Darlow, J.M. and Stott, D.I. (2005) V_H replacement in rearranged immunoglobulin genes. *Immunology 114*(2):pp. 155-165.

V_H replacement in rearranged immunoglobulin genes

Short title: V_H replacement

JOHN M. DARLOW & DAVID I. STOTT *Department of Immunology, Division of Immunology, Infection and Inflammation, University of Glasgow, Western Infirmary, Glasgow, G11 6NT, UK*

john.m.darlow@clinmed.gla.ac.uk

Keywords V_H replacement; secondary rearrangement; receptor revision; receptor editing; immunoglobulin genes

Abbreviations: AID, activation-induced cytidine deaminase; DSB, double-strand break; NHEJ, non-homologous end-joining; nt, nucleotide; RSS, recombination signal sequence(s); SHM, somatic hypermutation; TdT, terminal deoxynucleotidyl transferase.

SUMMARY

Examples suggesting that all or part of the V_H segment of a rearranged V_HDJ_H may be replaced by all or part of another V_H have been appearing since the 1980s. Evidence has been presented of two rather different types of replacement. One of these has gained acceptance and has now been clearly demonstrated to occur. The other, proposed more recently, has not yet gained general acceptance because the same effect can be produced by PCR artefact. We review both types of replacement including a critical examination of evidence for the latter. The first type involves RAG proteins and recombination signal sequences (RSS) and occurs in immature B cells. The second was also thought to be brought about by RAG proteins and RSS. However, it has been reported in hypermutating cells which are not thought to express RAG proteins but in which AID has recently been shown to initiate homologous recombination. Re-examination of the published sequences reveals AID target sites in V_H-V_H junction regions and examples that resemble gene conversion.

INTRODUCTION

Faced with selective pressure to change the specificity of a rearranged immunoglobulin gene, (or rescue itself from two non-functional rearrangements) a B cell may make a secondary rearrangement on the same chromosome[1-4]. Some authors divide this into 'receptor editing' when it occurs centrally (in the bone marrow), and 'receptor revision' when it occurs in the periphery (in germinal centres) the former thought to be largely tolerance-driven and the latter mainly diversity-driven. Most rearranged light-chain genes will have both unrearranged V_L gene segments upstream and unrearranged J_L segments downstream, so it is possible for an upstream V_L segment to recombine with a downstream J_L segment to form a new V_LJ_L exon with deletion of the original rearrangement. At the heavy-chain locus the situation is different because the rearranged gene contains a D segment and in humans all the unused D segments are deleted during the D-J_H and V_H-DJ_H rearrangements. Several possible varieties of secondary rearrangement at the heavy chain locus have been suggested, but the ones for which there are the most evidence are replacement of all or part of the primarily rearranged V_H segment by all or part of an upstream V_H gene segment, and these comprise the subject of this article.

Two quite different mechanisms of V_H replacement have been reported. We refer to these as Types 1 and 2. We review the evidence for both types of replacement, describe examples of confusion between the two, and consider what is the most likely explanation for the Type 2 sequences in the light of other recent findings and our own re-examination of the sequences. We begin with a summary of current understanding of primary V(D)J recombination.

V(D)J REARRANGEMENT AND NON-HOMOLOGOUS END JOINING

V(D)J recombination takes place in two stages. The first, recognition and cleavage of specific sequences within the Ig loci, is unique to V(D)J recombination. The second, the processing and joining of the cleaved ends, involves components of the ubiquitous non-homologous end-joining (NHEJ) system. Understanding of the details of primary V(D)J rearrangement is developing rapidly[5-9]. The functional V, D and J gene segments have recombination signal sequences (RSS), each consisting of a specific heptamer and

nonamer separated by a spacer of 12 bp (roughly one turn of the DNA helix) or 23 bp (roughly two DNA turns). These distances are important for protein binding, and most efficient cleavage and rejoining occurs when one of each type of RSS are brought together in a protein complex. This ensures recombination of the correct coding sequences to make a functional product since, at the heavy-chain locus, the V and J segments all have RSS with a 23 bp spacer and the D segments have 12 bp spacers on both ends**.** The heptamer sequences are immediately adjacent to the ends of the coding sequences that are to be joined and have the same consensus sequence but are in opposite orientations. The sequence of the heptamer is more critical than that of the nonamer and the most important sequence is the three bases next to the cleavage point, CAC on the sense strand in RSS at the 3' ends of coding segments, GTG in those at the 5' ends[10]. A complex formed by the products of the 'recombination activating genes', RAG-1 and RAG-2, and the high mobility group 1 (HMG1) protein bind**s** the RSS and nick**s** the DNA precisely at the interfaces between the coding sequences and the RSS. The free –OH group on the 3' end of the coding sequence at each nick then attacks the opposite DNA strand and the result is a closed hairpin loop on each coding sequence end and a blunt double-strand break at the end of each RSS.

The ends are then joined by NHEJ, present in all cells for repair of double-stranded DNA breaks. It frequently uses very short sequence-matches of typically 1 - 4 bp near the ends to be joined and frequently involves removal of nucleotides from the ends. In V(D)J recombination this imprecision is a benefit, conferring extra diversity to antibodies.

The proteins of the NHEJ system involved in V(D)J recombination include the Ku70-Ku86 heterodimer, Artemis, DNA-dependent protein kinase catalytic subunit, XRCC4 (X-ray cross-complementation group 4), DNA ligase IV, and others yet to be identified[5-7,11]. The cleaved ends bearing RSS are usually ligated together (signal-to-signal joint) by perfect blunt-ended ligation without loss or addition of bases. At the coding ends, the hairpins are opened by nicking and, if not central, this results in addition of palindromic (P) nucleotides. However, there is often excision of several nucleotides. Before joining there is usually nontemplated (N) addition of nucleotides to the 3' ends by terminal deoxynucleotidyl transferase (TdT). Complementary matches of one or a few bases at or near the ends of these extensions may then be used to pair the ends, non-matching bases are removed, gaps are filled in and nicks ligated. Where the coding segments joined are in the same orientation, as in the unrearranged human heavy-chain locus, ligation of the RSS causes the exclusion of the intervening DNA from the chromosome as a closed loop. Presentation of a functional B-cell receptor on the cell surface leads to down-regulation of the *RAG* genes, preventing further V(D)J rearrangement. It has yet to be established for certain whether *RAG* genes can be reactivated in mature cells *in vivo* or whether RAG-mediated secondary rearrangement only occurs in immature cells[12]. However, it has been shown that RAG expression and receptor revision can be induced in mature B cells *in vitro*[13].

TYPE 1 V_H REPLACEMENT

Type 1 V_H replacement is essentially the same mechanism as primary V(D)J rearrangement, using legitimate or 'cryptic' RSS in opposite orientations, usually involving deletion from the cleaved coding ends, addition of P and N nucleotides at the new junction, and sometimes a substantial change in length. It also frequently causes a change in frame, making a non-functional product from a functional one, or *vice-versa*.

Because the D segment of the rearranged V_H-D-J_H exon no longer has its own 5' RSS, cryptic RSS within the V_H are always involved in Type 1 V_H replacement. There are variations between RSS of different gene segments in the heptamer, spacer, and nonamer sequences[10,14] but the consensus sequence of the heptamer on the 5' ends of the D segments is CACTGTG. Close to the 3' end of nearly all of the functional human and mouse V_H segments is the sequence TACTGTG, only one base different from the consensus heptamer, and this can act as a surrogate for the original RSS on the 5' end of the D segment[15]. Within most V_H segments there are also more proximal cryptic RSS in both orientations, not as close to the RSS consensus as the 3' cryptic RSS. These appear to be used very rarely in Type 1 replacement. Separated by 12 bases upstream from the 3' cryptic heptamer (also known as 'the internal heptamer') is a nonamer sequence that varies a little between V_H genes. It does not resemble the consensus nonamer of the real RSS but, because of its conservation, Chen *et al.*[16] considered that it might act as a signal sequence. Nadel *et al.*[17] did not detect recombination in pre-B cells transfected with plasmids containing the complete nonamer-spacer-heptamer 'embedded RSS' present in the 3' end of several of the mouse V_H genes, and concluded that V_H replacement must be an infrequent event, unlikely to contribute significantly to receptor editing. However, numerous examples of V_H replacement continued to be observed in pre-B-cell lines (endogenous genes and introduced substrates), lymphomas and leukaemias[15,18-29], and in transgenic mice[16,30,31].

The reported Type 1 V_H replacements fall into three subtypes. (i) As in the references above, a complete upstream V_H segment may be recombined into a rearranged V_H-D-J_H through the agency of its own RSS and the 3' cryptic RSS of the rearranged V_H (Fig. 1a). This effectively increases the length of the V_H-D junction and has a (theoretically ⅓) chance of producing a functional heavy chain (depending on the reading frame). Most examples of Type 1 V_H replacement recorded so far are of this subtype. (ii) A complete V_H gene segment may be recombined into the V_H-D-J_H with its own RSS and a more 5' cryptic RSS in the already-combined V_H (Fig. 1b). In the one example to date[32], the incoming V_H was in a different frame from the original and this resulted in a stop codon being read in the original V_H. The nearest possible cryptic RSS 5' of the cleavage point in the original V_H (now known as V_H3-49) is T<u>A</u>TG<u>GTG</u> in codons 53 – 55. Only the four bases underlined are shared in common with the consensus RSS, which may explain why this is a rare event. (iii) Cryptic RSS in opposite orientations and different positions within both V_H might be involved (Fig. 1c). However, we have now shown (unpublished observation) that the three published examples[33] result from a duplication within one V_H.

Other types of RAG-mediated rearrangement, D→V_HDJ_H and V→D→V_HDJ_H may occur in artificial systems[16,34]. Taki *et al.*[34] found that of 53 RSS-like motifs present in the monoclonal mouse system that they were using, only two were used, suggesting that other sequences are involved in the targeting of RAG-mediated recombination in addition to the RSS.

The sequence between the 3' cryptic RSS and the 3' end of the V_H is 5 – 9 (usually 7) bases long in human V_H genes. This may be truncated from its 3' end in the original rearrangement and from its 5' end after the 3' cryptic RSS is cleaved off. Therefore the V_H-D junction, after Type 1(i) V_H replacement, will usually be longer than the original V_H-D junction with the same sequence towards its 3' end, new sequence at its 5' end, and possibly a remnant of the tail of the original V_H in the middle, as shown in Fig. 2(a).

This is easy to identify by comparing sequences from members of the same B-cell clone but almost impossible to identify in isolation.

Recently, Zhang *et al.*[35] demonstrated Type I V_H replacement conclusively in an acute lymphocytic leukaemia cell-line that undergoes continuous pro-B to pre-B to B-cell differentiation and serial V_H replacement. Different members of the cell line contained rearranged genes with the same D-J_H junction but different V_H segments, and remnants of previous V_H segments in some of the V_H-D junctions. The investigators demonstrated double-strand DNA breaks at 3' cryptic RSS, V_H-replacement excision circles, binding of co-purified recombinant RAG1/RAG2 to cryptic RSS, nicking at the heptamer ends, and coding-end hairpin formation. Furthermore, they found that the 3' cryptic heptamer enhanced hairpin formation with true RSS with both 12 bp and 23 bp spacers. They also found evidence for Type 1 V_H replacement in healthy human neonates and adults. A search of the V-D junctions of 343 functional IgH sequences from 18 subjects identified 16 (~5%) with 5 consecutive matches to a 3' V_H sequence. When they lowered the stringency to 6 bases with one mismatch, this increased to ~12%, and the true frequency of Type I replacement is likely to be much higher, because, as we mentioned earlier, there will frequently be little or no trace of the previous V_H segment.

TYPE 2 V_H REPLACEMENT

Type 2 V_H replacement has been described much more recently than Type 1[36-38]. The result gives the appearance of homologous recombination between two V_H gene segments (Fig. 2b). Because the exchange occurs in a stretch in which both segments have the same sequence, the exact point of exchange cannot be determined for certain; it is only delimited by the closest positions of difference between the two sequences in either direction. It has been proposed to occur at cryptic RSS but the recombination process is quite different from normal V(D)J rearrangement because the supposed 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[36]) no nucleotides are deleted from the gene sequence. The specificity of the antibody is changed, the reading frame of the gene is unchanged and the V_H-D junction is always unchanged. If this type of V_H replacement is guided by cryptic RSS, the joints formed must be signal-to-coding joints, known as hybrid joints. These have been shown to occur but, when they do, there is usually deletion of nucleotides from the coding end and addition of new nucleotides[39].

During PCR amplification of DNA prepared from many lymphocytes, it is possible for the reaction to produce a rearranged V_H-D-J_H sequence with a hybrid V_H just like the postulated Type 2 replacement (Fig. 3). Double strand breaks (DSBs), targeted to RGYW·WRCY motifs have been found to occur in unrearranged V gene segments and in rearranged genes in the absence of AID, a protein essential for hypermutation[40-42]. On AID expression, additional staggered breaks occur at the same target[43]. All these breaks will result in incomplete single-strand PCR products, ripe for hybrid artefact formation, whose likelihood may thus be increased in mutating cells. It is therefore necessary to establish that Type 2 replacements occur *in vivo*.

Wilson *et al.*[36] examined sequences from purified $IgD^+IgM^-CD38^+$ tonsil B cells. In one group of 65 clonally related sequences (with the same CDR3) there were 7 that all had the same partial V_H replacement but differed in their mutations. In a phylogeny of the complete sequences these hybrids would be bound to come together on the tree, but

these sequences all came on the same branch of a phylogenetic tree based only on the unreplaced segment, strongly suggesting that a single replacement event occurred, followed by further somatic mutation. There are at least three ways that this appearance might be produced by PCR artefact. Firstly, a single PCR hybridisation could have occurred, followed by polymerase errors in some of the copies. These sequences would come together on the tree because they only differed by a single nucleotide. Experience in our laboratory has shown that even with a high-fidelity polymerase system this can happen when many copies of the same template are cloned and sequenced (error frequency ~1/1200 bp in nested PCR or $\sim 1{\cdot}1 \times 10^{-5}$/changes/bp/cycle[44]). Discounting this, one might postulate that 7 separate PCR hybridsations had occurred between the same regions of copies of the same two rearranged genes but these would have to have mutation patterns so similar that the products came together on the tree despite the presence of many differently-mutated copies of these genes (and many other rearranged templates) in the reaction. The chance of this seems negligible. In another group of 77 clonally related sequences there were two with the same Type 2 V_H replacement and they also came side-by-side on their phylogenetic tree, apparently compounding the likelihood that this was *in vivo* recombination.

There is, however, another possibility. If, after the first PCR hybridisation, further hybridisation occurred not to native templates but to the hybrid template, these secondary hybridisations would be much more likely to occur with sequences having closely related patterns of mutation. On both trees there were nonhybrid sequences beyond hybrid ones on the same branch. This is clearly impossible. The trees were produced by parsimony, and the findings might be a demonstration of the inadequacy of parsimony alone in constructing phylogenetic trees, or due to use of incomplete sequences, but it could also be an indication that the results were artefactual. The investigators[36] took the presence of somatic mutations before as well as after the replacements to indicate that replacement must have taken place in the periphery, *i.e.* that they were examples of receptor revision, and this must surely be the case if their explanation is correct.

Four more hybrid V_H genes were found in the two clones already mentioned, and a further ten amongst >700 sequences from different B-cell subsets from tonsils of four other individuals, but they were all single examples of partial V_H replacements and so could not be distinguished from PCR artefacts. However, cryptic RSS were found in all nine different cross-over regions in the fifteen different examples of V_H exchange. As some of these regions were quite long (up to at least 43 bp) and some of the identified cryptic RSS were only CAC or GTG, there was a high probability of some of these being present by chance, but the presence of an RSS in all nine regions cannot be ignored.

The V_H in a rearranged V_HDJ_H can only be replaced by a V_H upstream on the same chromosome, because the downstream V_H segments have all been deleted in the original rearrangement, but in 8/15 of the hybrids the incoming V_H was downstream in the existing V_H locus maps, which again could indicate PCR artefact. The authors considered the most likely possibility to be that mapping of the V_H region of more individuals will reveal more differences in order since the only two complete maps differ considerably. Less likely, they considered, was the possibility that the incoming V_H sequence had come from the other chromosome. Zhang *et al.*[35] noted that 4 of the 16 possible V_H replacement products, that they found by searching human V-D junctions for matches to 3' ends of V gene segments, disagree with the gene order known so far,

but these were not proven V_H replacements, and the authors did not find violation of the order in their serially-V_H-replacing cell-line.

If the hybrids were formed by RAG-mediated recombination at cryptic RSS, one would expect recombinations between V_H segments from different families, just as with Type 1 V_H replacements. The PCRs were carried out with V_H Family 4 leader and C_μ, C_δ and C_γ primers and both parts of each hybrid were derived from V_H4 segments, as would be expected in PCR artefact. However, if the sequences do represent *in vivo* recombined genes, the finding strongly favours homologous recombination since this is much more likely to occur between members of the same family. Another argument against RAG-mediated recombination is that it should have produced some hybrids shorter and others longer than normal V_H segments by recombining at RSS at different positions in the two V_H segments. No such products were found, though this could have been because they were non-functional and the cells bearing them were thus not amongst the selected IgD^+IgM^- cells.

If the hybrids of Wilson *et al.*[36] were formed *in vivo* by RAG mediated recombination, then DNA double-strand breaks should be found at the proposed recombination junctions (from cleavage at cryptic RSS) as well as circular recombination products but all attempts to detect them failed or were inconclusive. The authors suggested that the Type 2 V_H replacements they found might be the tip of the iceberg of all receptor revision because complete (*i.e.* Type 1) functional replacements of V_H or J_H would be invisible and non-functional rearrangements would be inactivated, allowing rearrangement on the other chromosome, or eliminated.

Itoh *et al.*[37] amplified IgM, IgG, and IgA heavy-chain sequences with Family 1 V_H segments from synovial tissue samples from patients with rheumatoid arthritis. They too found hybrid sequences in some clones but no examples of the same hybrid with different mutations. Again, cryptic RSS were present in all the cross-over regions and again were always at the same position in the two recombining V_H segments. As with the other investigators' results, in half of their hybrids the 5' V_H was from downstream of the original V_H on existing genomic maps. The authors gave the following reasons why they thought that their results were genuine: (i) hybrid rearrangments were found in different tissues from the same patient and in different patients; (ii) they were found in PCRs from both cDNA and genomic DNA (using a consensus J_H primer); (iii) RSS-like sequences were always found in the cross-over regions; (iv) rearrangements were always in-frame and coded for translatable proteins; (v) two different sets of related sequences showed members with different partial V_H replacements to the same original VDJ either in the same heptamer-containing region or in different ones. Reasons (i), (ii) and (v) could all be accommodated by PCR artefact. Reason (iii) we have already discussed and will return to later. Reason (iv) could be explained by PCR hybridisation or homologous recombination, but would be less expected in a RAG-mediated process in which one might expect some diversification of the junctions. However, there is a better reason for believing that their results represented *in vivo* recombination. Though most of the receiving segments were of the same family as the incoming segments, there were two different hybrids in which a Family 1 segment was hybridised to the same 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, making the chances of generating artefactual hybrids with such a gene very small. Further support seemed to come from the following additional investigations.

Firstly, they looked for double strand breaks occurring at the cryptic RSS by blunt-end ligation followed by PCR, using human bone marrow cells as positive controls and fibroblasts as negative controls. This showed clearly that double strand breaks (DSBs) occurred in the synovial tissue samples and the positive control but not in the fibroblasts and there were two distinct bands corresponding to cleavage at two V_H heptamers, one of them the 3' cryptic RSS seen in most V_H segments. Sequences of the PCR products were said to show that the breaks were at or within heptamers. RAG-mediated cleavage should be precisely at the edge of a heptamer. A break within one might suggest that the DNA had been chewed back but none of the examples of recombination in this paper showed any deletion. However, DSBs have since been found to occur even in unrearranged V gene segments[40-42] which somewhat diminishes the significance of the finding, though the breaks tend to occur at RGYW·WRCY motifs. Secondly, Itoh *et al.* found evidence of *RAG* activity in several of their synovial tissue samples and the positive but not the fibroblast control. Taken together, these findings suggested that RAG-mediated recombination was occurring in the tissue but not whether it was Type 1 or Type 2. Type 1 V_H replacement with the 3' cryptic RSS would sometimes code for a functional product and the sequence would be identified as having secondary rearrangement only if a sequence with the original rearrangement were also cloned. Type 1 recombination with the more 5' cryptic RSS at which they saw cleavage would not be expected to have produced a functional antibody (see Fig. 1b. The product in this particular case would have been about 70 bp longer than the original). The authors considered the possibility that their hybrids could have been formed by RAG-mediated cleavage at cryptic RSS followed by homologous recombination.

Single-cell PCR avoids the problem of hybrid product formation that can occur when DNA or RNA are extracted from many cells together. In order to determine whether V_H replacement accompanies somatic hypermutation in the germinal centre Goossens *et al.*[45] looked for hybrid V_H segments in single-cell PCR products from 124 IgD-only and 49 IgD^-IgM^- peripheral blood B cells and found no examples of Type 2 or Type 1. However, recently Lenze *et al.*[38] have discovered very strong evidence in another way. They looked for Type 2 V_H replacement in two human mucosa-associated lymphoid tissue (MALT) lymphomas. Using DNA from many cells and nested PCR of rearranged heavy-chain sequences, they found a major clone of related VDJs (with identical CDR3 and J_H) and a polyclonal background of many unrelated sequences in each tumour. In each case, the related set contained members with hybrid V segments. However, 23/26 related sequences from three independent sets of PCRs from one tumour were V_H3-23(5')/V_H3-07(3') hybrids. 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. The remaining two members of the clone had V_H3-30 hybridised to the same original V_H3-07 rearrangement, and shared several mutations with the other hybrids that were not present in the sequence with the complete V_H3-07. These findings alone are very good evidence that this type of V_H replacement is a real *in vivo* event. Furthermore all the hybrids could be amplified using primers for the common CDR3, and the different CDR2s. In the other case, 49/53 related sequences had a complete V_H4-59. The other four sequences were hybrids with three different V_H segments. All the mutations in the complete V_H were also present in the hybrids so PCR artefact cannot be excluded on these grounds. However, both alleles of V_H4-61 were present in hybrids and therefore, if this occurred *in vivo*, there *must* have been exchange

between chromosomes. The other hybrid in this case was with V_H4-04, which is downstream of V_H4-59 and therefore should have been deleted in the initial VDJ recombination. The authors suggest that hybridisation with V_H4-04 was the initial trans-chromosomal event and that this was subsequently replaced by the V_H4-61 allele upstream on the same chromosome. This is borne out by the mutation pattern.

In both lymphomata, all the hybrids contained donor and acceptor V_H segments from the same family. This agrees with our conclusion from the earlier papers that homology is involved in Type 2 recombination. In the first case, the likely point of exchange does not contain a cryptic RSS. In the second case, there are several potential cryptic RSS, one of which is shared by all hybridising genomic segments. Thus this paper provides strong evidence that Type 2 V_H replacement is a real *in vivo* phenomenon, that it depends on sequence homology, and that, though cryptic RSS are frequently present in the cross-over region, this is not always the case.

Cleavage at RSS by RAG proteins is affected by the DNA sequence in the coding flank[46,47] and non-favourable sequences tend to receive a single-strand nick but not double-strand cleavage. Poor cryptic RSS away from the standard 3' position may therefore be more likely to be nicked. Though the RAG proteins direct DSBs to repair by NHEJ, these nicks are liable to engender repair by homologous recombination[48] so, if there were any RAG activity, cryptic RSS could well be sites of origin of Type 2 replacement by homologous recombination. However, most of the examples of Type 2 replacement come from mutating cells in which most investigators believe *RAG* genes are not expressed.

One of the hybrids of Itoh *et al.*[37] shows no mutation before or after recombination. Recently further examples of V_H-V_H hybrids have been published[49] in human immature B cells that have developed in NOD/SCID mice and show no somatic hypermutation. The five hybrid sequences in this paper are all between the same two family 4 V_H segments, with one exchange region in two and another in the other three, and again cryptic RSS have been suggested in each region. One of these has a heptamer GACTGTT which seems very unlikely to act as an RSS since it contains neither CAC nor GTG. Though the authors experimented with producing PCR hybrids and concluded that the published sequences were probably not artefacts, there is nothing in the sequences themselves that can verify this.

EXCEPTIONS OBSCURING THE DIFFERENCE BETWEEN TYPES 1 AND 2

Though there is usually deletion and addition of nucleotides in Type 1 V_H replacement and, in the common subtype, alteration of the V_H-D junction, there are exceptions. One such was seen in one of the first cases reported[50]. Two sequences from a mouse B-cell lymphoma line were found to have different V_H segments but to be identical from the beginnings of their V_H-D junctions onwards so that this might appear to be a case of Type 2 V_H replacement in which the cross-over was within the last 3 bp at the 3' ends of the V_H segments. The authors, however, believed that the replacement had occurred by the mechanism that we now call Type 1. In the V_H gene segments concerned, the distance between the 3' cryptic heptamer and the 3' end of the segments is 5 bp and there is a difference in the second base to the 3' of the heptamer. For the replacement to have occurred by the Type 1 mechanism described above, the new and old V_H would have had to have been modified and joined in such a way that the V_H was replaced seamlessly with neither frame-shift nor any change in the V_H-D junction. Later the same authors[19] published the same replacement and a replacement of yet another complete V_H

to the same original, again with no frame-shift nor change in the junction. The finding of two such replacements to the same substrate is very unlikely unless either they were Type 2 or some other mechanism was operating to ensure perfect match.

It is now known that there is another mechanism that can explain their observations. In mice, particularly fetal and neonatal ones[51,52], and in fetal and neonatal humans[53], pairs of coding ends to be joined in V(D)J rearrangement may be overlapped and spliced at regions of sequence identity as little as 1 - 5 bp in length. This is more frequent in the absence of TdT (refs in Ref [52]) or Ku86, which also seems to be involved in the addition of extra nucleotides[54]. In the light of knowledge of the place of non-homologous recombination in V(D)J rearrangement, the explanation is clear. After non-templated addition of nucleotides to the coding ends, the formation of the join by the matching of 'microhomologies' in these sequences is not apparent because these nucleotides are not present in the genomic sequences. However, when there is no addition of nucleotides, the matching of micohomologies is apparent because these sequences belong to the genomic segments. After cleavage at the internal heptamer in the V_H portion of a rearranged V_HDJ_H, the stump of the old V_H is effectively the new 5' end of the D that is to be joined to the new V_H so the substrate is there for the type of recombination described above. This mechanism could not account for most of the reported sequences that appear to be Type 2 V_H replacement because cleavage of the two V_H gene segments at equivalent positions would leave no homologous ends to overlap. However, it has also been reported that, in transfected V(D)J recombination substrates in Ku86 deficient mouse cell lines, hybrid joints are abundant, and that the majority of them have no nucleotide excision from either of the ends and no nucleotide addition[55].

THE TIMING OF EVENTS

Most of the examples proposed as Type 2 V_H replacement have been in cells undergoing somatic hypermutation (SHM). Therefore the contention that cryptic RSS are involved requires that RAG proteins are re-expressed in mature, mutating cells (unless there is another, unknown, protein that recognises RSS). Though Itoh *et al.*[37] found *RAG* mRNA in several of their synovial tissue samples, they did not demonstrate that it was expressed as protein or that it came from mutating cells, and Zhang *et al.*[35] concluded that Type 1 V_H replacement does not occur later than the immature B cell stage.

AID (activation-induced cytidine deaminase) is essential not only for SHM but also for immunoglobulin gene conversion and class-switching, and the initiating event in each case is believed to be the deamination of deoxycytosine to uracil[56]. Gene conversion is a non-reciprocal form of homologous recombination in which a section of a gene is replaced by copying from another without alteration of the template gene. This is the primary method of producing antibody diversity in birds and several investigated mammals but not primates and rodents[57]. Though gene conversion has not been observed in rearranged human V_H segments, 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[43]. Therefore, if Type 2 V_H replacement occurs by homologous recombination it could well appear in mutating cells and the mechanism might be gene conversion rather than reciprocal exchange.

By homologous recombination, one double- or single-stranded break in the rearranged V might be repaired by crossing over of strands, using as the template an unrearranged V segment on the same chromosome, the sister chromatid (in G2 phase of the cell cycle), or even on the homologous chromosome, in which case a downstream V segment might be used. Such a repair would not generate a deletion circle (as shown in Fig 2b) and this could be the reason why Wilson *et al.* failed to detect circular recombination products, though it is not the only possibility[36].

There are however two problems with this: 1. either the length of the repair must always be so long that only one cross-over point is within the V_H or there must be unrecognised cases in which both ends of the piece of new V-segment are within the old V segment, as in other species; 2. how to explain the fact that cryptic RSS are found in the cross-over region in almost all cases if they are not involved. We have therefore re-examined the published examples of Type 2 V_H replacement for AID recognition sites in the V-V cross-over regions that could explain replacement by AID-initiated homologous recombination, and discovered answers to both problems.

It has recently been shown[58] that AID prefers dA or dG at the –1 position to the targeted cytosine, and dA or dT at the –2 position. This gives WRC - reverse complement GYW - which corresponds to part of the previously identified RGYW/RGY hypermutation hotspot motif (reverse complement WRCY/RCY). Of the different V_H-V_H junctional regions shown by Wilson *et al.*[36] 5/9 have the complete WRCY·RGYW motif in the hybrid, or in the region of the germline sequence where hybridization occurred, and all the others contain WRC·GYW or RCY·RGW. Likewise, the sequences of Itoh *et al.*[37] and Lenze *et al.*[38] have AID target motifs in the exchange regions. Furthermore, those of Itoh *et al.*[37] include four examples suggestive of gene conversion, *i.e.* an internal section of the V_H resembles a different germline segment while both ends resemble the parent rearrangement. There is insufficient room for a detailed analysis here, and we shall discuss it more fully elsewhere, but give a brief description here of one of these examples. Two rearranged genes with the same HCDR3 were presented (their Figure 5B), one (ST1R G1) resembling a V_H1-02/V_H3-09 hybrid and the other (ST1R G2) a V_H1-69/V_H3-09 hybrid, which we shall refer to as G1 and G2. On closer inspection, both sequences resemble V_H1-69 at their 5' ends and the section of G1 that exclusively matches V_H1-02 is confined to nucleotides (nt) 148-222. G1 therefore seems to have resulted from a substitution of a section of V_H1-02 into the G2 rearrangement followed by further hypermutation. The matching parts of both sequences contain several mutations not present in the other, which seems to exclude the possibility of PCR artefact. This and the other three examples are therefore consistent with gene conversion.

RGYW motifs are common in V_H gene segments and the hybrid sequences of Kolar & Capra[49] have the motif in their exchange regions though the hybrids are from immature cells that would not be expressing AID. However, PCR artefact cannot be ruled out for these sequences, and could have occurred from template DNA with double strand breaks at RGYW motifs.

CONCLUDING REMARKS

There are many examples of recombination of V_H gene segments into previously existing VDJ sequences by a process that we have named Type 1 V_H replacement. Before its recent decisive investigation[35], this was already clearly a RAG-mediated process involving the RSS of the incoming V_H and a cryptic RSS within the 3' end of

the originally recombined V_H segment gene. It occurs in immature B cells in which the *RAG* genes are normally active. Evidence has also been presented[36-38,49] for V_H replacement by another mechanism, that we have called Type 2, of which most examples are from actively mutating B cells. It resembles homologous recombination between two V_H segments, and does not bear the hallmarks of NHEJ. It has not gained acceptance so far because similar hybrid sequences can be produced as PCR artefacts and it was believed to be mediated by RSS and RAG proteins, though *RAG* genes usually are suppressed in hypermutating cells. Those who would dismiss the evidence on the grounds that a single specific search[45] has failed to find this type of V_H replacement should remember that Type 1 V_H replacement was also previously dismissed on the basis of failure to find it in a specific search[17]. Detailed examination of the sequences shows that (a) some of them are extremely unlikely to have been produced by PCR artefact (b) AID target motifs can be found in the regions of exchange between the V_H segments (suggesting that the finding of cryptic RSS may have been a red herring), and (c) in some cases the region replaced does not reach either end of the original V_H. Therefore, though hybrids undoubtedly occur as an artefact during PCR, and may account for some of the published examples, we believe that Type 2 V_H replacement probably does occur *in vivo* and results from AID-mediated homologous recombination. Whether the exchange is reciprocal homologous recombination or gene conversion can only be determined experimentally but, in view of the results from other species, the most reasonable model is gene conversion with replacement often including the 5' end of the V_H.

ACKNOWLEDGEMENTS

This work was funded by the Scottish Executive Chief Scientist Office, Grant No. CZB/4/14, the Arthritis Research Council, Grant No. SO660, and Cancer Research UK Grant No. 4327.

REFERENCES

1 Aït-Azzouzene, Skog P, Retter M *et al.* Tolerance-induced receptor selection: scope, sensitivity, locus specificity, and relationship to lymphocyte-positive selection. Immunol Rev 2004; **197**:219-30.
2 Jankovic M, Casellas R, Yannoutsos N, Wardemann H, Nussenzweig MC RAGs and regulation of autoantibodies. Annu Rev Immunol 2004; **22**:485-501.
3 Monestier M, Zouali M Receptor revision and systemic lupus. Scand J Immunol 2002; **55**:425-31.
4 Nemazee D, Hogquist KA Antigen receptor selection by editing or downregulation of V(D)J recombination. Curr Opin Immunol 2003; **15**:182-9.
5 Grawunder U, Harfst E How to make ends meet in V(D)J recombination. Curr Opin Immunol 2001; **13**:186-94.
6 Gellert M V(D)J recombination: RAG proteins, repair factors, and regulation. Annu Rev Biochem 2002; **71**:101-32.
7 Lieber MR, Ma Y, Pannicke U, Schwarz K Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol 2003; **4**:712-20.
8 Schlissel MS Regulating antigen-receptor gene assembly. Nat Rev Immunol 2003; **3**:890-9.

9 Schatz DG V(D)J recombination. Immunol Rev 2004; **200**:5-11, introducing all other articles in this issue on aspects of the same topic.
10 Hesse JE, Lieber MR, Mizuuchi K, Gellert M V(D)J recombination: a functional definition of the joining signals. Genes Dev 1989; **3**:1053-61.
11 Roth DB Restraining the V(D)J recombinase. Nat Rev Immunol 2003; **3**:656-66.
12 Kouskoff V, Nemazee D Role of receptor editing and revision in shaping the B and T lymphocyte repertoire. Life Sci 2001; **69**:1105-13.
13 Nagafuchi H, Yoshikawa H, Takeba Y, Nara K, Miura K, Kurokawa MS, Suzuki N Recombination activating genes (RAG) induce secondary Ig gene rearrangement in and subsequent apoptosis of human peripheral blood circulating B lymphocytes. Clin Exp Immunol 2004; **136**:76-84.
14 Hassanin A, Golub R, Lewis SM, Wu GE Evolution of the recombination signal sequences in the Ig heavy-chain variable region locus of mammals. Proc Natl Acad Sci U S A 2000; **97**:11415-20.
15 Covey LR, Ferrier P, Alt FW VH to VHDJH rearrangement is mediated by the internal VH heptamer. Int Immunol 1990; **2**:579-83.
16 Chen C, Nagy Z, Prak EL, Weigert M Immunoglobulin heavy chain gene replacement: a mechanism of receptor editing. Immunity 1995; **3**:747-55.
17 Nadel B, Tang A, Feeney AJ V(H) replacement is unlikely to contribute significantly to receptor editing due to an ineffectual embedded recombination signal sequence. Mol Immunol 1998; **35**:227-32.
18 Reth M, Gehrmann P, Petrac E, Wiese P A novel VH to VHDJH joining mechanism in heavy-chain-negative (null) pre-B cells results in heavy-chain production. Nature 1986; **322**:840-2.
19 Kleinfield RW, Weigert MG Analysis of VH gene replacement events in a B cell lymphoma. J Immunol 1989; **142**:4475-82.
20 Wasserman R, Yamada M, Ito Y, Finger LR, Reichard BA, Shane S, Lange B, Rovera G VH gene rearrangement events can modify the immunoglobulin heavy chain during progression of B-lineage acute lymphoblastic leukemia. Blood 1992; **79**:223-8.
21 Shirasawa T, Miyazoe I, Hagiwara S, Kimoto H, Shigemoto K, Taniguchi M, Takemori T Heavy chain variable (VH) region diversity generated by VH gene replacement in the progeny of a single precursor cell transformed with a temperature-sensitive mutant of Abelson murine leukemia virus. J Exp Med 1992; **176**:1209-14.
22 Wasserman R, Felix CA, McKenzie SE, Shane S, Lange B, Finger LR Identification of an altered immunoglobulin heavy-chain gene rearrangement in the central nervous system in B-precursor acute lymphoblastic leukemia. Leukemia 1993; **7**:1294-9.
23 Steenbergen EJ, Verhagen OJ, van Leeuwen EF, dem Borne AE, van der Schoot CE Distinct ongoing Ig heavy chain rearrangement processes in childhood B-precursor acute lymphoblastic leukemia. Blood 1993; **82**:581-9.
24 Steenbergen EJ, Verhagen OJ, Nibbering CP, van den BH, van Leeuwen EF, Behrendt H, dem Borne AE, van der Schoot CE Clonal evolution of immunoglobulin heavy chain rearrangements in childhood B-precursor acute lymphoblastic leukemia after engraftment in SCID mice. Leukemia 1996; **10**:1471-8.

25 Steenbergen EJ, Verhagen OJ, van den BH, van Leeuwen EF, Behrendt H, Slater RR, dem Borne AE, van der Schoot CE 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 1997; **11**:1258-65.
26 Choi Y, Greenberg SJ, Du TL, Ward PM, Overturf PM, Brecher ML, Ballow M Clonal evolution in B-lineage acute lymphoblastic leukemia by contemporaneous VH-VH gene replacements and VH-DJH gene rearrangements. Blood 1996; **87**:2506-12.
27 Rosenquist R, Lindh J, Roos G, Holmberg D Immunoglobulin VH gene replacements in a T-cell lymphoblastic lymphoma. Mol Immunol 1997; **34**:305-13.
28 Rosenquist R, Thunberg U, Li AH *et al.* Clonal evolution as judged by immunoglobulin heavy chain gene rearrangements in relapsing precursor-B acute lymphoblastic leukemia. Eur J Haematol 1999; **63**:171-9.
29 Kobrin CB, Bendandi M, Kwak LW Novel secondary Ig V_H gene rearrangement and in-frame Ig heavy chain complementarity-determining region III insertion/deletion variants in de novo follicular lymphoma. J Immunol 2001; **166**:2235-43.
30 Cascalho M, Wong J, Wabl M VH gene replacement in hyperselected B cells of the quasimonoclonal mouse. J Immunol 1997; **159**:5795-801.
31 Golub R, Martin D, Bertrand FE, Cascalho M, Wabl M, Wu GE VH gene replacement in thymocytes. J Immunol 2001; **166**:855-60.
32 Deane M, Norton JD Immunoglobulin heavy chain gene rearrangement involving V-V region recombination. Nucleic Acids Res 1990; **18**:1652.
33 Brokaw JL, Wetzel SM, Pollok BA Conserved patterns of somatic mutation and secondary VH gene rearrangement create aberrant Ig-encoding genes in Epstein-Barr virus- transformed and normal human B lymphocytes. Int Immunol 1992; **4**:197-206.
34 Taki S, Schwenk F, Rajewsky K Rearrangement of upstream DH and VH genes to a rearranged immunoglobulin variable region gene inserted into the DQ52-JH region of the immunoglobulin heavy chain locus. Eur J Immunol 1995; **25**:1888-96.
35 Zhang Z, Zemlin M, Wang YH *et al.* Contribution of V_H gene replacement to the primary B cell repertoire. Immunity 2003; **19**:21-31.
36 Wilson PC, Wilson K, Liu YJ, Banchereau J, Pascual V, Capra JD Receptor revision of immunoglobulin heavy chain variable region genes in normal human B lymphocytes. J Exp Med 2000; **191**:1881-94.
37 Itoh K, Meffre E, Albesiano E *et al.* Immunoglobulin heavy chain variable region gene replacement As a mechanism for receptor revision in rheumatoid arthritis synovial tissue B lymphocytes. J Exp Med 2000; **192**:1151-64.
38 Lenze D, Greiner A, Knorr C, Anagnostopoulos I, Stein H, Hummel M Receptor revision of immunoglobulin heavy chain genes in human MALT lymphomas. Mol Pathol 2003; **56**:249-55.
39 Lewis SM, Hesse JE, Mizuuchi K, Gellert M Novel strand exchanges in V(D)J recombination. Cell 1988; **55**:1099-107.

40 Bross L, Muramatsu M, Kinoshita K, Honjo T, Jacobs H DNA double-strand breaks: prior to but not sufficient in targeting hypermutation. J Exp Med 2002; **195**:1187-92.
41 Faili A, Aoufouchi S, Gueranger Q, Zober C, Leon A, Bertocci B, Weill JC, Reynaud CA AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. Nat Immunol 2002; **3**:815-21.
42 Papavasiliou FN, Schatz DG The activation-induced deaminase functions in a postcleavage step of the somatic hypermutation process. J Exp Med 2002; **195**:1193-8.
43 Zan H, Wu X, Komori A, Holloman WK, Casali P AID-dependent generation of resected double-strand DNA breaks and recruitment of Rad52/Rad51 in somatic hypermutation. Immunity 2003; **18**:727-38.
44 Sims GP, Shiono H, Willcox N, Stott DI Somatic hypermutation and selection of B cells in thymic germinal centers responding to acetylcholine receptor in myasthenia gravis. J Immunol 2001; **167**:1935-44.
45 Goossens T, Brauninger A, Klein U, Kuppers R, Rajewsky K Receptor revision plays no major role in shaping the receptor repertoire of human memory B cells after the onset of somatic hypermutation. Eur J Immunol 2001; **31**:3638-48.
46 Cuomo CA, Mundy CL, Oettinger MA DNA sequence and structure requirements for cleavage of V(D)J recombination signal sequences. Mol Cell Biol 1996; **16**:5683-90.
47 Ramsden DA, McBlane JF, van Gent DC, Gellert M Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. EMBO J 1996; **15**:3197-206.
48 Lee GS, Neiditch MB, Salus SS, Roth DB RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. Cell 2004; **117**:171-84.
49 Kolar GR, Capra JD Immunoglobulin heavy-chain receptor editing is observed in the NOD/SCID model of human B-cell development. Scand J Immunol 2004; **60**:108-11.
50 Kleinfield R, Hardy RR, Tarlinton D, Dangl J, Herzenberg LA, Weigert M Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a Ly 1^+ B-cell lymphoma. Nature 1986; **322**:843-6.
51 Chukwuocha RU, Nadel B, Feeney AJ Analysis of homology-directed recombination in VDJ junctions from cytoplasmic Ig- pre-B cells of newborn mice. J Immunol 1995; **154**:1246-55.
52 Lewis SM The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. Adv Immunol 1994; **56**:27-150.
53 Shiokawa S, Mortari F, Lima JO *et al.* IgM heavy chain complementarity-determining region 3 diversity is constrained by genetic and somatic mechanisms until two months after birth. J Immunol 1999; **162**:6060-70.
54 Bogue MA, Wang C, Zhu C, Roth DB V(D)J recombination in Ku86-deficient mice: distinct effects on coding, signal, and hybrid joint formation. Immunity 1997; **7**:37-47.
55 Han JO, Steen SB, Roth DB Ku86 is not required for protection of signal ends or for formation of nonstandard V(D)J recombination products. Mol Cell Biol 1997; **17**:2226-34.

56 Neuberger MS, Harris RS, Di Noia J, Petersen-Mahrt SK Immunity through DNA deamination. Trends Biochem Sci 2003; **28**:305-12.
57 Flajnik MF Comparative analyses of immunoglobulin genes: surprises and portents. Nat Rev Immunol 2002; **2**:688-98.
58 Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J Mol Biol 2004; **337**:585-96.

Figure legends

Figure 1. Type 1 V_H replacement. (a) Usual subtype: recombination occurs after cleavage at the conventional RSS of a new (upstream) V_H and at the 3' cryptic RSS in the V_H portion of the recombined V_H-D-J_H. (b) Cleavage occurs at the conventional RSS of the incoming V_H but at an upstream cryptic RSS of the V_H of the VDJ. (c) Cleavage occurs at cryptic RSS in opposite directions and different positions in the two V_H genes. In this and all other figures, the conventional RSS are indicated by filled triangles and cryptic RSS by open triangles. Small arrows indicate cleavage positions and deletion products are omitted. '∼' indicates the typical random small deletions from either or both coding ends and added P and N bases.

Figure 2. Comparison of Types 1 and 2 V_H replacement. (a) Detail of the usual form of Type 1 involving the conventional RSS of the incoming V_H and the 3' cryptic RSS in the V_H portion of the pre-existing V_H-D-J_H. Diagram conventions are as in Fig. 1. Just as in normal V(D)J recombination, the new junction in the VDJ is a coding-to-coding junction and the deletion product has a signal-to-signal junction. (b) A Type 2 V_H replacement scheme to explain the published hybrid sequences by the investigators' hypothesis of RAG-mediated double-strand cleavage at cryptic RSS and rejoining without end modification. Reaction products shown are a V_HDJ_H with the hybrid V_H of the kind published and the deduced circular deletion product, both with cryptic RSS signal-to-coding junctions. Alternative mechanisms are homologous recombination following RAG-mediated single-strand nicking at a cryptic RSS and our preferred hypothesis of AID-initiated homologous recombination unrelated to RAG or cryptic RSS. Homologous repair in these cases need not involve any deletion.

Figure 3. Construction of a sequence resembling Type 2 V_H replacement by PCR artefact. The N and P bases between V_H and D and between D and J_H are omitted for simplicity. (a) Copies of cDNA of two rearranged immunoglobulin heavy-chain genes occurring at an early stage in the same reaction mixture. 'p' is a PCR product made with a specific family V_H leader primer and a specific constant-region primer. 'q' is another rearrangement with a different D and J_H. The V_H differs from that of p at various points indicated by the pink spots in p and blue spots in q. q may be another PCR product with the same leader sequence, or it may be a cDNA template with a different leader that would not have been amplified with the primers in use. (b) During one of the PCR cycles, 'n', a copy of p, is incomplete when melting occurs. (c) In the next cycle n anneals with q at a point at which q has the same sequence as p. A polymerase molecule attaches and copying continues, giving a product (the extended 'n') that has the leader and 5' end of the V_H the same as in p but the 3' end of the V_H and the VDJ junction the same as q. This new molecule will now be amplified in all following cycles.