



HAL
open science

Unequal V-H gene rearrangement frequency within the large V(H)7183 gene family is not due to recombination signal sequence variation, and mapping of the genes shows a bias of rearrangement based on chromosomal location

G. Stuart Williams, Alan Martinez, Alina Montalbano, Alan Tang, America Mauhar, Kisani M. Ogwaro, Denise Merz, Christophe Chevillard, Roy Riblet, Ann J. Feeney

► To cite this version:

G. Stuart Williams, Alan Martinez, Alina Montalbano, Alan Tang, America Mauhar, et al.. Unequal V-H gene rearrangement frequency within the large V(H)7183 gene family is not due to recombination signal sequence variation, and mapping of the genes shows a bias of rearrangement based on chromosomal location. *Journal of Immunology*, 2001, 167 (1), pp.257-263. hal-01593089

HAL Id: hal-01593089

<https://amu.hal.science/hal-01593089>

Submitted on 6 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Unequal V_H Gene Rearrangement Frequency Within the Large V_H7183 Gene Family Is Not Due to Recombination Signal Sequence Variation, and Mapping of the Genes Shows a Bias of Rearrangement Based on Chromosomal Location¹

G. Stuart Williams,* Alan Martinez,* Alina Montalbano,* Alan Tang,* America Mauhar,[†] Kisani M. Ogwaro,* Denise Merz,* Christophe Chevillard,^{2†} Roy Riblet,[†] and Ann J. Feeney^{3*}

Much of the nonrandom usage of V, D, and J genes in the Ab repertoire is due to different frequencies with which gene segments undergo V(D)J rearrangement. The recombination signal sequences flanking each segment are seldom identical with consensus sequences, and this natural variation in recombination signal sequence (RSS) accounts for some differences in rearrangement frequencies in vivo. Here, we have sequenced the RSS of 19 individual V_H7183 genes, revealing that the majority have one of two closely related RSS. One group has a consensus heptamer, and the other has a nonconsensus heptamer. In vitro recombination substrate studies show that the RSS with the nonconsensus heptamer, which include the frequently rearranging *8IX*, rearrange less well than the RSS with the consensus heptamer. Although *8IX* differs from the other *7183-I* genes at three positions in the spacer, this does not significantly increase its recombination potency in vitro. The rearrangement frequency of all members of the family was determined in μ MT mice, and there was no correlation between the in vitro recombination potential and V_H gene rearrangement frequency in vivo. Furthermore, genes with identical RSS rearrange at different frequencies in vivo. This demonstrates that other factors can override differences in RSS potency in vivo. We have also determined the gene order of all V_H7183 genes in a bacterial artificial chromosome contig and show that most of the frequently rearranging genes are in the 3' half of the region. This suggests that chromosomal location plays an important role in nonrandom rearrangement of the V_H7183 genes. *The Journal of Immunology*, 2001, 167: 257–263.

The diversity of the Ab and TCR repertoires is due in large part to the combinatorial association of the large numbers of V, D, and J segments to create thousands of different heavy and light, or α and β , chain molecules, respectively. However, this combinatorial diversity is constrained by the fact that V, D, and J segments recombine at very different frequencies. An example of unusually high frequency of rearrangement is observed with the *8IX* gene, which is the most 3' functional V_H gene in the murine V_H locus. Many laboratories have observed that this V_H gene family rearranges very frequently early in ontogeny and also in the adult bone marrow (1–9). The overuse of the *7183* family, especially *8IX*, gave rise to the chromosomal proximity hypothesis of Alt and colleagues, which proposed that the rearrangement mechanism preferentially rearranges genes that are in closer proximity in the chromosome (1, 10). Supporting the generality of this proximity observation, mice that lack the IL-7 receptor show de-

creased rearrangement of V_H genes, displaying a gradient of rearrangement throughout the entire V_H locus, ranging from almost normal levels of rearrangement of the V_H7183 family to essentially no rearrangement of the distal V_HJ558 and V_HJ606 families (11). Also, overuse of V-proximal $J\alpha$ genes has been observed in the *TCR α* loci in fetal and newborn thymi (12), and only later do the more distal J genes rearrange. Most strikingly, there is a strict chromosomally ordered gene rearrangement in the *V γ* loci (13). However, other loci, such as the human V_H loci, do not display this phenomenon of chromosomally ordered rearrangements (14). It has been shown that many of the V_H3 genes that are over-represented in rearrangements map in the middle or 5' portion of the locus (15). However, the most 3' human V_H gene, V_H6 , is over-represented in rearrangements early in ontogeny (16). A direct tracking mechanism as an explanation for this phenomenon has been deemed unlikely, but the reason for these several examples of positionally graded rearrangement frequencies is not clear.

An alternative, or additional, explanation is that each individual gene controls the frequency of its rearrangement to a large extent by virtue of its relatively unique recombination signal sequence (RSS).⁴ Most natural TCR and Ig RSS vary from the consensus RSS, and this variation could result in differences in rearrangement frequency (17). In these cases, the chromosomal location would be irrelevant, and thus could explain the inconsistent correlation of high frequency of rearrangement with proximal chromosomal location for only some, but not other, V and J loci. We and others,

*Department of Immunology, The Scripps Research Institute, La Jolla, CA 92037; and [†]Torrey Pines Institute for Molecular Studies, San Diego, CA 92121

Received for publication January 29, 2001. Accepted for publication April 16, 2001.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ This work was supported by National Institutes of Health Grants AI29672 (to A.J.F.) and AI23548 (to R.R.). A.Ma. and A.Mo. were supported by Training Grant T32GM08303. This is manuscript 13802-IMM from The Scripps Research Institute.

² Current address: Institut National de la Santé et de la Recherche Médicale, Unité 399, Immunology and Genetics of Parasitic Diseases, Faculty of Medicine, 27 boulevard J. Moulin, Marseilles 13385, France.

³ Address correspondence and reprint requests to Dr. Ann J. Feeney, Department of Immunology, IMM-22, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037. E-mail address: feeney@scripps.edu

⁴ Abbreviation used in this paper: RSS, recombination signal sequence(s); BAC, bacterial artificial chromosome; FR, framework.

particularly Wu and colleagues, have, in fact, shown that the natural variation in the RSS of V genes can affect rearrangement frequencies (18–23). Although the classic studies of Gellert and colleagues have shown that some positions in the heptamer and nonamer are critical for rearrangement, while others are less critical, those studies have not analyzed all possible changes in the RSS (24). We have shown that an allelic change involving a single substitution in the heptamer can lead to a 4.5-fold reduction in rearrangement frequency in a V_K gene critical for an effective Ab response to *Haemophilus influenzae* type b (21, 25). Furthermore, Navajos, who have the allele with the nonconsensus heptamer, have a greatly increased susceptibility to *H. influenzae* type b disease. Thus, variations in RSS leading to relatively modest differences in rearrangement frequencies can alter the composition of the Ab repertoire and can have severe biological ramifications (18).

It has been proposed that the reason that *81X* rearranges so often is due to its RSS (23). Using miniloci recombination substrates transiently transfected into pre-B cells, the relative frequency of rearrangement of *81X* vs V_HJ558 genes was determined, and *81X* rearranged ~15 times more often than V_HJ558 (22, 23). However, the choice of a V_HJ558 gene as a competitor was likely to show a bias in favor of V_H7183 , because V_HJ558 genes have a poor RSS. It stands to reason that if *81X* rearrangements comprise over half of the V_H7183 rearrangements, the other ~18 functional members of the V_H7183 family cannot all rearrange at that same high rate. Hence one would predict that *81X* would have a unique RSS if its high frequency of rearrangement were due to a particularly potent RSS.

Therefore, in this current study we have sequenced the RSS of 19 of 20 members of the V_H7183 family, and they fall into two similar groups. All but two genes share the same nonconsensus nonamer. Strikingly, over half of the V_H7183 genes have the same nonconsensus heptamer as *81X*, although *81X* has a unique spacer sequence. The other V_H7183 genes have RSS with a consensus heptamer and a slightly different spacer. Twelve of the 19 RSS are identical with one of the two prototypic RSS, and the other seven have a small number of changes. We determined the relative potential of the various RSS to support recombination in a recombination substrate assay, and, not surprisingly, the RSS with the consensus heptamer rearranged more often than RSS with the nonconsensus heptamers. Even the RSS of *81X*, with its unique spacer, supported less rearrangement than the V_H7183 RSS with the consensus heptamer.

We determined the chromosomal order of the V_H7183 genes from a bacterial artificial chromosome (BAC) contig covering the entire V_H7183 region of the IgH locus, and we also determined the relative frequency of rearrangement of each of the V_H7183 genes. All but one of the frequently rearranging genes were in the 3' half of the region, and the rearrangement frequency fell off in gradient fashion 5' of *81X*, with the exception of one poorly rearranging gene in the 3' portion. Thus, these data clearly show that variation in RSS potency is not the major determinant in the nonrandom V_H usage of the V_H7183 gene family or in the overuse of *81X* in rearrangements, and that the RSS variation is overridden by other factors. It appears that chromosomal location may play a more major role in the over-rearrangement of many of the 3' genes. Finally, despite identity in RSS and similarity in coding regions, similar genes rearrange at quite different frequencies in different parts of the locus, suggesting that the local region surrounding each V_H gene is a major determinant of the frequency of rearrangement.

Materials and Methods

Mice

μ MT mice (26) were obtained from The Jackson Laboratory (Bar Harbor, ME) and were bred at The Scripps Research Institute's animal facility.

PCR and sequencing

The following oligonucleotides were used, with engineered restriction enzyme sites underlined: 5' primers: AF191, 5'-CGGGCTCGAGTTGGTTTTCCTTGTCC-3' (leader); AF215, 5'-GAGGCTCGAGGAACTCTCCTGTG-3' (starts at bp 43 of the coding region); AF303, 5'-GGGGCTCGAGGAGTCTGGGGGA-3' (starts at bp 6 of the coding region); AM3, 5'-GTCCGAATTCCTCTCTGTGCAA-3' (starts at bp 48 of the 50.1 coding region); AM6, 5'-AATAGGAATTCCTGGTAGCAGGTAT-3' (starts at bp 154 of the 68–5N coding region); RR190, 5'-ATGCTGGTGGAGTCTGGG-3' (starts at bp 7 of the coding region); *Ox2* 5', 5'-GCTGAAGCTTTCAGGCAATGGCCA-3'; 3' primers: AF192, 5'-CTCCGCGGCCGCTGCTGGTCT-3' (starts 16 bp 3' of the RSS); AF13N, 5'-AAAAGCGGCCGCTTACCTGAGGAGACGGTCA-3' (J_H); AF269, 5'-GGTCGCGGCCGCTCTGCAGGAGTGT-3' (starts 5 bp from the end of the nonamer of the RSS); AM5, 5'-GCAGAAGCTTTAGTTGAGTCA-3' (starts in the middle of the spacer of the V_H7183 family RSS and extends through the nonamer); RR192, 5'-GACCTCAGACTGCTCATTTGC-3' (starts at bp 243 of the coding region); *Ox2* 3', 5'-GCGCGAATTCCTGATCATCTTTCTAGTCA-3'; and *E4.psi*.seq, 5'-ACTGATGGTAGTTTCA-3' (starts at bp 157 of *E4.psi* coding region).

All oligonucleotides were supplied by Genosys (The Woodlands, TX) or Life Technologies (Gaithersburg, MD). For the repertoire analysis, newborn liver DNA was prepared using the method described by Schlissel et al. (27). AF303 and AF13N were used to amplify VDJ rearrangements from newborn liver DNA. Thirty-five cycles of 30 s at 94°C, 30 s at 57°C, and 1 min at 68°C, followed by one cycle of 10 min at 68°C, were performed using an Ericomp cyclor (Ericomp, San Diego, CA). Elongase (Life Technologies) was used for all PCRs because it contains a proofreading polymerase. The resulting PCR fragments were digested with *XhoI* and *NotI* and cloned into pBluescript (Stratagene, La Jolla, CA). Individual colonies were minipreped and sequenced with a T7 primer using Sequenase version 2.0 (U.S. Biochemical Corp., Cleveland, OH) sequencing kit and [³⁵S]ATP. For the repertoire analysis, once repeat sequences (i.e., the same V, D, and J segments and the same junctional sequence) were identified coming from the same PCR, no further sequences from that PCR were obtained. Identification of the V_H genes was routinely based on >150 bp of V_H sequence information from the 3' half of the gene.

AF191 and AF192, AF191 and AF269, RR190 and RR192, or AF215 and AM5 were used for amplifying the genomic sequences from individual BAC DNA for the mapping analysis. Twenty-five cycles of 30 s at 94°C, 30 s at 55°C, and 1.5 min at 68°C, followed by one cycle of 10 min at 68°C, were performed using Elongase. The PCR fragments were digested with *XhoI* and *NotI*, or *EcoRI* and *HindIII*, depending upon the primers used. Clones were sequenced from both sides, and a minimum of two sequences from independent PCRs were determined for each V_H gene to ensure the absence of errors.

Additionally, in the assembly of the BAC contig (see below) each V_H7183 gene, except 7183.10, 7183.11, *81X*, *E4.psi*, *D6.96*, and 3:3.39, has recently been sequenced directly from BAC DNA without PCR amplification, confirming the sequences obtained by PCR and cloning. Direct sequencing was performed on agarose gel eluates purified with the Qiaquick kit (Qiagen, Valencia, CA). Cycle sequencing used BigDye terminators (PE Biosystems, Foster City, CA) for one cycle of 95°C for 5 min, followed by 100 cycles of 96°C for 30 s, 50°C for 15 s, and 60°C for 4 min.

Recombination substrate assay

The competition recombination substrate assay has been described previously (21). The RSS and ~100–200 bp of flanking regions were cloned into a competition recombination substrate plasmid 5' of the transcriptional stop, and an inverted *DSP2* segment was cloned into the 3' side of the transcriptional stop using the following oligonucleotides (restriction enzyme sites are underlined): P50a, 5'-GGGTACGCGTGCCTCCAGGGAAG-3'; P50b, 5'-CCTGGCGGCCGCGAGACCCCTGCA-3'; P58A2, 5'-CCCTGCGGCCGCAATGAGCAGTC-3'; P61b, 5'-CTCTCGTCGACCCCTGTGGTCTCT-3'; P76b, 5'-TCTGGTGCACAGAACACCCTGCA-3'; P82, 5'-TTCAGGCGGCCGCGCTGAGAGCTGAGG-3'; P97a, 5'-TGAAGCGGC CGCTGGACAGGGACTTG-3'; and P97b, 5'-TTCTGTCTGACTCAA TCCCAGTGC-3'.

P50a and P50b were used to amplify the $V_HS107/V1$ RSS for cloning into the external site; P82 and P76b were used for cloning the *V1* gene into the internal RSS cloning site. P58A2 and P61b were used to clone *81X*,

3.3.39 (representative 7183-I gene), and 37.1 (representative 7183-II gene) into the internal RSS cloning site. P97a and P97b were used to clone the V_HJ558 RSS into the internal RSS cloning site. The resulting constructs were sequenced to confirm that no mutations had been introduced during the PCR.

The RSS competition assay was performed as previously described (21). Briefly 18.8 Abelson-murine leukemia virus-transformed pre-B cells were transfected using a Bio-Rad GenePulser with 300 V and 960 μ F capacitance (Bio-Rad, Hercules, CA). Transfected cells were cultured for 48 h in the presence of 0.5–1.0 mM caffeine (ICN Pharmaceuticals, Costa Mesa, CA), then the plasmids were recovered using a standard alkaline lysis protocol. Recovered plasmid was digested with *SpeI*, transformed into Electrocompetent XL1 Blue (Stratagene), and plated on Luria Bertoni agar containing isopropyl β -D-thiogalactoside and chloramphenicol. The bacterial colonies were screened using a PCR-based assay as previously described to determine the relative rearrangement of the *DSP* RSS to the external vs internal V_H RSS (21).

Assembly of V_H7183 BAC contig

The BAC library of the mouse strain 129/Sv (129S3) from the embryonic stem cell line CJ7 distributed by Research Genetics (Huntsville, AL) was screened using PCR assays for V_HIII , D12Nds2 ($V_HS107/V1$), $V_HE4.psi$, and BAC ends CT7–19B1-T7 and CT7–057J13-Sp6 (C. Chevillard et al., manuscript in preparation). BAC ends were isolated by vector-hexameric PCR (28) and hybridized to Southern blots of *HindIII*-digested BAC DNAs to identify overlap patterns. High density BAC library membranes were probed with pooled V_H7183 PCR products from several 7183⁺ BACs, and additional BACs were identified. All BACs were digested with *EcoRI* and blotted; replicate blots were probed with amplified inserts of 81X and V_HQ52 plasmids. The gene content of each hybridizing fragment was identified from at least two representative BACs by excision of the fragment from agarose gel and sequencing its V_H7183 gene. At least one sequence in each case was obtained directly from eluted restriction fragment without PCR, with the exception of six genes, as noted above; confirmatory sequences were in some cases obtained from PCR amplification from eluted restriction fragments. Taken together, these data allowed the unambiguous ordering of BACS and V_H7183 genes.

Results

Sequence of the germline genes

The RSS of 81X is nonconsensus, but has been shown to be much more efficient at rearrangement than a V_HJ558 RSS (22). We wanted to determine the sequence of the RSS of each V_H7183 gene to determine the extent of diversity of RSS within this V_H gene family and to determine whether the relative frequency of recombination of the individual family members correlated with the sequence of their RSS. To determine the sequence of the RSS of each V_H7183 gene, we initially used a primer in the leader, and a primer 3' of the RSS. The leader sequence is often very conserved within members of a V_H family, and the sequence just 3' of the RSS is relatively conserved. Therefore, we used this primer set to amplify as many germline V_H7183 genes as possible. PCRs were performed on the BAC DNA from 129/Sv mice, which, like BALB/c, are of the *Igh*^a haplotype. We were able to amplify 13 genes in this way, with multiple independent sequences of those genes. Because there were >13 *EcoRI* bands on Southern blots that hybridized with the V_H7183 , and because some known genes were missing initially, we also amplified some of the BAC DNAs with framework 1 (FR1) primers and either AF269 (which overlaps the 3' end of the RSS by 5 bp) or AM5 (which contains the nonamer and half the spacer) in an attempt to choose regions more likely to be conserved. *E4.psi* did not amplify with the primers, so we used a sequencing primer in FR3 of the published sequence of *E4.psi* to sequence the RSS. In other cases direct sequencing of DNA obtained from *EcoRI* bands of BAC DNA excised from agarose gels was performed for genes that we were unable to amplify from the BAC.

The sequence of the RSS and the adjacent coding end sequence for all members of the family are shown in Fig. 1 (full-length sequences are available from GenBank under accession no.

		RSS			
		CACAGTG		ACAAAAACC	
7183 I	7183.9	---A--	AGGAAATGTTACTGTGAGCTCAA	--T-----	
	7183.9T	---A--		--T-----	
	3:3.39	---A--		--T-----	
	69.1	---A--		--T-----	
	D6.96	---A--		--T-----	
	68-5N	---A--		--T-----	
	50.1	---A--		--T-----	
	Ib	81x	---A--	--C--A	--T-----
	Ic	7183.13	---A--	--A--A	--T-----
	Id	7183.19	---A--		--T-----
Ie	7183.20	G--A--		--GT-----	
7183 II	B4.15/283	-----	--TG-	-----T	
	37.1	-----	--TG-	-----T	
	7183.14	-----	--TG-	-----T	
	98-3G	-----	--TG-	-----T	
	7183.10	-----	--TG-	-----T	
IIc	B4.psi	-----	--TC-	-----C	
	IIb	61-1P	-----	--G--A--T	
	IIid	7183.12	---T--	--TG-	
	7183.11	Unknown	RSS		
S107	V1	-----	--A-G-CG-C-T	-----C--G	
J558		-----	TTGT--CCAC-TCC	-----TGTG T--G-----	
	or:	-----	TTGT--CCAC-TCC	-----TGTGT CAG--C--	

FIGURE 1. RSS sequences of the V_H7183 family members. Genes with a 7183-I-type RSS are on the top; genes with a 7183-II RSS are on the bottom. The variants are separated from the genes possessing prototypic RSS. The RSS of $V_HS107/V1$ and V_HJ558 genes used in the competition recombination substrates are also shown. The RSS of *J558* is shown as if it has a 22-bp spacer on the top line, and the same RSS is written as though it has a 23-bp spacer on the second line. The full-length sequence of any previously unpublished sequences of these genes is available from GenBank under accession numbers AF290959–AF290972. Some of these sequences have previously been reported (1, 7, 29, 30).

AF290959–AF290972). The RSS of the V_H7183 genes fell into two major groups, which we called 7183-I and 7183-II. 7183-I RSS had the same nonconsensus heptamer as 81X (CACAATG; the one change underlined) and a nonconsensus nonamer (ACTA AAACC), while the 7183-II RSS had the same nonconsensus nonamer as the 7183-I genes, but had a consensus heptamer. The spacers were very similar, with the 7183-I and 7183-II prototypes differing only in three positions. Seven genes had the identical prototypic 7183-I RSS, and five had the identical prototypic 7183-II RSS. In addition to these genes, seven other genes were found to vary from the prototypes at a few positions in the RSS, predominantly in the spacer, including 81X, which differed from the 7183-I RSS by two changes in the spacer. Only two genes have heptamers that vary from these two prototypes, and they are classified as variants of 7183-I or -II based on spacer sequence.

Three new V_H7183 genes

During our analysis of the BAC contig we found a previously unknown functional member of the V_H7183 family by using FR1 and FR3 primers. It is identical with 7183.9 in the coding region, other than two changes at the 5' end and a single base pair change of a G to a T in FR3, so we call this gene 7183.9T. We have extended the sequence through the RSS, and it is also identical with 7183.9 in that region, but its leader/intron region shows four differences between the two genes. We also identified a second new, apparently functional, gene that we have provisionally called 7183.19. All these V_H genes will be renamed numerically once all the interspersed pseudogenes and V_HQ52 genes are identified. It is only 94% homologous to the most similar V_H7183 gene and has a 7183-I variant RSS. A new presumably nonfunctional gene of the V_H7183 family, called 7183.20, was also identified. Its RSS was

related to the *7183-I* RSS; however, it contains a C to G mutation at the first position in the heptamer, an invariant position essential for RSS function (24). Therefore, this RSS is highly likely to be nonfunctional.

Several years ago we reported the sequence of four new V_H7183 genes from BALB/c mice and also made a compilation of all the V_H7183 genes described at that time in the BALB/c strain (also *Igh^α*) (29). The sequences obtained in this current study enable us to refine that list. We have added three new sequences as just described. We also now eliminate *MOPC21* as a germline gene; rather, it is almost certainly a *61-IP* gene with three somatic mutations. Also, we previously listed a sequence from the literature, *V-BK*, which only differs from *7183.12* by 3 bp, two of which are a GA dinucleotide instead of an AG dinucleotide in FR3. We never observed *V-BK* in rearrangements, nor did we find it in the BAC contig, and we assume that it is the same as the germline *7183.12* gene. Also, we never observed the published sequence for 283 (30). The published coding sequence of 283 is almost identical with that of *E4.15*, but they vary in the intron (1). 283 is the only published V_H7183 gene that ends with . . .TA, so we conclude that the two published sequences, 283 and *E4.15*, really represent the same gene.

Analysis of the role of the RSS in the nonrandom V_H gene use

It has been shown that *81X* rearranges 7- to 30-fold more frequently than the V_HJ558 gene in a competition substrate (22, 23). However, the V_HJ558 genes have very poor RSS, and the nonamer is so different from the consensus that it is not clear whether they also have a shortened, less functional, 22-bp spacer (see Fig. 1) (24). Therefore, the first issue we wanted to address was whether the RSS of *81X* is much better than that of a more conventional RSS. We chose the V_HS107 gene *VI* as a representative gene. This gene has a consensus heptamer, and the nonamer has one change from consensus. The spacer length is the optimal 23 bp, as are all of the V_H7183 RSS. We wished to assess the relative recombination supported by *VI* vs *81X*, so we made competition recombination substrates comparing the relative frequency of recombination of these two RSS to recombine to a *DSP2* fragment. RSS fragments are made by PCR and are usually ~100–200 bp in size, containing the RSS as well as some of the 5'- and 3'-flanking regions. The substrates are transiently transfected into pre-B cells, and the recovered plasmid is plated on a chloramphenicol plate. Colonies are screened by PCR to determine the relative frequency of rearrangement of the internal vs the external V_H fragment (Fig. 2).

Fig. 2 shows that a control competition substrate containing *VI* in both internal and external positions demonstrates the minimal effect of position (internal vs external) of the two competing RSS fragments in the substrates. This is consistent with the controls we have previously published using this competition recombination system (18, 21). When we cloned a V_HJ558 RSS fragment in the competition substrate, it was vastly disfavored, with 91% of the rearrangements occurring to *VI* instead of *J558* (Fig. 2). Thus, we conclude that V_HJ558 genes do, in fact, have poor RSS, as predicted by inspection of their RSS. However, when the *J558* fragment is replaced by *81X* in the above substrate, the relative frequencies of recombination of *VI* and *81X* are similar. This shows that the *81X* RSS is only slightly better than that of a V_HS107 gene and suggests that *VI* should rearrange almost as often as *81X* in vivo if the RSS is the major determinant of rearrangement frequency.

Within the V_H7183 family, all V_H genes but two share an identical nonconsensus nonamer. Approximately one-third of the V_H genes have a consensus heptamer, while two-thirds have a non-consensus heptamer, identical with that reported for *81X*. Based on classical studies with recombination substrates, one would expect

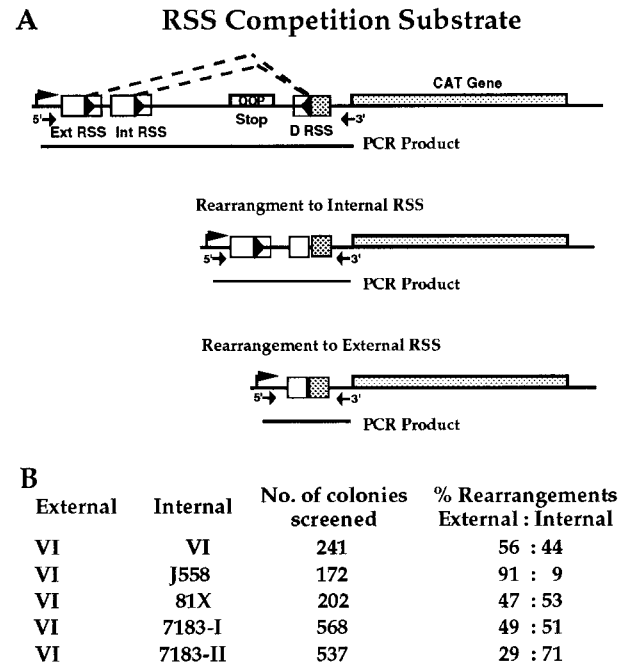


FIGURE 2. A, A linear map of the RSS competition substrate. The primers used to screen for internal vs external rearrangements are indicated. Amplified PCR products as the result of rearrangement to the internal or external RSS are shown. B, The results of the RSS competition substrate experiments are presented. For each competition recombination substrate, we describe the V gene/RSS in the external and the internal position, the total number of colonies assayed, and the percentage of rearrangements to internal and external RSS. The sequences of all RSS are given in Fig. 1.

that the *7183-II* RSS, with its consensus heptamer, might rearrange at a higher frequency than the *7183-I* RSS, with one change in the heptamer sequence (24, 31). *81X*, which rearranges at a very high frequency, has a *7183-I*-type RSS, but with a unique spacer sequence. It is possible that its variant spacer is the reason for its high frequency of recombination. Although the heptamer and nonamer are traditionally believed to be the most important regions of the RSS, we have previously shown that the spacer can also play a significant role in recombination frequency of V_K genes (18). In that study we also showed the surprising result that a RSS with a nonconsensus nonamer but an optimal spacer rearranges more frequently than a consensus RSS with a poorer spacer, thus underscoring the necessity to actually test the recombination potential of each RSS experimentally. Therefore, to determine the relative frequency of recombination of prototypic *7183-I* and *7183-II* RSS, we replaced the internal gene fragment in our control recombination substrate with that of *3.3.39* (a representative *7183-I* gene) or *37.1* (a representative *7183-II* gene). As predicted, the *7183-II* gene, with its consensus RSS, is better than *VI* or the *7183-I* genes for supporting rearrangement in the recombination substrates (Fig. 2). Seventy-one percent of the rearrangements were to the *7183-II* gene, rather than to *VI*. Importantly, the *7183-I* gene rearranged much less well than the *7183-II* gene, and *81X* rearranged at only a marginally higher frequency. Therefore, for *81X*, the two changes in the spacer sequence did not significantly affect its recombination frequency. Thus, if the RSS plays a major role in controlling the relative rearrangement frequency of this V_H family in vivo, then one would predict that the *7183-II* genes would rearrange most frequently, and that *81X* rearrangement would occur at approximately the same frequency as that of the other *7183-I* genes.

Relative frequency of rearrangement of individual V_H7183 genes

Therefore, we analyzed the initial rearrangement frequency of the V_H7183 gene family members, determining only the biases inherent to the VDJ recombination process and not subsequent biases imposed by preferential selection and clonal expansion. The μ MT mice (26), which have a targeted deletion of the cytoplasmic tail of μ , have such an unbiased repertoire. Because these mice cannot express surface Ig nor can they signal through the pre-B receptor, there can be no selection of developing pre-B cells, and furthermore, there is a resulting block of B cell development at the pre-B to B cell transition step. Thus, neither selection nor environmental bias is able to shape the observed repertoire in μ MT mice. As would be anticipated in such mice, two-thirds of the sequences are out-of-frame. Using these mice, we asked the following questions. Do the $7183-I$ vs $7183-II$ RSS competition substrate results reflect the rearrangement frequency of those genes in vivo? Do all prototypic $7183-I$ (or all $7183-II$) genes rearrange at the same frequency?

To determine the V_H7183 family repertoire we used a 5' pan- V_H7183 family-specific primer (AF303) located in FR1 in a region that is identical among all the V_H7183 family members, and a 3' J_H primer (AF13N). Rearranged V_H7183 genes were amplified from newborn liver DNA and cloned into pBluescript, and individual clones were sequenced to determine their identities. The data in Fig. 3 represent 126 sequences derived from nine independent PCRs.

There is a wide range of rearrangement frequencies among the 20 genes. $81X$ clearly dominates the V_H7183 repertoire, comprising 59% of the rearrangements. Of the others, 7183.9 , $3:3.39$, 50.1 , $D6.96$, and $61-1P$ rearrange the most frequently, while other genes rearrange at a lower range of frequencies. A comparison of the usage of the $7183-I$ and $7183-II$ genes clearly revealed the surprising observation that despite the fact that the $7183-II$ RSS is better than $7183-I$ RSS in the in vitro recombination substrate assay, the genes that rearrange the most in vivo were predominantly $7183-I$ genes. The seven genes with prototypic $7183-I$ RSS account for 62% of the non- $81X$ rearrangements, while the five genes with prototypic $7183-II$ RSS together account for only 21% of the non- $81X$ rearrangements. This argues against the hypothesis that the RSS plays a predominant role in the nonrandom recombination frequency of the V_H7183 genes. Also of importance, the range of rearrangement of genes with identical RSS (shown with one or two asterisks for $7183-I$ or $7183-II$ RSS, respectively, in Fig. 3) is quite large. Two of the $7183-I$ genes showed minimal if any rearrangement (69.1 and $68-5N$), while other $7183-I$ genes were among the

five most frequently rearranging non- $81X$ genes. Surprisingly, the prototypic $7183-II$ genes were all relatively poor at rearranging in vivo. Of the genes with variant RSS other than $81X$, only $61-1P$ rearranged at a high frequency.

Mapping of the chromosomal order of the V_H7183 genes

Because the frequency of recombination of individual genes did not correlate with variation in the RSS, we wished to determine whether there was a gradient of usage from 3' to 5' throughout the locus as predicted by the chromosomal location hypothesis. The ongoing assembly of a BAC contig of the *Igh* locus of strain 129 (P. Goebel, N. Janney, W. J. Romanov, C. Hurre, and A. J. Feeney, manuscript in preparation) enabled us to map all the V_H7183 genes and correlate their relative positions with their frequency of rearrangement.

V_H7183 genes were identified and ordered using two strategies. In one approach, each BAC in a tiling path across the V_H7183 region was analyzed by PCR amplification using sets of V_H7183 -specific primers (primarily AF191 (leader) and AF192 (3' of the RSS)), followed by cloning and sequencing multiple clones to determine the content of each BAC. This approach gave rise to most of the full-length sequences. In the second approach, hybridizing bands on the Southern blots were ordered by analysis of the overlapping BACs; the corresponding restriction fragments were isolated from agarose gel, PCR-amplified, and sequenced without cloning. This approach supplemented the first for genes that did not amplify with the external primers. Additionally, most of the genes were confirmed by direct sequencing from the agarose-isolated restriction fragment. Using all these approaches, we have identified almost all *EcoRI* bands that hybridize with 7183 probes, suggesting we have essentially completed this analysis. It is likely that most of the hybridizing bands that we have not been able to amplify represent pseudogenes, which may vary significantly from our primer sequences. A map of the BAC contig and the V_H7183 gene order is shown in Fig. 4.

We found that there was no strict clustering of the $7183-I$ or $7183-II$ genes, although there are more $7183-I$ genes than $7183-II$ genes at the 3' end, which may partially explain the higher frequency of recombination of $7183-I$ genes. All the genes with variant RSS are at the 5' end of the locus, with the exception of the two most 3' genes, which are also variant. By comparing the genomic map (Fig. 4) with our repertoire data (Fig. 3) which is displayed with the genes in the correct chromosomal order, we identified an approximate gradient of rearrangement from the 3' to the 5' within the locus. $3:3.39$, $D6.96$, 7183.9 , and 50.1 , all located near $81X$ at the 3' end of the locus, are all over-represented. However, $61-1P$, which rearranges at a relatively high rate is located at the 5' end of the locus, and 7183.14 , on the 3' side, seldom rearranges.

Mapping of V_HOx2

The large V_HQ52 family is interspersed with the V_H7183 genes. One V_HQ52 gene, *Ox2*, was shown to be greatly over-represented in rearrangements from fetal liver-derived Abelson-murine leukemia virus pro-B cell lines, and also from hybridomas from LPS-stimulated adult spleen cells (32). To determine whether this gene mapped near the frequently rearranging $81X$ at the 3' end of the locus, we mapped it using primers from the published sequence of *Ox2*, followed by cloning and sequencing. We mapped *Ox2* to the most 3' BAC, 167C1, which contains $81X$, $D6.96$, and *E4.Psi* (data not shown). This supports the hypothesis that V_H genes at the 3' end of the locus undergo frequent rearrangement.

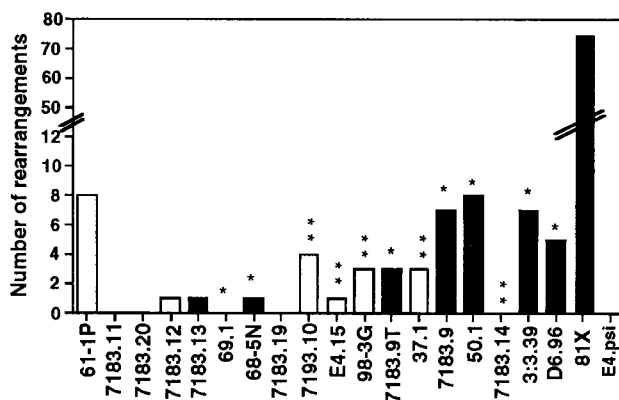


FIGURE 3. Repertoire analysis of rearranged V_H7183 genes in newborn liver. ■, genes possessing $7183-I$ RSS (prototypic and variant); □, those with $7183-II$ RSS (prototypic and variant). *, Genes with prototypic $7183-I$ RSS; **, those with a prototypic $7183-II$ RSS.

can its RSS. Thus, chromosomal location appears to bias the rearrangements in general, although location proximal to the *D* cluster is clearly not the only factor that affects recombination frequency.

All frequently rearranging genes other than *61-IP* were *7183-I* genes. These genes also are located in the 3' half of the locus, so their location may contribute to their overuse. However, *7183.14*, a *7183-II* gene, is located within this region of higher recombination, but rearranges at a much lower level, suggesting that perhaps *7183-I* genes may be more recombinogenic than *7183-II* genes. However, any difference in *7183-I* and *7183-II* use would clearly not be a result of the RSS per se, but could perhaps be due to some other commonality among the *7183-I* vs *7183-II* genes in a region outside of the portion that we have sequenced. Perhaps the *7183-I* genes have a common flanking DNA element that influences their recombination frequency. Further analysis of the sequence of the flanking regions surrounding these genes should yield insight into genomic control regions potentially located in the *V_H7183* locus. Also, because not all *7183-I* genes rearrange at high frequency, this again suggests that individual genes control their rearrangement frequency by some flanking DNA. In accord with this concept, $\gamma\delta$ T cells show strictly ordered rearrangement in fetal and adult thymus, and swapping of ~ 1 kb of 5'-flanking regions, including the promoters of the *V γ 3* and *V γ 2* genes, led to alteration of their recombination pattern (36). Also, we have recently shown that the transcription factors E2A and EBF show gene-specific preferences in their ability to induce rearrangements in transfected epithelial cells, suggesting that another factor influencing rearrangement frequency could be the presence and relative location of binding sites for transcription factors that promote accessibility to rearrangement.⁵ Perhaps the *7183* genes, which do not rearrange as often as other genes with the same RSS, such as *7183.14*, might have a deleterious change in the sequence of the binding sites for factors that control accessibility, and thus be less accessible for rearrangement.

In conclusion, we determined the relative rearrangement frequency of all members of the *V_H7183* family and showed that the individual genes rearranged with a wide range of different frequencies. We compared individual *V_H* gene family usage with RSS efficiency and chromosomal location, two factors that have been proposed to be major contributors to rearrangement frequencies of a given gene segment in vivo. We determined that the RSS was clearly not the predominant reason for preferential rearrangement, but, rather, that other factors must override the relative efficacy of the RSS. Because the mapping of all the individual genes showed that most frequently rearranging genes were located in the 3' half of the *V_H7183* region, this suggests that chromosomal location may have the most major influence on *V_H7183* gene rearrangement.

Acknowledgments

We acknowledge the excellent technical support of Elizabeth Kompfner, Paula Oliveira, and Noel Janney.

References

1. Yancopoulos, G. D., S. V. Desiderio, M. Paskind, J. F. Kearney, D. Baltimore, and F. W. Alt. 1984. Preferential utilization of the most *J_H* proximal *V_H* gene segments in pre-B-cell lines. *Nature* 311:727.
2. Perlmutter, R. M., J. F. Kearney, S. P. Chang, and L. E. Hood. 1985. Developmentally controlled expression of immunoglobulin *V_H* genes. *Science* 227:1597.
3. Malynn, B. A., G. D. Yancopoulos, J. E. Barth, C. A. Bona, and F. W. Alt. 1990. Biased expression of *J_H*-proximal *V_H* genes occurs in the newly generated repertoire of neonatal and adult mice. *J. Exp. Med.* 171:843.
4. Jeong, H. D., and J. M. Teale. 1988. Comparison of the fetal and adult functional B cell repertoires by analysis of *V_H* gene family expression. *J. Exp. Med.* 168:589.
5. Wu, G. E., and C. J. Paige. 1986. *V_H* gene family utilization in colonies derived from B and pre-B cells detected by the RNA colony blot assay. *EMBO J.* 5:3475.
6. Freitas, A. A., M.-P. Lembezat, and A. Coutinho. 1989. Expression of antibody V-regions is genetically and developmentally controlled and modulated by the B lymphocyte environment. *Int. Immunol.* 1:342.
7. Carlsson, L., C. Overmo, and D. Holmberg. 1992. Developmentally controlled selection of antibody genes: characterization of individual *V_H7183* genes and evidence for stage-specific somatic diversification. *Eur. J. Immunol.* 22:71.
8. Marshall, A. J., G. E. Wu, and C. J. Paige. 1996. Frequency of *V_H81x* usage during B cell development. *J. Immunol.* 156:2077.
9. ten Boekel, E., F. Melchers, and A. G. Rolink. 1997. Changes in the *V_H* gene repertoire of developing precursor B lymphocytes in mouse bone marrow mediated by the pre-B cell receptor. *Immunity* 7:357.
10. Yancopoulos, G. D., and F. W. Alt. 1986. Regulation of the assembly and expression of variable-region genes. *Annu. Rev. Immunol.* 4:339.
11. Corcoran, A. E., A. Riddell, D. Krooshoop, and A. R. Venkitaraman. 1998. Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor. *Nature* 391:904.
12. Thompson, S. D., J. Pelkonen, and J. L. Hurwitz. 1990. First T cell receptor alpha gene rearrangements during T cell ontogeny skew to the 5' region of the *J α* locus. *J. Immunol.* 145:2347.
13. Goldman, J. P., D. M. Spencer, and D. H. Raulet. 1993. Ordered rearrangement of variable region genes of the T cell receptor γ locus correlates with transcription of the unrearranged genes. *J. Exp. Med.* 177:729.
14. Schroeder, H. W., Jr., J. L. Hillson, and R. M. Perlmutter. 1987. Early restriction of the human antibody repertoire. *Science* 238:791.
15. van Dijk, K. W., L. A. Milner, E. H. Sasso, and E. C. B. Milner. 1992. Chromosomal organization of the heavy chain variable region gene segments comprising the human fetal antibody repertoire. *Proc. Natl. Acad. Sci. USA* 89:10430.
16. Berman, J. E., K. G. Nickerson, R. R. Pollock, J. E. Barth, R. K. B. Schuurman, D. M. Knowles, L. Chess, and F. W. Alt. 1991. *V_H* gene usage in humans: biased usage of the *V_H6* gene in immature B lymphoid cells. *Eur. J. Immunol.* 21:1311.
17. Ramsden, D. A., K. Baetz, and G. E. Wu. 1994. Conservation of sequence in recombination signal sequence spacers. *Nucleic Acids Res.* 22:1785.
18. Nadel, B., A. Tang, G. Escuro, G. Lugo, and A. J. Feeney. 1998. Sequence of the spacer in the RSS affects V(D)J rearrangement frequency and correlates with non-random *V κ* usage in vivo. *J. Exp. Med.* 187:1495.
19. Ramsden, D. A., and G. E. Wu. 1991. Mouse κ light-chain recombination signal sequences mediate recombination more frequently than do those of λ light chain. *Proc. Natl. Acad. Sci. USA* 88:10721.
20. Connor, A. M., L. J. Fanning, J. W. Celler, L. K. Hicks, D. A. Ramsden, and G. E. Wu. 1995. Mouse *V_H7183* recombination signal sequences mediate recombination more frequently than those of *V_HJ558*. *J. Immunol.* 155:5268.
21. Nadel, B., A. Tang, G. Lugo, V. Love, G. Escuro, and A. J. Feeney. 1998. Decreased frequency of rearrangement due to synergistic effect of nucleotide changes in the heptamer and nonamer of the recombination signal sequence of the *V κ* gene A2b, which is associated with increased susceptibility of Navajos to *Haemophilus influenzae* type b disease. *J. Immunol.* 161:6068.
22. Yu, C. C. K., M. Larijani, I. N. Miljanic, and G. E. Wu. 1998. Differential usage of *V_H* gene segments is mediated by *cis* elements. *J. Immunol.* 161:3444.
23. Larijani, M., C. C. K. Yu, R. Golub, Q. L. K. Lam, and G. E. Wu. 1999. The role of components of recombination signal sequences in immunoglobulin gene segment usage: a *V81x* model. *Nucleic Acids Res.* 27:2304.
24. Hesse, J. E., M. R. Lieber, K. Mizuuchi, and M. Gellert. 1989. V(D)J recombination: a functional definition of the joining signals. *Genes Dev.* 3:1053.
25. Feeney, A. J., M. J. Atkinson, M. J. Cowan, G. Escuro, and G. Lugo. 1996. A defective *V κ A2* allele in Navajos which may play a role in increased susceptibility to *Haemophilus influenzae* type b disease. *J. Clin. Invest.* 97:2277.
26. Kitamura, D., J. Roes, R. Kühn, and K. Rajewsky. 1991. A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin μ chain gene. *Nature* 350:423.
27. Schlissel, M. S., L. M. Corcoran, and D. Baltimore. 1991. Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. *J. Exp. Med.* 173:711.
28. Herring, C. D., C. Chevillard, S. L. Johnston, P. J. Wettstein, and R. Riblet. 1998. Vector-hexamer PCR isolation of all insert ends from a YAC contig of the mouse *Igh* locus. *Genome Res.* 8:673.
29. Chukwuocha, R. U., A. B. Hartman, and A. J. Feeney. 1994. Sequences of four new members of the *V_H7183* gene family in BALB/c mice. *Immunogenetics* 40:76.
30. Olo, R., J.-L. Sikorav, and F. Rougeon. 1983. Structural relationships among mouse and human immunoglobulin *V_H* genes in the subgroup III. *Nucleic Acids Res.* 11:7887.
31. Akamatsu, Y., N. Tsurushita, F. Nagawa, M. Matsuoka, K. Okazaki, M. Imai, and H. Sakano. 1994. Essential residues in V(D)J recombination signals. *J. Immunol.* 153:4520.
32. Lawler, A. M., P. S. Lin, and P. J. Gearhart. 1987. Adult B-cell repertoire is biased toward two heavy-chain variable-region genes that rearrange frequently in fetal pre-B cells. *Proc. Natl. Acad. Sci. USA* 84:2454.
33. Fanning, L., A. Connor, K. Baetz, D. Ramsden, and G. E. Wu. 1996. Mouse RSS spacer sequences affect the rate of V(D)J recombination. *Immunogenetics* 44:146.
34. Posnett, D. N., C. S. Vissinga, C. Pambuccian, S. Wei, M. A. Robinson, D. Kostyu, and P. Cancannon. 1994. Level of human TCRBV3S1 (*V β 3*) expression correlates with allelic polymorphism in the spacer region of the recombination signal sequence. *J. Exp. Med.* 179:1707.
35. Akira, S., K. Okazaki, and H. Sakano. 1987. Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. *Science* 238:1134.
36. Baker, J. E., D. Cado, and D. H. Raulet. 1998. Developmentally programmed rearrangement of T cell receptor *V γ* genes is controlled by sequences immediately upstream of the *V γ* genes. *Immunity* 9:159.

⁵ P. Goebel, N. Janney, J. R. Valenzuela, W. J. Romanow, C. Murre, and A. J. Feeney. 2001. Localized gene-specific induction of accessibility to V(D)J recombination induced by E2A and EBF in non-lymphoid cells. *Submitted for publication.*