Chapter 14
The Generation of Junctional Diversity by V(DJ) Recombination
RACHEL M. GERSTEIN

V(DJ) recombination is the process that assembles immunoglobulin (Ig) and T cell receptor (TCR) genes in developing lymphocytes. The coding exons for antigen receptor variable regions does not exist as such prior to rearrangement but is in fact the product of one (in the case of Ig α and κ light chains and TCR α and β) or two (in the case of Ig heavy chain, TCR β and δ) recombination reactions. These recombinant junctions fall within CDR3, the most hyper-variable region of antigen receptors [1]. Since the hyper-variable region regions are critical residues for the contact with antigen [2, 3], the diversity generated at CDR3 is a major determinant of the antigen receptor repertoire [1, 4]. This review will focus on the mechanisms by which V(DJ) recombination dictates CDR3 diversity throughout lymphocyte development. Of particular interest will be to discuss how widely different methodologies, despite each having intrinsic virtues and limitations, have yielded complementary information.

The complete variable region is assembled from smaller genomic elements known as variable (V), diversity (D), and joining (J) exons. Each coding element is flanked by a signal sequence that directs recombination. The recombination signal sequence (RSS) consists of an A/T rich nonamer and a palindromic heptamer, which are separated by a spacer region of 12 or 23 base pairs. Recombination occurs between pairs of elements; one with a 12-base pair signal and one with a 23-base signal (the 12/23 rule). The basic features of the reaction include 1) binding of proteins to the RSS, 2) synapses, 3) cutting, 4) base loss and addition, and 5) joining and repair. Cuts at the heptamer ends of the signals generate two signal ends and two coding ends. When two signal ends join, the product is termed a signal joint. The junction of two coding ends is called a coding joint. This review will focus on the formation of coding joints as particularly relevant to the generation of diverse CDR3 regions. The mechanism of signal joint formation and other mechanistic features of the V(DJ) recombination reaction are discussed in Lewis [5], Lieber [6], and Lieber et al [7].

Nucleotide loss and nucleotide addition generates diversity at coding ends during V(DJ) recombination. Nucleotide loss occurs frequently; as assessed by V(DJ) recombination substrates, 75% of coding ends have loss, and consequently over 90% of coding joints have loss from one of the two ends [8–10]. The mechanism of nucleotide loss is not yet known. An exonuclease activity has been postulated [11] and several mechanisms have been discussed, including either an exonuclease or endonuclease that acts on coding end termini, or base loss following alignment or ligation of coding ends [5]. Other activities that may be involved in nucleotide loss during V(DJ) recombination include a proposed single strand endonuclease activity involved in hairpin resolution [6] and the activity of FEN1 [7], an endonuclease that cleaves DNA flap structure intermediates (a DNA flap is a bifurcated structure composed of double-stranded DNA and a displaced single-strand). In addition to endonuclease activity, FEN-1 has a 5'-3' exonuclease activity which is specific for double-stranded DNA [12, 13]. It is not known at this point in time whether FEN1 is involved in V(DJ) recombination.

Nucleotides are added to coding junctions by at least two distinct mechanisms. Nucleotides are frequently added to coding joints in a template-independent manner. As discussed below, deoxynucleotidyl terminal transferase (TdT) is responsible for the majority of these additions, known as N regions. Another class of additions, known as P nucleotides, show template dependence. Usually one to three nucleotides in length, P regions generate an inverted repeat of the adjacent coding end and are observed adjacent to coding ends that have suffered no nucleotide loss [14–16]. The total potential sequence diversity for antigen receptors generated in adult mice has been estimated to be $2 \times 10^{18}$ different immunoglobulins, $10^{17}$ for TCR aβ, and $5 \times 10^{12}$ for TCR γδ [6]. Junctional diversity may be particularly important for the TCR repertoire, as less combinatorial diversity is available due to the smaller number of V genes and all TCR junctions have been observed to have N regions.

Traditional molecular cloning and sequencing

The initial understanding of the generation of junctional diversity during V(DJ) recombination came from genomic cloning and sequencing. This work has been reviewed by Tonegawa [17], Alt et al [18], and Davis and Bjorkman [19]. These studies defined the genomic organization of the Ig and TCR loci and established the general features of the recombination mechanism. These included the mechanisms that ensure junctional diversity and provided some information as to the range of different receptors. Large-scale surveys of the diversity of antigen receptor rearrangements were not feasible.

Genomic cloning and sequencing also led to the identification of N regions [20] and the observation that N region insertions were associated with Tdt activity [11] and that the incidence of N regions correlated with amounts of Tdt activity [21].

The molecular technology used to sequence the largest number of antibody gene sequences, prior to the widespread use of the polymerase chain reaction (PCR), was the direct sequencing of expressed immunoglobulin transcripts from hybridsomas with reverse transcriptase coupled with the Sanger method of chain-terminating deoxynucleotides [22, 23]. Sequencing of mRNA proved to be a useful way to determine the expressed Ig sequence without genomic or cDNA library construction. Many sequences of the expressed Ig genes from hybridsomas were obtained, which led to a wealth of information as to what antigens specificities were correlated with particular combinations of Ig sequences. This approach does not, however, necessarily generate an accurate picture of the diversity of junctions created by the recombination machinery alone. A good example of this potential inaccuracy is in
the cross-reactive idiotype (CRL4) elicited by phenylarsionate in AJ mice: a particular P nucleotide insertion, responsible for the invariant TCX serine codon at amino acid 99 is present at the V-D junction [24]. This could theoretically either be a frequent addition during V(DJ) recombination or an infrequent event that is amplified by antigen selection. It should therefore be noted that the picture of Ig gene variability that has been generated [25] is heavily influenced by the imprint of antigenic selection, as the majority of the IgH sequences used in this calculation are derived from antigen-specific hybridomas. In addition, the potential for an inaccurate representation of the repertoire is also inherent in the process of making hybridomas, where it is impossible to prove that all B cells have an equal chance of being stably fused with the myeloma partner. Evaluation of the repertoire from unimmunized mice using single-cell PCR [26–31], will allow the refinement of the Ig and TCR variability plots.

Integrated substrates

The first generation of V(DJ) recombination substrates were designed to allow transfection or retroviral infection of transformed cell lines, and the subsequent selection of stably integrated recombination substrates, based on constitutive expression of a gene for drug resistance [32–36]. Cell-harboring recombined V(DJ) substrates were characterized by the acquisition of a new phenotype conferred by rearrangement of the substrate. Usually a gene conferring resistance to a second drug that allows eukaryotic selection (distinct from the initial drug resistance gene used to select for stable integration of the substrate) was placed "upside down" in between the recombination signal sequences. Upon inversional V(DJ) recombination, the drug resistance gene would be placed in the correct transcriptional orientation. Therefore, selection allowed the isolation of cells bearing recombined substrates. A substrate that conferred LacZ expression upon recombination was also developed, which allowed assay and selection of V(DJ) recombients by FACS [37]. A more detailed comparison of different integrated substrates can be found in Lewis [5].

The integrated substrates yielded important information: since the substrates have defined, non-variable DNA sequence content, the complexity of assigning base loss and addition events was greatly reduced. Another variable was controlled by the ability to easily isolate multiple clones from particular cell lines. A powerful aspect of this approach is the ability to select for rare recombients with drug selection. This feature was useful for describing the scid defect [38–41] and essential for the cloning of the recombination activating genes RAG1 and RAG2 [42–44].

The problems associated with the use of integrated substrates include the following. 1) Drug-resistant clones arise after two selection periods (one to indicate integration and one for recombination), as the population under consideration may change and events may be lost; this also places a limitation on the timely testing of many substrates. 2) Recombination events may occur prior to integration that would give the same drug-resistant phenotype. 3) Position effects conferred by chromosomal context that may affect the reproducibility of results and require multiple clones to be analyzed. 4) Loss of expression of the selectable marker or screening marker over time [37]. 5) The problem of quantitation: how to look at a large number of independent events with a defined denominator; that is, the number of cells in the population available to be sampled.

Despite these limitations, important mechanistic inferences were first made using these types of substrates. Also, the use of integrated V(DJ) substrates extended the correlation between the level of Tdt activity and the incidence of N region insertion [36]. This approach added additional information: since an integrated substrate contains defined nucleotide sequence, the use of substrates confirmed that bases are definitely added de novo. In addition, the cloning of the Tdt gene allowed manipulation of Tdt expression using expression vectors [45]. Coupled with the use of integrated substrates, this work strengthened the evidence that more Tdt activity resulted in more frequent N insertions.

As this correlation between Tdt activity and N region addition was not absolute [8, 45], workers asked whether all non-templated insertions are the result of Tdt activity. Other mechanisms have been suggested after examining sequences derived from recovered recombination substrates [8, 46, 47]. Evidence for non-Tdt-mediated insertion was provided by the finding of non-templated nucleotides in V(DJ) recombients isolated from fibroblasts, thus occurring in the complete absence of Tdt [10, 48]. The most definitive evidence for additional mechanisms for the generation of nucleotide insertions in coding junctions derive from the Tdt knockout mice [49, 50] where non-templated nucleotide insertions were found, despite the inactivation of the Tdt gene. These insertions occur at low frequency (3%), suggesting that this is not a major source of junctional diversity.

Mice that have been engineered with transgenic recombination substrates have also been used to study the specificity and diversity of V(DJ) recombination [51–65]. This topic has been reviewed extensively elsewhere [5, 66, 67].

Polymerase chain reaction analysis of the antigen receptor repertoire

The widespread use of PCR has made possible the quick isolation, cloning, and sequencing of large numbers of recombined antigen receptor genes. PCR cloning has allowed large surveys of gene usage and the developmental regulation of junctional diversity. This approach revealed that the incidence of N region insertions varies greatly during murine ontogeny [16, 68–75]. For example, N regions in Ig heavy-chain locus (IGH) rearrangements were very rare in neonatal spleen and liver, and were not detected in fetal liver from mice on day 17 of development (d17) [76]. In contrast, 83% of Ig junctions from adult spleen had N regions. Another study found 12% of IgH junctions isolated by PCR from CDNA from fetal liver day 18 had N regions [69]. Most of these insertions consisted of one or two nucleotides and are likely to be either template-independent insertions not imparted by Tdt or P nucleotides.

These studies demonstrated that by sampling all the cells in a lymphoid organ, one could determine the frequency and range of junctional diversity within the population surveyed. An important implication emerged; the immune repertoire in fetal and neonatal B cells is strikingly less diverse than that of adults. These data regard the region incidence correlate well with the known pattern of Tdt expression [77–80]. Thus, the developmental regulation of Tdt has a large impact on the diversity generated by V(DJ) recombination. It was also shown that 63% of IgH junctions with no deletions had P nucleotides, suggesting that PCR would also be useful for tracking the in vivo incidence of this templated addition process [76].

The junctional diversity of the B1a and B1b cells (the B cell
untinages that arise earliest in marine ontogeny (the origin of B cell lineages is reviewed in Kantor and Herzenberg [81]), as well as conventional B cells, has been assessed [31]. Conventional B cells are IgM⁺, IgD⁺, and CD5-negative. Both Bia and BiB cells are IgM⁺, IgD⁺. Bia cells are CD5-positive and BiB cells are CD5-negative. By sorting single B cells and determining the expressed Ig sequence of these cells (see the preceding chapter of this section), Kantor and co-workers have recently determined that Bia cells isolated from the peritoneum of adult mice have more V(DJ)I junctions with no or few N region nucleotides than either BiB or conventional B cells; 5% of conventional B cells lack N regions, whereas this is true of 25% of Bia transcripts. The Bia cells found in adult mice that lack N insertions are likely to have arisen in fetal/neonatal ontogeny, undergone V(DJI) recombination in the absence of TdT, and persisted into adulthood by self-renewal. Since most of the (V(DJ)I transcripts from Bia cells that were sequenced contain N regions, it is likely that Bia cells undergo V(DJ)I rearrangement during both fetal development and also throughout the first six weeks of life.

It has been suggested that nucleotide base loss at IgH during V(DJ)I recombination is more evident in post-natal or adult samples than fetal [16, 69, 82, 83]. Other studies do not confirm this observation [60, 84-86]. The true extent of base loss, particularly during fetal developments, is probably confounded by the persistence and potential amplification by selection in vivo of coding joints where homology at the coding ends is used (discussed below). In this case, there is often no base loss from one of the coding ends. Junctional diversity of T cells has also been surveyed by PCR [75, 85, 87-93]. Adult levels of TdT are not reached in the thymus until one week of age in mice [77, 78, 80]. Accordingly, whether or not T cell γ receptor N regions depend on when they are generated; fetal γ receptors have very few N regions [16, 75, 87, 88]. Feeney surveyed α T cell receptor cells in the thymus [85]. PCR of T cell receptor cDNA in mice revealed that 16% of junctions had N regions at d18, 33% had N regions when less than 24 hours old, 76% had N regions when four days old, and 88% had N regions when five weeks old. It is of interest to note that 22% of neonatal junctions had N regions when DNA was amplified in this study.

In fact, when junctions were amplified from DNA, Feeney found N regions were less frequent in out-of-frame junctions than in in-frame junctions. It is likely that selection enriches somewhat for junctions with N regions within the expressed receptor population (33% N region-containing junctions were obtained from cDNA versus 22% from DNA). Lafaille and colleagues also found that in-frame coding joints were more homogeneous than out-of-frame [16]. This type of observation is also true of analyses of IgH junctions, as Feeney analyzed junctions amplified from cDNA from LPS-stimulated spleen cells and also from genomic DNA. This analysis showed that the expressed junctions had a higher N region incidence [76]. As valuable as these types of surveys are, these results, and others to be discussed below, reveal that the measurement of incidence of particular junctions or types of junctions by PCR is subject to biases due to cellular selection.

Several potential problems with PCR should always be kept in mind when evaluating the efficacy or validity of results obtained with this method. Contamination is a concern that can be addressed with appropriate safeguards, controls, and demonstration of the reproducibility of data. Taq polymerase is known to be error prone, so sequence is often obtained for multiple clones from the same PCR reaction and compared with clones from the same DNA preparation amplified separately. If antigen receptors from a diverse population of many cells is analyzed, there is no way to be certain that errors introduced by Taq are not present in some of the sequenced products. Another PCR artifact especially problematic for the analysis of populations of lymphocytes is the formation of chimeric molecules during PCR. In a study of this phenomenon, sequence analysis of immunoglobulin Λ regions revealed that 30% of the PCR products amplified from a population of germinal center B cells were chimeric molecules created during PCR, consisting of segments of the V4-1 and V2-2 genes. Furthermore, an amplification- and cloning-associated artifact of exchanged sequences between mutualional variants of V4 genes is another artifact that was observed [94].

By isolating single cells and performing PCR amplification of the DNA [26, 27] or cDNA [28-31] (see also Kantor et al in this section) from individual cells, many of the problems mentioned above are avoided. In addition, the contribution of unequal amounts of mRNA per cell to the pool of cDNAs is circumvented. Finally, by sequencing both recombinant IgH alleles from a cell, a new category of information is now accessible by PCR analysis [27].

Homology use in the resolution of coding joints during V(DJ)I recombination

The PCR approach

V(DJ)I recombination is biochemically noteworthy in that identical reactants yield different products at the junction of joined segments. Despite the obvious advantage conferred by the ability to generate different antigen receptors from the same genetic elements, there is a clear example in which the mechanism of V(DJ)I recombination is geared to limit diversity. The diversity of coding joints can be severely restricted when the coding ends participating in the reaction have short regions of homology. This section will review the experimental evidence for this restriction, and examine the different methodologies utilized.

An abundance of certain V(DJ)I junctions obtained by PCR amplification and sequencing has been noted [16, 68, 70-72, 74, 75, 95]. It was proposed that these junctions are favored because they occur where short regions of homology in participating coding ends might align preferentially [68, 84, 85, 96, 97]. While the potential to use homology at coding ends during V(DJ)I recombination had been suggested previously to this work [11, 20, 24, 98], its significance was made more apparent by the frequency of the clones isolated from PCR analysis. By examining T and B cells and their progenitors, these investigators made the important connection between a high incidence of the potential use of homology in V(DJ)I recombination, and the low incidence of N region addition during fetal B cell development and in B1 cells.

Like any V(DJ)I joining rearrangement, D4 joining can create junctions that, upon translation, could yield any of three potential reading frames. Since D4 can be translated in three potential reading frames (RF), it is possible to find expressed VDJ4 genes using any of the three RF of DH [68, 98, 99]. While bias of RF in hybrids towards RF1 (using the nomenclature of Ichihara et al [98]) had been noted previously [99], the observation of short sequence homologies in junctions using RF1 suggested a molecular basis, founded in the recombination mechanism, for this bias [16, 68, 70-72, 74, 75, 95, 99].
As seminal as these studies were, since the recombinant junctions were isolated from cells subject to multiple influences in vivo there is no way to exclude the contribution of cellular selection and clonal expansion to the frequency of finding a particular VDJH or DJH clone. Since these forces can influence the representation of particular coding joints in the repertoire, the true extent to which V(DJ) recombination alone can dictate junctional restriction due to usage of homology at coding ends could not be accurately assessed using PCR cloning of primary cell populations.

Cellular selection and clonal expansion operate on cells containing VDJH or DJH junctions such that representation of particular products becomes biased through a variety of mechanisms. First, several studies have sequenced the DJH junctions within a fully rearranged VDJH exon. Two of these studies clearly present evidence that productive VDJH junctions are over represented in their samples [71, 73]. Decker and colleagues provide evidence for clonal expansion of pre-B cells with productive VDJH exons [73]. Other studies have also demonstrated clonal expansion after VDJH joining [100]. In fact, the clonal expansion of B cell progenitors can depend on the nature of the VH region of the heavy chain [101-103].

Thus, VDJH exon clones isolated by PCR are not a random sample of independent recombination events, but rather the probability of isolating a particular clone will be determined by some combination of the frequency of the recombination event, the size of the proliferative burst of pre-B cells following VDJH assembly, as well as selective events. Selection is mediated by both endogenous and environmental antigens in the case of B cells [104, 105] and through the VDJH pseudo-light-chain complex in pre-B cells [70, 105-112]. The strongest evidence for selection prior to the expression of conventional light chains comes from mice in which Cμ has been disrupted by homologous recombination. Unlike the situation in normal mice, in which VDJH in RF2 are rare, RF2 is found in similar numbers as the other reading frames of DJH in the targeted alleles of splenic B cells of heterozygous Cμ+h− mice. This indicates that the targeted allele cannot mediate the signal that leads to the elimination of DJH in RF2 [113, 114]. Mice deficient for the A5 gene [110] also fail to eliminate RF2 junctions [115], implicating this gene product in the signaling process required for counter selection against RF2.

Expression of a particular DJH junction can also determine the fate of a developing B cell progenitor. Some studies of junctional diversity have sequenced DJH junctions from DJH clones [70, 72]. One paper [70] clearly indicates that DJH products can mediate selection. While in normal mice, DJH in RF2 are underrepresented, Cμ+h− mice have no bias with respect to DJH reading frame.

What is the mechanism of this selection against RF2? There are cis-acting elements that can serve as promoters positioned upstream of the DJH coding regions. Also, an ATG start codon is present at -53 for the DSF elements, -108 for DFL, and -120 for DQ52, which allows the initiation of translation in frame with JH when the DJH rearrangement is in RF2. The protein product of this type of rearrangement is a truncated H chain consisting of the DJH coding region and Cμ [116]. Rajewsky and colleagues have suggested that expression of DJH protein blocks further differentiation of progenitors [70]. Data obtained from PCR analysis of recombinant Ig genes from single B cell progenitors [27] supports this hypothesis, as DJH junctions encoding potential

Dα protein were only found in cells with DJH rearrangements on both alleles. This type of junction was not found in cells containing a VDJH junction at the other allele. RF2 is used at approximately the same frequency as RF1 and RF3 in cells with 2 DJH joints. This finding is similar to the observation that in early B cell progenitors isolated from fetal liver, RF2 is not suppressed [95, 115].

The finding that DJH in RF2 are not yet eliminated in cells with 2 DJH joints implies that either selection against RF2 is not a consequence of the immediate induction of cell death in cells that express Dα protein [27] or that Dα protein is necessary but not sufficient for the counter selection process. The distribution of reading frames is biased away from the use of RF2 only in VDJH rearrangements [27, 68, 70, 115]. The authors' interpretation [27, 70] is that V joining to DJH is inhibited in cells that express Dα protein, as earlier work had suggested [117]. When Dα protein associates with A5 and VpreB [109], it may deliver a signal identical to that given by the a/surrogate light-chain complex that mediates allelic exclusion [27] This would result in premature cessation of the B cell developmental program, as further V to DJH joining would be inhibited.

Cellular selection also is likely to be responsible for the Dα RF bias observed in pre-B and B cells in which N region insertion prevents the use of homology to align coding ends. When we consider cellular selection and clonal expansion as forces that bias the repertoire, it is difficult to determine the relative contribution of the bias in recombination outcomes in the context of several redundant influences that affect RF bias for VDJH. Since the majority of putative homology usage by DJH junctions is observed among productive rearrangements, it is important to ask what extent the in vivo data [16, 68, 70-72, 74, 75, 95] reflects independent recombinants, arising solely due to a mechanistic bias imposed by homology use during V(DJ) recombination, versus a survival advantage conferred on a cell that has a productive rearrangement and the clear survival disadvantage imposed upon pre-B cells that have DJH junctions where the D is in reading frame 2. The fact that some non-productive rearrangements could also be favored by the usage of homology does not obviate the concern that the enrichment for productive rearrangements via cellular selection may obscure the true extent of homology use in vivo.

**Determination of the use of homology using extrachromosomal substrates**

In order to measure the extent to which homology at coding ends can constrain junctional diversity in V(DJ) recombination, we used a system that is entirely free from cellular selection, namely extrachromosomal V(DJ) recombination substrates [10]. We found that the diversity of coding joints can be severely restricted when the coding ends participating in the reaction have short regions of homology. The frequency of coding end homology usage in V(DJ) recombination reactions was determined by analyzing recombinants generated on extrachromosomal substrates transfected into murine pre-B cell lines. The recombinant substrates used retain coding joints after undergoing V(DJ) recombination. One substrate contains DFL16.1 and Jα1 from the Ig heavy-chain locus. The 3' end of the DFL16.1 coding sequence and the 5' end of the Jα1 coding sequence share four nucleotides: CTAT. Among independent recombinant coding joints isolated
from a pre-B cell line expressing low levels of Tdt, a particular coding joint junctions with loss of four nucleotides (termed -4 junctions); the expected product if the maximal homology is used to align the coding ends is observed in 30 of 54 (55%) reaction products that lack N regions or P nucleotide insertions (termed NP"). These data represent the first assessment of the extent to which sequence homology can clearly determine reaction outcome.

DNL16.1-J3,1 coding joints were also isolated from a pre-B cell line expressing higher levels of Tdt. The frequency of the maximum homology coding joint is 4 of 32 (12.5%) junctions. This data indicates that the role of Tdt activity affects how often homology is apparent in the outcome at V(DJ) coding joints.

In order to examine further the extent of homology utilization in the presence or complete absence of Tdt, we co-transfected a hamster fibroblast line with substrate, RAG1 and RAG2 expression vectors, and with or without a Tdt expression vector. In the absence of Tdt, an incidence of 33% (16 of 48 NP- joints) maximum homology coding joints was observed. In contrast, only two of 39 coding joints isolated from fibroblasts co-transfected with Tdt were maximum homology coding joints (5%). This directly indicates that when Tdt is present homology at coding ends is not frequently used in coding joint resolution.

Recombinants from a substrate that contains DNL16.1 and J3,3, which have less homology at the coding ends, were also examined. This substrate has the potential for using one or two nucleotides of homology during coding joint formation. In both the low and high Tdt-expressing cell line, few recombinants were -4 coding joints. As examination of 250 coding joints shows there is no intrinsic bias to form -4 junctions; homology is the only basis for this bias in the DNL16.1-J3,3 junctions. In addition, the diverse array of coding joints observed among DNL16.1-J3,3 recombinants shows that one or two nucleotides of homology is not sufficient to substantially constrain joining.

The efficiency of recombination, as determined by the ratio of recombined substrate molecules to total recovered molecules, is very similar among substrates with more (DNL16.1-J3,3) or less (DNL16.1-J3,3) potential homology at the coding ends. This indicates that homology does not affect the efficiency of the reaction, just the diversity of the products. In fact, homology of even one nucleotide in V(DJ) coding ends is not required for efficient recombination (discussed below) [10].

When homology at coding ends is used during V(DJ) recombination in the absence of Tdt, the diversity of coding joints dramatically decreases. Because Tdt-mediated N region insertion is very rare at DJ3,3 joints formed early in murine ontogeny, these results suggest that homology accounts for over-representation of certain DJ3,3 combinations as suggested previously [68, 70, 72]. Our results also demonstrate that for other DJ4,3, V(DJ) recombination is not biased.

The frequent generation of important specificities may be ensured by generating a high proportion of productive DNL16.1-JH1 junctions. The DNL16.1-JH1 maximum homology joint is present in the majority of T15 positive antibodies [118-121]. Since T15 antibodies are optimally protective against infection with Streptococcus pneumoniae [122, 123], restricting DJ4,3 diversity seems to be one strategy to ensure protection against this and other phosphocholine-bearing pathogens. [124-126] Perhaps the diversity-increasing benefit of Tdt outweighs the conservative trend of diversity restriction only after the establishment of the first layer of immune specific recognition [127].

**Mice with altered genomes**

Mice in which the Tdt gene has been inactivated by homologous recombination [49, 50] provide the basis of a third system for assessing the mechanisms of restricting diversity of antigen receptor junctions. By using PCR analysis from thymus or spleen of Tdt-/- animals, Alt and colleagues noted 45% of the junctions had potential overlap of two nucleotides or more of DNA sequence (i.e., homology use during joining) [49]. This calculation is the result of examining 583 sequences from a variety of loci; note that the frequency of homology use is higher for particular pairs of coding ends such as Vγ3Jγ1 of the TCR (75%). Mathis and co-workers found 66% of the Vγ3Jγ1 junctions had overlap of greater than 1 base pair [50]. Of the DJ4x,3 junctions within Vγ783 or Vα558 VDJγ, regions cloned from the spleen, 54% had potential homology at the coding ends. While the junctions isolated from these Tdt-/- mice are likely to be subjected to the same possible selective forces as discussed above, these studies are an important contribution to the "homology-use" literature, since they document the range of potential homology use for a number of loci, in the complete absence and also the presence (in Tdt+ mice) of Tdt.

The authors point out that 25% of the junctions from Tdt-/- lack even 1 base pair of potential homology at the coding ends, indicating the existence of a pathway of V(DJ) recombination in which homology usage in coding joint formation is not used [49]. This suggestion is confirmed by our studies of extrachromosomal substrates with no potential for homology-based alignment of coding ends within the first eight nucleotides proximal to the coding ends in which the efficiency and structure of recombinants was not different than observed for other substrates [10]. Another study with substrates supports this result [128].

In addition, these studies definitively show that Tdt is the activity responsible for the majority of N region additions [49, 50]. Also, the frequency of P nucleotide addition is independent of Tdt expression. In the absence of Tdt, however, approximately 3% of junctions still contain non-templated nucleotide additions. As noted above, these types of inserts have previously been noted in non-immune system cells [47], and also within V(DJ) junctions isolated from fibroblasts, where the reaction occurs in the complete absence of Tdt [10]. Mathis and co-workers [50] consider several possibilities to account for these nucleotides; namely, that nucleotides are incorporated during joining, or by DNA polymerase after joining, as addition of extra nucleotides by polymerase has precedent [129].

Mice with C8 inactivated by homologous recombination are also informative with respect to the use of homology in V(DJ) recombination [130]. Of V61D2J2 junctions, 68% had overlapping sequence, as did 36% of Vγ3Jγ1 and 31% of Vγ6Jγ1. This analysis suggests that homology at coding ends can bias the rearrangement outcome in developing T cells in which cellular selection cannot directly influence the persistence or elimination of the recombinant allele, as the C8 mutation prevents expression of the T cell receptor. A similar conclusion was also reached by the use of a transgenic substrate bearing a mutated TCR γ gene in which the canonical Vγ3Jγ1 (that contains a 2-base pair potential overlap) formed at a high frequency (82%) in the
absence of the possibility of selection of the protein products of the rearrangement process [65].

Conclusion

In summary, a variety of approaches have been used to assess the diversity of antigen receptor gene rearrangements in mice and the mechanisms that expand or constrain this diversity. Genomic cloning laid the foundation for much of what we know about the structure of antigen receptor gene loci and the mechanism of V(D)J recombination. The incidence of particular junctions that arise in vivo has been extensively cataloged by PCR. Extrachromosomal V(D)J recombination substrates have been valuable for the analysis of many aspects of the mechanism of recombination; here I have focused on the ability to isolate independent events in order to accurately measure the extent of constrictive mechanistic biases. And mice with engineered genomes allow definitive attribution of the role of enzymatic activities, such as TdT, to the role that each activity plays in shaping the antigen receptor repertoire. Like a society with a diverse allocation of human talents into different professions, the society of scientists working to reveal the forces that shape the antibody repertoire have benefited greatly from looking at similar problems with different approaches and complementary information.

Acknowledgments

I thank Michael Lieber, whose contribution was essential to both the work done in his laboratory and to the critique of PCR analysis of in vivo V(D)J junctions. I also appreciate helpful discussions with Aaron B. Kantor, and his permission to include data prior to publication. Thanks to Katherine Seidl for critical comments on the manuscript. I am a Fellow of the Leverhulme Trust for Medical Research. This work was supported in part by National Institutes of Health Grant AI4762.

References

9. GERSTEIN RM, LIEBER MR: Coding end sequence can markedly affect the initiation of V(D)J recombination. Genes Dev 7:1469–1469, 1993
27. LAJOS PA, NOSAL GJ, SANDERSON RD, MEHYER-ZEWS-Williams MG: Functional and molecular characterization of single- (4-hydroxy-3-nitrophenyl)asetyl (NP)-specific, IgG1+ B cells from antibody-secreting and memory B cell pathways in the C57BL/6 immune response to NP. Eur J Immunol 22:3001–3011, 1992
31. LEWIS S, GIFFORD A, BALTIMORE D: DNA elements are asymmetrically joined during the site-specific recombination of kappa immunoglobulin genes. Science 228:677–685, 1985
32. LEWIS S, GIFFORD A, BALTIMORE D: Joining of kappa kappa to J kappa
gene segments in a retroviral vector introduced into lymphoid cells. Nature
36. BLACKWELL TK, ALT FW: Site-specific recombination between immunoglobulin D and JH segments that were introduced into the genome of a murine pre-B-cell line. Cell 37:105-112, 1984
42. LANDAU NR, SCHATZ DG, ROSA M, BALTIMORE D: Increased frequency of N-region insertion in a murine pre-B-Cell line infected with a terminal deoxynucleotidytransferase retroviral expression vector. Mol Cell Biol 7:3237-3243, 1987
45. KALENNBACH S, DOYEN N, FANTON D'ANDON M, ROUGEOY F: Three lymphoid-specific factors account for all junctional diversity characteristics of somatic assembly of T-cell receptor and immunoglobulin genes. Proc Natl Acad Sci USA 89:2799-2803, 1992
46. KOMORI T, ORAKU A, STEWART V, ALT FW: Lack of N regions in antigen receptor variable region genes of Tg(D)-deficient lymphocytes. Science 261:1171-1173, 1992

80. BOGLE M, CHIFFELA S, BENZIG C, MATHIS D: Regulation of NF kappa B activation in lymphocytes and their ontogeny in thymic differentiation and thymus ontogeny. Proc Natl Acad Sci USA 89:10101-10105, 1992


82. ROEBENGER E, TRECULLA D: Cytoloidal proliferation unlinked to terminal deoxynucleotidyl transferase synthesis in thymocytes of young mice. J Immunol 130:1627-1633, 1983


87. MILLER TS, NICHOLS EF, PATTEN PA, DAVIS MM, CHEN YH: The adult T-cell receptor delta-chain is diverse and distinct from that of fetal thymocytes. Nature 331:627-631, 1988

88. GUSTAVS A, SLYSENZAWSKA S: Selection of gamma delta T cells with canonical T-cell antigen receptors in fetal thymus. Proc Natl Acad Sci USA 87:7935-7938, 1990


123. BRIES DE, FORMAN C, HUDAK S, CLAFLIN JL: The effects of idiootype on the ability of IgG1 anti-phosphorylcholine antibodies to protect mice from fatal infection with Streptococcus pneumoniae. *Eur J Immunol* 14:1027-1030, 1984


