Review

Activation-induced cytidine deaminase-mediated hypermutation in the DT40 cell line

Hiroshi Arakawa and Jean-Marie Buerstedde*
Helmholtz Center Munich, German Research Center for Environmental Health, Institute for Molecular Radiobiology, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

Depending on the species and the developmental stage of B cells, activation-induced cytidine deaminase (AID) triggers immunoglobulin (Ig) gene diversification by gene conversion, hypermutation or switch recombination. The bursal B cell line DT40 usually diversifies its rearranged Ig light chain (IgL) gene by gene conversion, but disruption of the RAD51 gene paralogues or deletion of the ψV conversion donors induces hypermutation. Although not all aspects of somatic hypermutation can be studied in DT40, the compact size of the chicken IgL locus and the ability to modify the genome by targeted integration are powerful experimental advantages. We review here how the studies in DT40 contributed to understanding how AID initiates Ig gene diversification and how AID-induced uracils are subsequently processed by uracil DNA glycosylase, proliferating cell nuclear antigens and error-prone polymerases. We also discuss the on-going research on the Ig locus specificity of hypermutation and the possibility of using hypermutation for the artificial evolution of proteins and regulatory sequences in DT40.

Keywords: activation-induced cytidine deaminase; DT40; gene conversion; immunoglobulin; somatic hypermutation

1. IMMUNOGLOBULIN REPERTOIRE DEVELOPMENT IN THE CHICKEN
Chicken B cells diversify their immunoglobulin (Ig) gene by gene conversion within the bursa of Fabricius (Reynaud et al. 1987). A careful sequence comparison of all 25 pseudo variable (ψV) genes, the single germ line of V and joining (J) segments and diversified VJ segments from bursal B cells of the CB inbred strain established that the sequence differences between the germ line and the diversified VJ segments can be accounted for by stretches of ψV gene sequences. Gene conversion was initially believed to be limited to avian species, but it is now clear that many mammalian farm animals also diversify their Ig genes by gene conversion (Butler 1998).

Although, the analysis of diversified VJ sequences from bursal B cells indicated that the vast majority of sequence changes after VJ rearrangement were the result of ψV templated conversion tracts, a few single nucleotide substitutions could not be accounted for by ψV gene donors (Reynaud et al. 1987). Whether these apparently untemplated substitutions were equivalents of somatic hypermutation (SHM) or vestiges of error-prone polymerases involved in the synthesis of gene conversion tracts remained an open question. Interestingly, the number of untemplated single nucleotide substitutions increased in sequences from chicken germinal-centre B cells indicating that gene conversion and hypermutation cooperate outside the bursa during antigen-driven immune responses (Arakawa et al. 1996, 1998).

2. THE B CELL LINE DT40 AS A MODEL FOR IMMUNOGLOBULIN GENE CONVERSION
The DT40 cell line derived from bursal B cells continues to diversify its Ig gene by gene conversion during in vitro cell culture (Buerstedde et al. 1990). Cl18, a cell surface IgM negative (sIgM-) variant of DT40, carries a frameshift mutation in the Ig light chain V region that can be repaired by overlapping gene conversion events leading to re-expression of sIgM (figure 1a). The percentages of sIgM- cells within expanding subclones can be analysed by fluorescence-activated cell sorting (FACS) and used to quantify gene conversion activities by fluctuation analysis (Ig reversion assay).

Perhaps related to its proficiency in gene conversion, DT40 integrates transfected DNA constructs at high ratios into the endogenous gene loci (Buerstedde & Takeda 1991). This efficient gene targeting allows the easy disruption of genes to test whether they are required for Ig gene conversion (Takeda et al. 1992; Bezzubova et al. 1997). Years of work cumulated in the discovery that the activation-induced cytidine deaminase (AID) gene, previously shown to be essential for switch recombination and SHM in mice (Muramatsu et al. 2000) and humans (Revy et al. 2000), is also required for gene conversion (Arakawa et al. 2002; Harris et al. 2000).
murine and human studies. The rate of mutation for RGYW, known as a hypermutation hotspot from mutation in DT40 (Sale pool, provided the first credible evidence for hypermutation. Paralogues showed few gene conversion tracts and accounted for by the published A few single nucleotide substitutions that could not be accounted for by the published \( \psi ^{V} \) sequences of DT40 (Buerstedde et al. 1990; Kim et al. 1990). However, it is difficult to unambiguously classify these as untemplated mutations, since the \( \psi ^{V} \) gene sequences of the rearranged IgL of DT40 are unavailable. The discovery that IgL sequences from disruption mutants of RAD51 paralogues showed few gene conversion tracts and significantly increased frequencies of single-nucleotide substitutions with no equivalent in the \( \psi ^{V} \) sequence pool, provided the first credible evidence for hypermutation in DT40 (Sale et al. 2001). The untemplated mutations occurred preferentially in the sequence RGYW, known as a hypermutation hotspot from murine and human studies. The rate of mutation for the XRCC3 disruption mutant was approximately \( 0.4 \times 10^{4} \) bp\(^{-1}\) per generation, similar to the rate previously calculated for the human Ramos lymphoma line (Sale & Neuberger 1998). The vast majority of hypermutations occurred at G or C bases with C:G and G:C transversions being the most frequent types of substitution. Notably, the rate of hypermutation in the XRCC3 mutant was approximately 10 times higher than the rate of gene conversion in wild-type DT40.

The hypermutation activity in the RAD51 paralogue mutants manifested itself by the appearance of sIgM\(^{-}\) cells, presumably due to deleterious mutations in the rearranged Ig light and heavy alleles (Sale et al. 2001). The percentages of sIgM\(^{-}\) cells within growing cultures can be easily quantified by FACS to measure hypermutation activity by fluctuation analysis (Ig loss assay; figure 1b) similar to the way the Ig reversion assay is used for the measurement of gene conversion activity.

Deletion of the upstream \( \psi ^{V} \) locus not only abolished gene conversion, but also induced untemplated single-nucleotide substitutions in the rearranged IgL gene of DT40 (Arakawa et al. 2004). As observed for the RAD51 paralogue mutants, these mutations preferentially targeted hypermutation hotspots and occurred almost exclusively in G and C bases with the most common types of mutation being G:C and C:G transversions. The mutation rate as determined by sequencing and the Ig loss assay was even higher than in the RAD51 paralogue mutants. Most mutations were located between 150 and 500 bp downstream of the Ig light chain promoter. No sequence diversity apart from a few probable PCR artefacts was detected in the highly transcribed elongation factor 1\( \alpha \) and the BuI genes. Expression of AID was essential for the mutation activity induced by the \( \psi ^{V} \) locus deletion.

Together, these results indicate that a mutation activity can be induced in the IgL locus of DT40 that closely resembles SHM of murine and human germinal-centre B cells. Among the conserved features are a typical distribution of the mutations downstream of the IgL promoter, preference for hotspot motifs, Ig locus specificity and dependence on AID. Important differences to SHM in murine and human B cells are the lack of mutations at A or T bases and the predominance of G:C and C:G transversions.

3. IMMUNOGLOBULIN HYPERMUTATION IN DT40

A few single nucleotide substitutions that could not be accounted for by the published \( \psi ^{V} \) sequences of the CB strain were observed in IgL \( V^{\psi} \) sequences of DT40 (Buerstedde et al. 1990; Kim et al. 1990). However, it is difficult to unambiguously classify these as untemplated mutations, since the \( \psi ^{V} \) gene sequences of the rearranged IgL of DT40 are unavailable. The discovery that IgL sequences from disruption mutants of RAD51 paralogues showed few gene conversion tracts and significantly increased frequencies of single-nucleotide substitutions with no equivalent in the \( \psi ^{V} \) sequence pool, provided the first credible evidence for hypermutation in DT40 (Sale et al. 2001). The untemplated mutations occurred preferentially in the sequence RGYW, known as a hypermutation hotspot from murine and human studies. The rate of mutation for the XRCC3 disruption mutant was approximately \( 0.4 \times 10^{4} \) bp\(^{-1}\) per generation, similar to the rate previously calculated for the human Ramos lymphoma line (Sale & Neuberger 1998). The vast majority of hypermutations occurred at G or C bases with C:G and G:C transversions being the most frequent types of substitution. Notably, the rate of hypermutation in the XRCC3 mutant was approximately 10 times higher than the rate of gene conversion in wild-type DT40.

The hypermutation activity in the RAD51 paralogue mutants manifested itself by the appearance of sIgM\(^{-}\) cells, presumably due to deleterious mutations in the rearranged Ig light and heavy alleles (Sale et al. 2001). The percentages of sIgM\(^{-}\) cells within growing cultures can be easily quantified by FACS to measure hypermutation activity by fluctuation analysis (Ig loss assay; figure 1b) similar to the way the Ig reversion assay is used for the measurement of gene conversion activity.

Deletion of the upstream \( \psi ^{V} \) locus not only abolished gene conversion, but also induced untemplated single-nucleotide substitutions in the rearranged IgL gene of DT40 (Arakawa et al. 2004). As observed for the RAD51 paralogue mutants, these mutations preferentially targeted hypermutation hotspots and occurred almost exclusively in G and C bases with the most common types of mutation being G:C and C:G transversions. The mutation rate as determined by sequencing and the Ig loss assay was even higher than in the RAD51 paralogue mutants. Most mutations were located between 150 and 500 bp downstream of the Ig light chain promoter. No sequence diversity apart from a few probable PCR artefacts was detected in the highly transcribed elongation factor 1\( \alpha \) and the BuI genes. Expression of AID was essential for the mutation activity induced by the \( \psi ^{V} \) locus deletion.

Together, these results indicate that a mutation activity can be induced in the IgL locus of DT40 that closely resembles SHM of murine and human germinal-centre B cells. Among the conserved features are a typical distribution of the mutations downstream of the IgL promoter, preference for hotspot motifs, Ig locus specificity and dependence on AID. Important differences to SHM in murine and human B cells are the lack of mutations at A or T bases and the predominance of G:C and C:G transversions.

4. THE RELATIONSHIP OF GENE CONVERSION AND HYPERMUTATION

The induction of Ig hypermutation by blockage of Ig gene conversions supports a simple model explaining how the hypermutation and gene conversion pathways are initiated and regulated (figure 2). The first event common to both pathways is a modification of the rearranged \( V(D)J \) segment by AID. The default processing of this lesion in the absence of nearby donors or of high homologous recombination activity leads to Ig hypermutation in the form of a single-nucleotide substitution (figure 2, right side). However, if donor sequences are available, processing of the AID-induced lesion can be divided into a stage before strand exchange, when a shift to Ig hypermutation is still possible, and a stage after strand exchange, when the commitment towards Ig gene conversion has been made (figure 2, left side). Whereas, completion of the first stage requires the participation of the RAD51 paralogues, the second stage involves more downstream recombination factors, such as RAD54.

This difference in commitment explains why disruptions of the RAD51 paralogues not only decrease Ig gene conversion, but also induce Ig hypermutation (Sale et al. 2001), whereas disruption of the RAD54 gene (Bezzubova et al. 1997) only decreases Ig gene conversion. DT40 mutants of BRCA1 (Longerich et al. 2008) and BRCA2 (Hatanaka et al. 2005) show phenotypes similar to the RAD51 paralogue mutants, whereas mutants of the Fanconi anaemia pathway behave like the RAD54 mutant (Yamamoto et al. 2005) indicating that the encoded proteins participate in the first and the second stage, respectively. The model also predicts that low cellular homologous recombination activity prevents Ig gene conversion even in the presence of conversion donors. Such a low homologous-recombination activity might be the reason why human and murine B cells never use Ig gene conversion despite the presence of nearby candidate donors in the form of unarranged \( V \) segments and why chicken germinal-centre B cells have shifted the balance from Ig gene conversion to Ig hypermutation (Arakawa et al. 1998).
5. ACTIVATION-INDUCED CYTIDINE DEAMINASE DEAMINATES CYTIDINE WITHIN THE IMMUNOGLOBULIN LOCUS

AID was initially proposed to be an RNA-editing enzyme based on its homology to the RNA-editing enzyme APOBEC-1 (Muramatsu et al. 2000). The first evidence that the substrates of AID are cytidines within the Ig loci came from the expression of the uracil glycosylase inhibitor (UGI) in hypermutating variants of DT40 (Di Noia & Neuberger 2002). The authors expected that interference with uracil excision would alter the spectrum of mutations if AID produces uracils by the deamination of cytidines. This was indeed the case as the preference for transversion mutations in untransfected cells was replaced by a dominance of transition mutations in UGI-expressing cells. The C:T and G:A mutation bias in the presence of UGI can be explained if UGI prevents the removal of AID-induced uracils by uracil glycosylases and the unexcised uracils pair with adenines during replication. A similar shift towards transition mutations at the C/G base pair was subsequently found for hypermutations of uracil DNA glycosylase (UNG) knockout mice indicating that UNG is the glycosylase responsible for the excision of AID-induced uracils (Rada et al. 2002).

Transfection of the UGI transgene into wild-type DT40 reduced the gene conversion rate suggesting that the excision of AID-induced uracils is required for gene conversion (Di Noia & Neuberger 2004). This was confirmed by the analysis of the UNG knockout in DT40, which dramatically reduced gene conversion rates as measured by the Ig reversion assay and sequencing (Saribasak et al. 2006). Interestingly, a few gene conversion events are still detected in the UNG knockout background, suggesting that a few uracils are processed by an alternative UNG-independent pathway. As expected, failing to remove AID-introduced uracils, UNG-negative DT40 showed evidence of a high Ig mutation rate in the Ig loss assay. Sequencing the IgL V\j segments revealed that the mutation rate was seven times higher than in the V\j-deleted cells and that 98 per cent of the mutations were either C:T or G:A transitions. Assuming that the transition mutations in the UNG-negative cells reflect the total number of AID-induced uracils, approximately one in seven of these uracils, is processed into a mutation. This relatively frequent conversion of uracils into mutations suggests the presence of a specialized error-prone repair pathway that recognizes AID-induced uracils.

6. ERROR-PRONE POLYMERASES AND A CONNECTION TO TRANSELESION DNA SYNTHESIS

Error-prone polymerases are involved at later stages of the hypermutation pathway after the action of AID and UNG. In the mouse, Pol\eta seems to be responsible for mutations at A/T bases, since Pol\eta-deficient B cells lack most of these mutations (Zeng et al. 2001). In DT40, the error-prone polymerase REV1 is needed for most of the C:G and G:C transversion mutations that are strongly reduced in REV1 disruption mutants (Simpson & Sale 2003; Arakawa et al. 2006). Interestingly, hypermutation requires the deoxycytidyl transferase activity of REV1 that is dispensable for DNA damage repair (Ross & Sale 2006) suggesting that the transferase incorporates cytosine at the abasic sites produced by the combined action of AID and UNG.

Studies in yeast have implicated mono-ubiquitination of proliferating cell nuclear antigen (PCNA) at lysine 164 as a signal that recruits error-prone polymerases to the replication fork (Hoege et al. 2002). The role of PCNA ubiquitination for translesion DNA synthesis and Ig hypermutation were tested in the \psi\j-deleted DT40 by changing the evolutionary conserved codon 164 into arginine (Arakawa et al. 2006). The PCNA(K164R) single codon change not only renders cells sensitive to DNA-damaging agents, but also strongly reduces hypermutations in the IgL locus. The most affected types of mutation were the G:C and the C:G transversions indicating that DT40 exploited the PCNA-ubiquitin pathway for Ig hypermutation, most likely through the recruitment of REV1.

The relevance of PCNA ubiquitination for SHM has been confirmed in a PCNA(K164R) knock-in mouse, but the change in the spectrum of hypermutations

Figure 2. A model explaining the regulation of Ig gene conversion and Ig hypermutation.
differed from DT40 (Langerak et al. 2007). Whereas, mutations at A and T bases were significantly decreased, the total mutation rate was not changed due to compensatory mutation increases at C and G bases. Thus in mice, pol η seemed to be the main polymerase recruited by ubiquitinated PCNA, whereas polymerases responsible for mutations at C and G bases operated independent of PCNA ubiquitination at lysine 164. More research is needed to clarify species differences in the role of error-prone polymerases for hypermutation and to understand the signals leading to their recruitment.

RAD6/RAD18 is the only ubiquitin ligase for PCNA mono-ubiquitination in the yeast Saccharomyces cerevisiae. However, PCNA ubiquitination was decreased but not abolished in RAD18-knockout of DT40 (Arakawa et al. 2006) indicating that an alternative ligase is able to ubiquitinate PCNA lysine 164 in vertebrate cells. Consistent with this, the decrease of hypermutations in RAD18 disruption mutants is more modest compared to the PCNA(K164R) mutant (Arakawa et al. 2006; Bachl et al. 2006). A previous report found no obvious decrease of hypermutation in a RAD18 mutant of DT40 (Simpson & Sale 2005), but this may be due to the lack of sensitivity as the analysis was performed in gene conversion active DT40.

8. LOCUS CONTROL OF HYPERMUTATION
Another very interesting question is how AID is targeting the Ig loci. To explain the difference in the mutation rates of Ig and non-Ig genes, it had been postulated that cis-acting sequences in the Ig loci activate hypermutation possibly by recruiting AID. However, intense efforts did not succeed in unambiguously identifying these sequences for the murine and human Ig loci (reviewed in Odegard & Schatz 2006).

DT40 might offer an opportunity to clarify the molecular mechanism of this phenomenon. The ease of genome modifications and the compact composition of the chicken IgL locus offer unique advantages for experimentation. Furthermore, the locus specificity of hypermutation seems to be well preserved. Analysis of the elongation factor 1α and the Bcl1 genes from ψV-deleted, hypermutating DT40 indicated that AID-mediated sequence diversity was limited to the Ig genes (Arakawa et al. 2004). This was recently confirmed by showing that even long-term cultures of DT40 showed no genetic diversity in the VpreB3 and carboxic

7. GENETIC ANALYSIS OF ACTIVATION-INDUCED CYTIDINE DEAMINASE ACTION
Figure 3 summarizes the mentioned studies that contributed to our understanding of how AID initiates gene conversion and hypermutation in DT40. A number of intriguing questions still remain, for example: (i) why do almost no mutations occur at A/T bases in DT40, (ii) why are the abasic sites produced by the combined action of AID and UNG repaired in an unusual error-prone fashion, (iii) what triggers PCNA mono-ubiquitination at lysine 164, and (iv) what are the intermediates downstream of UNG leading to gene conversion?

Figure 3. DNA editing model of AID.
anhydrase genes, upstream and downstream neighbours of the IgL gene, respectively (Gopal & Fugmann 2008).

The deletion of a short sequence encompassing the only known enhancer of the IgL locus did not affect Ig gene conversion (Yang et al. 2006), but it was subsequently reported that the deletion of a larger sequence including neighbouring regions prevented gene conversion and untemplated mutations of the IgL gene (Kothapalli et al. 2008). Because the ψV locus is not required for IgL hypermutation in DT40 (Arakawa et al. 2004) the probable location of cis-regulatory elements for hypermutation is within the 10 kb sequence stretch separating the IgL transcriptional start site from the downstream carbonic anhydrase gene.

Although others reported transient mutation activity after random integration of non-Ig transgenes in DT40 (Yang et al. 2007), our laboratory was able to design a green fluorescent protein (GFP) transgene that is stably expressed when inserted into non-Ig loci, but crippled by mutations when inserted into the IgL locus of DT40. Using this transgene as a hypermutation reporter, we identified a sequence of the IgL locus that was necessary and sufficient to confer hypermutation activity in AID expressing DT40 (Blagodatski et al. submitted). Further attempts to define shorter sequence motifs responsible for the activation of AID-mediated diversification in neighbouring transcription units are now on-going in the laboratory.

Repeats of the E-box sequence motif are contained in various Ig enhancers and the coincidental insertion of two E-box sites may have caused increased hypermutation of a murine transgene (Michael et al. 2003). Disruption of the E2A gene in ψV-deleted DT40 decreased the frequency of Ig hypermutation three-fold (Schoetz et al. 2006) and a similar disruption in wild-type DT40 decreased the gene conversion frequency three-fold (Kitao et al. 2008). However, the latter study detected reduced histone H4 acetylation of the IgL locus in E2A-negative cells suggesting that the effect of E2A was not directly related to AID recruitment, but due to changes in the chromatin configuration of the IgL locus.

9. APPLICATION TO BIOTECHNOLOGY

Possibilities to use AID-mediated gene diversification for protein evolution in DT40 have been explored by various groups. For example, enhanced Ig gene conversion by trichostatin stimulation has been used to generate and select antigen-specific antibodies in DT40 (Seo et al. 2005). The on-going Ig gene conversion and hypermutation in RAD51 paralogue mutants have likewise been employed to optimize the antigen specificity of antibodies (Cumbers et al. 2002). Furthermore, a blue fluorescent protein gene together with a GFP pseudogene has been inserted into the IgL locus demonstrating that non-Ig transgenes can be diversified by gene conversion when inserted into the IgL locus together with a conversion donor sequence (Kanayama et al. 2006).

Based on the observation that the deletion of the nearby ψV genes induced hypermutation in the rearranged IgL gene, it was postulated that any transgene inserted into the Ig loci of DT40 in the absence of homologous donor sequences would be diversified by hypermutation. As an example, the eGFP gene was inserted into the IgL locus of DT40 and cells expressing desirable mutation were selected by preparative FACS sorting (Arakawa et al. 2008). Only three rounds of FACS sorting during two months of culturing generated new GFP variants that displayed more than three-fold higher fluorescence activity than the best GFPs currently available for bio-imaging of vertebrate cells. This artificial evolution system might be applicable for any gene or regulatory DNA sequence if a sensitive strategy for the selection of beneficial mutations can be implemented.

REFERENCES


Cumbers, S. J., Williams, G. T., Davies, S. L., Grenfell, R. L., Takeda, S., Batista, F. D., Sale, J. E. & Neuberger, M. S.


Hosang, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G. & Jentsch, S. 2002 RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* 419, 135–141. (doi:10.1038/nature00991)


Sale, J. E. & Neuberger, M. S. 1998 TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity* 9, 859–869. (doi:10.1016/S1074-7613(00)80651-2)


