



**UNIVERSITY**  
*of*  
**GLASGOW**

Darlow, J.M. and Stott, D.I. (2006) Gene conversion in human rearranged immunoglobulin genes. *Immunogenetics* 58(7):pp. 511-522.

<http://eprints.gla.ac.uk/3646/>

**John M. Darlow · David I. Stott**

# **Gene conversion in human rearranged immunoglobulin genes**

J. M. Darlow (✉) · D. I. Stott

Department of Immunology

Level 4

Glasgow Biomedical Research Centre

120 University Place

Glasgow, G12 8TA, UK

e-mail: [john.m.darlow@clinmed.gla.ac.uk](mailto:john.m.darlow@clinmed.gla.ac.uk)

Tel: +141 330 8417

Fax: +141 330 4297

**Abstract** Over the past 20 years, many DNA sequences have been published suggesting that all or part of the V<sub>H</sub> segment of a rearranged immunoglobulin gene may be replaced *in vivo*. Two different mechanisms appear to be operating. One of these is very similar to primary V(D)J recombination, involving the RAG proteins acting upon recombination signal sequences, and this has recently been proven to occur. Other sequences, many of which show partial V<sub>H</sub> replacements with no addition of untemplated nucleotides at the V<sub>H</sub>-V<sub>H</sub> joint, have been proposed to occur by an unusual RAG-mediated recombination with the formation of hybrid (coding-to-signal) joints. These appear to occur in cells already undergoing somatic hypermutation in which some authors are convinced that RAG genes are silenced. We recently proposed that the latter type of V<sub>H</sub> replacement might occur by homologous recombination initiated by the activity of AID (activation-induced cytidine deaminase) which is essential for somatic hypermutation and gene conversion. The latter has been observed in other species but not so far in human Ig genes. Here we present a new analysis of sequences published as examples of the second type of rearrangement. This not only shows that AID recognition motifs occur in recombination regions but also that some sequences show replacement of central sections by a sequence from another gene, similar to gene conversion in the immunoglobulin genes of other species. These observations support the proposal that this type of rearrangement is likely to be AID- rather than RAG-mediated and is consistent with gene conversion.

**Key words** Gene conversion - Genes, Immunoglobulin - Humans - Antibody Diversity

## **Introduction**

Evidence that the  $V_H$  segment of a rearranged immunoglobulin  $V_HDJ_H$  exon could be replaced *in vivo* by a secondary rearrangement was first presented in 1986 (Kleinfield et al. 1986; Reth et al. 1986), since which time many more examples have followed. The mechanism, recently conclusively demonstrated (Zhang et al. 2003), is similar to primary V(D)J recombination. It is mediated by the interaction of the RAG protein complex on the recombination signal sequence (RSS) of the incoming  $V_H$  and a cryptic RSS, in the opposite direction, within the 3' end of the rearranged  $V_H$ , which acts as a surrogate for the RSS lost from the 5' end of the D segment in the primary rearrangement. As in primary rearrangement, the RSS are cleaved off and there is a variable amount of nucleotide deletion and addition at the joined coding ends.

In 2000, evidence was presented for another type of  $V_H$  replacement (Itoh et al. 2000; Wilson et al. 2000), and a few more examples of this have followed. It looks rather different from the first type, resembling homologous recombination between the primarily rearranged  $V_H$  gene segment and a new  $V_H$ . As the exchange usually occurs in a region in which both segments have the same sequence, the exact point of recombination cannot be determined. It was proposed (Itoh et al. 2000; Wilson et al. 2000) that it occurs at cryptic RSS but, if so, the

recombination process would be quite different from normal V(D)J rearrangement because the sequences proposed to function as cryptic RSS are in the same orientation, one of them remains in the recombined product, there is no addition of nucleotides at the point of exchange, and [with a single exception (Wilson et al. 2000)] no nucleotides are deleted from the gene sequence. Another problem with this hypothesis is that most of the published examples show somatic hypermutation (SHM) and the evidence points to some of this having occurred before the secondary recombination. Thus the proposal that RAG-mediated recombination is responsible goes against evidence that RAG expression has ceased by the time that hypermutation begins. Finally, the sequences were amplified from DNA isolated from pools of lymphocytes, giving the possibility that the sequences could have arisen by hybridisation during PCR amplification.

In a recent review (Darlow and Stott 2005), we concluded that though hybrids undoubtedly occur as an artefact during PCR, and may account for some of the published examples of the second type of  $V_H$  exchange (which we called Type 2  $V_H$  replacement), some were much less likely to be due to PCR hybridisation. Firstly, two sequences of Itoh et al. (Itoh et al. 2000) obtained from a PCR using  $V_H$  Family 1 primers had a Family 1 segment hybridised to an original rearrangement with a Family 3  $V_H$  segment. Their  $V_H$  primer was specific for Family 1 and they did not amplify any rearranged genes that started with a Family 3  $V_H$  segment. (These sequences are illustrated in our Fig. 3). Secondly, 23/26 related sequences from three independent sets of PCRs from a tumour investigated by Lenze et al. (Lenze et al. 2003) were  $V_{H3-23(5')/V_{H3-07(3')}$

hybrids, and two were hybrids with another member of the family. Only a single related sequence had the complete parent  $V_{H3-07}$ , but this contained several mutations that were not present in the hybrids but were in the region shared in common with them. This showed that it could not have been a PCR template for a hybrid artefact. (These sequences are not illustrated in this paper). We considered other reasons put forward by the original authors but these were the ones that we found hardest to dismiss and led us to consider whether there could be another explanation that might include more of the sequences, and we suggested a possible alternative. AID, which is essential for SHM, is also required for immunoglobulin gene conversion and class-switching (Neuberger et al. 2003). Gene conversion is non-reciprocal homologous recombination in which a section of DNA is replaced by copying from another strand without alteration of the template (Wysocki and Geftter 1989). It is the primary mechanism producing antibody diversity in birds (Arakawa and Buerstedde 2004) and the rabbit (Becker and Knight 1990; Sehgal et al. 2000; Weinstein et al. 1994; Winstead et al. 1999). Birds have single functional  $V_H$  and  $V_L$  gene segments and many non-functional V segments upstream which act as the templates. The rabbit has approximately 100  $V_H$  gene segments but preferentially rearranges to the 3'-most segment,  $V_{H1}$ , in 80-90% of cases and uses the others as templates for gene conversion. Much more VJ recombinational diversity occurs in the light chains but these too exhibit gene conversion (Sehgal et al. 2000). There is also evidence for immunoglobulin gene conversion in the cow (Lucier et al. 1998; Parng et al. 1996), which appears to have limited recombinational diversity, and it may play a minor role in the

sheep (Gontier et al. 2005), but it is widely believed that it does not occur in primates. Strictly speaking these recombinations in mammalian species are referred to as 'gene-conversion-like' because the templates have not been isolated from the same cells and shown to be unchanged. Primates and rodents have multiple functional V, D and J gene segments and continue gene rearrangement throughout life, affording extensive recombinational possibilities. In contrast, species that use gene conversion as a major mechanism not only have only one or a few functional segments at the heavy and/or light chain loci, or restricted segment-usage, but cease V(D)J recombination early in life, resulting in very limited recombinational diversity. All species further increase diversity by SHM but to varying extents. Primates and rodents appear to have less need of gene conversion but there is good evidence that it can occur in mice, though it does not appear to be frequent (D'Avirro et al. 2002; D'Avirro et al. 2005; David et al. 1992; Dildrop et al. 1982; Krawinkel et al. 1983; Selsing et al. 1996; Tsai et al. 2002).

Though gene conversion at the Ig loci has not been recognised in Man, it has recently been shown that the expression of AID in human B cells is associated with the appearance of double-strand breaks with single-stranded protruding ends which are repaired by homologous recombination (Zan et al. 2003). This suggests that if Type 2 V<sub>H</sub> replacement occurs by homologous recombination it could well occur in mutating cells and the mechanism might be gene conversion rather than reciprocal exchange. One of the main targets of somatic hypermutation has been shown to be the RGYW motif, reverse complement WRCY (Foster et al. 1999;

Rogozin and Kolchanov 1992) [now broadened to DGYW, complement WRCH (Rogozin and Diaz 2004)] and it has recently been shown (Beale et al. 2004) that AID prefers dA or dG at the -1 position to the targeted cytosine, and dA or dT at the -2 position which gives WRC - reverse complement GYW\*. We therefore re-examined the published sequences of Type 2 replacement for these motifs in the V<sub>H</sub>-V<sub>H</sub> junction regions and not only found that they were present but also that some examples resemble gene conversion as seen in the Ig genes of other species.

## Materials and methods

The cloned sequences examined are those published by Wilson and co-workers (Wilson et al. 2000) and Itoh and co-workers (Itoh et al. 2000). Alignments between cloned and germline sequences were checked with V Base/DNAPlot available at <http://vbase.mrc-cpe.cam.ac.uk/> and the JoinSolver alignment program available at <http://joinsolver.niams.nih.gov/index.htm>. Germline sequences in figures were checked against the database sequences available from the same websites and those cloned hybrid sequences that are available were also checked. Of the submitted sequences of Wilson and co-workers, accession numbers AF26069-AF262209, available from EMBL/GenBank/DDBJ, those appearing in Fig. 1 are as follows: Wilson et al. sequence name CP1-HB1-1 has accession number AF262202, CP2-HB1-1 is AF262141, CP2-HB3 is AF262144, CP1-HB2 is AF262209 and CP2-HB2 is AF262143. BBEEdit Lite and DNAPlot

---

\* R = A/G, Y = C/T, W = A/T, D = A/G/T, H = C/T/A.



software were used in making the alignments shown in Figs 2 and 3. The accession numbers of sequences found in the databases are, in order of appearance in the figures: ST1R M26, AF308556; ST1R M31, AF308557; ST1R G1, AF308551; ST1R G2, AF308552. However, the sequences labelled in the databases as ST1R M10 (AF308554) and ST1R M6 (AF308559) are entirely unrelated to those labelled as ST1R M10 and M6 in Fig. 5 A of Itoh et al. so the sequences shown in our Fig. 2 were taken from the paper only.

## Results

All the sequences published as examples of Type 2  $V_H$  replacement were presented as having recombinations at a single point between two  $V_H$  segments and we did not set out with any presupposition that they contained second unseen recombination points within or upstream of the sequenced regions. To investigate the possibility that Type 2  $V_H$  replacement might be due to homologous recombination (at a single point) initiated by DNA breaks at AID recognition motifs, rather than by RAG mediated recombination at cryptic RSS, we examined the  $V_H$ - $V_H$  recombination regions for AID recognition motifs. Wilson and co-workers (Wilson et al. 2000) showed the presence of RSS-like sequences in all their recombination regions. Fig. 1 is a redrawing of their Figure 3 showing that AID recognition motifs are also present in all recombination regions, though this could be by chance as the motifs are so numerous.

We then examined the sequences of Itoh and co-workers (Itoh et al. 2000) for AID motifs in recombination regions and found that some sequences resemble examples of Ig gene conversion in other species because both ends of the replaced section are within the  $V_H$ . Four rearranged genes had been identified (ST1R M26, M10, M6, and M31) which share the same HCDR3 sequence but in which the V segments all have overall matches to different germline V sequences (Fig.2a). Because M26 matches the same germline  $V_H$  throughout, and all the others resemble it at the  $V_H$ -D junction, Itoh and co-workers concluded that M26 was the original rearrangement and that three separate  $V_H$  replacement events had occurred at different cryptic RSS to produce the other three sequences. The best matches for the four  $V_H$  segments were IGHV1-58, 1-69, 1-18, and 1-24 respectively. Alignment of all eight sequences confirms that all four rearranged genes are identical at the 3' ends of their V segments (resembling the IGHV1-58\*01 germline sequence) but also shows that, despite giving best overall matches to different germline V segments, M10, M6, and M31 all match IGHV1-69 at their 5' ends (Fig. 2a). Further inspection shows that, though M10 appears to have resulted from a replacement of IGHV1-69 into IGHV1-58 with a single region of recombination in the  $V_H$  segment, M6 and M31 appear to be examples of different gene replacement events into the M10 rearrangement.

When aligned with M26, M10 can be divided into two sections (Fig. 2a, b). In the first it matches the germline gene segment IGHV1-69\*01 and in the second the rearranged gene M26 including its HCDR3. The dividing line appears to run precisely between nucleotides 220 and 221 (Fig. 2b). It does not correspond either

with the proposed cryptic RSS of Itoh et al. or with any AID recognition motif (the closest of which are shown on either side), nor with a section in which the germline genes are identical, but it is consistent with being the 3' end of a gene conversion of which the 5' end is upstream of the start of the sequenced part of the  $V_H$ . (The sequences of Itoh et al. only start at nt 41 of the  $V_H$  segment.) M10 has numerous mutations on both sides of the division but M26 has a single mutation, and this is upstream, so we cannot tell whether there had already been any mutation when the recombination occurred. However, the presence of mutations in the region of M10 matching M26 precludes the possibility that recombination as shown in Fig. 2b occurred by hybridisation to M26 during PCR. An alternative hypothesis that an uncloned mutated version of the M26 rearrangement (from a clonally related cell) hybridised with a mutated rearranged IGHV1-69 gene (also uncloned in its entirety) seems unlikely in view of the lack of any homology at the point of recombination.

The alignment of M6 with M10 (Fig. 2a, c) may be divided into five sections: end sections (1 and 5), in which both rearranged genes share the same germline sequence matches and CDR3, and a middle section (3), in which M6 has a different germline  $V_H$  match from M10, with sections either side of it (2 and 4, the boxes in Fig. 2a) in which the germline sequences are the same. As we have seen above, M10 does not match the same  $V_H$  throughout, but first matches IGHV1-69 and then 1-58. M6 appears to represent a gene conversion into the M10 rearrangement in which a central portion of IGHV1-18 has replaced the region of recombination of IGHV1-69 with IGHV1-58. Thus, in section 1, M6 and M10

both match IGHV1-69 while in section 5 they both match the same D and J<sub>H</sub> with identical V<sub>H</sub>-D and D-J<sub>H</sub> junctions. In section 3 M6 matches IGHV1-18 while M10 matches first IGHV1-69 and then IGHV1-58. In section 2 the sequences of IGHV1-69 and 1-18 are identical, and in section 4 those of IGHV1-58 and 1-18 are identical. Section 2, the postulated cross-over region, upstream of the converted section, contains two complete RGYW motifs and a WRC motif and in section 4 there are two DGYW motifs and a WRCY motif. (Note: D includes R).

The alignment of M31 with M10 (Fig. 2a, d) may likewise be divided into five sections (sections 1 & 5 of M31 resembling the M10 rearrangement, section 3 this time resembling IGHV1-24, and sections 2 and 4 (boxed in Fig. 2a) being regions where the sequences of the respective germline genes are the same). However, the sections are of different lengths, 1 and 2 being shorter and 3, 4 and 5 longer than in the previous comparison. As the sequences only start at nt 41 of the V<sub>H</sub> segment, we cannot confirm that the first 40 nt of M31 resemble the same germline sequence as M10, but the fact that ST1R M10, M6 and M31 all resemble IGHV1-69 at their 5' ends suggests that this is probably so. There are AID recognition motifs precisely at both ends of section 3 as well as others within the possible recombination regions, sections 2 and 4.

Fig. 3 shows a new presentation of the sequences ST1R G1 and G2 of Itoh and co-workers (Itoh et al. 2000). The authors observed that they had the same HCDR3 but that their V<sub>H</sub> segments gave best matches to IGHV1-02 and 1-69 respectively, and that the 3' ends of both V<sub>H</sub> segments matched IGHV3-09. From this they concluded that the two cloned rearranged genes represented Type 2

replacements by the two different family 1 V<sub>H</sub> segments into an original rearrangement with the family 3 V<sub>H</sub> segment but we noticed that both rearranged genes matched the same V<sub>H</sub> germline segment at their 5' ends. Fig. 3 shows that the V segments may be divided into five sections. In section 5, both rearranged genes give a best match to IGHV3-09. G2 gives a best match to IGHV1-69 throughout sections 1-4, whereas G1 matches IGHV1-69 in section 1 but IGHV1-02 in section 3. In sections 2 and 4 the two germline gene segments IGHV1-02 and 1-69 are identical except at two adjacent positions in section 4 at both of which G1 resembles IGHV1-69. The analysis thus supports the conclusion that the original rearrangement was with IGHV3-09 and that G2 represents a replacement by IGHV1-69 into the original rearrangement, but suggests that G1 is the result of a gene conversion event in which section 3 of a less-mutated precursor of G2 has been replaced by the corresponding region of IGHV1-02. Both rearranged genes (G1 and G2) contain numerous mutations not present in the other, so there is no possibility that G1 could have been derived from G2 as a PCR artefact. The inference is that both sequences acquired further mutations after the gene conversion event. There is a WRC motif and an RGYW motif in section 2. At the 5' end of section 4, there are 4 bp of homology (the second box on G1 in Fig. 3a, nt 223-226 in Fig. 3b) before the bases matching IGHV1-69 but no AID recognition motif except one created by mutations in ST1R G2. At the beginning of section 5 there are 3 bp of homology between IGHV1-69 and 3-09 (the third box on G1 in Fig. 3a, nt 248-250 in Fig. 3b) and motifs that could lead

to cleavage at either end of this 3 bp stretch and another within it created by a mutation present in both rearranged genes.

## Discussion

We examined the sequences presented as examples of a type of  $V_H$  replacement (that we called ‘Type 2’) because we were not convinced that RAG-mediated recombination was the most likely explanation for them. We considered PCR artefact, as did the original authors, but found that for some examples were inconsistent with this explanation and, as we mentioned in the introduction, there is evidence that AID can engender homologous recombination in human immunoglobulin genes, which suggests that gene conversion might occur. As AID recognition motifs are present in high numbers in immunoglobulin variable region gene segments, their occurrence in the  $V_H$ - $V_H$  recombination regions of the  $V_H$  hybrids of Wilson et al. (Wilson et al. 2000) does not necessarily imply that they are involved in the recombination process but their presence in all cases suggested to us that an AID-dependent mechanism might be an alternative to the hypothesis of RSS involvement and encouraged us to examine the sequences of Itoh et al. (Itoh et al. 2000).

A single point of homologous recombination between two  $V_H$  segments could have two possible origins – either it is a single recombination or there was another point of recombination that was outside the sequenced region – and each of these possibilities has two subdivisions (Fig. 4).

Amongst the sequences of Itoh et al. (Itoh et al. 2000) we found examples that had two points of recombination within the sequenced part of the V segment and therefore resembled gene conversion. Matching of whole V<sub>H</sub> segments to germline gene segments had revealed genes with the same CDR3 sequences but different V<sub>H</sub> matches, but we noticed that some of these matched the same germline segment at the 5' end. The alignments of M6 (Fig. 2a, c), and M31 (Fig. 2a, d), with M10 both showed the same germline matches at both ends but matches to different germline genes in the centre with regions on either side of the differently-matching region in which the germline sequences were identical and contained AID recognition motifs. M6 has no mutations and M31 has only one base-substitution and that is within section 3 so, if gene conversion from the M10 rearrangement is the explanation, it must have occurred before any SHM, at least in the parts of M10 that are retained in the other sequences. This is consistent with the finding in the domestic hen that gene conversion occurs early in the germinal centre reaction and is later down-regulated while SHM continues (Arakawa and Buerstedde 2004).

Alternative explanations of the ST1R M6 and M31 sequences would require two PCR hybridisations in each case since each contains matches to parts of three genes (including the same HCDR3 as ST1R M26). On the other hand, two gene conversions in the same rearranged gene are frequently seen in the domestic hen and rabbit.

In comparison of two clonally related rearranged genes, one of which had undergone gene conversion, one would expect that the donor and recipient

germline genes might show regions of identity at either end of the converted section, *i.e.* places where homologous recombination could occur. However, examination of gene conversion in chicken immunoglobulin genes shows that, though there is always homology between donor and recipient at the 5' end, this is not always so at the 3' end, suggesting a 5' to 3' polarity in the mechanism (McCormack and Thompson 1990).

Returning to the alignment of M31 with M10, of the 18 nucleotide changes from germline in M10, eight are in positions at which the germline sequences differ and it is interesting that in all seven of those that are within section 3 the nucleotide has changed so as to be the same as is found in M31 and IGHV1-24 (Fig. 2d). This seems too many to be explained by chance. There are several potential mechanisms of repair of a double-stranded break by gene conversion (Fig 5a-c). All these possibilities can occur in mammalian cells and heteroduplex formation can be detected in the majority of cases (Elliott and Jasin 2001). Recombination is repressed proportionally to the degree of difference between the sequences but this repression is overcome in mismatch-repair (MMR)-deficient cells. The mechanism of gene conversion in rearranged immunoglobulin genes is unknown but it can be deduced that there must be some means of overcoming the repression of recombination between diverged sequences in the presence of a normal MMR system, and that invasion may occur by a single 3' end only since the process appears to be uni-directional (McCormack and Thompson 1990). It is thus highly likely to involve heteroduplex formation and in this case the MMR system might not be able to distinguish between the original and copied strand,



resulting in some corrections on one strand and some on the other (Martin and Scharff 2002), *i.e.*, some towards the original sequence and some to the template sequence (Fig. 5b, c). This partial gene conversion might explain the mutations in section 3 of ST1R M10 (Fig 2d, Fig. 5d).

In view of the improbability that a rearranged gene containing a complete IGHV3-09 would have been amplified by the family-1-specific V<sub>H</sub> primer used by Itoh et al., it is unlikely that either ST1R G1 or G2 (Fig. 3) were produced by PCR artefact, especially G1, which would have required at least two PCR hybridisation events. However, there is one concern about these sequences. On their own, they are too short at the 3' end for recognition of a J<sub>H</sub> segment but nt 303-328 are identical to nt 302-327 of ST1R M26, M10, M6, and M31 (Fig. 2) showing that they have the same D, J<sub>H</sub> and D-J<sub>H</sub> junction, while the V<sub>H</sub>-D junctions are different in sequence and length. Since both groups of cells come from the right knee of the same patient (ST1), it is possible that the genes come from cells with a common precursor having only a DJ<sub>H</sub> rearrangement and derive from daughter cells that made different V<sub>H</sub>→DJ<sub>H</sub> rearrangements [a commonly observed phenomenon (Choi et al. 1996; Kobrin et al. 2001; Rosenquist et al. 1999; Steenbergen et al. 1993; Steenbergen et al. 1996; Steenbergen et al. 1997; Wasserman et al. 1992)]. The V<sub>H</sub>→DJ<sub>H</sub> rearrangement of ST1R G1 and G2 is out of frame, yet the cells have apparently survived long enough to undergo two further recombination events and numerous mutations. It seems likely therefore that there is an error in the V<sub>H</sub>-D junction sequence. It is unlikely that the sequences come from cells that had made a successful rearrangement on the other

chromosome since they were amplified with an IgG constant region primer and, if the original rearrangement had been out of frame, they would not have switched class.

We noted earlier that the conversion-like events shown in Figs 2 (c) and (d) appear to have occurred before any significant hypermutation. The same appears to apply to ST1R G1 and G2. The mutations at positions 250 and 261 could have occurred after the recombination of IGHV1-69 with the original rearranged gene bearing IGHV3-09 and must have occurred before the gene conversion of a section of IGHV1-02 into IGHV1-69 since they occur in both of the cloned genes, but the latter gene conversion must have occurred very soon after these mutations because none of the other mutations in G1 are present in G2 and *vice versa*. This is again consistent with gene conversion occurring early in germinal centre diversification.

The finding that the sequences described here could be explained by gene conversion suggests that other cases of Type 2 V<sub>H</sub> replacement in Man might also be examples of the same process but are not recognised because the 5' end of the converted section is upstream of the beginning of the sequenced region (Fig. 4b(ii)). This has been reported in a mouse heavy chain from a hybridoma cell-line variant with altered antigen specificity from that of the parent line (Dildrop et al. 1982; Krawinkel et al. 1983). As this example was discovered by amino-acid and DNA sequencing without the use of PCR, there can be no question of PCR artefact in this case, and it therefore appears to be the result of either gene conversion or reciprocal homologous recombination.

It has now been shown in a transgenic mouse (D'Avirro et al. 2005) that sequence transfer between a promoterless VDJ exon and a functional one is due to gene conversion because the donor sequence was found to be unchanged. This has demonstrated conclusively that mouse B-cells do possess the machinery to effect gene conversion of rearranged immunoglobulin genes. It is suggested by the authors that the undoubtedly reduced frequency of such events in some species, such as mouse and man, is due to the regulation of chromosomal accessibility of the upstream potential donor V segments, the chromatin being 'open' during rearrangement and 'closed' again by the time AID is expressed. However, the evidence from the hybridoma, which was not transgenic, suggests that gene conversion can occur with normal chromosome architecture.

If gene conversion always occurs from 5' to 3' of the sense strand, as appears to be the case in the domestic hen, there may be no more significance in the finding of AID motifs at the junctions that represent the 3' ends than of the finding of possible cryptic RSS in the human sequences presented. However, the lack of homology often seen at the 3' ends of converted sections (McCormack and Thompson 1990) means that gene conversion could explain the sharp junction between the two sections of ST1R M10 (Fig. 2b) and the minimal homology at the 3' junctions of IGHV1-02 to IGHV1-69 and 1-69 to 3-09 in ST1R G1 and G2 (Fig. 3). There are also often insertions or deletions of nucleotides at the 3' ends of converted regions in chicken immunoglobulin genes so, if the hybridisation regions of Wilson et al. (Wilson et al. 2000) (Fig. 1) do represent such 3' ends, this could explain the deletion found in CP1-HB2 (alignment 10). Homologous

recombination appears to be a much better explanation of the sequences discussed here than RAG-mediated recombination. The possibility that they were produced as PCR artefacts cannot be completely excluded, and template DNA broken at AID motifs could facilitate this process. However, our findings suggest that gene-conversion occurs in human immunoglobulin genes as it does in other mammalian species, though it may be rare.

**Acknowledgements** The authors wish to acknowledge the support of the Scottish Executive, Chief Scientist Office, Grant No. CZB/4/14 and Cancer Research UK, Grant No. 4327.

## References

- Arakawa H, Buerstedde JM (2004) Immunoglobulin gene conversion: insights from bursal B cells and the DT40 cell line. *Dev. Dyn.* 229:458-64
- Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS (2004) Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J. Mol. Biol.* 337:585-96
- Becker RS, Knight KL (1990) Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. *Cell* 63:987-97
- Choi Y, Greenberg SJ, Du TL, Ward PM, Overturf PM, Brecher ML, Ballow M (1996) Clonal evolution in B-lineage acute lymphoblastic leukemia by contemporaneous VH-VH gene replacements and VH-DJH gene rearrangements. *Blood* 87:2506-12
- D'Avirro N, Truong D, Luong M, Kanaar R, Selsing E (2002) Gene conversion-like sequence transfers between transgenic antibody V genes are independent of RAD54. *J. Immunol* 169:3069-75

- D'Avirro N, Truong D, Xu B, Selsing E (2005) Sequence Transfers between Variable Regions in a Mouse Antibody Transgene Can Occur by Gene Conversion. *J. Immunol* 175:8133-7
- Darlow JM, Stott DI (2005) V<sub>H</sub> replacement in rearranged immunoglobulin genes. *Immunology* 114:155-65
- David V, Folk NL, Maizels N (1992) Germ line variable regions that match hypermutated sequences in genes encoding murine anti-hapten antibodies. *Genetics* 132:799-811
- Dildrop R, Bruggemann M, Radbruch A, Rajewsky K, Beyreuther K (1982) Immunoglobulin V region variants in hybridoma cells. II. Recombination between V genes. *EMBO J.* 1:635-40
- Elliott B, Jasin M (2001) Repair of double-strand breaks by homologous recombination in mismatch repair-defective mammalian cells. *Mol. Cell Biol.* 21:2671-82
- Foster SJ, Dorner T, Lipsky PE (1999) Somatic hypermutation of V<sub>kappa</sub>J<sub>kappa</sub> rearrangements: targeting of RGYW motifs on both DNA strands and preferential selection of mutated codons within RGYW motifs. *Eur. J. Immunol* 29:4011-21
- Gontier É, Ayrault O, Godet I, Nau F, Ladevèze V (2005) Developmental progression of immunoglobulin heavy chain diversity in sheep. *Vet. Immunol Immunopathol.* 103:31-51
- Itoh K, Meffre E, Albesiano E, Farber A, Dines D, Stein P, Asnis SE, Furie RA, Jain RI, Chiorazzi N (2000) Immunoglobulin heavy chain variable region gene replacement as a mechanism for receptor revision in rheumatoid arthritis synovial tissue B lymphocytes. *J. Exp. Med* 192:1151-64
- Kleinfield R, Hardy RR, Tarlinton D, Dangl J, Herzenberg LA, Weigert M (1986) Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a Ly 1<sup>+</sup> B-cell lymphoma. *Nature* 322:843-6
- Kobrin CB, Bendandi M, Kwak LW (2001) Novel secondary Ig V<sub>H</sub> gene rearrangement and in-frame Ig heavy chain complementarity-determining region III insertion/deletion variants in de novo follicular lymphoma. *J. Immunol* 166:2235-43
- Krawinkel U, Zobelein G, Bruggemann M, Radbruch A, Rajewsky K (1983) Recombination between antibody heavy chain variable-region genes: evidence for gene conversion. *Proc. Natl. Acad. Sci. U. S. A* 80:4997-5001

- Lenze D, Greiner A, Knorr C, Anagnostopoulos I, Stein H, Hummel M (2003) Receptor revision of immunoglobulin heavy chain genes in human MALT lymphomas. *Mol. Pathol.* 56:249-55
- Lucier MR, Thompson RE, Waire J, Lin AW, Osborne BA, Goldsby RA (1998) Multiple sites of V lambda diversification in cattle. *J. Immunol* 161:5438-44
- Martin A, Scharff MD (2002) AID and mismatch repair in antibody diversification. *Nat Rev. Immunol* 2:605-14
- McCormack WT, Thompson CB (1990) Chicken IgL variable region gene conversions display pseudogene donor preference and 5' to 3' polarity. *Genes Dev.* 4:548-58
- Neuberger MS, Harris RS, Di Noia J, Petersen-Mahrt SK (2003) Immunity through DNA deamination. *Trends Biochem. Sci.* 28:305-12
- Parg C-L, Hansal S, Goldsby RA, Osborne BA (1996) Gene conversion contributes to Ig light chain diversity in cattle. *J. Immunol* 157:5478-86
- Reth M, Gehrman P, Petrac E, Wiese P (1986) A novel VH to VHDJH joining mechanism in heavy-chain-negative (null) pre-B cells results in heavy-chain production. *Nature* 322:840-2
- Rogozin IB, Diaz M (2004) Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J. Immunol* 172:3382-4
- Rogozin IB, Kolchanov NA (1992) Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim. Biophys. Acta* 1171:11-8
- Rosenquist R, Thunberg U, Li AH, Forestier E, Lonnerholm G, Lindh J, Sundstrom C, Sallstrom J, Holmberg D, Roos G (1999) Clonal evolution as judged by immunoglobulin heavy chain gene rearrangements in relapsing precursor-B acute lymphoblastic leukemia. *Eur. J. Haematol.* 63:171-9
- Sehgal D, Schiaffella E, Anderson AO, Mage RG (2000) Generation of heterogeneous rabbit anti-DNP antibodies by gene conversion and hypermutation of rearranged VL and VH genes during clonal expansion of B cells in splenic germinal centers. *Eur. J. Immunol* 30:3634-44

- Selsing E, Xu B, Sigurdardottir D (1996) Gene conversion and homologous recombination in murine B cells. *Semin. Immunol* 8:151-8
- Steenbergen EJ, Verhagen OJ, Nibbering CP, van den Berg H, van Leeuwen EF, Behrendt H, dem Borne AE, van der Schoot CE (1996) Clonal evolution of immunoglobulin heavy chain rearrangements in childhood B-precursor acute lymphoblastic leukemia after engraftment in SCID mice. *Leukemia* 10:1471-8
- Steenbergen EJ, Verhagen OJ, van den Berg H, van Leeuwen EF, Behrendt H, Slater RR, dem Borne AE, van der Schoot CE (1997) Rearrangement status of the malignant cell determines type of secondary IgH rearrangement (V-replacement or V to DJ joining) in childhood B precursor acute lymphoblastic leukemia. *Leukemia* 11:1258-65
- Steenbergen EJ, Verhagen OJ, van Leeuwen EF, dem Borne AE, van der Schoot CE (1993) Distinct ongoing Ig heavy chain rearrangement processes in childhood B- precursor acute lymphoblastic leukemia. *Blood* 82:581-9
- Tsai HF, D'Avirro N, Selsing E (2002) Gene conversion-like sequence transfers in a mouse antibody transgene: antigen selection allows sensitive detection of V region interactions based on homology. *Int. Immunol* 14:55-64
- Wasserman R, Yamada M, Ito Y, Finger LR, Reichard BA, Shane S, Lange B, Rovera G (1992) VH gene rearrangement events can modify the immunoglobulin heavy chain during progression of B-lineage acute lymphoblastic leukemia. *Blood* 79:223-8
- Weinstein PD, Anderson AO, Mage RG (1994) Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms. *Immunity* 1:647-59
- Wilson PC, Wilson K, Liu YJ, Banchereau J, Pascual V, Capra JD (2000) Receptor revision of immunoglobulin heavy chain variable region genes in normal human B lymphocytes. *J. Exp. Med* 191:1881-94
- Winstead CR, Zhai SK, Sethupathi P, Knight KL (1999) Antigen-induced somatic diversification of rabbit IgH genes: gene conversion and point mutation. *J. Immunol* 162:6602-12

- Wysocki LJ, Geftter ML (1989) Gene conversion and the generation of antibody diversity. *Annu. Rev. Biochem.* 58:509-31
- Zan H, Wu X, Komori A, Holloman WK, Casali P (2003) AID-dependent generation of resected double-strand DNA breaks and recruitment of Rad52/Rad51 in somatic hypermutation. *Immunity* 18:727-38
- Zhang Z, Zemlin M, Wang YH, Munfus D, Huye LE, Findley HW, Bridges SL, Roth DB, Burrows PD, Cooper MD (2003) Contribution of V<sub>H</sub> gene replacement to the primary B cell repertoire. *Immunity* 19:21-31



## Figure legends

**Fig. 1** The hybrid  $V_H$  DNA sequences of Wilson et al. (Wilson et al. 2000) (modified from their Fig. 3) displayed to show that AID recognition motifs  $WRC(Y)\cdot(R)GYW$  and  $WRCH\cdot DGYW$  are related to all recombination regions. All AID motifs are underlined and in bold but only the relevant ones are marked with  $WRCH$  etc.. These have cleavage points at or within the borders of the recombination regions (arrowed) within which the germline genes are identical. Newer gene nomenclature has been substituted for the old gene names. '-' indicates same nucleotide as top line. The first three alignments have been extended beyond those shown by Wilson et al. so as to include ends of recombination regions. Alignment 10 has not been extended because the recombination is presumed to occur at the border of the 24 bp deletion indicated by asterisks. Because of mutations in the hybrid, the exact position of the deletion is uncertain. Three possible positions are shown. The top one is that suggested by Wilson et al., with its 5' end corresponding to the edge of a suggested cryptic RSS. In the third, the 5' end corresponds with the target C of a  $WRCY$  motif, TACT that may have existed in the hybrid sequence before the deletion, the T (present in both germline sequences) being lost in the deletion. We have rearranged the sequences to follow the pattern, used in the other alignments, of having the germline sequence that matches the 3' portion above the hybrid and that matching the 5' portion below, and show some differences between the germlines not shown by Wilson et al.. No limits for a recombination region are shown on alignment 14 because the hybrid sequence resembles the lower germline sequence throughout the illustrated portion. We were

not able to check this sequence because it has not been submitted to a database but the G of an RGYW motif lies precisely at the proposed cleavage point related to a proposed cryptic RSS. Some corrections have been made in sequences, sequence labelling and nucleotide numbering.

**Fig. 2** A new analysis of the sequences of Fig. 5 A from Itoh et al. (Itoh et al. 2000). In all parts of the figure, the D segment is shown in blue, the J in mustard and the V-D and D-J non-genomic sections in red. The  $V_H$  segments are colour-coded as marked. (a) Diagrammatic alignment of the four rearranged genes with the respective germline  $V_H$  sequences that give the closest matches overall. In ST1R M6 and M31 the sections between the dashed boxes are the minimum lengths that match exclusively to the  $V_H$  coded by their colour while the distance between the outer ends of the boxes indicates the maximum match. Within each box the sequence of the  $V_H$  of the central section is the same as that coded by the colour within the box. The sections in black were not sequenced and are not shown in b-d. In parts b-d, the progenitor  $V_H$  gene sequences are highlighted in continuous colour and the cloned rearranged genes are highlighted where they are the same as one of the germline genes but not another and AID recognition motifs are marked as in Fig. 1. (b) Alignment of the rearranged genes ST1R M26 and ST1R M10 with germline sequences to show the division of ST1R M10 into two sections. The division does not correspond exactly with the supposed cryptic RSS of Itoh et al. (dark red bar). Bases of ST1R M10 highlighted in mauve or grey indicate matches to the germline  $V_H$  gene segments. Here, and in subsequent parts

of this figure, base substitutions are highlighted in green. (c) Alignment of the rearranged genes ST1R M10 and ST1R M6 with germline sequences. Here ST1R M10 and IGHV1-18 are highlighted to show that ST1R M6 may have been derived from the ST1R M10 rearrangement by gene conversion of a central section from IGHV1-18. Sections 2 and 4 correspond to the first and second boxes in Fig. 2(a). (d) Alignment of the rearranged genes ST1R M10 and ST1R M31 with germline sequences. ST1R M10 and IGHV1-24 are highlighted to show that ST1R M31 may have been derived from the ST1R M10 rearrangement by gene conversion of a central section from IGHV1-24. Again, sections 2 and 4 correspond to the first and second boxes in Fig. 2(a).

**Fig. 3** A new presentation of the sequences of Fig. 5 B of Itoh et al. (Itoh et al. 2000). Same conventions as in Fig. 2. (a) Diagrammatic alignment of the genes as in 2a. (b) Sequence alignment. Both cloned genes are highly mutated. However, in section 3, the two germline V gene sequences differ in 19 nucleotides. Of these, G1 resembles IGHV1-02 in 17 positions while G2 resembles IGHV1-69 in 15 positions.

**Fig. 4** Four different possible explanations for a finding of a single point of homologous recombination between a rearranged gene (green-blue-orange = VDJ) and another V segment (brown). Either there really is only one point of recombination (a) or there are two but the other point is outside the sequenced region (b). In case (a), either the genes are on different chromosomes (i), in which

case chromosome ends are exchanged, or they are on the same chromosome (ii), in which case a deletion occurs. In other species another possibility is inversion, but this case does not arise in Man because all the V gene segments are in the same orientation. In case (b), either there is reciprocal exchange (i) or gene conversion (ii) in which case the template is not altered. For mechanisms of the latter, see Fig. 5.

**Fig. 5** Potential mechanisms of repair of a double-stranded break by gene conversion. (a) exonucleolytic digestion of both ends, followed by invasion of the template by both 3' broken ends to copy the missing region, resulting in perfectly complementary strands to replace the missing region; (b, c) invasion of the template by both 3' ends (b) or just one end (c) after little or no exonucleolytic cleavage, resulting, after repair of the broken strand, in stretches of heteroduplex DNA (having slightly different sequences in the two strands due to differences between the template and the broken gene), which could then be corrected by mismatch repair (MMR). The MMR system might not be able to distinguish between the original and copied strand, resulting in some corrections on one strand and some on the other, a partial gene conversion. (d) Section 3 of ST1R M10 from Fig. 2d redisplayed to highlight the fact that in seven positions (arrowed) it resembles IGHV1-24. This might be explained by partial gene conversion.