Nucleic Acids Research, 1994, Vol. 22, No. 10 1785–1796

Conservation of sequence in recombination signal sequence spacers

Dale A.Ramsden+, Kristin Baetz and Gillian E.Wu*
Department of Immunology, University of Toronto, Toronto, Ontario M5S 1A8 and The Wellesley Hospital Research Institute, 160 Wellesley Street East, Toronto, Ontario M4Y 1J3, Canada

Received October 21, 1993; Revised and Accepted April 13, 1994

ABSTRACT

The variable domains of immunoglobulins and T cell receptors are assembled through the somatic, site specific recombination of multiple germline segments (V, D, and J segments) or V(D)J rearrangement. The recombination signal sequence (RSS) is necessary and sufficient for cell type specific targeting of the V(D)J rearrangement machinery to these germline segments. Previously, the RSS has been described as possessing both a conserved heptamer and a conserved nonamer motif. The heptamer and nonamer motifs are separated by a 'spacer' that was not thought to possess significant sequence conservation, however the length of the spacer could be either 12 +/– 1 bp or 23 +/– 1 bp long. In this report we have assembled and analyzed an extensive data base of published RSS. We have derived, through extensive consensus comparison, a more detailed description of the RSS than has previously been reported. Our analysis indicates that RSS spacers possess significant conservation of sequence, and that the conserved sequence in 12 bp spacers is similar to the conserved sequence in the first half of 23 bp spacers.

INTRODUCTION

The adaptive immune response in vertebrates combats environmental pathogens by the use of a vast repertoire of antigen specific receptors (immunoglobulins and T cell receptors). The diversity of this repertoire is resident in the variable domain, which is assembled though a somatic, cell type specific process involving the site specific recombination of germline V, D and J segments [V(D)J rearrangement].

Adjacent to the coding sequence of all V, D and J segments that are capable of V(D)J rearrangement is a conserved non-coding sequence that functions as a targeting signal for recombination, termed the recombination signal sequence (RSS) [1, 2] Recombination substrates have demonstrated that RSS are both necessary and sufficient for targeting of V(D)J rearrangement to lymphoid cell types [3, 4]. RSS were originally defined through alignment and comparison of multiple examples, resulting in a definition of two classes of RSS, both possessing identical conserved seven bp (heptamer) and nine bp (nonamer) motifs. One class has an approximately 12 bp spacer of non-conserved sequence separating the heptamer and nonamer motifs, while the other class has an approximately 23 bp spacer [5]. V(D)J rearrangement occurs efficiently only between a 12 bp spacer RSS and a 23 bp spacer RSS.

The RSS spacers, as previously discussed, are generally assumed to lack conserved sequence. The overall sequence composition of the spacer was considered as potentially significant, however, as early mechanisms of V(D)J rearrangement suggested that a recombination intermediate required melting of RSS DNA [1]. Two experiments, involving complete substitution of spacer sequence with GC base pairs, attempted to address this question, and have conflicting results. Experiments by Sakano and colleagues suggested that GC substitution of an RSS spacer resulted in an impairment of recombination frequency [3], while in experiments by Lieber and colleagues the authors suggest that GC substitution of RSS spacers made no significant difference [6]

Previous consensus analysis of RSS have concentrated on the heptamer and nonamer [7] We have used the considerable increase in the number of sequenced RSS present in the data bases to analyse the spacer sequences. We have constructed a large database of aligned, functional RSS from different species and different loci, classified according to the size of the RSS spacer. A comprehensive RSS consensus, based upon classification by RSS spacer size and including heptamer and nonamer motifs as well as the spacer, is presented. In contrast to previous definitions of RSS, we observed significant conservation of sequence in RSS spacers. Moreover, the conserved sequence for 12 bp spacers is similar to the conserved sequence in the first half of 23 bp spacers.

MATERIALS AND METHODS

RSS analysis

Alignment of RSS. We have obtained 453 examples of RSS, from different species and different loci (see Tables 1 and 2). Sequences were retrieved from GenBank (release 67.0) or the specified references using 'lineup', a Genetics Computer Group [8] (GCG)

*To whom correspondence should be addressed

+Present address: National Institutes of Health, NIDDK, Room 312, Building 5, Bethesda, MD 20892, USA

program that allows visual alignment of multiple sequences. As our primary goal is to relate sequence conservation to function, we exclude RSS that are associated with a pseudo-gene segment (as defined by the ability to contribute to a functional, mature protein), and classified the RSS solely on the length of the spacer. Care was taken to include only one example of a given gene segment's RSS when multiple versions of the same gene segment were present in GenBank. We note that while considerable effort has been expended to ensure this database is comprehensive, it is not complete. We define here an abbreviation to aid in future description of the RSS: RSS derived from 12 bp spacer RSS will be referred to as 12 RSS, while RSS derived from 23 bp spacer RSS will be referred to as 23 RSS.

Sequences were aligned using 'pileup', a GCG program that aligns groups of sequences based on comparison of the closest related pairs, and introduces gaps to promote optimal alignment. As experiments suggest that RSS function efficiently only if the first three nucleotides of the heptamer are fixed at CAC, and the heptamer and nonamer are separated by a spacer with variation in length of 11–13 bp or 22–24 bp only [7], gaps were inserted for optimal alignment based on these criteria. This was achieved using the pileup parameters 'gap weight' set at three, and 'gap length weight' set at 0.2. (see Tables 3 and 4). As similar sequences are often grouped together in these tables, gap position may occasionally appear somewhat idiosyncratic when limited portions of the database are observed. Moreover, while the gap weight and gap length parameters applied resulted in largely 12 and 23 bp spacers, as hoped, RSS that appear to have longer spacers than 12 or 24 bp cannot accommodate extensive gaps without high penalty, and thus may appear misaligned.

Consensus determination. Consensus sequences were determined using the 'plurality' rule [9] (Tables 5 and 6). This method determines a consensus result with varying degrees of ambiguity, such that for each position the degree of ambiguity is related to the significance of the observed nucleotide conservation. A consensus result may consist of only one nucleotide and thus be unambiguous, indicating a highly conserved position, or may be ambiguous for up to all four nucleotides. A position with a consensus result ambiguous for all four nucleotides has a nucleotide distribution indistinguishable from random. Analysis of the properties of this rule indicate that when there are at least 100 sequences in a database (both 12 RSS 23 RSS sets have over 100 sequences), the probability that a randomly generated database would produce a consensus result ambiguous for less than four nucleotides is less than 1% [10]. We further define consensus results ambiguous for more than one nucleotide by reporting the nucleotides in order of the frequency that they are observed, from the most frequent to the least frequent.

RESULTS

Alignment of RSS

255 examples of 12 RSS were obtained, largely derived from IgH D and Igκ V loci (Table 1). As described in the Materials and Methods, gaps were inserted for optimal alignment, although gaps were rarely required for the alignment of 12 RSS (Tables 3 and 5). 198 examples of 23 RSS were obtained (Table 2). Gaps were introduced such that there are 24 positions between the heptamer and the nonamer (see Tables 4 and 6). 80% of RSS (159/198) contained a single one base pair gap, and therefore possessed 23 bp spacers. There were 20 (10%) sequences with 22 bp separating the heptamer and nonamer, two (1% with 21 bp separating the heptamer and nonamer, and 17 (9%) with 24 bp separating the heptamer and nonamer.

The results of the sequence analysis of these alignments will refer to the positions in each alignment as belonging to one of the three elements (heptamer, nonamer, and spacer), and the 5′ terminus of each element will be referred to as the first position of each element.

Conservation of sequence in the heptamer and nonamer

The consensus sequence for all positions of the heptamer, for both 12 RSS and 23 RSS, was unambiguous (Tables 5a and 6a). The first three nucleotides of the heptamer were almost perfectly

Table 1. Sources of 12 RSS

	Species										
Locus	Mus	Hum	Chk	Rab	Hef	Xel	Rat	Bov	Shp	Duk	totals
IGHD	20	32	16	12	12	-	2	-	-	-	94
IGκV	30	17	-	5	2	3	-	-	-	-	57
IGλJ	3	4	1	-	-	-	1	-	1	1	11
TcRαJ	46	5	-	-	-	-	-	-	-	-	51
TcRβJ	12	13	-	-	-	-	-	-	-	-	25
TcRβD	2	2	-	-	-	-	-	-	-	-	4
TcRγJ	2	2	-	-	-	-	-	1	-	-	5
TcRδD	2	2	-	-	-	-	-	-	-	-	4
TcRδJ	2	2	-	-	-	-	-	-	-	-	4
totals	119	79	17	17	14	3	3	1	1	1	255

Abbreviations: IG; immunoglobulin. TcR; T cell receptor. H; heavy chain. κ; light chain of the kappa isotype. λ; light chain of the lambda isotype. Isotype classification for Xel and Hef chains is not clear, however they are grouped with whatever light chain isotype has the same sized RSS spacer for the purposes of these tables. α; T cell receptor alpha chain. β; T cell receptor beta chain. γ; T cell receptor gamma chain. δ; T cell receptor delta chain. V; variable gene segment. D; diversity gene segment. J; joining gene segment. Species are Mus; Mouse (*Mus musculus*). Hum; Human (*Homo sapiens*). Xel; Frog (*Xenopus laevis*). Shp; Sheep (*Ovis aries*). Hef; Horned shark (*Heterodontus franciscus*). Rab; Rabbit (*Oryctolagus cuniculus*). Chk; Chicken (*Gallus gallus*). Bov; Cow (*Bos taurus*). Rat; Rat, (*Rattus norvegicus*). Duk; Muscovy duck.

Table 2. Sources of 23 RSS

	Species									
Locus	Mus	Hum	Xel	Shp	Hef	Rab	Chk	Rat	Duk	totals
IGHV	32	25	16	-	4	5	1	-	-	83
IGHD	-	-	-	-	4	-	-	-	-	4
IGHJ	4	5	-	-	4	4	1	1	-	19
IGκJ	4	5	-	-	2	1	-	5	-	17
IGλV	3	9	-	14	-	-	1	1	2	30
TcRαV	9	2	-	-	-	-	-	-	-	11
TcRβV	8	8	-	-	-	-	-	-	-	16
TcRβD	2	2	-	-	-	-	-	-	-	4
TcRγV	3	5	-	-	-	-	-	-	-	8
TcRδV	-	2	-	-	-	-	-	-	-	2
TcRδD	2	2	-	-	-	-	-	-	-	4
totals	67	65	16	14	14	10	3	7	2	198

Abbreviations: IG; immunoglobulin. TcR; T cell receptor. H; heavy chain. κ; light chain of the kappa isotype. λ; light chain of the lambda isotype. Isotype classification for Xel and Hef chains is not clear, however they are grouped with whatever light chain isotype has the same sized RSS spacer for the purposes of these tables. α; T cell receptor alpha chain. β; T cell receptor beta chain. γ; T cell receptor gamma chain. δ; T cell receptor delta chain. V; variable gene segment. D; diversity gene segment. J; joining gene segment. Species are Mus; Mouse (*Mus musculus*). Hum; Human (*Homo sapiens*). Xel; Frog (*Xenopus laevis*). Shp; Sheep (*Ovis aries*). Hef; Horned shark (*Heterodontus franciscus*). Rab; Rabbit (*Oryctolagus cuniculus*). Chk; Chicken (*Gallus gallus*). Bov; Cow (*Bos taurus*). Rat; Rat, (*Rattus norvegicus*). Duk; Muscovy duck.

Table 3. Alignment of 12 bp spacer RSS

Species	Locus	Segment	Heptamer	Spacer	Nonamer	Reference
MUS	IGH	D Q52	CACTGTG	GTGCTCCGCTTA	GTCAAAACC	[17]
		D Q52	CACGGTG	ACGCGTGGCTCA	ACAAAAACC	[17]
		D SP2-2	CACAGTA	GTAGATCCCTTG	ACAAAAATC	[18]
		D SP2-2	CACAGTG	ATATATCCAGCA	ACAAAAACC	[18]
		D SP2-3	CACAGTA	GTAGATCCCTTG	ACAAAAATC	[19]
		D SP2-3	CACAGTG	ATATATCCAGCA	ACAAAAACC	[19]
		D SP2-4	CACAGTA	GTAGATCCCTTG	ACAAAAATC	[19]
		D SP2-4	CACAGTG	ATATATCCAGCA	ACAAAAACC	[19]
		D SP2-5	CACAGTA	GTAGATACCTTG	ACAAAAATC	[19]
		D SP2-5	CACAGTG	ATATATCCAGCA	ACAAAAACC	[19]
		D SP2-6	CACAGTA	GTAGATCCCTTG	ACAAAAATC	[19]
		D SP2-6	CACAGTG	ATATATCCAGCA	ACAAAAACC	[19]
		D SP2-7	CACAGTA	GTAGATCCCTTG	ACAAAAATC	[19]
		D SP2-7	CACAGTG	ATATATCCAGCA	ACAAAAACC	[19]
		D SP2-8	CACAGTA	GTAGATCCCTTG	ACAAAAATC	[19]
		D SP2-8	CACAGTG	ATATATCCAGCA	ACAAAAACC	[19]
		D FL16.1	CACAGTA	GTAGATCCCTTC	ACAAAAAGC	[19]
		D FL16.1	CACAGTG	CTATATCCATCA	GCAAAAACC	[19]
		D FL16.2	CACAGTA	GTAGATCCCTTC	ACAAAAAGC	[19]
		D FL16.2	CACAGTG	CTATATCCAGCA	ACAAAAATC	[19]
	IGκ	V 18.1	CACAGTG	ATGCAGACCCTA	ACAAAAACA	[20]
		V K1A5	CACAGTG	ATACAGACCCTA	ACAAAAATA	[20]
		V 5.1	CACAGTG	ATACAGACCCTA	ACAAAAATA	[20]
		V K24C	CACGGTG	ATACAGCCCTGA	ACAAAAACC	[21]
		V K24A (Pa)	CACAGTG	ATACAAACCTGA	ACAAAAACC	[21]
		V K24.1	CACATTG	ATACTGCACTGG	ACAAAAACC	[21]
		V -Ser	CACAGTG	CTTCAGCCTCCT	ACACAAACC	[22]
		V 167	CACAGTG	ATAGAGCCCTGA	ACAAAAACC	[23]
		V MOPC173b	CACAGTG	ATACAAATCACA	ACATAAACC	[24]
		V K41	CACAGTG	ATACAAATCATA	ACATAAACC	[25]
		V K2	CACAGTG	ATTCAAGCCATG	ACATAAACC	[26]
		V K1.6 (21x)	CACAGTG	CTCCAGGGCTGA	ACAAAAACC	[27]
		V K21E	CACAGTG	CTCCAGGGCTGA	ACAAAAACA	[27]
		V K21B	CACAGTG	CTCCAGGGCTGA	ACAAAAACC	[27]
		V K21C	CACAGTG	CTCCAGGGCTGA	ACAAAAACC	[27]
		V K18	CACAGTG	CTCCAGGGCTGA	ACAAAAACC	[27]
		V K24A	CACAGTG	ATGCAGCCCTGA	ACAAAAACC	[28]
		V K24B	CACACTG	ATACAGCCCTGA	ACAAAAACA	[28]
		V 1B	CACAGTG	ATACAGACCCTA	ACAAAAATA	[29]
		V 1C	CACAGTG	ATACAGACCCTA	ACAAAAATA	[29]
		V R11	CACAGTG	ATACAGGCTGGA	ACAAAAAC.	[30]
		V R1	CACAGTG	CTACATACTGAA	ACAAAAACA	[30]
		V L8	CACAGTG	CTACAGACTGGA	ACAAAAACA	[30]
		V H6	CACAGTG	ATACAGACTGGA	ACAAAAACC	[30]
		V H1	CACAGTG	CTACAGACTAGA	ACAAAAACC	[30]
		V H4	CACAGTG	ATACAGACTGGA	ACAAAAACC	[30]
		V H9	CACAGTG	ATACAGACTGGA	ACAAAAACC	[30]
		V R9	CACAGTG	ATACAGACTGGA	ACAAAAACC	[30]
		V H13	CACAGTG	ATACAGACTGGA	ACAAAAACC	[30]
		V H3	CACAGTG	ATACAGACTGGA	ACAAAAACC	[30]
MUS	IGλ	J 1	CACTGTG	ATATAGACTCAT	GCAAAAAA.	[31]
		J 2	CACAATG	ACTAAAACCCAA	CCCAAAACC	[31]
		J 3	CACAGTG	ACTGAAACCCAA	CCCTAAACC	[31]
	TCRα	J TA65	CACTGTG	ACAATAACCTCA	ACAAAAACC	[32]
		J new2	CACAGCA	AATCAACCCTTT	ACAAAAAAC	[33]
		J TA91	CACACGT	CTCTTCGTGAGA	AGACACTGT	[33]
		J C5A	CACTGTA	ACACGGGCCTTT	ACAAAAACA	[33]
		J new1	CACAGCC	TGGGGAGGCTTT	ACAAAAACA	[33]
		J 2b4A	CACAATG	ACACGGGACTCT	ACAAAAACT	[33]
		J TA27	CACACCC	ACACACTGCCTT	ACAAATACT	[33]
		J TA1	CACACTG	CACTGAAGGGCT	TTGCAAAAA	[33]
		J 45	CACACTG	CACTGAAGGGCT	TTGCAAAAA	[34]
		J BDFLII	CACAGTG	ATTTGTCCTGTG	ACAAAATGG	[33]
		J PHDS	CACAGTG	GCTGACTCTACA	ACAAAAACT	[32]
		J TA 84	CACAGTG	ATCTCTTCCACC	ACAAAAACT	[32]
		J T2C	CACAGTG	ATATCATGTTCT	ACAAAAACC	[35]
		J TA31	CACAGTG	TGCCAAGCCATT	ACAAAATCC	[33]
		J new 3	CACTGTC	TCCAATAACAGC	ACAGAAAAC	[33]
		J TA80	CACCCTG	AGGCAAGCCTTG	ACATAAACC	[32]
		J TA46	CACTGTG	AGACACTCCATA	TCAGAAACC	[33]
		J new4	CACAGTA	ATACACACTCTA	ACAAAAACT	[33]
		J new5	CACAGTC	ATTTGGGGCCTT	ACAATAACC	[33]
		J TA19	CACAGTG	TTCTGTGTCTCT	ACATAAACC	[32]
		J TA37	CACAGTG	ATCTCCAGCTCA	GCAAAAACC	[33]
		J NAT1	CACAGTT	ATAGAGAGCTTT	ACAGAAATG	[33]
		J TA57	CACCCCA	ATGCTGCACTTT	ACAAAAACT	[33]
		J new6	CACAGTG	ATATCATGTTCT	ACAAAAACC	[33]
		J new7	CACAGAC	ACAAAAACCTTA	ACAAAAACA	[33]
		J TT11	CACAGCC	CTGCAGAGCCTT	ACAATAACT	[32]
		J TA20	CACATCA	TCTCTTGCCTTT	ACTGAAACC	[33]

Table 3. (cont.)

Species	Locus	Segment	Heptamer	Spacer	Nonamer	Reference
		J BM10-37	CACACTG	TGATTGGGACCA	TACCCAAAA	[33]
		J new10	CACAGTG	ATCTGAAGCCAA	GCAAAAACA	[33]
		J b12	CACAGTG	CCAGCCCCCTTT	ACACAAATC	[33]
		J 14-4	CACAATG	GTTAGCACCATT	ACAGAAAGC	[33]
		J TA28	CACTGTG	ATTTGCTCAACA	ACAAGAACC	[33]
		J BM2T3-1	CACTGTG	TTACATACCCTG	TCAAAAACA	[33]
		J new9	CACACTG	TGACAAACACGT	CTACAAAAT	[33]
		J 112-2	CACAGTG	GGTTTCCTCTTA	GCAAAAACT	[33]
		J TA61	CACAGTG	CTCCGTGCTATT	GCAATAACC	[33]
		J new8	CACAGAA	TTTCCTTTCTTT	GCAAAAACT	[33]
		J TA26	CACTGCA	GGTGACACCTTT	ACAGAAGCC	[33]
		J new14	CACAGTA	GAAAGGTGCTTT	ACAAGAATT	[33]
		J new13	CACAGTG	AGGAAAGCCTTT	GATGAAACC	[33]
		J new12	CACTCTG	AGTAAGTGCTTC	ACAAAAACG	[33]
		J TA72	CACAGTG	ATTTGTCCTGTG	ACAAAATGG	[33]
		J new11	CACAGCA	GCAAACCTCTCC	ACAAAAATG	[33]
		J TA39	CACTGTA	AGTGAGGTCTTT	ACAAAATGG	[33]
		J DK1	CACAGTG	AAACGAGGCCCT	GCAAATTCT	[33]
		J LB2A	CACAGTG	CCAGCCCCCTTT	ACACAAATC	[33]
MUS	TCRβ	J 1.1	CACAGTG	CCATAGGATGAG	GAGAAAAAT	[36]
		J 1.2	CACATCA	GAATACAGATAC	TGCAATATG	[36]
MUS	TCRβ	J 1.3	CACAGCC	TCCCGGGTTCAC	TTCAAAACC	[36]
		J 1.4	CACAACA	TTAAAGCCTAGT	GGTAAAACT	[36]
		J 1.5	CACAGTA	CAACATGAGGGT	GACAAACTC	[36]
		J 1.6	CACAGCT	GCAGGTGACCTT	GGTAAAACC	[36]
		J 2.1	CACAGCA	GAAAAGGGCTAC	CAAGAATTC	[37]
		J 2.2	CACAGTC	TTGGAAATGCTG	GCACAAACC	[37]
		J 2.3	CACAGCC	TCCAGGCTCAGG	ACAAAAACT	[37]
		J 2.4	CACAGCC	TCTTGGTACAGG	ACAAAAACT	[37]
		J 2.5	CACAGCC	CCAGAACCCAAC	ACAAAAACT	[37]
		J 2.7	CACAGTG	GCTCAACCCCAC	ACACAAACC	[37]
	TCRβ	D 1-1	CACAATG	TTACAGCTTTAT	ACAAAAAAG	[38]
		D 2-1	CACAATG	TTACATCGTGAT	ACAAAAAAG	[38]
	TCRγ	J 1	CACAGTG	CTCACAGCTTCT	ACAAAAATC	[39]
		J 2	CACAGTG	CTCACAGCTTCT	ACAAAAATC	[39]
	TCRδ	D 2	CACGGTG	CTACAGAGCTTT	GCAAAAACC	[40]
		D 1	CACAGTG	AAACACAGCCGT	ACAAAAACA	[40]
	TCRδ	J 2	CACGTTA	TAATCTTGCTTT	GCAGATAAC	[40]
		J 1	CACAGCT	ACTGAGGCCCAT	TCCAAAAAC	[40]
HUM	IGH	D HQ52	CACAGTG	ATTGGCAGCTCT	ACAAAAACC	[41]
		D HQ52	CACAGTG	GTTCTCAGCTCA	GCCAAAAAC	[41]
		D LR1	CACAGTG	ACACAGCCCCAT	TCCCAAAGC	[42]
		D LR1	CACAGTG	ACACGAGCCCCC	ACAAAATCC	[42]
		D LR2	CACAGTG	ACACGAGCCCCC	ACAAAATCC	[42]
		D LR2	CACAGTG	ACACAGACCCAT	TCCCAAAGC	[42]
		D LR3	CACAGTG	ACACAACCCCAT	TCCTAAAGC	[42]
		D LR3	CACAGTG	ACACGAGCCCCC	ACAAAATCC	[42]
		D LR4	CACAGTG	ACACGAGCCCCC	ACAAAATCC	[42]
		D LR4	CACAGTG	ACACAGCCCCAT	TCCCAAAGC	[42]
		D XP4	CACAGTG	ACACAGACCTCA	CCCCAAACC	[43]
		D XP4	CACAGTG	TCACAGAGTCCA	TCAAAAACC	[43]
		D XP1	CACAGTG	ACACAGACCTCA	CCCCAAACC	[43]
		D XP1	CACAGTG	TCACAGAGTCCA	TCAAAAACC	[43]
		D XP'1	CACAGTG	ACACAGACCTCA	CCCCAAACC	[43]
		D XP'1	CACAGTG	TCACAGAGTCCA	TCAAAAACC	[43]
		D A1	CATAGTG	ATGAACCCAGTG	GCAAAAACT	[43]
		D A1	CACAGCA	GGAGGGCCCTTC	ACAAAAAGC	[43]
		D A4	CACAGTG	ATGAACCCAGCA	GCAAAAACT	[43]
		D A4	CACAGTA	GGAGGACCCTTC	ACAAAAAGC	[43]
		D K4	CACAGTG	GTGCTGCCCATA	GCAGCAACC	[43]
		D K4	CACAGTC	TGACACCCCCTG	ACAATAACC	[43]
		D K1	CACAGTG	GTGCCGCCCATA	GCAGCAACC	[43]
		D K1	CACAGTC	TGACATCGCCTG	ACAATAACC	[43]
		D N4	CACAGTG	ACACTCGCCAGG	CCAGAAACC	[43]
		D N4	CACTGTG	ACACAGACACCT	TCAGAAACG	[43]
		D N1	CACAGTG	ACACTCACCCAG	CCAGAAACC	[43]
		D N1	CACAGTG	ACACAGACACCT	TCAGAAACC	[43]
		D M1	CACTGTG	AGAAAAGCTTCG	TCCAAAACG	[43]
		D M1	CACTGTG	ACTCGGGGCTGT	TCAGAATCC	[43]
		D M2	CACTGTC	AGAATAGCTACG	TCAAAAACT	[43]
		D M2	CGCTGTG	ACTCGGGGCTGT	TCGGAATCC	[43]
	IGκ	V 321	CACAGTG	ATTCAGCTTGAA	ACAAAAACC	[44]
HUM	IGκ	V 305	CACAGTG	ATTCAGCTTGAA	ACAAAAACC	[44]
		V 328-h2	CACAGTG	ATTCAACATGAA	ACAAAAACC	[45]
		V 328	CACAGTG	ATTCAACATGAA	ACAAAAACC	[45]
		V b	CACAGTG	TTACCAACCCGA	ACATAAACC	[46]
		V b'	CACAGTG	TTACCAACCCGA	ACATAAACC	[46]
		V HK101	CACAGTG	TTACACACCCAA	ACATAAACC	[47]
		V HK102	CACAGTG	TTACACACCCGA	ACATAAACC	[47]

Table 3. (cont.)

Species	Locus	Segment	Heptamer	Spacer	Nonamer	Reference
		V HK146	CACAGTG	TTACACACCCAA	ACATAAACC	[48]
		V HK137	CACAGTG	TTACACACCCAA	ACATAAACC	[48]
		V HK166	CACAGTG	TTACACACCCAA	ACATAAACC	[48]
		V HK189	CACAGTG	TTACACACCCAA	ACATAAACC	[48]
		V a'	CACAGTG	TTACAAACCCGA	ACATAAACC	[46]
		V d	CACAGTG	TTACAAACCTGA	ACATAAACC	[46]
		V e	CACAGTG	TTACACACCCAA	ACAAAAACC	[46]
		V g	CACAGTG	ATTCCACATGAA	ACAAAAACC	[49]
		V-h	CACAGTG	ATTCAACATGAA	ACAAAAACC	[49]
	IGλ	J 1	CACAGTG	ACTGAGGCTCAG	ACCAAAACC	[50]
		J 2	CACTGTG	ACACAGGCTCAT	ACAAAAACC	[50]
		J 3	CACTGTG	ACACAGGCTCAT	ACAAAAACC	[50]
		J 7	CACAGTG	ACACAGCCCCAC	ACACAAACC	[51]
	TCRα	J C	CACTATG	ATTTGCTCAACA	ACAAAACCA	[52]
		J B	CACAGTG	TTTCTTAGTCAG	TCAAAAACA	[52]
		J AB	CACAGTG	ATACTGAGATCT	ACAAAAACC	[53]
		J RP	CACTGTG	AGATGCTTCATA	ACAGAAACC	[53]
		J AA	CACAGTG	TTATGTGTCTCT	ACATAAACC	[53]
	TCRβ	J 1.1	CACAGTG	ACAGGGGTCAAG	GTGAAAATC	[54]
		J 1.2	CACATAA	GAATATAGCCAC	TCTAAAAGG	[54]
		J 1.3	CACAGCC	TCCCAGGGCCAC	TTCAAAACC	[55]
		J 1.4	CACAACA	TTAAAGACTGGA	AGGAAAACC	[55]
		J 1.5	CACAGTG	CATCATGAGTGT	GGCAAACCC	[55]
		J 1.6	CACAGCT	GCAGAGGCTTAG	ATAAAACCC	[55]
		J2.1	CACAGTG	GGAAGGGGCTGC	CCAGAATTC	[56]
		J 2.2	CACAGCC	CTGGGGACCCTG	GCGCAAACC	[56]
		J 2.3	CACAGCC	TGGAGGCCCAGG	ACAAAAACC	[56]
		J 2.4	CACAGCC	CCGAGACGCGGC	ACAGAAACT	[56]
		J 2.5	CACGGCC	CCCGAGCCCCGC	ACAAAAACC	[56]
		J 2.6	CACAGCC	CGGGGACTCCCC	GCAAAAACC	[56]
		J 2.7	CACGGAG	GTGCACCCCCGC	ATGCAAACC	[56]
	TCRβ	D 1.1	CACAATG	TTACAGCTTTGT	ACAAAAACA	[55]
		D 2.1	CACAATG	TTACACCATGAT	ACAAAAATG	[55]
	TCRγ	J 1	CACAGTG	ATTCAGTCCATA	TCAAAAACT	[57]
		J 2	CACAGTG	ATTCAGTCCATA	TCAAAAACT	[57]
	TCRδ	D 1	CACAATG	AAACACATCAGT	ATAAAAACC	[58]
		D 2	CACAGTG	CTACAGAGCTTT	ACAAAAACT	[58]
HUM	TCRδ	J 2	CACATTA	TGACAGTGCCTC	ACAGGTAAC	[59]
		J 1	CACAGCA	CTTGAGGACGTT	CCAAAAACC	[59]
CHK	IGH	D 1	CACGGTG	CTCCATCCCATA	ACAAAAACC	[60]
		D 1	CACAGTG	ATACAACGTTGA	CCAAAATCC	[60]
		D 2	CACGGTG	CTCCATCCCATA	ACAAAAACC	[60]
		D 2	CACGGTG	ACACGACGTTGA	CCAAAATCC	[60]
		D 3	CACGGTG	ATCCATCCCATA	ACAAAAACC	[60]
CHK	IGH	D 3	CACGGTG	ACACAACGTTGA	CCAAAATCC	[60]
		D 4	CACAATG	CTCCATCCCATA	ACAAAAACC	[60]
		D 4	CACGGTG	ACACAACGTTGA	CCAAAATCC	[60]
		D 5	CACGGTG	CTCCATCCCATA	ACAAAAACC	[60]
		D 5	CACGGTG	ACACAACGTTGA	CCAAAATCC	[60]
		D 6	CACGGTG	CTCCATCCCATA	ACAAAAACC	[60]
		D 6	CACGGTG	ACACAACGTTGA	CCAAAATCC	[60]
		D 7	CACGGTG	CTCCATCCCATA	ACAAAAACC	[60]
		D 7	CACAGTG	ATACAACGTTGA	CCAAAATCC	[60]
		D 8	CACAATG	CTCCATCCCATA	ACAAAAACC	[60]
		D 8	CACGGTG	ACACAACGTTGA	CCAAAATCC	[60]
	IGλ	J	CACAGTG	ATACGGAGCAAT	GCAAAAACC	[61]
RAB	IGH	D 1a	CACGGTG	GGTTGGCCCTTC	ACAAAATCC	[62]
		D 1a	CACAGTG	GTGCA.CCCAGC	ACAAAAACC	[62]
		D 1b	CACGGTG	GGTCGGCTCTTC	ACAAAATCC	[62]
		D 1b	CACAGTG	GTGCA.CCCAGC	ACAAAAACC	[62]
		D 1c	CACGGTG	GGTCGGCCCTTC	ACAAAATCC	[62]
		D 1c	CACAGTG	GTGCA.CCCAGC	ACAAAAACC	[62]
		D 1d	CACGGTG	GGTCGGCCCTTC	ACAAAATCC	[62]
		D 1d	CACAGTG	GTGCA.CCCAGC	ACAAAAACC	[62]
		D 2a	CACCATG	CTGCAGACCAGT	ACAAAATCC	[62]
		D 2a	CACAGTG	CCTCA.GGCCTC	ACATAAAAC	[62]
		D 2b	CACTGTG	TCTCAGACCAGC	ACAAAATCC	[62]
		D 2b	CACAGTG	CCTCA.GGCCTC	ACATAAAAC	[62]
	IGκ	V 20	CACAGTG	ATACAAGCCCTA	ACAAAAACC	[63]
		V 18a	CACAGTG	ATACAAGCCCTT	ACAAAAACC	[64]
		V 18b	CACAGTG	TTAGAAGCCCTA	ACAAAACCA	[64]
		V 19a	CACAGTG	TTCCAAGCCCTA	ACAAAAACC	[64]
		V 19b	CACAGTG	TTCCAAGCCCTA	ACAACTCCC	[64]
HEF	IGH	D 2 1403	CACAGCA	GTTACTGTCAGT	ACAAAAAGT	[65]
		D 2 2807	CACAGCA	GTTACTGTCAAT	ACAAAAAGC	[65]
		D 1 1113	CACAGTG	AGACACACCGTG	TCAAATACT	[65]
		D 1 1113	CACTGTG	ACACGAACCCGC	ACAAATACT	[65]
		D 1 2807	CACAGTG	ACACGAACCAGC	ACAAATACT	[65]
		D 1 1403	CACAGTG	GACTTCAAAGCT	GTACAAATA	[65]
		D 1 1315	CACAGTG	ACACGAACCTGC	ACAAATACT	[65]

Table 3. (cont.)

Species	Locus	Segment	Heptamer	Spacer	Nonamer	Reference
		D 2 2807	CACAGTG	AGACACACCGTG	TCAAATACC	[65]
		D 2 1403	CACAGTG	AGACAAACCGTG	TCAAATACT	[65]
		D 2 1315	CACAGTG	AGACAAACCGTG	TCAAATACT	[65]
		D 2 1315	CACAGCA	GTTACTGTCAAT	ACAAAAACT	[65]
		D 2 1113	CACAGCA	GTTACTGTCAAT	ACAAAAAGT	[65]
	IGL	V 122	CACAGTG	AGACAGGGCAAT	ACAAAAACT	[66]
		V 141	CACAGTG	AGACAGGGCAAT	ACAAAAACT	[66]
XEL	IGκ	V 1	CACAGTG	ATACAGAGCTGA	ACAAAAACC	[67]
		V 2	CACAGTG	ATACAGAGCTGA	ACAAAAACC	[67]
		V 3	CACAGTG	ATACAGAGCTGA	ACAAAAACC	[67]
RAT	IGH	D	CACAGTG	ACTTGTGGCTCA	ACAAAAACC	[68]
		D	CACAGTG	ATGCTTTGCTTA	GTCAAAACC	[68]
	IGλ	J 2	CACAGTG	ACTGAGACTCAA	CCCAAAACC	[69]
BOV	TCRγ	J	CACAGTG	ATTCAAGTCATA	TCAAAAACT	[70]
SHP	IGλ	J	CACAGTG	ACACAGGCTTGC	ACAAAAACC	[71]
DUK	IGλ	J	CACAGTG	ATACAGGGCCAT	GCAAAAACC	[72]

Abbreviations: IG; immunoglobulin. TcR; T cell receptor. H; heavy chain. κ; light chain of the kappa isotype. λ; light chain of the lambda isotype. Isotype classification for Xel and Hef chains is not clear, however they are grouped with whatever light chain isotype has the same sized RSS spacer for the purposes of these tables. α; T cell receptor alpha chain. β; T cell receptor beta chain. γ; T cell receptor gamma chain. δ; T cell receptor delta chain. V; variable gene segment. D; diversity gene segment. J; joining gene segment. Species are Mus; Mouse (*Mus musculus*). Hum; Human (*Homo sapiens*). Xel; Frog (*Xenopus laevis*). Shp; Sheep (*Ovis aries*). Hef; Horned shark (*Heterodontus franciscus*). Rab; Rabbit (*Oryctolagus cuniculus*). Chk; Chicken (*Gallus gallus*). Bov; Cow (*Bos taurus*). Rat; Rat, (*Rattus norvegicus*). Duk; Muscovy duck. Periods in sequences designate a gap inserted for best alignment.

conserved. While this high conservation is derived to some degree from alignment considerations (see Materials and Methods), it is consistent with a previous analysis, which indicated that these positions were both highly conserved and critical for efficient function of RSS [7].

The fifth and sixth positions of the nonamers of 12 RSS are also almost perfectly conserved (Table 5c). The sixth position is required for efficient RSS function, however the fifth position is not [7]. In 23 RSS, only this functionally important sixth position is highly conserved (Table 6c). The nonamer appears to have much more variability in the degree to which individual positions are conserved as in both 12 RSS and 23 RSS the first position, the fourth position, and the ninth positions of the nonamer are relatively poorly conserved. This is particularly true of the fourth position of 23 bp spacer nonamers, where the most frequently observed nucleotide (A) is found in only 56% of the aligned 23 RSS. The functional consequences of consensus substitution at these relatively poorly conserved positions has not been evaluated.

Conservation of sequence in RSS spacers

Analysis of aligned RSS revealed significant conservation of sequence in both 12 and 23 bp spacers (Tables 5b and 6b). Of particular significance is an A located at the fifth position 3′ of the heptamer in both spacers (this position is hereafter referred to as A^5). An A is observed at this position in 67% of all 12 bp spacers and in 64% of all 23 spacers. Both spacers often have a G at this position whenever there is not an A. The plurality rule therefore returns a consensus result for this position that is ambiguous for either purine, A or G. 12 bp spacers and 23 bp spacers have a G at this position in 19% and 25% of spacers, respectively.

The most frequently occurring nucleotide is the same for 12 bp spacers and the heptamer proximal half of 23 bp spacers at several other positions as well. In 12 bp spacers the most frequently observed nucleotides are, from the first base 3′ of the heptamer, A, T, A, C, and A ('A^5'), found in 50%, 56%, 58%, 62%, and 67% of 12 RSS respectively. The next two positions possess a more random distribution of nucleotide composition than the preceding positions. C is the most frequently observed nucleotide at the following two positions, the eighth and ninth positions, in 59% and 68% of 12 RSS, respectively, 23 bp spacers have almost the same pattern, however the extent of conservation is much lower. The most frequently observed nucleotides from the first base 3′ of the heptamer in 23 RSS are A, T, G, C and A (A^5), found in 45%, 44%, 40%, 36%, and 64% of 23 RSS, respectively. At the eighth and ninth positions the most frequently observed nucleotide is again C, found in 44% and 38% of 23 RSS respectively.

The spacers of 12 RSS and 23 RSS therefore maintain significant sequence conservation. Surprisingly, the 12 bp spacer and the first half of the 23 bp spacer possess six positions where the most conserved nucleotide is the same. In 12 bp spacers the most conserved nucleotides 3′ of the heptamer are, from 5′ to 3′, ATACA--CC; the most conserved nucleotides at the analogous positions in 23 bp spacers are ATGCA--CC.

The latter half of 23 bp spacer possess a high frequency of TG and AG dinucleotides, often tandemly repeated, as well as occasional runs of Cs or Gs (4−5 bp long). This results in a number of positions (the 14th, 16th, and 19th through to the 22nd positions) where the consensus results are ambiguous for two nucleotides.

DISCUSSION

In this report, we have used the considerable increase in size of the available database of RSS to redefine the RSS consensus, particularly with respect to spacer sequences. We found that: 1) The consensus heptamer and nonamer was the same for both the types of RSS (the 12 bp spacer RSS and the 23 bp spacer RSS); 2) There is a significant sequence conservation in both the 12

Table 4. Alignment of 23 bp spacer Recombination Signal Sequences

Species	Locus	Segment	Heptamer	Spacer	Nonamer	Reference
MUS	IGH	V AR100	CACAGTG	TTCTAA.CCACATCCTGAGTGTGT	.CAGAAACC	[73]
		V H16	CACAGTG	GTGCAA.CCACATCCCGACTGTGT	.CACAAACC	[74]
		V H124	CACAGTG	TTGTAA.CCACATTCTGAGAGTGT	.TAGAAACC	[75]
		V PJ14	CACAGTG	AGGGAAGTCCAATGTGAGCCT.GC	ACAAATACC	[76]
		V 108A	CACAGTG	TTACAA.ACACATCCTGAGTGTGT	.CAGAAACC	[77]
		V 108B	CACAGCG	TTGTAA.CCACAGGCTGAGTGTGT	.CAGAAACC	[77]
		V H441	CACAGTG	AGGAAATCTCAGTTTGTACCCA.G	ACATGAACC	[78]
		V H4A-3	CACAGTG	TTGCAA.CCACATCCTGAGTGTGT	.CAGAAACC	[79]
		V H 30	CACAGTG	GTGCAA.CCACATCCCGACTGTGT	.CACAAACC	[74]
		V Hid11	CACAGTG	TTTTAA.CCACATCCTGAGTGTGT	ACAGAAACC	[80]
		V H101	CACAGTG	AGGGAAGTCCATTGTGAACTT.GA	ACAAAAATT	[81]
		V A1/A4	CACAGTG	TTGTAA.CCACATCCTGAGTGTGT	.CAGAAACC	[82]
		V H104A	CACAGTG	TTGTAA.CCACATCCTGAGTGTGT	.CAGAAACC	[83]
		V H10	CACAGTG	TTGCAA.CCACATCCTGAGCGTGT	.CAGAAACC	[79]
		V 1	CACAGTG	AGAGGACGTCATTGTGAGCCCA.G	ACACAAACC	[5]
		V 13	CACAGTG	AGGGTACTTCAGTGTGAGCCTA.G	ACACAAACC	[84]
		V 11	CACAGTG	AGGGTACTTCAGTGTGAGCCTA.G	ACACAAACC	[84]
		V H2B-3	CACAGTG	TTGCAA.CCACATCCTGAGAGTGT	.CAGAAACC	[79]
		V 186-1	CACAGTG	TTGCAA.CCACATCCTGAGAGTGT	.CAGAAAAC	[85]
		V 186-2	CACAGTG	TTGCAA.CCACATCCTGAGAGTGT	.CAGAAAAC	[85]
		V 145	CACAGTG	TTGCAA.CCACATCCTGAGAGTGT	.CAGAAACC	[85]
		V 23	CACAGTG	TTGCAA.CCACATCCTGAGAGTGT	.CAGAAAAC	[85]
		V 6	CACAGTG	TTGCAA.CCACATCCTGAGAGTGT	.CAGAAAAC	[85]
		V 3	CACAGCG	TTGTAA.CCACATCCTGAGAGTGT	.CAGAAACC	[85]
		V H102	CACAGTG	TTGTAA.CCACATCCTGAGAGTGT	.CAGAAACC	[85]
		V 81X	CACAATG	AGCAAAAGTTACTGTGAGCTCA.A	ACTAAAACC	[86]
		V 283	CACAGTG	AGTGAATGTTACTGTGAGCTCA.A	ACTAAAACC	[87]
		V 5A	CACAGTG	AGGGGAGGTCAGTGTGAGCCCA.G	ACACAAACC	[74]
		V RV10	CACAGTG	AGGGGCCCTCAGGC.GAGTCCT.G	ACACAAACC	[74]
		V 105	CACAGTG	TTGTAA.CCACATCCTGAGTGTGT	.CAGAAACC	[83]
		V H26-6	CACAGTG	TTGCAA.CCACATCCTGAGTGTGT	.CAGAAATC	[74]
		V DFL.1	CACAGTG	TTGCAA.CCACATCCTGAGAGTGT	.CAAAAATA	[88]
	IGH	J 4	CACAATA	GTGGGTTTTTCCTCTGTACCC..G	ACAAAAACC	[76]
		J 3	CACATTG	TGACAACAATGATTAGACCCCTGA	CAATAAATG	[76]
		J 2	CACACTA	TCATAGACCCCTTTAGTGGGTG.T	ACAAAAACC	[76]
		J 1	CACAGT.	CTCTGTTCTGCCTCTGTTCCTA.T	ACTAAAACT	[76]
	IGκ	J 5	CACAGTG	AGGACTATGACA.TGCCCCTCTCT	ACAAAAACC	[2]
		J 4	CACAGTG	ATTCATATCACTGCGCCCCCCTTT	ACAAAAACC	[2]
		J 2	CACACTG	GTGTCCCTTCAC.TCAACCCCCAT	ACAAAAACT	[2]
		J 1	CACAGTG	GTAGTACTCCAC.TGTCTGGCTGT	ACAAAAACC	[2]
	IGλ	V 1	CACAATG	ACATGTGTAGATGGGGAAGTAG.A	ACAAGAACA	[89]
		V 2	CACAATG	ACATGTGTAGATGGGGAAGTAG.A	ACAAGAACA	[90]
		Vx	CACAGTA	ACGGAGATAAAGGAGGAAGCAG.G	ACAGAAACT	[91]
	TCRα	V 5H	CACAGTG	..TCCCAGACAC.CTGCAGCCTGT	ATGTAAACC	[32]
		V 1-8.2	CACAGTG	..TCCCCAGGAC.CTGCAGCCTGC	ACCTAAACC	[92]
		V 1-8.1	CACAGTG	CTCTCCAGGCAC.CTGCAGGCTGC	ACCCAAACC	[92]
		V 2C	CACAGTG	TGTGGGGCTGCAGGGGGAGCTG.A	ACACAAACA	[35]
		V F3.2	CACAGTG	AGGGAGACTGCAGGGGAAGCTG.C	ACATGAACC	[93]
		V F3.3	CACAGTG	AGGGAGACTGCAGGGGAAGCTG.C	ACATGAACC	[93]
		V F3.4	CACAGTG	AGGGAGACTGCAGGGGAAGCTG.C	ACATGAACC	[93]
	TCRα	V F3.5	CACAGTG	AGGGAGACTGCAGGGGAAGCTG.C	ACATGAACC	[93]
		V F3.6	CACAGTG	AGGGAGACTGCAGGGGAAGCTG.C	ACATGAACC	[93]
	TCRβ	V 8.3	CACAGTG	ATGTGTGG.CTTCCTTCCCTTTGC	ACAGAAAGT	[94]
		V 8.2	CACAGTG	ATGTGGGG.TTTCCTCCCCTCTGC	ACAGAAAGG	[94]
		V 8.1	CACAGTG	ATGTGTGG.CTTCCTTCACTCTGC	ACAGAAAGG	[94]
		V 18	CACAGTG	CTGG..TTCAAGGGAGAAATCTCA	GCGAGAACT	[95]
		V 19	CACAGTG	GTGACTACT..GGCTTTTCTCAGA	ACACAAACT	[95]
		V 10-8	CACAGTG	GTGCAGAGTCA.CTGTTTCCCTGT	GCACAAACC	[92]
		V 5.1	CACAGCC	TTACAGAGCTACTGGCTTTCTGTA	ACTTAATC.	[94]
		V 5.2	CACAGCC	TTACAAAGCTACTGGCTTTCTGTA	ACTTAATC.	[94]
		D 2	CACAATG	ATTCAACT.GGAAGAGGTGCTTTT	ACAAAAAGC	[38]
		D 1	CACGGTG	ATTCAATT.CTATGGGAAGCCTTT	ACAAAAACC	[38]
	TCRγ	V 108A	CACAACA	TTAGAGCCTCTAGACT.AGCCTGC	ATAAGAACC	[39]
		V 108B	CACAACA	TTAGAGCCTCTAGACT.AGCCTGC	ATAAGAACC	[39]
		V 4	CACTCTA	TCAAGAT.ACTGCACTGTTAACAA	ACAAAACCC	[96]
	TCRδ	D 1	CACAGGT	TGAAGTAT.ATTAAACCTCTGTTC	AGAAACACT	[40]
		D 2	CACAGTG	TTGCAAAC.CCCATAGGGACCTGT	ACAAAAACT	[40]
HUM	IGH	V 251	CACAGTG	AGAGAAACCAGCCCCGAGCCC.GT	CTAAAACCC	[97]
		V 12G-1	CACAGTG	AGGGGAGGTGAGTGTGAGCCCA.G	ACACAAACC	[98]
		V 2-1	CACAGTG	AGGGGAGGTGAGTGTGAGCCCA.G	ACACAAACC	[98]
		V 79	CACAGTG	AGGGGAGGTGAGTGTGAGCCCA.G	ACACAAACC	[98]
		V 7-2	CACAGTG	TGAAAACCCACATCCTGAGACCGT	.CAGAAACC	[99]
		V 35	CACAGTG	TGAAAACCCACATCCTGAGGGTGT	.CAGAAACC	[100]
		V 71-4	CACAGTG	AGGGGAGGTGAGTGTGAGCCCA.G	ACAAAAACC	[101]
		V 58	CACAGTG	AGGGGAGGTGAGTGTGAGCCCA.G	ACAAAAACC	[98]
		V 71-2	CACAGTG	AGGGGAGGTGAGTGTGAGCCCAGG	ACACAAACC	[101]
		V H26	CACAGTG	AG.GGAAGTCATTGTGAGCCCA.G	ACACAAACC	[102]
		V H52	CACAGTG	AG.GGAAGTCAGTGTGAGCCCA.G	ACACAAACC	[102]
		V H11	CACAGTG	AG.GGAAGTCAATGTGAGCCCA.G	ACACAAACC	[103]
		V H105	CACAGTG	AGGGGAGGTCAGTGTGCGCCCA.G	ACACAAACC	[74]

Table 4. (cont.)

Species	Locus	Segment	Heptamer	Spacer	Nonamer	Reference
		V 6-1G1	CACAGTG	AGGGGAAGTCAGTGTGAGCCCA.G	ACACAAACC	[99]
		V 1.9II	CACAGTG	AGGGGAGGTGAGTGTGAGCCCA.G	ACACAAACC	[99]
		V 1.9III	CACAGTG	AGGGGAAGTCATTGTGCGCCCA.G	ACACAAACC	[99]
		V 9-1	CACAGTG	AGGGGAGGTCAGTGTGAGCCCG.G	ACACAAACC	[99]
		V 12-2	CACAGCG	AGGGGAGGTCAGTGTGAGCCCG.G	ACACAAACC	[99]
		V 13-2	CACAGTG	AGGGGAAGTCAGTATGAGCCCA.G	ACACAAACC	[99]
		V 8-1B	CACAGTG	AGGGGAGGCCATTGTGCGCCCA.G	ACACAAACC	[99]
		V 15-2B	CAGAGTG	AGGGGAAGTCAGTGAGAGCCCAGG	.CACAAACC	[99]
		V 22-2B	CACAGTG	AGGGGAAGTCAGTGTGAGCCCA.G	ACACAAACC	[99]
		V HG3	CACAGTG	TGAGAAACCACATCCTCAGA.TGT	.CAGAAACC	[104]
		V 21-2	CACAGTG	TGAGAAACCACATCCTCAGAGTGT	.CAGAAACC	[99]
		V 3-1	CACAGTG	TGAGAAACCACATCCTCAGAGTGT	.CAGAAACC	[99]
	IGH	J 6	CACAATG	GCAGAATGTCCATCCTCACCC..C	ACAAAAACC	[41]
		J 5	CACATTG	TGACAACAATG.CCAGACCCCGAC	AAAGAACCG	[41]
		J 4	CACATTG	TGGGAGGCCCCATTAAGGGGTG.C	ACAAAAACC	[41]
		J 3	CACAGGG	ACACAGTCCGTTCCTAGACCCA.G	ACACAAACC	[41]
		J 2	CACAGTC	CTCTGCCCTCCTGCTTCTCCCA.T	ACAAAAACA	[41]
	IGκ	J 5	CACAGTG	TTAACT.TAATTACTTTCCCCTTA	ACAAAAATC	[105]
		J 4	CACAGTG	AGGGATCTCACC.CTTTCCCCTCA	ACAAAAACC	[105]
		J 3	CACAGTG	ATTCGTCTTAA..CTTTTCCCTTT	ACAAAAACC	[105]
		J 2	CACAATG	GTTCCTCTTAAC.TTCCCTCCTAT	ACAAAAACT	[105]
HUM	IGκ	J 1	CACAGTG	AGAGCTCTCCAT.TGTCTTGCTGA	ACAGAAACC	[105]
	IGλ	V 3S1	CACAGTG	ACACAGACAGATGGGGAAGTGA.G	ACAGAAACC	[106]
		V 7.1	CACAGTG	ACAGACTCATAAGAGGAACCAA.G	ACATAAACC	[107]
		V 117	CACAGTG	CTCCAGCCCAATGGGGAACTGA.G	ACAAGAACC	[108]
		V 119	CACAGTG	CTCCAGCCCAATGGGGAACTGA.G	ACAAGAACC	[108]
		V 3S2	CACAGTG	CTCAGGCCGGGGTGGGAACTGA.G	ACAAGAACC	[109]
		V 2.1	CACAGTG	GTCCAAGTTCATGGGGAACTGA.G	ACCAAAACC	[110]
		V 418	CACAGTG	ACACAGACAGATGGGGAAGTGA.G	ACAGAAACC	[111]
		V 318	CACGGTG	ACACAGGCAGATGAGGAAGTGA.G	ACAAAAACA	[111]
		V 1.1	CACAGTG	CTCCAGGCCAATGGGGAACTGA.G	ACAAGAACC	[107]
	TCRα	V pY14.2	CACAGTG	CCTGAGACTGCAGGAG.AGCTG.A	ACACAAACC	[52]
		V 13.1	CACAGTG	CTCCCCAGGCAC.CTGAAGCCTGT	ACCCAAACC	[59]
	TCRβ	V 16	CACAGTG	CTTCACAGTCGTGC.CCTTGCTGT	GCAAAACCA	[112]
		V 8.1	CACAGCG	CTGCAGAATCA.CCCCTTTCCTGT	GCAGAAAAC	[113]
		V 8.2	CACAGCG	CTGCAGAATCA.CCCCTTTCCTGT	GCAGAAACC	[113]
		V 8.3	CACAGCG	CTGCAGAATCA.CCTGCTCCCTGT	GCAGAAACC	[113]
		V M3-2	CACAGCG	CTGCAGAATCA.CCCCTTTCCTGT	GCAGAAACC	[54]
		V MT1-1	CACAGCG	CCAGGAGGGGA.TCAGACACCGCG	GCAAGAACC	[114]
		VATL12-2	CACAGCA	TGGCACAGTTG.CCTCCTTCCTGT	TCACAAACC	[114]
		V ATL 2-1	CACAGTG	CTTCTTGG.CCACCTGCTCTCTAC	ACAGAAAGA	[114]
	TCRβ	D 2.1	CACGATG	ATTCAGGT.AGAGGAGGTGCTTTT	ACAAAAACC	[55]
		D 1.1	CACAATG	ATTCAACT.CTACGGGAAACCTTT	ACAAAAACC	[55]
	TCRγ	V 2	CACAGTG	ATTCAGATCCGCCCTACACCACAC	TGAAAATC.	[115]
		V 3	CACAGTG	ATTCAGACCTGTCCTACACCACAC	TGAAAATC.	[115]
		V 8	CACAGTG	ATTCAGACCTGTGCTACACCACAC	TGAAAATC.	[115]
		V 9	CACAGCA	GCAGACAGTTTGAGCCATCCCATT	TCAATAAAA	[116]
		V 10	CACATAC	TAGAA.CTGTTGAAACAACATGC.	ACAAAATCC	[116]
	TCRδ	V DS6	CACAGTG	ACAGAACTGTCGGAGGGAGGTG.T	ACAAAAGCC	[117]
		V 1	CACAGTG	TTTGAAGTGATAGTAAAAGCAA.A	ACAAAAACC	[59]
	TCRδ	D 1	CACACAG	GTTGGAGT.GCATTAAGCCTTTGT	CCAAAAACA	[58]
		D 2	CACAGTG	CTACAAAA.CCTACAGAGACCTGT	ACAAAAACT	[58]
XEL	IGH	V LL3.1	CACAGTG	GGACATAT.ATTGTGAAAACATGT	ATAAAAACA	[118]
		V LL3.4	CACAGTG	GGAAATAT.ATTATGAAAACATGT	ATAAAAACA	[118]
		V LL3.5	CACAGTG	GGACAAAT.ATTAAGAAAGCCTGT	GTAAAAACA	[118]
		V LG2.1	CACAGTG	ACAGAAGAGAATGAGGAAGTCA.G	ACAATAACT	[118]
		V LG2.2	CACAGTG	ACAGAAGAGAATGAGGAACTCA.G	ACAATATCT	[118]
		V LG2.4	CACAGTG	ACTAAATATACTGAGGAAGTGA.G	ACAATAACA	[118]
		V LG2.7	CACAGTG	ACAGAAAAAAATAAGGAGGTCA.G	ACAATATCA	[118]
		V LG2.8	CACAGTG	ACAGAACAAAATAAGGAAGTCA.G	ACAATATCA	[118]
		V LL1.1	CACAGTG	ACAAATAGTCTCAGAGCAGTGC.A	ACAAAAACA	[118]
		V LL1.2	CACAGTG	ACAAATAGTCTCAGAGCAGTGC.A	GCAAAAACA	[118]
		V LL1.3	CACAGTG	ACAAATAGTCTCAGAGCAGTGC.A	ACAAAAACA	[118]
		V LL1.4	CACAGTG	ACAAAGAATCCCAGAGTCATGT.A	GAAAATACA	[118]
		V LL1.6	CACAGTG	ACAAAGAAACACAGAGCAGTGC.A	ACAAAAACA	[118]
		V LL1.7	CACACTG	ACAAATAGTCTCAGAGCAGTGC.A	ACAAAAACA	[118]
		V LL1.8	CACAGTG	ACAAATAGTCTCAGAGCAGTGC.A	GCAAAAACA	[118]
		V LL1.9	CACAGTG	ACAAATAGTCTCAGAGCAGTGC.A	ACAAAAACA	[118]
SHP	IGλ	V 6.2	CACACTG	GTTCAAGTTCATGGGGAAGTGA.C	GCCAAAACC	[71]
		V 12.2	CACAGTG	CTCCAGGCCAGGGGGGAAGCGA.C	ACAAAAACC	[71]
		V 4.2	CACGGTG	CTCCAGGCCAGGGGGGAAGTGA.C	GCGAAACCC	[71]
		V 5.2	CACGGTG	CTCCAGGCCAGGGGGGAAGTGA.C	ACCAAAGCC	[71]
		V 17	CACGGTG	CTCCAGGCCAGGGGGGAAGTGA.C	ACCAAAGCC	[71]
		V 10	CACGGTG	CTCCAGGCCAGGTGGGAAGTGA.C	ACCAAAGCC	[71]
		V 9	CACGGTG	CTCCAGGCCAGGTGGGAAGTGA.C	ACCAAAGCC	[71]
		V 18	CACAGTG	CTCCAGGTCAAGGGGGAAGTGA.C	ACAAAAACC	[71]
		V 26.3	CACAGTG	CTCCAGGCCAAGGGGGAAGTGA.C	ACAAAAACC	[71]
		V 3	CACAGTG	CTCCAGGCCAGGGGGGAAGTGA.C	ACAAAAACC	[71]
		V 4.1	CACAGTG	CTCCAGGCCAGCGGGGAAGTGA.C	ACAAAAACC	[71]
		V 16.1	CACAGTG	CTCCAGGCCAGGGGGGAAGTGA.C	ACAAAAACC	[71]

Table 4. (cont.)

Species	Locus	Segment	Heptamer	Spacer	Nonamer	Reference
SHP	IGλ	V 26.1	CACAGTG	CTCCAGGCCAGGGGGGAAGCGA.C	ACAAAAACC	[71]
		V 5.1	CACAGTG	CTCCAGGCCAGGGGGGAAGTGA.C	ACAAAAACC	[71]
HEF	IGH	V 1113	CACTGCC	ACCCAAGCAAATCCTGGGCTCG.T	ACAGAAACA	[65]
		V 2807	CACAATG	AGAGGAACCAGGGCTGGACCC.GT	ACAAGAACA	[65]
		V 1403	CACAGCG	AGAGGAACCAGGGCTGGACCC.GT	ACAAGAACA	[65]
		V 1315	CACAACG	AGAGGAACCAGGGCTGGACAT.GT	ACAATAACA	[65]
		D 2 1113	CACGGTA	CTGTACAGAGCGAGTTT.CTTA.T	ACAAAAACC	[65]
		D 1 2807	CACGGTG	CTGTACAGAACGAGTTC.CTCA.T	ACAAAAACC	[65]
		D 1 1403	CACGGTG	CTGTACAGAGCGAGTTC.CTCA.T	ACAAAAACC	[65]
		D 1 1315	CACGGTG	CTGTACAGAGCGAGATC.TTCA.T	ACAAAAACC	[65]
		J 1315	CACAGTG	TTACATTCCCTGGGCTGGGTCA.C	ACAATAACC	[65]
		J 1403	CACAGTG	TTACATTCCCCGGGCTGGTTCA.G	ACAATAACC	[65]
		J 2807	CACAGTG	TTACATTCCCTGGGCTGGGTCA.G	ACAATAACT	[65]
		J 1113	CACAGTG	TTACATTCCCTGGGCTGGGTCA.C	ACAATAACC	[65]
	IGL	V 122	CACAGTG	CAGTGTTTTAAATGGGACGGGTCA	CTTAAAACC	[66]
		V 141	CACAGTG	CAGTGTTTTAAATGGGACGGGTCA	CTTAAAACT	[66]
RAB	IGH	VH1 (a3)	CACAGTG	AGGGGCCCTCAGGCTGAGCCCA.G	ACACAAACC	[119]
		VH3 (a3)	CACAGTG	AGGGTCCCTCAGGCTGAGCCCA.G	ACACAAACC	[119]
		VH4 (a3)	CACAGTG	AGGGGCCCTAGGGCTGAACCCA.G	ACACAAACC	[119]
		VH6 (a3)	CACAGTG	AGGTG.CCTCAGGCTGAGCCCA.G	ACACAAACC	[119]
		V 832	CACAGTG	AGGGGCCCTAGGGCGCA..CCTAG	ACACAAACA	[120]
		J 2	CACAGGG	GCACA.TCCCCTGTTGCTGCCCAG	ACACAAACC	[121]
		J 3	CACTGTG	ACGACCGTGCCAGGACCCCCGGCA	AGAACCGGT	[121]
		J 4	CACATTG	CTGTAGACACCTT..AGGGGGCGT	GCAAAAACC	[121]
		J 5	CACATTG	TGATGACCGTGCCAGGACCCCA.G	GCAAGAACC	[121]
	IGκ	J 2	CACAGTG	GTTCCTCCTAAC.CTCCCTCCTGT	ACAAAAACT	[122]
CHK	IGH	V	CACGGTG	ACACCGATCCCCAGCACGGTGG.C	ACAAAACCC	[60]
		J	CACAATG	CCCCAAAATCCGCCTTTTTTCA.C	CCAAAAACT	[60]
	IGλ	V	CACGGTG	ACACAAAGCAATGGGGAAATGA.T	ACAAAAACC	[61]
RAT	IGH	J	CACAGT.	CTCTGTTCTGCCACTGTTCCT.GT	ACTAAAACT	[68]
	IGκ	J 1	CACAGTG	GTAGTTCTCCAT.TGTCTGGCTGT	ACAAAAACC	[123]
		J 2	CACACTG	GTATCCCTTGACTCACCACCGA.T	ACAAAAACT	[123]
		J 2a	CACACTG	GTTTCCCTTGACTCACCCCCCA.T	ACAAAAACT	[123]
		J 3	CACAGTG	ATTCATGTCAAAGC.CCCCC.TTT	ACAAAAACC	[123]
		J 4	CACAGTG	AAGACTC.TGACATATGCACCTCT	ACAAAAACC	[123]
	IGλ	V	CACAATG	GCATGT.CA.GATGAGGAAGTAGG	ACAAAAACC	[69]
DUK	IGλ	V L5	CACAGTG	ACACAGAGC.AATGGGGAAGTGAT	ACAAAAACC	[72]
		V L1	CACAGTG	ACACAAAGC.AATGGGGAAGTGAT	ACAAAAACC	[72]

Abbreviations: IG; immunoglobulin. TcR; T cell receptor. H; heavy chain. κ; light chain of the kappa isotype. λ; light chain of the lambda isotype. Isotype classification for Xel and Hef chains is not clear, however they are grouped with whatever light chain isotype has the same sized RSS spacer for the purposes of these tables. α; T cell receptor alpha chain. β; T cell receptor beta chain. γ; T cell receptor gamma chain. δ; T cell receptor delta chain. V; variable gene segment. D; diversity gene segment. J; joining gene segment. Species are Mus; Mouse (*Mus musculus*). Hum; Human (*Homo sapiens*). Xel; Frog (*Xenopus laevis*). Shp; Sheep (*Ovis aries*). Hef; Horned shark (*Heterodontus franciscus*). Rab; Rabbit (*Oryctolagus cuniculus*). Chk; Chicken (*Gallus gallus*). Bov; Cow (*Bos taurus*). Rat; Rat, (*Rattus norvegicus*). Duk; Muscovy duck. Periods in sequences designate a gap inserted for best alignment.

and the 23 bp spacers; and, 3) The conserved sequence in the 12 bp spacer is similar to the conserved sequence in the heptamer proximal half of the 23 bp spacer.

Upon separating the RSS on the basis of whether they were derived from 12 or 23 RSS, we determined that the heptamers and nonamers appear equivalent, irrespective of which type of RSS they are derived from. They possess the same consensus profile, in that equivalent positions are conserved to an equivalent degree, regardless of classification by size of spacer. For example, the first, fourth, and ninth positions of the nonamer are relatively poorly conserved in both 12 and 23 RSS. This observation provides further support for the hypothesis that the heptamer and nonamer function in an identical manner (e.g., serving as recognition sites for the same protein) for both 12 RSS and 23 RSS.

Previous examinations of the RSS consensus have concentrated on the heptamer and nonamer, relying on early studies that suggested that the intervening sequence is truly a spacer, conserved in length (12 of 23 bp), but not sequence (reviewed in [11]). The inclusion of spacer sequences in our analysis of RSS has yielded conservation that has not previously been observed.

To best assess the relevance of sequence conservation, we made our consensus determination using the plurality rule. The plurality rule returns a result for all positions analyzed in an aligned set of sequences. The result, however, may have ambiguity varying from one (highly conserved) to all four (indistinguishable from random) possible nucleotides [9] There is less than a 1% chance of randomly getting a plurality rule result with ambiguity for less than all four nucleotides at any one position, in the databases observed here [10]. It is with some surprise, therefore, that most positions in 12 and 23 bp spacers demonstrate consensus results with ambiguity for less than all four nucleotides.

Some of the sequence conservation observed in this manner may be due to the inclusion of many members of a gene segment family that has been expanded only recently in evolution. The fact that the conserved sequence motif (heptamer proximal) in 12 bp spacers is similar to the conserved sequence motif in the

Table 5. Consensus sequences for 12 RSS

a

Position[2]	1	2	3	4	5	6	7
Consensus[1]	C	A	C	A	G	T	G
%G	0	0	0	9	88	0	77
%A	0	100	0	81	5	2	13
%T	0	0	0	9	2	86	2
%C	100	0	100	1	4	12	8

b

Position[3]	1	2	3	4	5	6	7	8	9	10	11	12
Consensus[1]	A/T/C/G	T/C	A/T	C/T	A/G	G/A/T/C	C/A/G	C/G	C/T	T/C/A/G	T/G/C/A	A/T/C/G
%G	16	13	10	13	19	38	26	25	2	16	26	13
%A	50	5	58	10	67	25	33	6	8	20	19	43
%T	18	56	20	14	7	21	7	10	22	36	35	27
%C	16	26	13	62	7	15	34	59	68	28	20	16
%gap	0	0	0	0	0	2	0	0	0	0	0	0

c

Position[4]	1	2	3	4	5	6	7	8	9
Consensus[1]	A	C	A	A	A	A	A	C	C
%G	12	2	3	8	1	0	0	7	5
%A	68	2	86	76	95	95	87	6	9
%T	12	5	2	8	2	5	11	10	14
%C	8	91	9	8	2	0	2	77	71
%gap	0	0	0	0	0	0	0	0	2

a. Consensus[1] and nucleotide frequencies in 12 RSS heptamers
b. Consensus and nucleotide frequencies in 12 RSS spacers
c. Consensus and nucleotide frequencies in 12 RSS nonamers
[1]Consensus as determined by the plurality rule (see text and reference [9]). We further define consensus results ambiguous for more than one nucleotide by reporting the nucleotides in order of the frequency that they are observed, from the most frequent to the least frequent.
[2]bases numbered beginning at the first base of the heptamer
[3]bases numbered beginning at the first base 3′ of the last base of the heptamer
[4]bases numbered beginning at the first base 3′ of the last base of the spacer
A period in place of a nucleotide code represents a gap

Table 6. Consensus sequences for 23 RSS

a

Position[2]	1	2	3	4	5	6	7
Consensus[1]	C	A	C	A	G	T	G
%G	0	0	1	7	85	2	91
%A	0	100	0	91	8	1	5
%T	0	0	0	2	3	89	1
%C	100	0	99	0	4	9	3

b

Position	1	2	3	4	5	6	7	8	9	10	11
Consensus	A/T/C/G	T/G/C	G/A/C	C/G/T	A/G	A/G/T/C	A/G/C/T	C/G/T/A	C/T/A	A/C/G/T/.	A/C/T/G
%G	11	29	40	31	25	25	27	26	7	16	15
%A	45	2	31	11	64	45	39	9	11	36	42
%T	23	44	13	15	3	19	11	21	38	7	13
%C	20	18	14	36	8	11	21	44	38	34	32
%gap	1	8	2	8	0	1	2	1	5	7	1

Table 6b cont.

Position	12	13	14	15	16	17	18	19	20	21	22	23	24
Consensus	G/A/T/C	T/G/A/C	G/C	T/G/C/A	G/T	A/G/C/T	A/G/T/C	G/C	C/T	C/G	A/T	./G	T/G/C/A
%G	29	32	45	29	57	24	21	44	8	33	12	34	27
%A	29	13	9	15	7	43	53	7	9	6	38	6	14
%T	21	38	10	32	25	9	13	8	35	14	37	6	39
%C	16	10	37	22	11	23	11	42	48	47	8	5	20
%gap	5	7	1	2	0	2	3	1	0	0	5	50	1

c

Position	1	2	3	4	5	6	7	8	9
Consensus	A	C	A	A/G/C	A	A	A	C	C
%G	8	3	2	19	3	0	3	3	3
%A	73	2	90	56	89	98	89	3	15
%T	3	5	4	6	5	1	5	3	11
%C	3	91	5	19	3	1	3	91	71
%gap	14	0	0	0	0	0	0	0	3

a. Consensus[1] and nucleotide frequencies in 23 RSS heptamers
b. Consensus[1] and nucleotide frequencies in 23 RSS spacers
c. Consensus[1] and nucleotide frequencies in 23 RSS nonamers
[1]Consensus as determined by the plurality rule (see text and reference [9]). We further define consensus results ambiguous for more than one nucleotide by reporting the nucleotides in order of the frequency that they are observed, from the most frequent to the least frequent.
[2]bases numbered beginning at the first base of the heptamer
[3]bases numbered beginning at the first base 3′ of the last base of the heptamer
[4]bases numbered beginning at the first base 3′ of the last base of the spacer
A period in place of a nucleotide code represents a gap

first half of 23 bp spacers argues that the origin of this motif is distinct from a recent expansion of gene segment families, however.

The conserved sequence common to both RSS spacers could be derived from two possible sources. Firstly, both 12 and 23 RSS may have a common ancestral origin. For example, early RSS may have all possessed 12 bp spacers. A requirement for the directed joining of one type of segment (e.g. a V segment) to a second type of segment (e.g. a J segment) might have resulted in an adaptation of this early version of the V(D)J rearrangement machinery to include a 12/23 rule, and an accompanying change of the spacer length of one type to 23 bp.

A second, more likely possibility is that this sequence is conserved because it contributes to RSS function. In support of this hypothesis, we found, using extra chromosomal recombination substrates, that a single substitution of the most conserved position (replacement of the conserved A at the 5th position of a 12 bp spacer with a C) resulted in a significant, though modest (approx. 15%) drop in the frequency with which the substituted RSS mediated recombination (unpublished results). This observation appears to contradict a previous report from Lieber and colleagues, where the authors concluded that complete replacement of a spacer with GC base pairs did not appear to influence the frequency with which the substituted RSS mediated recombination [6]. In the study by Lieber and colleagues, the substituted and unsubstituted RSS were tested in separate substrates, rather than in a competitive substrate as was used in our experiment, and thus subtle differences in recombination frequency may have been less readily observable. We note, however, that the Lieber report does demonstrate a significant loss of recombination frequency ($p < 0.05$, using a two tailed Mann–Whitney test) when comparing a substrate where both the 12 and 23 bp spacers were substituted with GC base pairs to a substrate with unsubstituted RSS spacers [6]. Thus the available data are consistent with the conclusion that differences in RSS spacer sequence contribute to minor differences in the efficiency with which the RSS mediates recombination.

We suggest two possible methods by which the conserved sequence in RSS spacers could contribute to RSS function. One possibility is that this sequence may represent a functional extension of the heptamer. Thus while the element of recombinase that recognizes RSS likely makes critical contacts with the highly conserved heptamer, contact with the RSS may extend into the spacer. This possibility is supported by the fact that conservation in both the 12 and 23 bp spacers peak at the fourth and fifth positions 3′ of the heptamer, which is located approximately one turn of a B DNA helix from the critical first three nucleotides of the heptamer. Moreover, in 12 bp spacers this position is also one turn of a B DNA helix from the beginning of the nonamer. Thus the sequence recognition component of recombinase could lie along one face of the 12 RSS DNA helix, making sequence

specific contacts at the heptamer, the fourth and fifth positions of the spacer, as well as the nonamer.

Alternatively, the observed conserved sequence may induce functionally important structural changes in RSS DNA. In 12 bp spacers the most conserved positions are C and A, four and five bp 3′ of the heptamer, respectively. Polymeric CA sequences have been linked with sequences active in recombination and transcription [12]. This has been attributed to the fact that CA tracts cause unusual perturbations in DNA structure, including the de-stacking of bases and the formation of non-Watson–Crick base pairs [13, 14], as well as a reduced electrophoretic mobility associated with helical kinking [15]. It is unknown if the structural alterations described above would necessarily be associated with a single CA dinucleotide, in the context of the 12 bp spacer. It is worth noting, however, that 23 bp spacers are generally rich in CA, as well as the complementary dinucleotide, TG (data not shown).

We have demonstrated here that, contrary to previous analysis, the RSS spacer does possess significant conservation of sequence. The degree of conservation, as well as experiments using recombination substrates, suggests that, though significant, conserved sequences in RSS spacers are not as critical to RSS function as the heptamer and nonamer motifs. As has been demonstrated with kappa and lambda RSS, however, multiple, 'non-critical' substitutions can still result in a dramatic reduction in recombination frequency [16] Thus differences in the sequence of RSS spacers may also make a significant contribution to the frequencies with which endogenous gene segments rearrange. The possibility that portions of the RSS spacer could aid in RSS function, through direct sequence specific contacts or through DNA structural effects, warrants continued investigation with extra chromosomal constructs.

ACKNOWLEDGEMENTS

We would like to thank Helen Garrioch for assistance in completing the recombination assay. We would also like to thank Dr Marc Shulman and Matthew Binney for critical reading of the manuscript. This work was supported by grants from the National Cancer Institute of Canada and the Medical Research Council of Canada. G.E.Wu is a Medical Research Council of Canada Scholar, and D.A.Ramsden is the recipient of a National Cancer Institute of Canada Steve Fonyo Studentship.

REFERENCES

1. Sakano, H., K. Huppi, G. Heinrich and S. Tonegawa. (1979) *Nature*, 280, 288–294.
2. Max, E. E., J. G. Seidman and P. Leder. (1979) *Proc. Natl. Acad. Sci. USA*, 76, 3450–3454.
3. Akira, S., K. Okazaki and H. Sakano. (1987) *Science*, 238, 1134–1138.
4. Hesse, J. E., M. R. Lieber, M. Gellert and K. Mizuuchi. (1987) *Cell*, 49, 775–783.
5. Early, P., H. Huang, M. Davis, Calame and L. E. Hood. (1980) *Cell*, 19, 981–992.
6. Wei, Z. and M. R. Lieber. (1993) *J. Biol. Chem.*, 268, 3180–3183.
7. Hesse, J. E., M. R. Lieber, K. Mizuuchi and M. Gellert. (1989) *Genes Dev.*, 3, 1053–1061.
8. Devereux, Haeberli and Smithies. (1984) *Nucleic Acids Research*, 12(1), 387–395.
9. McMorris, F. R. and W. H. E. Day. (1992) *Nucleic Acids Res.*, 20, 1093–1099.
10. Day, W. H. E. and F. R. McMorris. (1992) *Math. Biosci.*, 111, 231–247.
11. Gellert, M. (1992) *Annu. Rev. Genet.*, 22, 425–446.
12. Barber, A. M. and V. B. Zhurkin. (1990) *J. Biomol. Struct. Dyn.*, 8, 213–232.
13. Patel, D. J., L. Shapiro and D. Hare. (1987). In R. D. Wells and S. C. Harvey (ed.), Unusual DNA structures. Springer publishing, New York, USA.
14. Timsit, Y., E. Vilbois and D. Moras. (1991) *Nature*, 354, 167–170.
15. Bolshoy, A., P. McNamara, R. E. Harrington and E. N. Trifonov. (1991) *Proc. Natl. Acad. Sci.*, 88, 2312–2316.
16. Ramsden, D. A. and G. E. Wu. (1991) *Proc. Natl. Acad. Sci. USA.*, 88, 10721–10725.
17. Sakano, H., Y. Kurosawa, M. Weigert and S. Tonegawa. (1981) *Nature*, 290, 562–565.
18. Kurosawa, Y., H. Von Boehmer, W. Haas, H. Sakano, A. Traunecker and S. Tonegawa. (1981) *Nature*, 290, 565–570.
19. Kurosawa, Y. and S. Tonegawa. (1982) *J. Exp. Med.*, 155, 201–218.
20. Corbet, S., M. Milili, M. Fougereau and C. Schiff. (1989) *J. Immunol.*, 138, 932–939.
21. Jouvin-Marche, E. and S. Rudikoff. (1986) *Immunogenetics*, 24, 191–201.
22. Boyd, R. T., M. M. Goldrick and P. D. Gottlieb. (1986) *Proc. Natl. Acad. Sci. U.S.A.*, 83, 9134–9138.
23. Selsing, E. and U. Storb. (1981) *Cell*, 25, 47–58.
24. Max, E. E., J. G. Seidman, H. Miller and P. Leder. (1980) *Cell*, 21, 793–799.
25. Seidman, J. G., E. E. Max and P. Leder. (1979) *Nature*, 280, 370–375.
26. Seidman, J. G., A. Leder, M. H. Edgell, F. Polsky, S. M. Tilghman, D. C. Tiemeier and P. Leder. (1978) *Proc. Natl. Acad. Sci. U.S.A.*, 75, 3881–3885.
27. Heinrich, G., A. Traunecker and S. Tonegawa. (1984) *J. Exp. Med.*, 159, 417–435.
28. Joho, R., H. Gershenfeld and I. L. Ceissman. (1984) *EMBO J.*, 3, 185–191.
29. Ng, K. H., A. Lavigueur, L. Ricard, M. Bouvrette, S. MacLean, D. Cloutier and D. M. Gibson. (1989) *J. Immunol.*, 143, 638–648.
30. Milstein, C., J. Even, J. M. Jarvis, A. gonzalez-Fernandez and E. Gherardi. (1992) *Eur. J. Immunol.*, 22, 1627–1634.
31. Miller, J., E. Selsing and U. Storb. (1982) *Nature*, 295, 428–430.
32. Winoto, A., S. Mjolsness and L. Hood. (1985) *Nature*, 316, 832–836.
33. Wilson, R. K., B. F. Koop, C. Chen, N. Halloran, R. Sciammis and L. Hood. (1992) *Genomics*, 13, 1198–1208.
34. Arden, B., J. L. Klotz, G. Siu and L. E. Hood. (1985) *Nature*, 316, 783–787.
35. Hayday, A. C., D. J. Diamond, G. Tanigawa, J. S. Heilig, V. Folsom, H. Saito and S. Tonegawa. (1985) *Nature*, 316, 828–832.
36. Gascoigne, N. R. J., Y. Chien, D. M. Becker, K. J. and M. M. Davis. (1984) *Nature*, 310, 387–390.
37. Chien, Y., N. R. J. Gascoigne, J. Kavaler, N. E. Lee and M. M. Davis. (1984) *Nature*, 309, 322–326.
38. Siu, G., M. Kronenburg, E. Strauss, R. Haars and T. W. Mak. (1984) *Nature*, 311, 344–350.
39. Hayday, A. C., H. Saito, S. D. Gillies, D. M. Kranz, G. Tanigawa, H. N. Eisen and S. Tonegawa. (1985) *Cell*, 40, 259–269.
40. Elliot, J. F., E. K. Rock, P. A. Patten and M. M. Davis. (1988) *Nature*, 331, 627–631.
41. Ravetch, J. V., U. Siebenlist, S. Korsmeyer, T. Waldmann and P. Leder. (1981) *Cell*, 27, 583–591.
42. Siebenlist, U., J. V. Ravetch, S. Korsmeyer, T. Waldmann and P. Leder. (1981) *Nature*, 294, 631–635.
43. Ichihara, Y., H. Matsuoka and Y. Kurosawa. (1988) *EMBO J.*, 7, 4141–4150.
44. Chen, P. P., K. Albrandt, N. K. Orida, V. Radoux, E. Y. Chen, R. Schrantz, F.-T. Liu and D. A. Carson. (1986) *Proc. Natl. Acad. Sci. U.S.A.*, 83, 8318–8322.
45. Liu, M.-F., D. L. Robbins, J. J. Crowley, S. Sinha, F. Kozin, T. J. Kipps, D. A. Carson and P. P. Chen. (1989) *J. Immunol.*, 142, 688–694.
46. Pech, M., H.-R. Jaenichen, H.-D. Pohlenz, P. S. Neumaier, H.-G. Klobeck and H. G. Zachau. (1984) *J. Mol. Biol.*, 176, 189–204.
47. Bentley, D. L. and T. H. Rabbitts. (1980) *Nature*, 288, 730–733.
48. Bentley, D. L. and T. H. Rabbitts. (1983) *Cell*, 32, 181–189.
49. Pech, M. and H. G. Zachau. (1984) *Nucleic Acids Res.*, 12, 9229–9236.
50. Udey, J. A. and B. Blomberg. (1987) *Immunogenetics*, 25, 63–70.
51. Vasicek, T. J. and P. Leder. (1990) *J. Exp. Med.*, 172, 609–620.
52. Yoshikai, Y., S. P. Clark, S. Taylor, U. Sohn, B. I. Wilson, M. D. Minden and T. W. Mak. (1985) *Nature*, 316, 837–940.
53. Baer, R., T. Boehm, H. Yssel, H. Spits and T. H. Rabbitts. (1988) *EMBO J.*, 7, 1661–1668.
54. Siu, G., S. P. Clark, Y. Yoxhikai, M. M., Y. Yanagi, E. Strauss, T. W.

Mak and L. Hood. (1984) *Cell*, 37, 393–401.
55. Toyonaga, B., Y. Yoshikai, V. Vadasz, B. Chin and M. T.W. (1985) *Proc. Natl. Acad. Sci. USA*, 82, 8624–8628.
56. Tunnacliffe, A. and T. H. Rabbitts. (1985) *Nucleic Acids Res.*, 13, 6651–6671.
57. LeFranc, M.-P., A. Forster and T. H. Rabitts. (1986) *Nature*, 319, 420–422.
58. Hockett, R. D., J. de Villartay, K. Pollock, D. G. Poplack, D. I. Cohen and S. J. Korsmeyer. (1988) *Proc. Natl. Acad. Sci. USA*, 85, 9694–9698.
59. Satyanarayana, K., S. Hata, P. Devlin, M. G. Roncarlo, J. E. De Vries, H. Spits, J. L. Strominger and M. S. Krangel. (1988) *Proc. Natl. Acad. Sci. USA*, 85, 8166–8170.
60. Reynaud, C.-A., A. Dahan, V. Anquez and J.-C. Weill. (1989) *Cell*, 59, 171–183.
61. Renaud, C.-A., A. Anquez, A. Dahan and J.-C. Weill. (1985) *Cell*, 40, 283–291.
62. Becker, R. S., M. Suter and K. L. Knight. (1990) *Eur. J. Immunol.*, 20, 397–402.
63. Leiberman, R., L. Emorine and E. E. Max. (1984) *J. Immunol.*, 133, 2753–2756.
64. Heidmann, O. and F. Rougeon. (1984) *Nature*, 311, 74–76.
65. Kokubu, F., R. Litman, M. J. Shamblott, K. Hinds and G. W. Litman. (1988) *EMBO J.*, 7, 3413–3422.
66. Shamblott, M. J. and G. W. Litman. (1989) *EMBO J.*, 8, 3733–3739.
67. Zezza, D. J., S. E. Stewart and L. A. Steiner. (1992) *J. Immunol.*, 149, 3968–3977.
68. Brueggemann, B., J. Free, A. Diamond, J. Howard, S. Cobbold and H. Waldmann. (1986) *Proc. Natl. Acad. Sci. U.S.A.*, 83, 6075–6079.
69. Steen, M.-L., L. Hellman and U. Pettersson. (1987) *Gene*, 55, 75–84.
70. Takeuchi, N., N. Ishiguro and M. Shinagawa. (1992) *Immunogenetics*, 35, 89–96.
71. Renaud, C.-A., C. R. Mackay, R. G. Muller and J.-C. Weill. (1991) *Cell*, 64, 995–1005.
72. McCormack, W. T., L. M. Carlson, L. W. Tjoelker and C. B. Thompson. (1989) *Int. Immunol.*, 1, 332–341.
73. Blankenstein, T., F. Bonhomme and U. Krawinkel. (1987) *Immunogenetics*, 26, 237–248.
74. Schroeder, H. W., J. L. Hillson and R. M. Perlmutter. (1989) *Int. Immunol.*, 2, 41–50.
75. Cohen, J. B. and D. Givol. (1983) *EMBO J.*, 2, 2013–2018.
76. Sakano, H., R. Maki, Y. Kurosawa, W. Roeder and S. Tonegawa. (1980) *Natue*, 286, 676–683.
77. Givoli, D., R. Zakut, K. Effron, G. Rechavi, D. Ram and J. B. Cohen. (1981) *Nature*, 292, 426–430.
78. Ollo, R., C. Auffray, J.-L. Sikorav and F. Rougeon. (1981) *Nucleic Acids Res.*, 9, 4099–4109.
79. Schiff, C., M. Milili, I. Hue, S. Rudikoff and M. Fougereau. (1986) *J. Exp. Med.*, 163, 573–587.
80. Siekevitz, M., S. Y. Huang and M. L. Gefter. (1983) *Eur. J. Immunol.*, 13, 123–132.
81. Kataoka, T., T. Nikaido, T. Miyata, K. Moriwaki and T. Honjo. (1982) *J. Biol. Chem.*, 257, 277–285.
82. Yancopoulos, G. D. and F. W. Alt. (1985) *Cell*, 40, 271.
83. Cohen, J. B., K. Effron, G. Rechavi, Y. Ben-Nenah, R. Zakut and D. Givol. (1982) *Nucleic Acids Res.*, 10, 3353–3370.
84. Crews, S., G. J., H. Huang, K. Calame and L. E. Hood. (1981) *Cell*, 25, 59–66.
85. Bothwell, A. L., M. Parking, M. Reth, T. Imanishi-kari, K. Rajewsky and D. Baltimore. (1981) *Cell*, 24, 625.
86. Yancopoulos, G. D., S. V. Desiderio, M. Paskind, J. F. Kearney, D. Baltimore and F. W. Alt. (1984) *Nature*, 311, 727–733.
87. Ollo, R., J.-L. Sikorav and F. Rougeon. (1983) *Nucleic Acids Res.*, 11, 7887–7897.
88. Ucker, D. S., Y. Kurwawa and S. Tonegawa. (1985) *J. Immunol.*., 135, 4204.
89. Bernard, O., N. Hozumi and S. Tonegawa. (1978) *Cell*, 15, 1133–1144.
90. Tonegawa, S., A. M. Maxam, R. Tizard, O. Bernard and W. Gilbert. (1978) *Proc. Natl. Acad. Sci. U.S.A.*, 75, 1485–1489.
91. Sanchez, P., P. N. Marche, D. Rueff-Juy and P.-A. Cazenave. (1990) *J. Immunol.*, 144, 2816–2820.
92. Hirama, T., S. Takeshita, Y. Matsubayashi, M. Iwashiro, T. Masuda, K. Kuribayashi, Y. Yoshida and H. Yamagishi. (1990) *Eur. J. Immunol.*, 21, 483–488.
93. Chou, H. S., M. A. Behlke, S. A. Godambe, J. H. Russell, C. G. Brooks and D. Y. Loh. (1986) *EMBO J.*, 5, 2149–2155.
94. Chou, H. S., S. J. Anderson, M. C. Louie, S. A. Godambe and M. R. Pozzi. (1987) *Proc. Natl. Acad. Sci. U.S.A.*, 84, 1992–1996.
95. Louie, M. C., C. A. Nelson and D. Y. Loh. (1989) *J. Exp. Med.*, 170, 1987–1998.
96. Garman, R. D., P. J. Doherty and D. H. Raulet. (1986) *Cell*, 45, 733–742.
97. Shen, A., C. Humprhies, P. Tucker and F. Blattner. (1987) *Proc. Natl. Acad. Sci. U.S.A.*, 84, 8563–8567.
98. Lee, K. H., F. Matsuda, T. Kinashi, M. Kodaira and T. Honjo. (1987) *J. Mol. Biol.*, 195, 761–768.
99. Berman, J. E., S. J. Mellis, R. Pollock, S. C.L., H. Suh, B. Heinke, C. Kowal, U. Surti, L. Chess, C. R. Cantor and F. W. Alt. (1988) *EMBO J.*, 7, 727–738.
100. Matsuda, F., K. H. lee, S. Makai, T. Sato, M. Kodaira, S. Q. Zong, H. Ohno, S. Fukuhara and T. Honzo. (1988) *EMBO J.*, 7, 1047–1051.
101. Kodaira, M., T. Kinashi, I. Umemura, F. Matsuda, T. Noma, Y. Ono and T. Honjo. (1986) *J. Mol. Biol.*, 190, 529–541.
102. Matthyssens, G. and T. H. Rabbitts. (1980) *Proc. Natl. Acad. Sci. USA*, 77, 6561–6565.
103. Rechavi, G., B. Bienz, D. Ram, Y. Ben-Neriah, J. B. Cohen, R. Zakut and D. Givol. (1982) *Proc. Natl. Acad. Sci. USA*, 79, 4405–4409.
104. Rechavi, G., D. Ram, L. Glazer, R. Zakut and D. Givol. (1983) *Proc. Natl, Acad. Sci, USA*, 80, 855.
105. Heiter, P. A., J. Maizel J.V. and P. Leder. (1982) *J. Biol. Chem.*, 257, 1516–1522.
106. Frippiat, J. P., P. Chuchanan, F. Bernard, L. Buluwela, G. Lefranc and M. P. Lefranc. (1990) *Nucleic Acids Res.*, 18, 7134–7134.
107. Alexandre, D., P. Chuchana, F. Brockly, G. Lefranc and M.-P. Lefranc. (1989) *Nucleic Acids Res.*, 17, 3975–3975.
108. Siminovitch, K. A., V. Misener, P. C. Kwong, Q.-L. Song and P. P. Chen. (1989) *J. Clin. Invest.*, 84, 1675–1678.
109. Bernard, F., P. Chuchana, J. P. Frippiat, L. Buluwela and M. P. Lefranc. (1990) *Nucleic Acids Res.*, 18, 7139–7139.
110. Brockley, F., D. Alexandre, P. Chuchana, S. Huck, G. Lefranc and M.-P. Lefranc. (1989) *Nucleic Acids Res.*, 17, 3976.
111. Daley, M. D., H. Peng, V. Misener, X. Liu, P. P. Chen and K. A. Siminovitch. (1992) *Mol. Immunol.*, 29, 1515–1518.
112. Smith, W. J., A. Tunnacliffe and T. H. Rabitts. (1987) *Nucleic Acids Res.*, 15, 4991.
113. Siu, G., E. C. Strauss, E. Lai and L. E. Hood. (1986) *J. Exp. Med.*, 164, 1600–1614.
114. Ikuta, K., T. Ogura, A. Shimizu and T. Honjo. (1985) *Proc. Natl. Acad. Sci. U.S.A.*, 82, 7701–7705.
115. Lefranc, M.-P., A. Forster, R. Baer, M. A. Stinson and T. H. Rabbitts. (1986) *Cell*, 45, 237–246.
116. Huck, S., P. Dariavach and M.-P. Lefranc. (1988) *EMBO J.*, 7, 719–726.
117. Guglielmi, P., F. Davi, L. d'Auriol, J.-C. Bories, J. Dausset and A. Bensussan. (1988) *Proc. Natl. Acad. Sci. U.S.A.*, 85, 5634–5638.
118. Schwager, J., N. Burckert, M. Courtet and L. D. Du Pasquier. (1989) *EMBO J.*, 8, 2989–3001.
119. Knight, K. L. and R. S. Becker. (1990) *Cell*, 60, 963–970.
120. McCormack, W. T., S. M. Laster, W. F. Marzluff and K. H. Roux. (1985) *Nucleic Acids. Res.*, 13, 7041–7045.
121. Becker, R. S., S. K. Zhai, S. J. Currier and K. L. Knight. (1989) *J. Immunol.*, 142, 1351–1355.
122. Heidmann, O. and F. Rougeon. (1983) *Cell*, 34, 767–777.
123. Gutman, G. A., R. M. Besta, M. B. Frank and P. R. Baverstock. (1987) *Immunogenetics*, 26, 14–20.