A new break in V(D)J recombination

Data on the structure of DNA ends in thymocytes are providing new insights into the genetic recombination process that generates the antigen receptor genes of lymphocytes.

Vertebrates are able to mount specific immune responses because they possess a vast array of antigen-specific B and T cells. In B cells, the antigen receptor is a membrane-bound form of immunoglobulin, whereas in T cells the antigen-specific component of the T cell receptor (TCR) is composed of a heterodimer of either TCR α and β gene products or TCR γ and δ gene products. In the germline, the loci encoding each of the subunits of these receptors contain pools of individual gene segments organized into variable (V), diversity (D), and joining (J) gene segment families [1]. During development, a functional exon that encodes the antigen-binding domain is created by joining individual members of these families in a lineage-specific manner by a process referred to as V(D)J recombination. For example, in an individual murine B cell, the generation of an immunoglobulin heavy chain gene is initiated by the recombination of one of the D gene segments with one of the several J gene segments present within the locus. A second V(D)J recombination subsequently joins one of several hundred V gene segments to this DJ unit. An essentially identical process is required for the generation of the genes encoding the TCR proteins. Present evidence suggests that a common recombinase complex is used to perform these rearrangements.

In recent years, the process of V(D)J recombination has been an area of intense study. Not only is it central to the development of antigen receptor diversity, it is also the only form of site-specific recombination that has so far been found in vertebrates. Although the products of immunoglobulin and TCR rearrangement have been examined in great detail, a molecular understanding of the process by which V(D)J recombination is achieved has remained elusive, though clues abound. It is known that V(D)J recombination involves evolutionarily conserved recombination signal sequences (RSS) immediately flanking the V, D and J gene segments. These RSSs include a palindromic heptamer sequence and an AT-rich nonamer sequence, which are separated by a spacer of either 12 or 23 base pairs. As a result of V(D)J recombination, two rearrangement products are formed (Fig. 1). One product results from the precise joining of the heptamer elements in the two RSS elements producing a signal joint, whereas the other results from the joining of the two coding elements producing a coding joint.

Unlike other forms of site-specific recombination, which serve to create novel gene products during development and/or differentiation in non-vertebrate organisms, the position of joining of the two coding elements during V(D)J recombination is imprecise. The coding sequences of V, D and J gene elements undergoing recombination

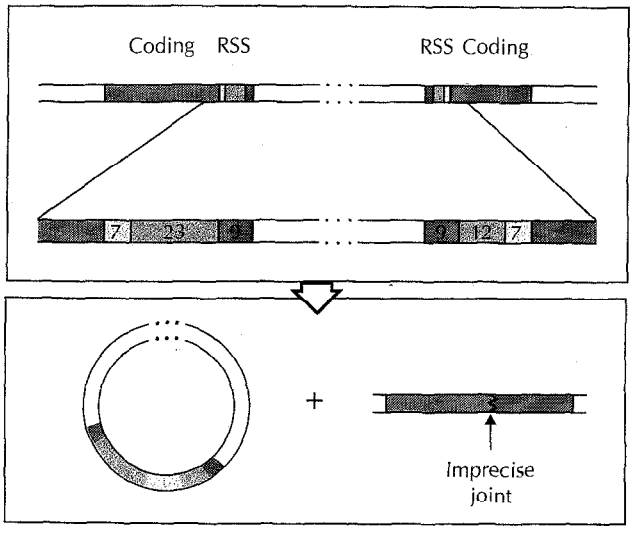

Fig. 1. V(D)J recombination mediated by the recombination signal sequences (RSS) flanking the coding sequences (blue and green). The joint between the coding sequences is imprecise.

frequently display nucleotide deletions and/or additions of up to 10–14 base pairs at their recombination junctions. This variability results in a high degree of inefficiency in the generation of a functional end product. The recombination joins coding elements at random with respect to their reading frames; thus, in two out of three events, the two coding elements are joined out-of-frame. For such a degree of inefficiency to be evolutionarily tolerated, it must also confer some benefit. The most likely advantage is that variations in the position and sequences that occur at the V(D)J joint contribute to the diversity of the antigen-binding portions of the receptor. In fact, the ability to generate a TCR repertoire in some species appears to be almost entirely dependent on variations generated in the coding sequences at the points of V(D)J recombination. Based on the crystal structure of immunoglobulin and modeling of TCR structure, the loop formed by the V(D)J recombination products defines the major antigen-binding domain of these molecules. Therefore, variation in the coding joint appears to play a central role in the production of receptor diversity.

An early model of V(D)J rearrangement proposed that the two recombination products were formed sequentially [2]. In this model, recombination is initiated by double-stranded breaks introduced between the conserved RSS heptamer sequences and the last base of the coding regions. Consistent with this proposal, endonucleases capable of cutting both strands at RSS sequences have been

detected in cell extracts from lymphoid tissues. Next, the RSS elements of each gene segment are immediately ligated to each other. The direct ligation of the RSS elements would account for the lack of nucleotide deletions and additions observed at signal joints. The two coding ends are thought then to remain exposed for a period of time during which they undergo modification, including deletions and random nucleotide addition, before their ligation. Deletions could result from exonucleases acting on the free ends generated following strand breakage. Consistent with the proposal that addition of the random nucleotides (N) to coding ends is catalysed by terminal deoxynucleotidyl transferase, cell lines expressing little or no transferase activity exhibit an increased number of N region additions when transfected with a construct that expresses the enzyme.

Until recently, the above model had been adequate to explain the known features of V(D)J recombination. However, it came under closer scrutiny with the discovery of an unexpected form of non-random base-pair addition first observed in the coding joints of avian species [2]. These non-random base pairs, now often referred to as P nucleotides [3], are found only at the ends of full-length coding segments and are always the reverse complement of the coding segment nucleotides next to the RSS. Since the original description of P nucleotides, many other examples of P nucleotide addition have been described [4].

The identification of P nucleotides led to the consideration of a new model of RSS-mediated recombination. As first proposed by Lieber [1], this model suggests that V(D)J recombination occurs by a process similar to that identified in the RNA self-splicing mechanism identified in *Tetrahymena* [5]. In this model (Fig. 2), rearrangement is initiated by the binding of a recombinase molecule or complex, introducing a single-stranded nick between a coding sequence and its RSS. The base affected by this bond is activated and attacks its complementary base on the other strand, generating a hairpin intermediate and creating a double-stranded break between the coding region and RSS. Resolution of this hairpin could occur through the action of a single-strand endonuclease, which can cut between any two bases in the hairpin structure. If the cut occurs at the base of the loop on the antisense strand, a sense strand with a 3′ overhang is created containing a short,inverted repeat (the P nucleotides). If this 3′ end is then used to ligate the coding elements, the stretch of P nucleotides will be incorporated into the product. In contrast, if the loop is opened at its base on the sense strand then the sense strand is shortened. If this shortened 3′ end is used to ligate the coding element, a product containing a coding segment deletion will be observed.

Preliminary experimental support for this model has now been achieved. By careful examination of TCR δ gene recombination, David Roth and colleagues [6,7] have been able to detect the existence of both RSS sequences ending in double-strand breaks and coding ends with hairpin loops. Roth *et al.* first looked for these intermediates at the TCR δ locus of normal neonatal mice, which undergoes preferential rearrangement of one particular DJ

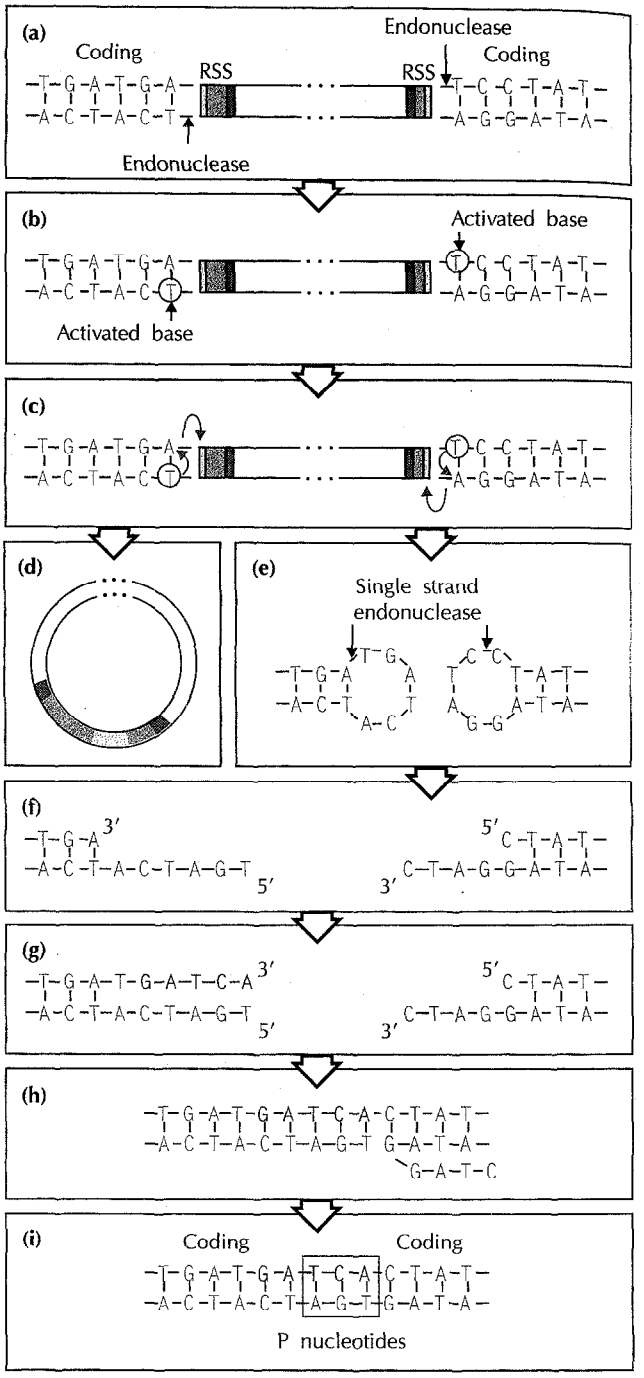

Fig. 2. A model for V(D)J recombination. **(a)** The recombinase introduces a single-strand nick between the RSS heptamer and the coding region. **(b)** The nucleotide bound to the recombinase is activated and **(c)** attacks its complementary base. This results in a hairpin structure formation at the coding ends, generating double-strand breaks between the RSS elements and the coding gene segments. **(d)** The RSS elements are ligated, generating a DNA circle which is lost. **(e)** A single-strand endonuclease nicks between any two bases exposed as a result of hairpin formation. **(f)** Bases are filled in, after which **(g)** the base-paired ends are ligated and **(h)** unpaired nucleotides are removed. **(i)** The end result in this example is the presence of three pairs of P nucleotides and a loss of two base pairs from the end of coding region 2.

pair, and detected double-strand breaks near RSS in DNA of the thymus [6]. These structures appear to be surprisingly stable, considering the high DNase content of

thymic tissue. This suggests that the free ends may be protected either through binding of a component of the recombinase system or through a novel DNA intermediate. However, Roth *et al.* were unable to detect any form of coding ends, which, they suggest, must be processed so rapidly that any intermediates in the V(D)J recombination process are present too transiently to be detected in their experiments.

Roth and colleagues then turned their attention to studying the same recombinations in SCID mice, which are severely immunodeficient and exhibit defects in both V(D)J recombination and general DNA repair. Although these mutant mice are defective in V(D)J rearrangement, the block occurs after initiation of the process. SCID mice have been shown to be defective in their ability to form coding junctions, even though they are able to form signal junctions. The researchers reasoned that these mice might be blocked in their ability to process coding end intermediates and, therefore, might accumulate these intermediates to detectable levels. Consistent with this hypothesis, Roth *et al.* [7] were able to identify in the thymocytes of SCID mice coding ends with hairpin structures that are resistant to exonuclease digestion. Using two-dimensional Southern blotting of DNA from thymocytes (where TCR gene rearrangements occur), hairpin structures of predicted size were detected by means of probes for TCR δ D and J coding regions. Interestingly, hairpin structures associated with the RSS could not be detected. These RSS ends apparently end in double-strand breaks which, at least in purified DNA samples, are susceptible to exonuclease digestion.

The presence of excessive P nucleotide addition in the rare coding joints generated in SCID mice [8,9] further suggests that defective processing of hairpin structures may underlie the SCID defect. Perhaps such hairpins are not detectable in normal mice because they are too efficiently processed to accumulate. Alternatively, as SCID mice are incapable of forming coding joints, hairpin formation may be a byproduct of the defective recombination process. Identification of hairpin structures in normal mice will be required to distinguish between these two possibilities.

The new data, while not conclusive, support several aspects of Lieber's model of V(D)J recombination. First, the initial cleavage mediated by the V(D)J recombinase seems to occur on a single strand. Second, the recombinase itself may play a role in protecting the RSS sequence from modification, perhaps by binding to it. Finally, deletions from coding ends may be determined by where the coding-end hairpin loop is opened to initiate recombination, rather than as a result of exonuclease digestion. Although by no means proven, this new model should be helpful for testing the role of the individual components of the V(D)J recombinase that have been identified and cloned. Three potential components of the recombinase have so far been identified; the products of the RAG-1, RAG-2 and SCID genes [10,11]. It will be interesting to determine whether any of these proteins have the ability to nick DNA and/or covalently bind to an RSS sequence or open a closed hairpin loop.

References

1. LIEBER MR: **Site-specific recombination in the immune system.** *FASEB J* 1991, **5**:2934–2944.
2. MCCORMACK WT, TJOELKER LW, CARLSON LM, PETRYNIAK B, BARTH CF, HUMPHRIES EH, THOMPSON CB: **Chicken IgL gene rearrangement involves deletion of a circular episome and addition of single nonrandom nucleotides to both coding segments.** *Cell* 1989, **56**:785–792.
3. LAFAILLE JJ, DECLOUX A, BONNEVILLE M, TAKAGAKI Y, TONEGAWA S: **Junctional sequences of T-cell receptor γδ genes: implications for γδ T-cell lineages and for a novel intermediate of V-(D)-J joining.** *Cell* 1989, **59**:859–870.
4. MEIER JT, LEWIS SM: **P nucleotides in V(D)J recombination: a fine-structure analysis.** *Mol Cell Biol* 1992, in press.
5. ZAUG AJ, GRABOWSKI PJ, CECH TR: **Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction.** *Nature* 1983, **301**:578–583.
6. ROTH DB, NAKAJIMA PB, MENETSKI JP, BOSMA MJ, GELLERT M: **V(D)J recombination in mouse thymocytes: double-strand breaks near T-cell receptor δ rearrangement signals.** *Cell* 1992, **69**:41–53.
7. ROTH DB, MENETSKI JP, NAKAJIMA PB, BOSMA MJ, GELLERT M: **V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes.** *Cell* 1992, **70**:983–991.
8. SCHULER W, RUETSCH NR, AMSTER M, BOSMA MJ: **Coding joint formation of endogenous T-cell receptor genes in lymphoid cells from scid mice: unusual P nucleotide additions in VJ-coding joints.** *Eur J Immunol* 1991, **21**:589–596.
9. KIENKER LJ, KUZIEL WA, TUCKER PW: **T-cell receptor γ and δ gene junctional sequences in scid mice: excessive P nucleotide insertion.** *J Exp Med* 1991, **174**:769–773.
10. OETTINGER MA, SCHATZ DG, GORKA A, BALTIMORE D: **RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination.** *Science* 1990, **248**:1517–1522.
11. BOSMA GC, CUSTER RP, BOSMA MJ: **Severe combined immunodeficiency mutation in the mouse.** *Nature* 1983, **301**:527–530.

Stacy E. Ferguson and Craig B. Thompson, Howard Hughes Medical Institute and Departments of Internal Medicine and Microbiology/Immunology, University of Michigan Medical Center, Ann Arbor, Michigan 48109-0650, USA.