dimer so as to insert a cytosine into the active site. The Zn-binding residues C87, C90, and H56, as well as the putative catalytic residue E58 and Y28, which stacks with the target base, are shown in the wire-frame representation (oxygen atoms are in red; sulfur atoms are in yellow).

Table 2. Sequence Preference of AID[a]

	hot spots		cold spots	
	sequence	no. of C to U	sequence	no. of C to U
	ATA**C**GC	34	TGC**C**CG	0
	ATG**C**TT	34	TCC**C**GA	0
	CAG**C**TA	33	CGC**C**TT	0
	CAA**C**TT	32	CGC**C**AG	0
	ATG**C**AG	31	GGC**C**GA	0
	AAA**C**CC	31	TtT**C**AC	1
	TTA**C**CC	31	CGC**C**TC	1
	AAA**C**CA	29	TCa**C**AC	1
	TAG**C**TG	28	ACa**C**AG	1
	AcG**C**AA	28	GGC**C**GT	1
	AAA**C**CG	28	CGT**C**GT	2
	TTG**C**AG	28	Taa**C**AA	2
	CAG**C**AC	28	TGg**C**CG	3
	AAA**C**AG	26	CGC**C**CA	3
	TTA**C**GA	26	TCT**C**AC	3
			CGa**C**AG	3
			CGT**C**GT	4
consensus:	WR**C**		SY**C**	

[a] W is A or T, R is purine, Y is pyrimidine, and S is G or C. Adapted with permission from Table 1 in ref 69. Copyright 2003 Nature Publishing Group.

domains that are in contact with the antigen, or sequences in DNA that reflect the DNA sequence specificity of one or more enzymes involved in SHM. Although it is likely that all these factors contribute to the observed sequence preferences in SHM, the last of these potential causes may make the largest contribution.

When SS M13 DNA was used as the substrate, AID converted multiple cytosines in each substrate molecule in a 230 nt *lacZ*α segment to uracil.[68,69] Many of the same cytosines were found mutated in multiple independent clones while some cytosines were rarely targeted by AID. A sequence analysis of the mutants revealed that while the hot spots had the consensus WRC, the cold spot consensus was SYC (S is G or C; ref 69 and Table 2). These data show that targeting of DNA by AID is based largely on two bases 5′ to the substrate cytosine and that its selectivity (or avoidance) of the target is a synergistic effect of selectivity at each of the two sites.

The consensus target for AID is very similar to the consensus sequence of the C:G hot spots in SHM,[66,70] and hence, it is likely that the former causes the latter. This correlation between targeting by AID and SHM hot spots is yet another piece of evidence that supports the idea that AID acts directly on Ig gene DNA rather than on an RNA. As AID is required for all hypermutations, it is also required for mutations at T:A pairs. However, the role played by AID in promoting mutations at T:A pairs is less clear, and this is discussed in section 6.

5.2. Processivity of AID

When an SS DNA substrate was used for AID *in vitro* and the DNA was subsequently introduced into *ung E. coli*, a large number of clustered mutations were observed (10 to 70 per clone in a 230 nt segment; ref 69). This preponderance of multiple closely spaced mutations is due either to multiple interactions of the substrate and AID, or processivity of the protein on DNA. If AID is processive, it may explain a subclass of SHMs that are clustered.[71] However, most SHMs are not clustered, and hence, the biological significance of this observation remains unclear. Furthermore, some additional considerations cast doubt on the idea that AID acts processively in SHM. One of those considerations is that the likely target for AID *in vivo* is Ig genes undergoing transcription and not SS DNA (see below).

When the *lacZ* gene fragment undergoing transcription from a T7 RNA polymerase promoter was used as the target for AID, the average number of mutations per clone was only ∼3 and about 50% of the LacZ$^-$ mutants had single mutations.[68] This is in stark contrast with the high degree of multiple clustered mutations reported when the SS DNA form of the same substrate was used (see above). Additionally, we have never observed multiple mutations in a genetic reversion assay where a transcribing *kan* gene was the target for AID (ref 72 and M.S. and A.S.B., unpublished results). This genetic system is capable of detecting revertants with multiple mutations including those mutants in which adjacent cytosines have been converted to thymines. Furthermore, when the same DS DNA is transcribed using T7 RNA polymerase, the fraction of revertants with multiple mutations appears to be restricted to a minority subpopulation of DNAs that are arrested during transcription (C. Canugovi and A.S.B., unpublished results). This contrasts with an average of 10 to 70 mutations per clone observed by Pham *et al.*,[69] who used a SS DNA substrate. Thus, the use of a nonphysiological substrate, SS DNA, may be responsible for the observation of apparent processivity by AID. It may not act in a processive manner on actively transcribing Ig genes. However, the reported processive action of AID[69] has been incorporated into certain models of SHM and is discussed further in section 9.

6. Role of DNA Repair in SHM

6.1. Uracil Excision Repair

Early evidence for the involvement of uracil excision in modulating the mutagenicity of AID was obtained by comparing wild-type *E. coli* with *ung* cells (phenotype: UDG$^-$). The Rif^R frequency was 9-fold higher in *ung* cells compared to *ung*$^+$ cells, suggesting that AID causes the conversion of cytosines in the chromosome to uracils.[16] In a related study using mice lacking UDG, SHM was affected by the UDG defect. Although the study did not determine overall mutation frequencies, the percent of mutations among hypermutated Ig genes that were C to T was 31% in UDG$^{+/+}$ mice and 52% in UDG$^{-/-}$ mice.[73] Interestingly, the distribution of mutations within the intronic region that was sequenced was similar in the two genetic backgrounds, suggesting that UDG was involved in determining the type of base substitutions found in SHM, but not their local distribution.[73] Similar results were also obtained in a chicken cell line where UDG was inhibited by expressing a specific inhibitor of the enzyme, UGI.[74] In this case, the frequency of C to T mutations increased from 38% of the total to 86% when UGI was expressed in the cell line.[75] Both the studies point to an important role for UDG in SHM and suggest that an intermediate in the SHM pathway is DNA containing uracils.

A couple of additional points should be made here. When UDG$^{-/-}$ mice were first described,[76] no phenotype could be attributed to the mutation. In fact, unlike *E. coli*, murine UDG$^{-/-}$ cells did not have a significant increase in mutation frequency. This was explained by the investigators as a consequence of the activity of a backup uracil−DNA glycosylase, SMUG1.[76,77] However, subsequent studies not only revealed altered SHM spectra and reduced CSR in these mice,[73] but also the presence of B-cell lymphomas and a slightly shortened life span.[78] These mice consistently showed lymphoproliferation and developed macroscopic hyperplasia of spleen and lymph nodes at 22 times the rate of the WT mice.[78] Although this study did not assay for mutations directly, it is reasonable to conclude that UDG must be the principal uracil removal enzyme in lymphatic cells and in its absence mutations and/or other genetic rearrangements occur in the cells at a much higher frequency. Thus, the load of uracils in DNA must be particularly high in these cells, supporting the hypothesis that AID (which is expressed only in lymphatic tissue) converts cytosines in DNA to uracils. These conclusions were recently confirmed in a separate study that investigated the relative importance of UDG and SMUG1 in the removal of uracil from DNA in lymphoid tissue.[79]

Some HIGM patients (HIGM5) have been found to have mutations in the UDG gene (refs 80 and 81, and Table 1). Three of the four UDG mutations found in these patients contain deletions that result in premature termination and a substantial shortening of the protein. It is reasonable to assume that the truncated proteins expressed in these cells are completely defective in uracil excision. The remaining patient was homozygous for the mutation F251S, and somewhat surprisingly, the mutant protein, when purified from *E. coli*, was fully active.[79] However, it was defective in transport to the nucleus, and as a result, uracil excision activity was substantially reduced in nuclear extracts from B lymphocyte cell lines derived from this patient. The extracts had 0.4% residual activity compared to extracts from $UDG^{+/+}$ individuals.[79] These results support an important role for UDG in CSR and have generally been interpreted to mean that UDG is required for the formation of DSBs in the switch regions that precede CSR.[73]

Begum *et al.*[82] have questioned such a role for UDG in CSR, indirectly questioning the importance of the catalytic ability of AID to convert cytosines in DNA to uracil. They found that the formation of γH2AX (i.e. phosphorylation of the minor histone H2AX) required AID but not UDG. H2AX is phosphorylated in response DNA strand breaks and is used as a readout for DSBs that occur during CSR. These investigators expressed UGI, a specific inhibitor of UDG, and found that γH2AX foci could still be observed in response to AID expression.[82] When any of the murine UDG single mutants, D145N, N204V, H268L, or F242S (equivalent to the human UDG mutant F251S mentioned above), were expressed in $UDG^{-/-}$ B-cells, CSR was normal, as evidenced by the titer of IgG. However, neither of the double mutants tested, D145N−N204V or H268L−D145N, could complement this defect. This apparent requirement for UDG in CSR was interpreted as “structural” rather than a catalytic requirement.[82]

It has been pointed out (ref 83 and G. Baldwin, personal communication) that the three single mutants, D145N, N204V, and H268L, of UDG are very powerful catalysts and can excise uracil from duplex DNA with a half-life of about 1 min. Thus, the CSR observed in the presence of these mutant proteins may simply be due to residual catalytic activities of these mutants. As mentioned above, the human equivalent of F242S mutation is catalytically active and the overexpression of this mutant from a retroviral vector is likely to result in nuclear accumulation of the active enzyme restoring CSR.[79] In contrast, the double mutants of UDG should be substantially more defective in catalytic activity than the single mutants and may have inadequate activity to promote CSR. These and other considerations suggest[79,83] that the results of Begum *et al.*[82] are, in fact, consistent with a role for both AID and UDG in the formation of DSBs that trigger CSR.

6.2. Translesion Synthesis DNA Polymerases

A number of DNA polymerases capable of synthesis across a variety of DNA lesions (translesion synthesis DNA polymerases or TLS Pols) have been implicated in SHM.[84] In particular, the role of the so-called Y-family DNA polymerases (Pol η, Pol ι, Pol κ, and Rev1) has been investigated most thoroughly because of their propensity to perform synthesis across bulky lesions and abasic sites. None is absolutely required for SHM, and eliminating some of the TLS Pols only modulates the hypermutation spectrum. It is likely that multiple polymerases participate in the steps that lead to SHM and may be able to compensate for each other.

The DNA polymerase whose involvement in SHM is best understood is Pol η. This polymerase is capable of synthesis across cyclobutane pyrimidine dimers and is missing in Xeroderma pigmentosum variant (XP-V) patients.[85,86] This synthesis is relatively “error-free”. However, pol η can also insert nucleotides across from an abasic site causing mutations,[87] and this may be its role in SHM (see below). The XP-V patients do not show HIGM syndrome, and in XP-V cells the frequency of hypermutations is normal.[88] Interestingly, however, the targeting of hypermutations changes in these cells. In the absence of Pol η, the mutations at T:A pairs were reduced from 54% to only 18%.[88] A similar bias toward mutations at G:C pairs was also observed during SHM in Pol η knockout mice.[89,90] In these animals, 79%[90] or 85%[89] of the total mutations occurred at G:C pairs. These studies also confirmed that a lack of Pol η does not strongly affect overall hypermutation frequencies. These studies identify Pol η as a major player in the targeting of T:A hot spots during SHM.

Additional support for this role comes from work in which *in vitro* copying of the κ light chain gene using Pol η gave rise to mutations that were consistent with the T:A hot-spot mutations.[67,91] The same data also suggest that Pol η may be preferentially copying the transcriptional template strand of the Ig genes to cause these hypermutations.[91] Finally, there are data that suggest a linkage between the roles of DNA mismatch repair and Pol η, and this will be discussed in section 6.3.

TLS Pol κ, μ, and λ have been shown not to play an essential role in SHM.[92−94] The other TLS Pols with a connection with SHM are Pol ι, ζ, and θ. However, there are contradictory data regarding the roles of these polymerases in the literature, and their precise role in SHM remains unclear. It was reported that when Pol ι gene was knocked out in a Burkitt’s lymphoma cell line in which SHM can be induced, SHM was eliminated.[95] This suggested that Pol ι must be required for SHM. However, this conclusion was contradicted by the observation that, in mice with a nonsense mutation in the Pol ι gene, the frequency and distribution of SHM were normal.[96] Furthermore, a mouse lacking both Pol η and Pol ι underwent hypermutations, and the mutation spectrum was similar to that in a Pol $\eta^{-/-}$ mouse.[89] These data cast further doubt about a role for Pol ι in SHM.

In an earlier study, Pol ζ transcripts in human B-cells were reduced by the use of Pol ζ-specific antisense oligonucleotides and this resulted in a reduction in SHM by a factor of up to 3.[97] A similar decrease in hypermutation frequency was also observed in mice expressing anti-Pol ζ anti-sense RNA.[98] Curiously, both the studies found that the hypermutation spectrum remained unchanged in Pol ζ-deficient cells.[97,98] This would suggest that Pol ζ plays a major role in causing hypermutations, but a mechanism with a mutational specificity similar to that for Pol ζ acts as a backup in SHM. However, this conclusion is inconsistent with data that suggest that Pol θ may play a major role in determining the SHM frequency and/or spectrum.

Zan *et al.*[99] reported that, in Pol θ knockout mice, the frequency of SHM decreased 2.6- to 5.0-fold without changing the ratio of mutations at C:G and A:T pairs. Thus, the SHM phenotypes of cells deficient in Pol θ and Pol ζ are quite similar to each other. It should be noted that Pol θ and Pol ζ belong to different DNA polymerase families, class B and class A, respectively. Further complicating our understanding of the role of TLS Pols in SHM is a recent study by Masuda *et al.*[100] This study found that a different mouse knockout of Pol θ exhibited only a slight reduction in overall SHM frequency (0.8% compared to 1.0% in WT mice). They also found a 41% reduction in mutation frequencies at C:G pairs (0.28% vs 0.48% in WT mice). It is also of interest to note that *PolQ*, the gene that encodes Pol θ, is specifically expressed in lymphoid tissues and abundant *polQ* transcripts are detected in germinal center B-cells, the target cells for both SHM and CSR.[101]

It should be clear from the discussion above that the role of TLS Pols in SHM is poorly understood at this time. This reflects partly our lack of a good understanding of the physiology of these enzymes, and as to when and how they participate in DNA synthesis. This is particularly true when it comes to the role of these enzymes in DNA synthesis during BER or mismatch repair (see the next section) as opposed to replicative DNA synthesis. For example, it is possible that two or more of these enzymes form a complex and the absence of one enzyme disrupts the whole complex. This would explain the reports of similar mutational phenotypes in cell lines missing different TLS Pols. Additionally, different Pols may partially compensate for each other, preventing a "clean" knockout phenotype. Finally, different TLS Pols may be involved in causing mutations at C:G and A:T sites, making the analysis of mutation spectra difficult.

6.3. DNA Mismatch Repair

All organisms possess the ability to correct replication errors that have been overlooked by the proofreading ability of DNA polymerases. Because the correction occurs on two normal DNA bases that are incorrectly paired together, it is referred to as mismatch repair (MMR). The molecular steps of MMR will be described below in a cartoon fashion, and the reader is referred to a specialized review (see ref 102) for further details.

In human cells, MMR is initiated by the binding of heterodimer of MSH2 and MSH6 at the mismatch (Figure 7). A second heterodimer containing MSH2 and MSH3 can initiate the repair of short extrahelical loops and is probably not relevant to antibody maturation. The mismatch-bound MSH2/MSH6 heterodimer undergoes an ATP-dependent conformational change, which converts it to a sliding clamp capable of translocating along the DNA. The MSH2/MSH6•ATP•DNA complex is bound by a second heterodimer, composed of MLHl and PMS2 in a second ATP-dependent step. This complex can translocate in either direction, in search of a strand discontinuity (Figure 7). A key requirement of MMR is that it must replace the base from the newly synthesized strand and not the "old" strand. The only known mechanism for this discrimination in eukaryotes is the gap between Okazaki fragments on the lagging strand, or the 3′-terminus on the leading strand. *In vitro*, MMR can be reconstituted using DNA substrates that are not actively undergoing replication but contain nicks or gaps, and hence, it is plausible that this could also occur in antibody maturation. EXO1, a 5′ to 3′ exonuclease, is stimulated by the traveling MSH2/MSH6•MLH1/PMS2 complex and can start from a nick situated 5′ from the mispair and travel toward the mispair, creating a gap. The region of single-stranded DNA is stabilized by replication protein A (RPA) (Figure 7). A 3′ to 5′ exonuclease activity (probably within EXO1 itself) can similarly travel from a nick 3′ to the mismatch, creating a gap in the other direction and allowing bidirectional MMR. Other proteins known to play a role in MMR are RFC and PCNA (Figure 7).

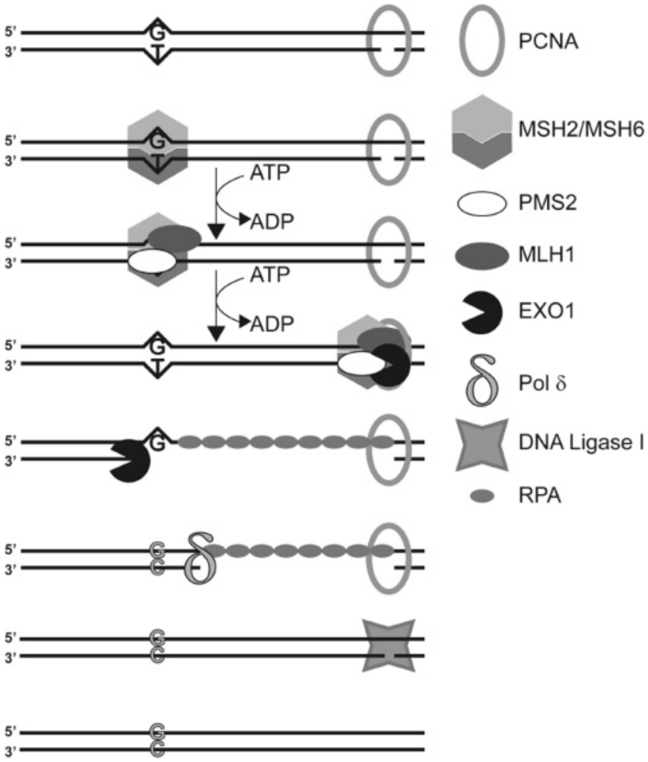


Figure 7. Principal steps during mismatch repair. The repair of a T•G mispair generated as a result of replication error is shown. The proteins involved in this repair are shown in a cartoon fashion and identified in the figure. The DNA substrate contains a strand discontinuity (bottom strand) presumably due to a gap between two Okazaki fragments. For details, see the text. Adapted from a figure in ref 102.

A series of papers a few years ago showed that mouse knockouts of MSH2, MSH6, MLH1, or PMS2 genes continue to undergo SHM, but at a reduced frequency.[103−108] Although there are some differences in the reported decreases in SHM levels, the more interesting observation is that, in the MMR deficient animals, SHM showed a stronger bias toward mutating C:G pairs than T:A pairs. In one report,[107] hypermutations at C:G pairs increased from 42% in WT to 91% in MSH2$^{-/-}$ mice. Similarly, in MSH6$^{-/-}$ mice the mutations at C:G were 87% of the total compared to 46% in WT animals.[103] A smaller increase in mutations at C:G pairs was reported for MLH1$^{-/-}$ and PMS2$^{-/-}$ mice in one study,[105] while another report did not find any significant increases in PMS2$^{-/-}$ or MLH1$^{-/-}$ animals.[106,107] Although there are some inconsistencies in the data, it is clear that the mismatch recognizing proteins MSH2 and MSH6 have a stronger effect on the mutation spectra than MLH1 and PMS2.

Another interesting observation regarding mutations in mice deficient in MSH2 or MSH6 is that they occurred at the same WRCY hot spots found in WT animals, and the hot spots tended to get hotter.[103,107,108] This observation led Rada *et al.*[108] to suggest that there are two phases in targeting of mutations in Ig genes. In the first phase, mutations were targeted at C:G pairs within WRCY sequence motifs by an unknown factor, and in the second phase, the mutations at C:G pairs were suppressed by MMR while increasing mutations at T:A pairs at the same time.[108] It is likely that the first phase of mutational targeting is performed by AID by its ability to deaminate cytosines within WRCY sequences in DNA.[69] The mechanism by which MMR increases mutations at T:A sites in phase II is less well understood.

The studies of SHM and CSR in EXO1$^{-/-}$ mice also confirm a role for MMR in antibody maturation. These mice are defective in CSR compared to WT and their heterozygote siblings.[109] SHM was also affected in these animals. While the frequency of hypermutations remained unchanged, the mutations shifted to C:G pairs. Furthermore, mutations within WRCY hot spots also increased. Thus, the effects of EXO1 defects and MSH2 defects are very similar, suggesting that they both affect the same mutational subpathways that contribute to SHM and CSR.

As noted above, cells deficient in Pol η also show a bias toward C:G targeted mutations. This suggests that MMR may use DNA synthesis by Pol η in extending the mutational process that begins at a C:G pair (presumably within a WRCY context) to T:A base pairs. Evidence has also been presented that both the MSH2•MSH6 and MSH2•MSH3 complexes bind Pol η, but not Pol ι.[110] Furthermore, MSH2•MSH6, but not MSH2•MSH3, stimulated primer extension by Pol η. The rate enhancement by MSH2•MSH6 was ~6-fold, while K_M also increased by 2.6-fold.[110] Thus, MSH2•MSH6 increased the catalytic efficiency of Pol η by a factor of only 2.3-fold. MSH2•MSH6 also had little effect on the processivity of the polymerase or its fidelity. Thus, the overall effect of the MSH2•MSH6 complex on Pol η activity is small, raising concerns about whether such an interaction has a significant effect on antibody maturation.

When an MSH2 defect in mice is combined with a UDG defect, SHM alters in an interesting way. The overall hypermutation frequency remains unchanged, but now essentally all the mutations are targeted at C:G pairs and are C to T.[111] This contrasts the case of the MSH2$^{-/-}$ mice, where 26% of the mutations were still at T:A pairs. This dramatic shift in the mutation spectrum in the double mutant suggests that MSH2•MSH6 is not the only complex that directs mutations to T:A base pairs. It is likely that, in the absence of MMR, processing of U•G mismatches by BER somehow results in shifting some mutations to T:A pairs. The near complete absence of non-C-to-T hypermutations in MSH2$^{-/-}$ UDG$^{-/-}$ mice also suggests that no DNA glycosylase is available to process U•G mismatches in activated B lymphocytes and the uracil is accurately copied by replication polymerases. Also, the fact that not all the mutations are C to T in UDG$^{-/-}$ MMR^{+} mice[51] says that the MMR process finds a way of using the U•G mismatches created by AID for error-prone repair that shifts mutations away from the U•G mispair.

Despite a number of interesting observations and some tantalizing clues, the role of MMR in SHM remains far from clear. The principal difficulty in understanding this role is that MMR acts during SHM in ways that are contrary to its perceived function—avoidance of replication errors. It appears to actively promote misincorporations in DNA by recruiting error-prone DNA polymerases such as Pol η. This would create a paradoxical and potentially explosive situation where MMR promotes creation of mismatches which it then must try to repair! This futile cycle cannot be sustained and in other situations, such as repair of O6-methyl-G:C pairs, is known to cause cell death.[112] Clearly, MMR cannot work this way during SHM.

There are other gaps in our understanding of how MMR is involved in SHM. One concerns the possibility that the MSH2•MSH6 complex may bind U•G pairs generated by AID or an abasic site arising from it. This is the simplest explanation for the mutational spectrum in MSH2$^{-/-}$ UDG$^{-/-}$ mice[111] but is puzzling. About 80 U•G mismatches are created in the human genome per generation from nonenzymatic hydrolytic deamination of cytosines.[113] These occur throughout the cell cycle and are unrelated to DNA replication. It is clear that although MSH2—MSH6 can bind a U•G mispair *in vitro* (it is really not that different from a T•G mispair it must frequently repair) MMR is thought not to interact with these nonenzymatically generated mismatches and they are handled exclusively by UDG and other BER enzymes. In fact, interference by MMR in BER of these U•G pairs would be disastrous, as MMR does not have any intrinsic discrimination between a U and a G. The repair of these nonenzymatically generated mismatches by MMR would create about 40 C to T mutations (one-half of 80) per generation. It is believed that the rate of mutations in human cells is about 50 times lower (less than one mutation per cell per generation; ref 114). Thus, the possibility that MMR routinely "repairs" most U•G mispairs is inconsistent with the observed mutation rate in human cells. If MMR does not repair U•G pairs obtained from nonenzymatic deaminations, how can it act on AID-generated U•G pairs? In other words, what factor(s) target MMR proteins to Ig gene undergoing SHM?

Another problem is that, in contrast to the case of prokaryotes, MMR in eukaryotes is incapable of generating the free 3′-OH needed for EXO1 action and must rely on preexisting nicks or gaps in DNA. What is the source of these nicks in the Ig genes for MMR to function? Processing of uracils by BER does generate transient nicks and gaps, but this also eliminates the U•G mismatches that MSH2—MSH6 may need to bind to participate in SHM. Additionally, a UDG$^{-/-}$ mouse has a different hypermutational spectrum than a UDG$^{-/-}$ MSH2$^{-/-}$ mouse,[73,111] suggesting that MMR does affect the SHM spectrum in mice defective in UDG. In principle, it is possible that there is a yet undiscovered U•G mismatch-specific endonuclease that nicks either DNA strand and helps initiate MMR. However, it would have to be B-cell-specific, as it would otherwise interfere with BER of U•G pairs elsewhere. Another solution to this problem may lie with the reported processivity of AID.[69] If AID generates a large number of uracils in Ig genes, some may be partially repaired by BER, while others may remain unrepaired. In such a situation, MMR may step in and initiate repair of U•G mispairs that have not been repaired by BER and use the nearby nicks generated by the partial repair of other U•G mispairs by BER to initiate DNA synthesis. A similar model for the role of MMR in CSR has recently been proposed by Schrader, Stavnezer, and colleagues.[115,116]

Finally, MMR is thought to be a "long patch" repair process, and this is not compatible with the low processivity of the translesion synthesis DNA polymerases. In other words, most scenarios for the involvement of TLS Pols and MMR in SHM force the latter to either become a "short patch" repair process or require a switch to a high fidelity polymerase such as Pol δ after one or two nucleotide incorporations. One observation that lends support to "short patch" repair by MMR during SHM is the relatively modest effects of MLH1 and PMS2 mutations on SHM (see above). As the binding of MLH1/PMS2 dimer to the MSH2•MSH6 complex is believed to precede the translocation of the latter molecule along DNA (Figure 7 and ref 102), the absence of the former dimer may keep the latter complex near the mismatch it binds to. However, many biochemical details including the logistics of a polymerase switch during DNA synthesis are poorly understood at this time.[117,118] In summary, we know that the MSH2•MSH6 complex plays a key role in shaping the SHM spectrum, especially at A:T pairs, that it may act without the aid of the MLH1/PMS2 dimer, and that it probably acts through a direct interaction with Pol η (and/or some other TLS Pols). However, a conceptual (or experimental) breakthrough is needed before a detailed molecular model for this process can be constructed.

7. Mutagenesis by AID

7.1. AID and C to T Hypermutations

The simplest explanation for the C to T mutations within SHM and switch region mutations is that they result from unrepaired uracils generated by AID. In this model, a certain fraction of uracils created by AID through deamination of cytosines escape repair by UDG, and these are eventually replicated to create C to T mutations (Figure 4). In *E. coli*, an *ung* mutant has a ~10-fold higher frequency of C to T mutations than its WT parent, suggesting that 9 out of 10 uracils in chromosomal DNA resulting from cytosine deamination are excised by UDG. If UDG has a similar efficiency in B lymphocytes, AID must deaminate ~10 times as many cytosines as there are C to T hypermutations. Typically, C to T mutations are ~25% of all SHMs, and hence, AID may generate ~2.5 times as many uracils in DNA as there are SHMs. Alternately, uracil repair in B lymphocytes could be much less efficient than in *E. coli*, and most uracil generated by AID may ultimately result in SHM. Which one of these two models is correct can be determined if the amount of uracil generated by AID in the variable segment of Ig genes could be quantified. This is a technically challenging goal where the presence of uracil must be determined in a specific 0.00003% (~1000 bp out of 3×10^9 bp) of the genome at a sensitivity of ~1 in 300 nt or better. Although this has never been done before, a complete understanding of SHM cannot be achieved without it.

7.2. AID and Non-C-to-T Hypermutations

While it is easier to understand how uracils generated by AID in DNA may cause C to T mutations, the origin of all other base substitutions and frame-shift mutations (hereafter referred to as non-C-to-T mutations) is much less clear. One possibility is that incomplete repair of uracils in DNA may generate non-C-to-T mutations. As mentioned above, even when MMR is absent, a significant fraction of the mutations occur at T:A pairs. The likely mechanism for these mutations is incomplete BER that leaves a nick which is converted to a gap by exonucleases, and the filling-in of these gaps by Pol η or other TLS Pols creates mutations at T:A pairs (Figure 8). Thus, repair of U•G can have three consequences: (1) complete, accurate BER resulting in no mutations; (2) incomplete BER resulting in mutations at C:G as well as T:A; and (3) no repair resulting in C to T mutations (Figure 8).

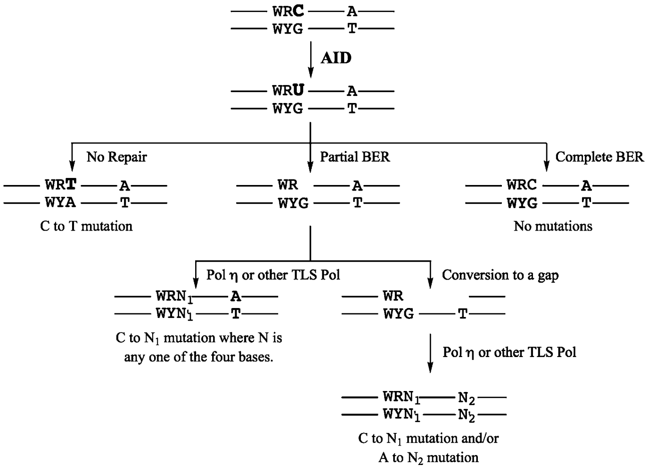


Figure 8. Processing of U•G mismatches generated by AID. The three possible pathways by which U•G mispairs created by AID may be processed are shown. The mutational consequences, if any, are also indicated in each case.

However, many more non-C-to-T mutations are created because of the involvement of MMR. It is also clear that TLS Pols, especially Pol η, play key roles in this process. Unfortunately, no plausible detailed molecular model for the involvement of MMR in SHM exists currently, and hence, the non-C-to-T mutations in SHM cannot be satisfactorily explained.

8. Role of Transcription in SHM

It has been recognized for some time that transcription of the rearranged Ig gene is essential for both SHM and CSR.[14,119,120] Recently, several lines of evidence have converged to highlight the connection between transcription and SHM and CSR. Immunoprecipitation experiments have found that AID associates with a complex containing RNA polymerase II (RNAP II).[121] Other experiments have found that there is a quantitative correlation between the level of expression of the target gene for mutations and the frequency of SHM within it. In B-cells this requirement for high transcription is met by the presence of enhancers in or near Ig genes, but many experiments have shown that the effect is not specific for the V(D)J promoter or the enhancers.[122] For example, a defective GFP gene expressed from a tetracycline-controlled promoter in a hypermutation-active pre-B-cell line accumulated mutations at a rate that was proportional to the level of transcription of the GFP gene.[123] Similar results were also obtained in a fibroblast cell line transfected with the AID gene.[124] Similarly, CSR is also stimulated by transcription of the switch region[122] and the directionality of transcription may be important for this effect.[125]

Ramiro *et al.*[35] and Sohail *et al.*[36] showed that when AID is expressed in *E. coli* from a native promoter, its mutagenicity is enhanced 20- to 50-fold by the transcription of the target gene. Further, Chaudhuri *et al.*[44] and Sohail *et al.*[36] showed that the same was true *in vitro*. When AID partially purified from human cells[44] or from *E. coli*[36] was used in an *in vitro* transcription reaction involving T7 RNA polymerase (T7 RNAP), the cytosine deaminations caused by AID increased 10- to ∼1000-fold. The fact that AID acts in a transcription-dependent manner when the target gene is transcribed by either the *E. coli* or T7 RNAP suggests that AID recognizes some feature of the transcription bubble rather than a specific RNAP.

8.1. Strand Bias in AID Action

A remarkable property of the transcription-dependence of AID action is its strand bias. Both in *E. coli* and *in vitro*, AID preferentially deaminates cytosines in the nontranscribed strand (nontemplate strand; NTS) compared to the transcribed strand (template strand; TS). Consequently, when an *ung* (i.e. UDG-deficient) host is used, AID promotes C to T mutations in *E. coli* preferentially in the NTS of the target gene. *In vivo* the cytosines are 20 to 50 times more frequent targets for deamination when they are in the NTS compared to the TS.[35,36] Recently, Martomo *et al.*[39] confirmed this observation biochemically and showed that uracils accumulate preferentially in the NTS of a gene in *E. coli* expressing AID. Their results differed somewhat from the results of genetic assays in that the biochemical assays found only a 2-fold difference in the accumulation of uracils in the NTS compared to the TS.[39] The reasons for this discrepancy between the genetic and biochemical assays for the magnitude of the strand bias in AID action are unclear. However, when DNA being transcribed *in vitro* is treated with AID, the bias in favor of converting cytosines in the NTS is at least 10-fold and may be as high as 100-fold,[36,44] suggesting that the bias is likely to be much greater than 2-fold.

We have previously shown that the NTS in transcribed genes of *E. coli* is much more accessible to reactive chemicals and acquires more DNA damage.[126,127] Specifically, nonenzymatic conversion of cytosines to uracil by water and of guanine to 8-oxoguanine by reactive oxygen species occurs at 6 to 40 times higher frequency in NTS than in TS.[126−131] Thus, AID shows the same transcriptional strand bias as seen with simple reactive chemicals in *E. coli*. The preference of AID for SS DNA may partially explain this strand bias. However, this cannot be the whole story, because not all chemicals can access NTS as well as water and reactive oxygen species (M. Sanath Kumar and A.S.B., unpublished results). As mentioned earlier, it is likely that AID somehow recognizes the transcription bubble itself and not just the NTS.

This observed strand bias of AID was not anticipated because a strand bias in C to T (G to A) mutations is not found in SHM. In other words, WRCY sequences in either DNA strand can be hot spots for hypermutations. Thus, either the observed strand bias of AID is an aberration of the experimental system or the original bias in AID action is somehow “lost” during subsequent DNA processing. This contradiction between the action of AID in *E. coli* and *in vitro* and the absence of bias in SHM has led to several alternate models for the involvement of transcription in AID action (see section 9).

8.2. Roles of Phosphorylation and RPA in AID Action

Alt and colleagues have struck a somewhat different theme[50,132] regarding the propensity of AID to act on genes undergoing transcription. They report that this activity is regulated by the phosphorylation of AID on Ser-38.[132] Another residue, Tyr-184, is also phosphorylated in B-cells, but the significance of this phosphorylation to AID activity is unclear.[132] Both the phosphorylations are performed by protein kinase A (PKA), and this creates the physiologically active form of the protein. Thus, PKA and a phosphatase modify AID to respectively turn it on and off.[132]

They also make a distinction between the activity of AID on SS DNA and that on the presumed physiological substrate, DS DNA, undergoing transcription (DS-T DNA). They find that the partially purified phosphorylated form of AID (AID-P) deaminates cytosines from both SS and DS-T DNAs; the unphosphorylated form (AID-UP) acts only on SS DNA.[50] Furthermore, when the AID-P is purified to apparent homogeneity from B-cells, it loses its ability to act upon DS-T DNA. This activity is restored when the single-strand DNA-binding protein, RPA, is added to the reaction.[50] Co-immunoprecipitation and other biochemical assays have been used to show that the 32 kDa subunit of RPA interacts with AID-P but not AID-UP. Thus, in this view of how AID finds transcriptionally active Ig genes, RPA plays a critical role.

These results are not consistent with data presented by other research groups.[36,68,69] In the latter studies, AID purified from insect cells[68,69] or *E. coli*,[36] which is likely to be unphosphorylated, acts robustly on genes actively transcribed *in vitro*. Additionally, the ability of AID to act on *E. coli* genes *in vivo* depends strongly on the transcription of the genes.[35,36] It has been suggested[44,50] that some genes form R-loops when transcribed and the SS DNA within the R-loop may be targeted by AID-UP, explaining the difference between the two sets of results. However, we find no correlation between the presence of R-loops and AID activity on DS-T DNA (C. Canugovi and A.S.B., unpublished results). Furthermore, there is little support in the transcription factor literature that RPA is part of the transcription elongation complex. It is still possible that the differences between the two sets of results reflect some subtle differences in the biochemical assays employed and that they can be reconciled.

9. Models for How AID May Target Transcribing Genes

Some earlier models regarding the role of transcription in SHM and CSR included the involvement of specific transcription factors or the RNAP II itself in recruiting AID to specific promoters. While such interactions cannot be ruled out, the work with AID in *E. coli* and *in vitro* strongly suggests that they are not a requirement for the transcription dependence of SHM. Subsequently, several other models have been proposed to explain either the dependence of SHM and CSR on transcription or other specific properties of SHM. These properties include strand bias in mutations (or the lack thereof), clustering of mutations, and acquisition of non-C-to-T mutations. These models for the involvement of transcription in SHM are outlined below (Figure 9). (It should be noted that these models are not mutually exclusive; two or more mechanisms may be active in causing AID-promoted mutations in SHM.)

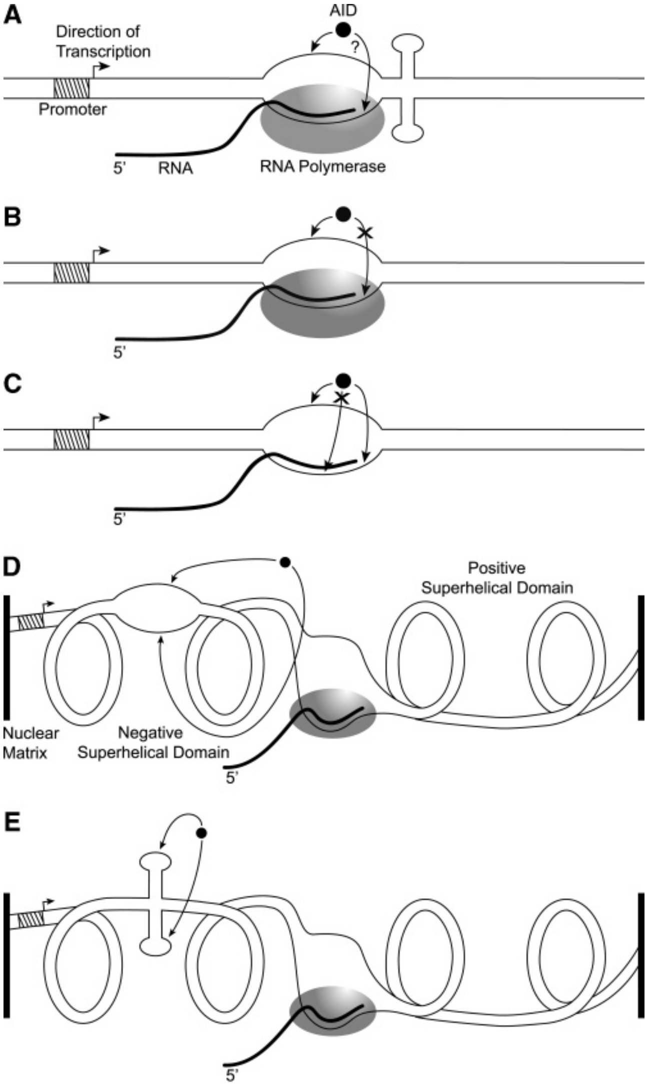


Figure 9. Models for the role of transcription in SHM. Various models regarding the role of transcription in SHM are presented (A through E). The positions of the melted segment of DNA in part D and of the cruciform structure in part E are arbitrary. These structures can form anywhere between the point of attachment of the DNA upstream of the promoter and the transcription bubble. See the text for additional details.

9.1. Transcriptional Pause Model

RNAP often pauses at certain sequences and arrests at others. Several years ago, it was proposed[14,133] that a "mutator factor" (now believed to be AID) would act at pause sites. In the original formulation, the mutations were attributed to faulty transcription-coupled repair, but they must now be ascribed to the action of AID itself. The NTS in the bubble at a pause site is more accessible than the TS, and hence, this model would predict that more mutations arise due to damage to cytosines in the NTS than in the TS (Figure 9A). It may also provide an extended time during which AID can repeatedly act within the bubble. This could create clustered mutations observed in some studies of SHM[71,134] and create multiple U•G mismatches needed for some models of SHM.

9.2. Bubble-Access Model

We have argued for some time[126] that the NTS in an elongation complex is accessible to chemicals and that this is true even *without* any pausing or arrest of the RNAP (Figure 9B). The action of AID in *E. coli*[35,36] and *in vitro*[36] is consistent with this idea. This model differs from the transcriptional pause model above in that it does not predict clustering of hypermutations. This is because AID catalysis is unlikely to keep pace with the speed of the transcription bubble, ~30 nt/s. We have pointed out that[72] most sequence-specific DNA-binding enzymes such as DNA methyltransferases and restriction endonucleases turn over at a rate of ~1 per min. If AID is similarly slow in its catalytic turnover, it would fall off DNA as soon as the transcription bubble passes it by. No studies of the interaction of AID with a stable elongation complex have been reported.

9.3. R-Loop Model

It has been suggested that the pre-mRNA resulting from transcription of the Ig gene may not be removed, thus creating R-loops. These have been specifically observed in Ig genes undergoing CSR[135] but have also been suggested in the case of SHM.[44] In this case, a large section of the NTS would be accessible to AID for a prolonged period of time (Figure 9C). The TS in the R-loop should largely be protected by the RNA, but it may also be available to AID at the edges of the bubble. Recent studies of the interaction of AID with an artificial R-loop show that the DNA in NTS is indeed accessible to AID.[136] This model would also predict multiple U•G mismatches in the Ig gene.

9.4. Superhelical Domain Model

Transcription of a gene creates a wave of positive supercoiling ahead of the RNAP and negative supercoiling behind it. This "twin-domain" model of transcription-induced topological changes in DNA[137] suggests that the DNA near the 5′ end of a gene (or upstream of a gene) tends to be underwound and can "breathe" more easily. Shen and Storb suggest[138] that heavy transcription of rearranged Ig genes underwinds the 5′ end of the gene and transient opening of DNA here makes it accessible to AID (Figure 9D). It is possible that AID, once bound to SS DNA in underwound regions, can travel some distance before falling off and creating multiple cytosine deaminations.

An attractive feature of this model is that negative superhelicity exposes both the DNA strands to AID and hence there is no strand bias in the resulting cytosine deaminations. While this is consistent with the mutational spectra in SHM, it is inconsistent with the data obtained using *E. coli* and *in vitro* transcription. As noted earlier, uracils accumulate in a strand-biased fashion in *E. coli* and *in vitro* when the target gene is transcribed.[35,36,39,44]

Another point to note is that transcription-driven superhelical domains should not be restricted to the gene being transcribed. They can extend both upstream and downstream of the gene. This would predict that the 5′ edge of SHM could be upstream of its promoter. However, the SHM data clearly show that hypermutations rarely occur 5′ of the promoter.[10,139] Some modifications to this model may be necessary to accommodate this fact.

9.5. Stem–Loop Structure Model

This is a variation on the superhelical domain model. Wright has pointed out that[140,141] if certain sequences in Ig genes contain inversely repeated sequences, they would tend to form stem−loop structures (SLS) when the DNA becomes underwound. The stability of these structures would be different from structure to structure based on the length of the stem, the length of the loop, G + C content, etc. If such structures are moderately stable, the cytosines in their loops would be accessible to AID (Figure 9E). She has found some correlation between the stability of the potential SLS and the occurrence of hypermutation hot spots.[141] Most of the predictions of this model are similar to those of the superhelical domain model discussed above, and the two models have many of the same strengths and weaknesses.

10. Antibody Maturation and Cancer

Malignant transformation is frequently associated with genomic instability and chromosome translocations. In particular, lymphomas often contain translocations involving the immunoglobulin (Ig) genes and oncogenes such as *c-myc* and *bcl2*.[142] One such translocation between IgH and *c-myc* is induced by IL6 and was studied in Balb/c mice expressing an IL6 transgene.[143] Two types of mice were used in this study; one type was defective in AID (genotype AID$^{-/-}$), and these mice did not undergo antibody maturation. When lymphatic hyperplasia were studied, the control group (genotype AID$^{+/-}$), but not their AID$^{-/-}$ siblings, contained translocations between IgH and *c-myc* genes.[143] Thus, the DNA rearrangements initiated during SHM and CSR may sometimes lead to chromosomal translocations that activate protooncogenes and contribute to tumorigenesis.

Another link between antibody maturation and cancer was demonstrated by Okazaki *et al.*[144] by studying a mouse with an AID transgene. They found that mice expressing an AID transgene constitutively had substantially shorter life spans and enlarged lymphoid organs. These mice developed T cell lymphomas and micro-adenomas or dysgenetic lesions of respiratory bronchiole.[144] Although the lymphomas did not contain higher than normal frequency of translocations, the *c-myc* gene did accumulate high levels of mutations in the region encompassing the exon 1 and intron 1, the hot spot of mutations and breakpoints of translocations in the B-cell. These and other studies show that if the DNA processing steps required for antibody maturation occur ectopically, they can lead to mutagenesis and cancer.

11. Antibody Maturation and Molecular Evolution

A bedrock principle of biology for the past 100 years has been that organisms try to maintain a low rate of mutations. This is because most mutations are harmful to the organism. Consequently, cells make extensive efforts to prevent DNA damage and to repair any damage that escapes the preventive mechanisms. Affinity maturation of antibodies in higher vertebrates is an exception to this rule. Here a class of cells, B lymphocytes, in the body introduce damage to DNA *in a programmed fashion* leading to mutations in one part of one (or a few) chromosomal gene. This creates a population of cells with different levels of Darwinian "fitness" for combating an infection. Furthermore, the process that acts upon these mutant cell populations and selectively expands the clones to make better antibodies uses polypeptides and other molecules derived from the agent that caused antibody maturation in the first place. Thus, the infectious agent itself causes the evolution of B-cells, making them more adept at the destruction of the agent. This is Lamarckian evolution at work—something that does not occur during the evolution of whole organisms. These cellular events have no clear parallels elsewhere in biology and are of intrinsic interest for understanding the interplay between mutations and selection that is inherent to biological evolution that has been ongoing for millions of years.

12. Concluding Remarks

In the past five years, considerable progress has been made in understanding how higher eukaryotes alter the antibody proteins so that the circulating antigens fit well within their binding pockets. Many key proteins required for this unique mutational pathway have been identified, and the role of some of these proteins in SHM and CSR is fairly well understood. In particular, the discovery of AID, an enzyme essential for antibody maturation, and the demonstration that it actively damages DNA have been exciting developments. However, there are some areas of chemistry and enzymology where important challenges lie ahead. These include the following: (1) Very little is directly known about the structure and reaction mechanism of AID. There is no detailed kinetics of this enzyme published, nor have inhibitors of the enzyme been validated. Considering the potential role of this family of enzymes in promoting mutations in oncogenes, designing and testing inhibitors for them should be an important goal. (2) What directs AID, UDG, MSH2, and other proteins to the rearranged transcribing Ig genes? Clearly, high transcription of the Ig genes plays a role in this selectivity, but additional factors should be involved in preventing genome-wide mutations during antibody maturation. These factors are likely to be associated with chromatin-modifying enzymes to allow greater access to the Ig locus. (3) Major hurdles remain in understanding the role of TLS Pols and MMR in antibody maturation. The richness of the TLS Pol families in mammalian cells in itself makes it difficult to present testable models for their role in SHM. This role will become clearer only when we have a better understanding of the biochemical interactions of these enzymes with other components of the DNA replication apparatus and MMR.

Finally, is antibody maturation the only process in biology that performs programmed genetic rearrangements that depend on enzymatically damaging DNA? Given its success in mammals and other eukaryotes, it would be reasonable to assume that it has evolved in other contexts where a high degree of structural variability is needed. Possible biological systems where such a program of DNA damage and mutations may be useful are receptors that recognize a large number of structurally similar chemicals, and pathogens that must evade host attack based on proteins such as antibodies. It would be exciting if we were to find that other biological systems have also used this inherently risky path of controlled mutagenesis to their advantage.

13. Acknowledgments

This work was supported by funds from NIH Grants GM57200 (to A.S.B.) and CA97899 (to A.S.B. and J.M.B.). We would like to thank Tami Zellner for help with some of the figures. We would like to thank all the reviewers for providing thoughtful comments on the manuscript and would like to particularly single out reviewer no. 1 for his/her detailed insightful critique.

14. References

(1) Janeway, C. A.; Travers, P.; Walport, M.; Shlomchik, M. *Imunobiology*, 5th ed.; Garland Publishing: London, 2001.

(2) Stavnezer, J.; Kinoshita, K.; Muramatsu, M.; Honjo, T. In *Molecular Biology of B Cells*; Honjo, T., Alt, F. W., Neuberger, M. S., Eds.; Elsevier Academic Press: London, 2004.
(3) Gellert, M. *Annu. Rev. Genet.* **1992**, *26*, 425.
(4) Sekiguchi, J.; Alt, F. W.; Oettinger, M. In *Molecular Biology of B Cells*; Honjo, T., Alt, F. W., Neuberger, M. S., Eds.; Elsevier Academic Press: London, 2004.
(5) Burnet, F. M. *The Clonal Selection Theory of Acquired Immunity*; Cambridge University Press: Cambridge, England, 1959.
(6) Talmage, D. W. *Science* **1959**, *129*, 1643.
(7) Maclennan, I. C. M.; Hardie, D. L. In *Molecular Biology of B Cells*; Honjo, T., Alt, F. W., Neuberger, M. S., Eds.; Elsevier Academic Press: London, 2004.
(8) Papavasiliou, F. N.; Schatz, D. G. *Cell* **2002**, *109*, S35.
(9) Berek, C.; Milstein, C. *Immunol. Rev.* **1988**, *105*, 5.
(10) Lebecque, S. G.; Gearhart, P. J. *J. Exp. Med.* **1990**, *172*, 1717.
(11) Rada, C.; Yelamos, J.; Dean, W.; Milstein, C. *Eur. J. Immunol.* **1997**, *27*, 3115.
(12) Maizels, N.; Scharff, M. D. In *Molecular Biology of B Cells*; Honjo, T., Alt, F. W., Neuberger, M. S., Eds.; Elsevier Academic Press: London, 2004.
(13) Neuberger, M. S.; Milstein, C. *Curr. Opin. Immunol.* **1995**, *7*, 248.
(14) Storb, U. *Curr. Opin. Immunol.* **1996**, *8*, 206.
(15) Muramatsu, M.; Kinoshita, K.; Fagarasan, S.; Yamada, S.; Shinkai, Y.; Honjo, T. *Cell* **2000**, *102*, 553.
(16) Petersen-Mahrt, S. K.; Harris, R. S.; Neuberger, M. S. *Nature* **2002**, *418*, 99.
(17) Muramatsu, M.; Sankaranand, V. S.; Anant, S.; Sugai, M.; Kinoshita, K.; Davidson, N. O.; Honjo, T. *J. Biol. Chem.* **1999**, *274*, 18470.
(18) Muto, T.; Muramatsu, M.; Taniwaki, M.; Kinoshita, K.; Honjo, T. *Genomics* **2000**, *68*, 85.
(19) Revy, P.; Muto, T.; Levy, Y.; Geissmann, F.; Plebani, A.; Sanal, O.; Catalan, N.; Forveille, M.; Dufourcq-Labelouse, R.; Gennery, A.; Tezcan, I.; Ersoy, F.; Kayserili, H.; Ugazio, A. G.; Brousse, N.; Muramatsu, M.; Notarangelo, L. D.; Kinoshita, K.; Honjo, T.; Fischer, A.; Durandy, A. *Cell* **2000**, *102*, 565.
(20) Arakawa, H.; Hauschild, J.; Buerstedde, J. M. *Science* **2002**, *295*, 1301.
(21) Harris, R. S.; Sale, J. E.; Petersen-Mahrt, S. K.; Neuberger, M. S. *Curr. Biol.* **2002**, *12*, 435.
(22) Dunnick, W.; Wilson, M.; Stavnezer, J. *Mol. Cell Biol.* **1989**, *9*, 1850.
(23) Nagaoka, H.; Muramatsu, M.; Yamamura, N.; Kinoshita, K.; Honjo, T. *J. Exp. Med.* **2002**, *195*, 529.
(24) Mehta, A.; Kinter, M. T.; Sherman, N. E.; Driscoll, D. M. *Mol. Cell Biol.* **2000**, *20*, 1846.
(25) Goossens, T.; Klein, U.; Kuppers, R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2463.
(26) Bross, L.; Fukita, Y.; McBlane, F.; Demolliere, C.; Rajewsky, K.; Jacobs, H. *Immunity* **2000**, *13*, 589.
(27) Papavasiliou, F. N.; Schatz, D. G. *Nature* **2000**, *408*, 216.
(28) Sale, J. E.; Neuberger, M. S. *Immunity* **1998**, *9*, 859.
(29) Kong, Q.; Maizels, N. *Genetics* **2001**, *158*, 369.
(30) Bemark, M.; Sale, J. E.; Kim, H. J.; Berek, C.; Cosgrove, R. A.; Neuberger, M. S. *J. Exp. Med.* **2000**, *192*, 1509.
(31) Bross, L.; Wesoly, J.; Buerstedde, J. M.; Kanaar, R.; Jacobs, H. *Eur. J. Immunol.* **2003**, *33*, 352.
(32) Bransteitter, R.; Pham, P.; Scharff, M. D.; Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4102.
(33) Dickerson, S. K.; Market, E.; Besmer, E.; Papavasiliou, F. N. *J. Exp. Med.* **2003**, *197*, 1291.
(34) Beale, R. C.; Petersen-Mahrt, S. K.; Watt, I. N.; Harris, R. S.; Rada, C.; Neuberger, M. S. *J. Mol. Biol.* **2004**, *337*, 585.
(35) Ramiro, A. R.; Stavropoulos, P.; Jankovic, M.; Nussenzweig, M. C. *Nat. Immunol.* **2003**, *4*, 452.
(36) Sohail, A.; Klapacz, J.; Samaranayake, M.; Ullah, A.; Bhagwat, A. S. *Nucleic Acids. Res.* **2003**, *31*, 2990.
(37) Mayorov, V. I.; Rogozin, I. B.; Adkison, L. R.; Frahm, C. R.; Kunkel, T. A.; Pavlov, Y. I. *BMC Immunol.* **2005**, *6*, 10.
(38) Poltoratsky, V. P.; Wilson, S. H.; Kunkel, T. A.; Pavlov, Y. I. *J. Immunol.* **2004**, *172*, 4308.
(39) Martomo, S. A.; Fu, D.; Yang, W. W.; Joshi, N. S.; Gearhart, P. J. *J. Immunol.* **2005**, *174*, 7787.
(40) Duncan, B. K.; Weiss, B. *J. Bacteriol.* **1982**, *151*, 750.
(41) Krokan, H.; Wittwer, C. U. *Nucleic Acids Res.* **1981**, *9*, 2599.
(42) Krokan, H. E.; Drablos, F.; Slupphaug, G. *Oncogene* **2002**, *21*, 8935.
(43) Lindahl, T. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3649.
(44) Chaudhuri, J.; Tian, M.; Khuong, C.; Chua, K.; Pinaud, E.; Alt, F. W. *Nature* **2003**, *422*, 726.
(45) Navaratnam, N.; Morrison, J. R.; Bhattacharya, S.; Patel, D.; Funahashi, T.; Giannoni, F.; Teng, B. B.; Davidson, N. O.; Scott, J. *J. Biol. Chem.* **1993**, *268*, 20709.
(46) Petersen-Mahrt, S. K.; Neuberger, M. S. *J. Biol. Chem.* **2003**, *278*, 19583.
(47) Conticello, S. G.; Thomas, C. J.; Petersen-Mahrt, S. K.; Neuberger, M. S. *Mol. Biol. Evol.* **2005**, *22*, 367.
(48) Ta, V. T.; Nagaoka, H.; Catalan, N.; Durandy, A.; Fischer, A.; Imai, K.; Nonoyama, S.; Tashiro, J.; Ikegawa, M.; Ito, S.; Kinoshita, K.; Muramatsu, M.; Honjo, T. *Nat. Immunol.* **2003**, *4*, 843.
(49) Xie, K.; Sowden, M. P.; Dance, G. S.; Torelli, A. T.; Smith, H. C.; Wedekind, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8114.
(50) Chaudhuri, J.; Khuong, C.; Alt, F. W. *Nature* **2004**, *430*, 992.
(51) Rada, C.; Jarvis, J. M.; Milstein, C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7003.
(52) Ito, S.; Nagaoka, H.; Shinkura, R.; Begum, N.; Muramatsu, M.; Nakata, M.; Honjo, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1975.
(53) Brar, S. S.; Watson, M.; Diaz, M. *J. Biol. Chem.* **2004**, *279*, 26395.
(54) McBride, K. M.; Barreto, V.; Ramiro, A. R.; Stavropoulos, P.; Nussenzweig, M. C. *J. Exp. Med.* **2004**, *199*, 1235.
(55) Reynaud, C. A.; Aoufouchi, S.; Faili, A.; Weill, J. C. *Nat. Immunol.* **2003**, *4*, 631.
(56) Barreto, V.; Reina-San-Martin, B.; Ramiro, A. R.; McBride, K. M.; Nussenzweig, M. C. *Mol. Cell* **2003**, *12*, 501.
(57) Shinkura, R.; Ito, S.; Begum, N. A.; Nagaoka, H.; Muramatsu, M.; Kinoshita, K.; Sakakibara, Y.; Hijikata, H.; Honjo, T. *Nat. Immunol.* **2004**, *5*, 707.
(58) Betts, L.; Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W., Jr. *J. Mol. Biol.* **1994**, *235*, 635.
(59) Harris, R. S.; Petersen-Mahrt, S. K.; Neuberger, M. S. *Mol. Cell* **2002**, *10*, 1247.
(60) Navaratnam, N.; Fujino, T.; Bayliss, J.; Jarmuz, A.; How, A.; Richardson, N.; Somasekaram, A.; Bhattacharya, S.; Carter, C.; Scott, J. *J. Mol. Biol.* **1998**, *275*, 695.
(61) Kurowski, M. A.; Bujnicki, J. M. *Nucleic Acids Res.* **2003**, *31*, 3305.
(62) Kosinski, J.; Cymerman, I. A.; Feder, M.; Kurowski, M. A.; Sasin, J. M.; Bujnicki, J. M. *Proteins* **2003**, *53* (Suppl. 6), 369.
(63) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. *J. Mol. Biol.* **1997**, *268*, 209.
(64) Ko, T. P.; Lin, J. J.; Hu, C. Y.; Hsu, Y. H.; Wang, A. H.; Liaw, S. H. *J. Biol. Chem.* **2003**, *278*, 19111.
(65) Zaim, J.; Kierzek, A. M. *Nat. Immunol.* **2003**, *4*, 1153; author reply 1154.
(66) Rogozin, I. B.; Kolchanov, N. A. *Biochim. Biophys. Acta* **1992**, *1171*, 11.
(67) Rogozin, I. B.; Pavlov, Y. I.; Bebenek, K.; Matsuda, T.; Kunkel, T. A. *Nat. Immunol.* **2001**, *2*, 530.
(68) Bransteitter, R.; Pham, P.; Calabrese, P.; Goodman, M. F. *J. Biol. Chem.* **2004**, *279*, 51612.
(69) Pham, P.; Bransteitter, R.; Petruska, J.; Goodman, M. F. *Nature* **2003**, *424*, 103.
(70) Rogozin, I. B.; Diaz, M. *J. Immunol.* **2004**, *172*, 3382.
(71) Gearhart, P. J.; Bogenhagen, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3439.
(72) Bhagwat, A. S. *DNA Repair* **2004**, *3*, 85.
(73) Rada, C.; Williams, G. T.; Nilsen, H.; Barnes, D. E.; Lindahl, T.; Neuberger, M. S. *Curr. Biol.* **2002**, *12*, 1748.
(74) Wang, Z.; Mosbaugh, D. W. *J. Biol. Chem.* **1989**, *264*, 1163.
(75) Di Noia, J.; Neuberger, M. S. *Nature* **2002**, *419*, 43.
(76) Nilsen, H.; Rosewell, I.; Robins, P.; Skjelbred, C. F.; Andersen, S.; Slupphaug, G.; Daly, G.; Krokan, H. E.; Lindahl, T.; Barnes, D. E. *Mol. Cell* **2000**, *5*, 1059.
(77) Nilsen, H.; Haushalter, K. A.; Robins, P.; Barnes, D. E.; Verdine, G. L.; Lindahl, T. *EMBO J.* **2001**, *20*, 4278.
(78) Nilsen, H.; Stamp, G.; Andersen, S.; Hrivnak, G.; Krokan, H. E.; Lindahl, T.; Barnes, D. E. *Oncogene* **2003**, *22*, 5381.
(79) Kavli, B.; Andersen, S.; Otterlei, M.; Liabakk, N. B.; Imai, K.; Fischer, A.; Durandy, A.; Krokan, H. E.; Slupphaug, G. *J. Exp. Med.* **2005**, *201*, 2011.
(80) Imai, K.; Slupphaug, G.; Lee, W. I.; Revy, P.; Nonoyama, S.; Catalan, N.; Yel, L.; Forveille, M.; Kavli, B.; Krokan, H. E.; Ochs, H. D.; Fischer, A.; Durandy, A. *Nat. Immunol.* **2003**, *4*, 1023.
(81) Lee, W. I.; Torgerson, T. R.; Schumacher, M. J.; Yel, L.; Zhu, Q.; Ochs, H. D. *Blood* **2005**, *105*, 1881.
(82) Begum, N. A.; Kinoshita, K.; Kakazu, N.; Muramatsu, M.; Nagaoka, H.; Shinkura, R.; Biniszkiewicz, D.; Boyer, L. A.; Jaenisch, R.; Honjo, T. *Science* **2004**, *305*, 1160.
(83) Stivers, J. T. *Science* **2004**, *306*, 2042; author reply 2042.
(84) Diaz, M.; Lawrence, C. *Trends Immunol.* **2005**, *26*, 215.
(85) Masutani, C.; Araki, M.; Yamada, A.; Kusumoto, R.; Nogimori, T.; Maekawa, T.; Iwai, S.; Hanaoka, F. *EMBO J.* **1999**, *18*, 3491.
(86) Masutani, C.; Kusumoto, R.; Yamada, A.; Dohmae, N.; Yokoi, M.; Yuasa, M.; Araki, M.; Iwai, S.; Takio, K.; Hanaoka, F. *Nature* **1999**, *399*, 700.
(87) Zhang, Y.; Yuan, F.; Wu, X.; Rechkoblit, O.; Taylor, J. S.; Geacintov, N. E.; Wang, Z. *Nucleic Acids Res.* **2000**, *28*, 4717.
(88) Zeng, X.; Winter, D. B.; Kasmer, C.; Kraemer, K. H.; Lehmann, A. R.; Gearhart, P. J. *Nat. Immunol.* **2001**, *2*, 537.

(89) Delbos, F.; De Smet, A.; Faili, A.; Aoufouchi, S.; Weill, J. C.; Reynaud, C. A. *J. Exp. Med.* **2005**, *201*, 1191.
(90) Martomo, S. A.; Yang, W. W.; Wersto, R. P.; Ohkumo, T.; Kondo, Y.; Yokoi, M.; Masutani, C.; Hanaoka, F.; Gearhart, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8656.
(91) Pavlov, Y. I.; Rogozin, I. B.; Galkin, A. P.; Aksenova, A. Y.; Hanaoka, F.; Rada, C.; Kunkel, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9954.
(92) Bertocci, B.; De Smet, A.; Flatter, E.; Dahan, A.; Bories, J. C.; Landreau, C.; Weill, J. C.; Reynaud, C. A. *J. Immunol.* **2002**, *168*, 3702.
(93) Schenten, D.; Gerlach, V. L.; Guo, C.; Velasco-Miguel, S.; Hladik, C. L.; White, C. L.; Friedberg, E. C.; Rajewsky, K.; Esposito, G. *Eur. J. Immunol.* **2002**, *32*, 3152.
(94) Shimizu, T.; Shinkai, Y.; Ogi, T.; Ohmori, H.; Azuma, T. *Immunol. Lett.* **2003**, *86*, 265.
(95) Faili, A.; Aoufouchi, S.; Flatter, E.; Gueranger, Q.; Reynaud, C. A.; Weill, J. C. *Nature* **2002**, *419*, 944.
(96) McDonald, J. P.; Frank, E. G.; Plosky, B. S.; Rogozin, I. B.; Masutani, C.; Hanaoka, F.; Woodgate, R.; Gearhart, P. J. *J. Exp. Med.* **2003**, *198*, 635.
(97) Zan, H.; Komori, A.; Li, Z.; Cerutti, A.; Schaffer, A.; Flajnik, M. F.; Diaz, M.; Casali, P. *Immunity* **2001**, *14*, 643.
(98) Diaz, M.; Verkoczy, L. K.; Flajnik, M. F.; Klinman, N. R. *J. Immunol.* **2001**, *167*, 327.
(99) Zan, H.; Shima, N.; Xu, Z.; Al-Qahtani, A.; Evinger Iii, A. J.; Zhong, Y.; Schimenti, J. C.; Casali, P. *EMBO J.* **2005**, *24*, 3757.
(100) Masuda, K.; Ouchida, R.; Takeuchi, A.; Saito, T.; Koseki, H.; Kawamura, K.; Tagawa, M.; Tokuhisa, T.; Azuma, T.; J, O. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13986.
(101) Kawamura, K.; Bahar, R.; Seimiya, M.; Chiyo, M.; Wada, A.; Okada, S.; Hatano, M.; Tokuhisa, T.; Kimura, H.; Watanabe, S.; Honda, I.; Sakiyama, S.; Tagawa, M.; J, O. W. *Int. J. Cancer* **2004**, *109*, 9.
(102) Stojic, L.; Brun, R.; Jiricny, J. *DNA Repair* **2004**, *3*, 1091.
(103) Martomo, S. A.; Yang, W. W.; Gearhart, P. J. *J. Exp. Med.* **2004**, *200*, 61.
(104) Cascalho, M.; Wong, J.; Steinberg, C.; Wabl, M. *Science* **1998**, *279*, 1207.
(105) Kim, N.; Bozek, G.; Lo, J. C.; Storb, U. *J. Exp. Med.* **1999**, *190*, 21.
(106) Phung, Q. H.; Winter, D. B.; Alrefai, R.; Gearhart, P. J. *J. Immunol.* **1999**, *162*, 3121.
(107) Phung, Q. H.; Winter, D. B.; Cranston, A.; Tarone, R. E.; Bohr, V. A.; Fishel, R.; Gearhart, P. J. *J. Exp. Med.* **1998**, *187*, 1745.
(108) Rada, C.; Ehrenstein, M. R.; Neuberger, M. S.; Milstein, C. *Immunity* **1998**, *9*, 135.
(109) Bardwell, P. D.; Woo, C. J.; Wei, K.; Li, Z.; Martin, A.; Sack, S. Z.; Parris, T.; Edelmann, W.; Scharff, M. D. *Nat. Immunol.* **2004**, *5*, 224.
(110) Wilson, T. M.; Vaisman, A.; Martomo, S. A.; Sullivan, P.; Lan, L.; Hanaoka, F.; Yasui, A.; Woodgate, R.; Gearhart, P. J. *J. Exp. Med.* **2005**, *201*, 637.
(111) Rada, C.; Di Noia, J. M.; Neuberger, M. S. *Mol. Cell* **2004**, *16*, 163.
(112) Kaina, B.; Ochs, K.; Grosch, S.; Fritz, G.; Lips, J.; Tomicic, M.; Dunkern, T.; Christmann, M. *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *68*, 41.
(113) Lindahl, T. *Nature* **1993**, *362*, 709.
(114) Loeb, L. A. *Cancer Res.* **1991**, *51*, 3075.
(115) Stavnezer, J.; Schrader, C. E. *Trends Genet.* **2005**, *22*, 23.
(116) Schrader, C. E.; Linehan, E. K.; Mochegova, S. N.; Woodland, R. T.; Stavnezer, J. *J. Exp. Med.* **2005**, *202*, 561.
(117) Friedberg, E. C.; Lehmann, A. R.; Fuchs, R. P. *Mol. Cell* **2005**, *18*, 499.
(118) Prakash, S.; Johnson, R. E.; Prakash, L. *Annu. Rev. Biochem.* **2005**, *74*, 317.
(119) Harriman, W.; Volk, H.; Defranoux, N.; Wabl, M. *Annu. Rev. Immunol.* **1993**, *11*, 361.
(120) Peters, A.; Storb, U. *Immunity* **1996**, *4*, 57.
(121) Nambu, Y.; Sugai, M.; Gonda, H.; Lee, C. G.; Katakai, T.; Agata, Y.; Yokota, Y.; Shimizu, A. *Science* **2003**, *302*, 2137.
(122) Honjo, T.; Kinoshita, K.; Muramatsu, M. *Annu. Rev. Immunol.* **2002**, *20*, 165.
(123) Bachl, J.; Carlson, C.; Gray-Schopfer, V.; Dessing, M.; Olsson, C. *J. Immunol.* **2001**, *166*, 5051.
(124) Yoshikawa, K.; Okazaki, I. M.; Eto, T.; Kinoshita, K.; Muramatsu, M.; Nagaoka, H.; Honjo, T. *Science* **2002**, *296*, 2033.
(125) Shinkura, R.; Tian, M.; Smith, M.; Chua, K.; Fujiwara, Y.; Alt, F. W. *Nat. Immunol.* **2003**, *4*, 435.
(126) Beletskii, A.; Bhagwat, A. S. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13919.
(127) Klapacz, J.; Bhagwat, A. S. *J. Bacteriol.* **2002**, *184*, 6866.
(128) Beletskii, A.; Bhagwat, A. S. *Biol. Chem.* **1998**, *379*, 549.
(129) Beletskii, A.; Bhagwat, A. S. *J. Bacteriol.* **2001**, *183*, 6491.
(130) Beletskii, A.; Grigoriev, A.; Joyce, S.; Bhagwat, A. S. *J. Mol. Biol.* **2000**, *300*, 1057.
(131) Klapacz, J.; Bhagwat, A. S. *DNA Repair* **2005**, *4*, 806.
(132) Basu, U.; Chaudhuri, J.; Alpert, C.; Dutt, S.; Ranganath, S.; Li, G.; Schrum, J. P.; Manis, J. P.; Alt, F. W. *Nature* **2005**, *438*, 508.
(133) Storb, U.; Klotz, E. L.; Hackett, J., Jr.; Kage, K.; Bozek, G.; Martin, T. E. *J. Exp. Med.* **1998**, *188*, 689.
(134) Michael, N.; Martin, T. E.; Nicolae, D.; Kim, N.; Padjen, K.; Zhan, P.; Nguyen, H.; Pinkert, C.; Storb, U. *Immunity* **2002**, *16*, 123.
(135) Yu, K.; Chedin, F.; Hsieh, C. L.; Wilson, T. E.; Lieber, M. R. *Nat. Immunol.* **2003**, *4*, 442.
(136) Yu, K.; Roy, D.; Bayramyan, M.; Haworth, I. S.; Lieber, M. R. *Mol. Cell Biol.* **2005**, *25*, 1730.
(137) Liu, L. F.; Wang, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7024.
(138) Shen, H. M.; Storb, U. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12997.
(139) Rothenfluh, H. S.; Taylor, L.; Bothwell, A. L.; Both, G. W.; Steele, E. J. *Eur. J. Immunol.* **1993**, *23*, 2152.
(140) Wright, B. E. *J. Bacteriol.* **2000**, *182*, 2993.
(141) Wright, B. E.; Schmidt, K. H.; Minnick, M. F. *Genes Immun.* **2004**, *5*, 176.
(142) Kuppers, R.; Dalla-Favera, R. *Oncogene* **2001**, *20*, 5580.
(143) Ramiro, A. R.; Jankovic, M.; Eisenreich, T.; Difilippantonio, S.; Chen-Kiang, S.; Muramatsu, M.; Honjo, T.; Nussenzweig, A.; Nussenzweig, M. C. *Cell* **2004**, *118*, 431.
(144) Okazaki, I. M.; Hiai, H.; Kakazu, N.; Yamada, S.; Muramatsu, M.; Kinoshita, K.; Honjo, T. *J. Exp. Med.* **2003**, *197*, 1173.
(145) Zhu, Y.; Nonoyama, S.; Morio, T.; Muramatsu, M.; Honjo, T.; Mizutani, S. *J. Med. Dent. Sci.* **2003**, *50*, 41.
(146) Quartier, P.; Bustamante, J.; Sanal, O.; Plebani, A.; Debre, M.; Deville, A.; Litzman, J.; Levy, J.; Fermand, J. P.; Lane, P.; Horneff, G.; Aksu, G.; Yalcin, I.; Davies, G.; Tezcan, I.; Ersoy, F.; Catalan, N.; Imai, K.; Fischer, A.; Durandy, A. *Clin. Immunol.* **2004**, *110*, 22.
(147) Minegishi, Y.; Lavoie, A.; Cunningham-Rundles, C.; Bedard, P. M.; Hebert, J.; Cote, L.; Dan, K.; Sedlak, D.; Buckley, R. H.; Fischer, A.; Durandy, A.; Conley, M. E. *Clin. Immunol.* **2000**, *97*, 203.
(148) Imai, K.; Zhu, Y.; Revy, P.; Morio, T.; Mizutani, S.; Fischer, A.; Nonoyama, S.; Durandy, A. *Clin. Immunol.* **2005**, *115*, 277.
(149) Fiorini, C.; Jilani, S.; Losi, C. G.; Silini, A.; Giliani, S.; Ferrari, S.; Notarangelo, L. D.; Plebani, A.; Sfar, T.; Helal, A. *Eur. J. Pediatr.* **2004**, *163*, 704.

CR040496T