© 1993 Oxford University Press
Nucleic Acids Research, 1993, Vol. 21, No. 3 365–372

DNA sequences at immunoglobulin switch region recombination sites

Wesley Dunnick, Gerald Z.Hertz[1], Lori Scappino[2] and Christine Gritzmacher[2,3]
Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109-0620, [1]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO, [2]Medical Biology Institute, La Jolla, CA and [3]Scripps Clinic and Research Foundation, La Jolla, CA, USA

Received October 22, 1992; Revised and Accepted December 28, 1992

ABSTRACT

The immunoglobulin heavy chain switch from synthesis of IgM to IgG, IgA or IgE is mediated by a DNA recombination event. Recombination occurs within switch regions, 2–10 kb segments of DNA that lie upstream of heavy chain constant region genes. A compilation of DNA sequences at more than 150 recombination sites within heavy chain switch regions is presented. Switch recombination does not appear to occur by homologous recombination. An extensive search for a recognition motif failed to find such a sequence, implying that switch recombination is not a site-specific event. A model for switch recombination that involves illegitimate priming of one switch region on another, followed by error-prone DNA synthesis, is proposed.

INTRODUCTION

Single immunoglobulin-producing lymphocytes (B cells) have the capability to produce one type of immunoglobulin (usually IgM) early in their development and another type (IgG, IgE, or IgA) after antigen-induced differentiation. All the types of immunoglobulin produced by the progeny of a single B cell have the same light chain and variable (VH) region of the heavy chain (excepting somatic mutations-ref. 1), but differ in the constant (C) region of the heavy chain. The switch in heavy chain C region is brought about by a DNA deletion that moves the VH gene from its position about 8 kb 5′ of the Cμ gene to a similar location 5′ of Cα, Cϵ, or Cγ genes. (There are four Cγ genes in mice, called γ1, γ2a, γ2b, and γ3.) The DNA deletion begins and ends in switch (S) regions, 2–10 kb segments that are found 5′ of each CH gene except Cδ. S regions are composed of simple sequences repeated in tandem. Murine Sμ, Sϵ, and Sα are composed of variations of pentamers such as GGGGT, GAGCT, and GGGCT; the Sγ regions are composed of repeats of a 49 or 52 bp sequence (2–7). Human S regions include similar sequences, but with a more irregular repetition pattern (8).

The switch deletion can also be thought of as a recombination event between the donor S region (usually Sμ) and the acceptor switch region (usually Sγ, Sϵ, or Sα). Switch recombination is a reciprocal event, at least to a first approximation. Not only is a breakpoint in the donor S region joined to a breakpoint in the acceptor S region, but also the ends of the deleted DNA are ligated together (9–11). The mechanism of switch recombination at the level of the chromosome and the heavy chain locus has been reviewed extensively, as has the regulation of switch recombination (12–14). We will discuss what has been learned about the mechanism of switch recombination from evaluation of DNA sequences at recombination sites.

DEFINING SWITCH RECOMBINATION SITES VERSUS SECONDARY, NON-SWITCH, DNA REARRANGEMENTS

More than 150 recombination sites in switch regions have been sequenced (Table 1). A continuing controversy has been the validity of subsets of switch recombination data. S regions might undergo secondary rearrangements, which are not related to switch recombination. These secondary rearrangements would delete the true switch recombination site and create a new site, obscuring the nature of true switch recombination sites. As discussed below, it is difficult to define characteristics that differentiate switch recombination sites from other illegitimate recombination events which are not B cell specific. Hence, it is not possible to distinguish switch recombination from secondary recombination by sequence evaluation.

Rearrangements of S regions during growth of bacteriophage clones is one type of secondary rearrangement that is known to occur (2, 49). In a few studies, the sequenced fragment has been shown to be the same size as the corresponding fragment in genomic DNA. However, for the majority of the switch region recombination sites, secondary rearrangement during recombinant clone growth or during PCR amplification is a formal possibility.

Available data suggest that secondary recombination within lymphoid cells is relatively rare. The SμSγ2b, Sμ-3′c-*myc*, and Sμ-5′ c-*myc* sites from MOPC21 have been cloned and sequenced by two or three laboratories (29, 35, 56). The three cell lines from which these recombination sites were cloned had been separated from one another by hundreds of generations. Nevertheless, the independent determination of each of the above recombination sites are identical in sequence, demonstrating that secondary rearrangments did not occur in cell lines derived from

Table 1. Complilation of DNA Sequences in Switch Regions

No.	Name	Sequence			Mutations?	Ref.
DONOR SEQUENCES:						
B cells:						
1.	RH20	GCGTGTATACAATTGTCTGG	AATTATTTCAGTTAAGTGTA	μ-μ		15
2.	GAM3-2	GGCTGAGCTGGACTCAGATG	TGCTAGACTGAGCTGTACTG	μ-γ3	7 ins.	15
3.	GAM3-8	AGTAAGCGAGGCTCTAAAAA	GCATGGCTGAGCTGAGATGG	μ-γ3		15
4.	T2-26 R	GCTGGGGTGAGCTGGGCTGA		μ-γ3		16
5.	L32-21 R		GCTGAGCTGGGGTGAGCTTG	μ-γ3		11
6.	pCS29 R	GATCGAGCTGAGCTGAGCTG	GGTGAGCTGAGCTGAGCTGA	μ-γ3		17
7.	pCS35 R	TGTGTGAGCTGAGCTGGGGT	CAGCTGAGCAAGAGTGAGTA	μ-γ3		17
8.	pCS28 R	CTGAGCTGAGCTGAGCTGGG	GTGAGCTGGGCTGAGCTGAG	μ-γ3		17
9.	pCS23 R		GAGTTGGGGTGAGCTGAGCT	μ-γ3		17
10.	RH4	CCACCCAGACCTGGGAATGT	ATGGTTGTGGCTTCTGCCAC	μ-γ1		15
11.	RH20	GAGCTGAGCTGAGCTGGGGT	GAG	μ-γ1		15
12.	RH21	TGAGGTACTGATGCTGTCTC	TACTTCAGTTATACATGTGG	μ-γ1		15
13.	pCS7 R		GGGGTGAGCTGGGCTGAGCT	μ-γ1		10
14.	pCS6 R		GCTGAGCTGAGCTGGGGTGA	μ-γ1		10
15.	pCS5 R		TGGGGTGAGCTGAGCTGGGG	μ-γ1		10
16.	pCS8 R		GCTGGGGTGAGCTGAGCTGA	μ-γ1		10
17.	pCS3 R		AGCTGAGCTGGGGTGAGCTG	μ-γ1		10
18.	pCS9 R		CTGAGCTGAGCTGAGCTGAG	μ-γ1	1 ins.	10
19.	pCS13 R		AGCTGAGCTGAGCTGGGGTG	μ-γ1		10
20.	N5-2 R	TACTTCGTTATACATGTGGG	TTTGAATTTTGAATCTATTC	μ-γ1		16
21.	N5-4 R		GTGGGGTGAGCTGGGCTGAG	μ-γ1		16
22.	148-3 R		TGAGATCTGAGCTGGGCTGA	μ-γ1		16
23.	148-21 R		CTGGGCTGAGCTGGGGTGAG	μ-γ1		16
24.	N52-1 R	GAGCTGAGTGAGCTGAGCTG		μ-γ1		16
25.	p3 R		GGGCTGAGCTGGGGTGAGCT	μ-γ2b		9
26.	pCS24 R	GAGCTGGGGTGAGCTGAGCT	GAGCTGAGCTGAGCTGGGGT	μ-γ2b		17
27.	pCS22 R	GAGCTGAGCTGAGCTGGGGT	GAGCTGAGCTGAGCTGGGGT	μ-γ2b	1 ins.	17
28.	N52-1 R		CTGAGCTGAGCTGAGCTGGG	μ-ε		16
29.	T2-11 R		AGCTGGGGTGAGCTGAGCTG	μ-α		11
30.	T2-26 R		CTGAGCTGAGCTGAGCTGGC	μ-α		16
31.	pCS18 R		TGGGCTGAGCTGGGTGAGCT	μ-α		17
32.	L32-52 R	CCAGGCTGGACGGCTCTGGG	GGTAGCTAGGGTAAGTGAGG	γ3-γ2b		11
33.	T2-7 R	AGGCTGAGCAGCTCTCAGGG	AGCTGGGGAGGTGGAGCTAT	γ3-α		11
34.	T2-2 R	GTAGTTGGGGGTGTGGGGAC	TAGGCTGGGCAGCTCTGGGG	γ3-α		11
35.	pCS19	GACTAGGTTGGGCAGCTACA	GGTGAGCTGGGTTGGATGGA	γ3-α		17
36.	N5-7 R	TGTGGGGACCCTGTAGGGTA	GCTGTAGGGAAATCAGGACA	γ1-ε		16
37.	1412-3		AGGGAGCCAGGACAGGTAGA	γ1-ε		16
38.	1412-4		AGGGAGGACAGGTAGAAGTG	γ1-ε		16
39.	T2-4 R	CAGCTAGGAGGGAGCTGGGG	CAGGTGGGAGTGTGAGGGAC	γ2b-α		11
40.	pCS43 R	ATGTGGAGGACCAGACCTAA	CAGCTAGGAGGGAGCTGGGG	γ2b-α		17
41.	pCS44 R		GGGACCTGACAGTACATCTG	γ2b-α		17
42.	pCS20 R	GAGGGACCAGTCTCAGCAGC	TAGGAGGGAGCTGGGGCAGG	γ2b-α	1 ins.	17
43.	pCS41 R	AATGTGGAGACCAGATCTA	GCAGCTATAGGGGAGCAGGG	γ2b-α	1/23	17
44.	pCS32 R	CTAGCAGCAGTGGGTGACTT	AGGAATGTTGGAAATGTGAG	γ2b-α		17
45.	pCS31 R	GCCCTGGCCTAAGTAGACTG	GGCTAGGCTGAGCAAATCTA	α-α		17
Lymphomas/leukemias:						
46.	18.81A2	GATTACTCTGAGGTAAGCAA	AGCTGGGCTTGAGCCAAAAT	μ-μ		18
47.	I.29 '	GAACTGAGCTGGGGTGAGCT	GAGCTGGGG	μ-γ3		19
48.	18.81A2	GAGCTGGGGTGAGTTGAGCT		μ-γ2b		18
49.	18.81 R	GCTAAGAATAGACTACCTGA	ATTGTGCCAAACTGGGCTGG	μ-γ2b	1/22	20
50.	300-18	ACTTAAGTTTATCGACTTCT	AAAATGTATTTAGAATTCAT	μ-γ2bI		21
51.	I.29(40)	GGGGTGAGCTGAGCTGGGCT	GGGCTGAGCTGGGCTGAGCT	μ-α		22
52.	BFO.3	TGAGCTGAGCTAGGGTGAGC	TGAGCTGGGGTGAGCTGAGC	μ-α	1/32	22
53.	I.29(Jα)	TGAGCTGGGGTGAATTGGGG	TGGGGTGAGCTGAGCTGAGC	μ-α	6/300	22
54.	300-18		GCTGAAGATGGTAGGAATGT	γ2bI-γ2b	6/378	21
55.	300-18	GTGAGGTACCAGTTCTAGCA	GCTATGGGGGAGCTGGGGAT	γ2b-γ2b		21
56.	300-18	ATATGAGGGACCAGTCTCAG	CAGCTATGGAGGAGCTGGGG	γ2b-γ2b		21
Leukemias (on retroviral vectors):						
57.	NB32	ATTACTCTGAGGTAAGCAAA	GCTGGGCTTGAGCCAAAATG	μ-μ		23
58.	NB3	CTACTGCCTACACTGGACTG	TTCTGAGCTGAGATGAGCTG	μ-γ2b		23
59.	NB27	AGATTGCGCTGGGATCAGCT	GTACTCAGATGAGCTGGGAT	μ-γ2b		23
60.	NB7	CTCAGCTATGCTACGCGTGT	TGGGGTGAGCTGATCTGAAA	μ-γ2b		23
61.	NB1	GGGCTGAGCTGGACTGAGCT	G	μ-γ2b		23
62.	NB29	TGGGGTAAGATGGGATGAGC	TGTGGTGAGGGGAGCTGGAT	μ-γ2b		23
63.	NB32	ATGAGCTGAACTGGGGTAAG	ATGGGATGAGCTGTGGTGAG	μ-γ2b		23
64.	NB32	CAGTCCTAGAAGCTATGGGG	GATCTGGAATAGGTAGTAGA	γ2b-γ2b		23
Hybridomas:						
65.	180.2B2	AACTCTCCAGCCACAGTAAT	GACCCAGACAGAGAAGGCCA	μ-μ		25
66.	KWD1	GAAAAACTAGTAAAAGAAAA	ATGTTGCCTGTTAACCAACC	μ-δ		24
67.	KWD2	GGGCCAAAGGTCTGAGACCA	GGCTGCTGCTGGGTAGGCCT	μ-δ		24
68.	KWD4	GCTGAGCTGAGCTGAGCTGG	GCTGAGCTGAGTTGAGCTGG	μ-δ		24
69.	KWD5	CTAAACTCTACTGCCTACAC	TGGACTGTTCTGAGCTGAGA	μ-δ		24
70.	KWD6	CAACTCAATGTGGTTTAATG	AATTTGAAGTTGCCAGTAAA	μ-δ		24
71.	KWD8	GCTGGGGTGAGCTGAGCTGG	GGTGAGCTGGACTGAGCTGA	μ-δ		24
72.	KWD9	TTCCTTGAAAAACTAGTAAA	AGAAAAATGTTGCCTGTTAA	μ-δ		24
73.	180.2B2	GCTGAGCTGGGGTGAGCTGA	GCTGGG	μ-γ3		26
74.	59.6C5	AAACTGAGGTGATTACTCTG	AGGTAAGCAAAGCTGGGCTT	μ-γ3		25
75.	198.5C8	GCTGGGGTGAGCTGAACTGA		μ-γ3		19
76.	470	GCAGCTGGGGTAAGCTGGAT		μ-γ3		26
77.	137.5G6	AGCTGAGCGAGCTGGGGTGA		μ-γ1		26
78.	3B12	ACTCAGTCAGTCAGTGGCGT	GAAGGGCTTCTAAGCCAGTC	μ-γ1		This pub.
79.	HB137	AACTTCATTAATCTAGGTTG	AATAGAGCTAAACTCTACTG	μ-γ1	3/341	This pub.
80.	ε-1	GGCTGAGCTGAGATGGGTGG	GCTTCTCTGAGCGCTTCTAA	μ-γ2b		7
81.	TIB141	TGGATACGCAGAAGGAAGGC	CACAGCTGTACAGAATTGAG	μ-γ2a	6/1313	This pub.
82.	2B4	TATGGATACGCAGAAGGAAG	GCCACAGCTGTACAGAATTG	μ-γ2a	3/96	27
83.	198.5C8	TGAGCACCTACAGTAGAGCT	GGGGCAGCTCTGGGGGATCT	γ3I-γ3		19
84.	470	GGGGACCAGGCTGGACAGCT	CTCGGGGGAGCTAGGGTAGG	γ3-γ1		26
85.	470	GGGGAGCCAGGACAGGTGGA		γ1-γ1		19
86.	HB137	GGAGCCAGGACAGGTGGAAA	TGTGGTGACCCAGGCAGAGT	γ1-γ1	3/279	This pub.
87.	3B12	AGGACAGGTGGAAATGTGGT	GACCCAGGCAGAGTAGCTCT	γ1-ε	2/413	This pub.
88.	HB137	AGTGCAGAGATCCAAGCTGA	GCAGCTCCAGCTTAGCTGTA	γ1-ε	4/632	This pub.
89.	26.82	TGGGATAGCTATGTGGGGAG	ACCAGGTTAAGCAAACAGTG	γ1-ε	13/148	This pub.
90.	ε-1	GCAGCTAGGAGGGAGCTGGG	GCAGGTGGGAGTGTGAGGGA	γ2b-ε		7
91.	TIB141	AGGGGCACATGGGGTCCACA	GAAACTCTAGAAACTTAGGG	γ2a-ε	7/547	This pub.
Myelomas:						
92.	MPC11	GACTCAGATGTGCTAGACTG	AGCTGTACTGGATGATCTGG.	μ-μI		28
93.	McPC603	AACAAGGTTGAGAGCCCTAG	TAAGCGAGGCTCTAAAAAGC	μ-μ		4
94.	TPC1033	GTAAGCCAGAGGCGCCACAG	CTGTGGCTGCTGCTCTTAAA	μ-δ		24
95.	TPC1017	ACAGAGAAGGCCAGACTCAT	AAAGCTTGCTGAGCAAAATT	μ-δ		24
96.	J606	AGCTGGGGTGAGCTGAGCTG	G	μ-γ3		7
97.	MOPC21	GGTGAGCTGAGCTGAGCTGA		μ-γ1		26
98.	MOPC21	TGGGGTGAGCTGAGCTGGGG	T	μ-γ2b		29
99.	MOPC141	ACTTCCTGGTTGTTAAAGAA	TGGTATCAAAGGACAGTGCT	μ-γ2b		2
100.	MPC11	TTAATCTAGGTTGAATAGAG	CTAAACTCTACTGCCTACAC	μ-γ2b		30
101.	TEPC15	TGTTAAAGAATGGTATCAAA	GGACAGTGCTTAGATCCAAG	μ-α		4
102.	McPC603	GCTGGGCTGAGCTGGGCTGA		μ-α		4
103.	MC101	GCGCTACAGCTGAGGTGATT	ACTCTGAGGTAAGCAAAGCT	μ-α	4/86	31
104.	J558	TGAGCTGATCTGAAATGAGC	TACTCTGGAGTAGCTGAGAT	μ-α		32
105.	MOPC167	AACAAGGTTGAGAGCCCTAG	TAAGCGAGGCTCTAAAAAGC	μ-α		33
106.	ABPC45	AAGAAAAGATGTTTTTAGTT	TTTATAGAAAACACTACTAC	μ-αI		34
107.	J558	GGCACCGCAAATGGTAAGCC	AGAGGCAGCCACAGCTGTGG	μ-c-myc		35
108.	MOPC21	TGGGGTGAGCTGAGCTGGGG	TG	μ-c-myc		29
109.	MOPC21		AGAGCTGAGGTGAGCTGAGC	μ-c-myc		29
110.	MPC11	GAGCACCTACAGTAGAGCTG	GGGCAGCTCTGGGGGATCTG	γ3-γ2b		5
111.	MC101	GCTGGTGTGTGCTAGGCTGG	TCTGAGCTGAGCTGGAATGA	α-γ1		36
112.	MOPC167	GGCTGAGCTGAGCTGAGCTG	AGCTGAGCTGAGCTGAGCTG	α-α	1/33	33
Hybridomas (in vitro); T cells:						
129.	9.9.2.1	GTCAAGGGAGAAAGGCATCT	AGCCTCGGTCTCAAAAGGGT	μ-γ2a		37
130.	9.7.1	CAACCCCTGTCCTCCATGCA	AGGAGTGTCACAAATGCCCA	γ2b-γ2a		37
131.	ICD137	AGCTGAGCTGAGCTGGGGTG		μ-JH		38
132.	ICD116	GTGGATTGAACTGAGCTGTG	TGAGCTGAGCTGGGGTCAGC	μ-JH		38
133.	ICD144	AGCTGAGCTGAGCTGGGGTG		μ-JH		38
134.	ICD162	CTCAGCTATGCTACGCTGTG	TTGGGGTGAGCTGATCTGAA	μ-JH		38
135.	ICD171	GCTGAGCTGAGCTGGGGTGA		μ-JH		38
136.	ICD172	GAGCTGAGCTGGGGGGGGTC		μ-JH		38
137.	ICD165	GCTGAGCTGAGCTGTGCAGG		μ-JH		38
Human cells:						
138.	SKS251	TGGGCTGAGCTAACCTGGGC	AGGGCTGAGCTGGGCTGAGC	μ-μ		39
139.	2C10	GAGCTGGGCTGGGCTGGGCT	GGGCTGAGCGGTCTAGCGGG	μ-γI	4/50	40
140.	U266	TTGAGCTTAGCTGGGCTGAG	TAACCTGGGCAGGGCTGAGC	μ-ε	33 ins.	39
141.	2C4	CGTAGGCTGGGCTGAGCTGG	GCTGAGCTAGGCTGGGCTGG	μ-ε	1/55	40
142.	SKS251	AGCTGGGCTGAGCTAACCTG	GGCTGGGCTGAGCTAACCTG	μ-ε		39
143.	SKS252	AAGCTGGGCTGAGCTGGGCA	GGGCTGGGCTGAGCTGAGCT	μ-ε		39
144.	SKS271	GAGCTGGGCTGAGCTGGGCT	GAGCAAGGCTAGGCTGAGCT	μ-ε		39
145.	SKS274	CAGGGCTGGGCTGAGCTGAG	CTGGGCTGGGCTGAGCTGGG	μ-ε	3/43	39
146	SKS275	CTTTCAGAAATGGACTCAGA	TGGCTAAACTGAGCCTAAGC	μ-ε		39
147.	SKS278	GCTGGGCTTGGCTGCACTAA	GCTGGGCTGAGCTGGGCAGG	μ-ε		39
148.	SKS281	GGCTGAGCTGGGCAGGGCTG	GGCTGAGCTGAGCTGGGCTG	μ-ε		41
149.	SKS282	AGTTGAACTGGGTTGAGCTG	AGCTGAGCTGAGCTGGGCTA	μ-ε		41
150.	SKS283	GCTGAGTTGAACTGGGTTGA	GCTGAGCTGAGCTGAGCTGG	μ-ε		41
151.	SKS285	TTCAGAAATGGACTCAGATG	GCTAAACTGAGCCTAAGCTG	μ-ε		41
152.	SKS286	ACTGGGTTGAGCTGAGCTGA	GCTGAGCTGGGCTAAGTTGC	μ-ε		41
153.	SKS287	GAGACAAAAGATGGAAGCCA	GCCTGGCTGTGCAGGAACCC	μ-ε		41
154.	SKS288	GCTGAACTGGGCTGAGTTGA	ACTGGGTTGAGCTGAGCTGA	μ-ε		41
155.	2C10	GGAAGCTCCTGGAGCTCAGA	GA	γ-ε		40
ACCEPTOR SEQUENCES:						
B cells:						
2.	GAM3-2	AGGCTGGGCAGCCTGGGGAG	CTAGGGTAGGTGGGATGTGG	μ-γ3	1/46	15
3.	GAM3-8	AGCTCTCAGGGAGCTGGGGT	GGGTGGGGTTGTGAGGACCA	μ-γ3	2/274	15
4.	T2-26 R	GGGAACCAGGCTGGACAGCT	CTGGGGAAACTGGGGTACAT	μ-γ3		16
5.	L32-21 R	CAGGCTGGACGGCTCTGGGG	GTAGCTAGGGTAAGTGAGGA	μ-γ3		11
6.	pCS29 R	TGAGCACCTACAGTAGAGCT	GGGGCAGCTCTGGGGATCT	μ-γ3	1/20	17
7.	pCS35 R	GGTAGGTGGGATGTGGGGAC	CAGGCTGGGGAGCAGCTCTG	μ-γ3		17
8.	pCS28 R	GCTGTGGGGGTGCTGGGGTA	GGTTAGAGCATGGGAACCAG	μ-γ3		17
9.	pCS23 R	GAGATATGTGGGGTTGTGGG	GAACAGGTTGGACAGCTCTG	μ-γ3		17
10.	RH4	AGCTATAGGGGAGCCAGGAC	AGGTGGAAGTGTGGTGACCC	μ-γ1		15
11.	RH20	TTCCAAACAGAAGAGCTACA	GAGGAGCCAAGACAACTAGA	μ-γ1	2/78	15
12.	RH21		GAGCCAGGACAGGTGGAAGT	μ-γ1		15
13.	pCS7 R	AGGAAAGGTGGAAGTGTGGG	GCTCCAGGCAGAGCAGTTAT	μ-γ1		15
14.	pCS6 R	AGGACAGGTGGAAGTGTGGT	GACCCAGGCAGAGCAGCTAT	μ-γ1		10
15.	pCS5 R	CCAGGCAGAGCAGCTATAGG	TTAGCCAGGACAGGTGGAAT	μ-γ1	1/24	10
16.	pCS8 R	TGTGGTGACCCAGGCAGAGC	AGCTATAGGGGAGCCAGGAC	μ-γ1		10
17.	pCS3 R	GGAAGTGTGGGGAGCCAGGC	AGAGCAGCTCCAGGGCAGCC	μ-γ1		10
18.	pCS9 R	CCAGGCAGAGCAGCTCCAGG	GCAGCCAGGACAGGTGGAAA	μ-γ1		10
20.	N5-2 R	GCAGCCAGGACAGGTGGAAG		μ-γ1		16
21.	N5-4 R	GATCCAGGCTGAGCAGCTCC	AGCTTAGCTGGGTAGGTGGA	μ-γ1		16
22.	148-3 R	CAGCTCCAGGGCAGCCAGGA	CAGGTGGAAGTGTGGTGACC	μ-γ1		16
23.	148-21 R	TGGAAGTGTGGTGACCCAGG	CAGAGCAGCT	μ-γ1		16
24.	N52-1 R	CAGGCAGAGTAGCTATAGGG	AGCCAGGACAGGTGGAAGTG	μ-γ1		16
25.	p3 R	AAGGACCAGAACTAGCAGCT	GTGGGAGAGCTGGGGATGGT	μ-γ2b		9
26.	pCS24 R	GGCAGCTGGGGATGGTAG	GAATGTGGAGGACCAGACCT	μ-γ2b		17
27.	pCS22 R	GATAGGAATGTGGAGGACCA	GTCCTAACAGCTAGGAGGGA	μ-γ2b	1 ins.	17
32.	L32-52 R	GATGATAGGAATGTGGAGGA	CCAGTCCTAACAGCTAGGAG	γ3-γ2b		11
28.	N52-1 R	CTGAGCTGAGTCAAGATGGT	CTGAGTTGATTTGAGTTGGC	μ-ε		16
36.	N5-7 R	AAGTTAGGAGGGACTTGGCT	TGGCTTAGCTGGGCCAAGCT	γ1-ε		16
37.	1412-3	GCTGTATGAGCTGTCTATCC		γ1-ε		16
38.	1412-4	TGGTATGAGCTGGTCTAATC		γ1-ε		16
29.	T2-11 R	GGCTAGGCTGAGCTGAGCTG	GGAATGAGCTGGGAT	μ-α		11
30.	T2-26 R	ACTTGGGTGAGCTAACTAGT	CTGATGGAGTACTGA	μ-α		16
31.	pCS18 R	TGAGCTGGAATGAGCTGGGA	TGGGCTGAACTAGGCTGGAA	μ-α		17
33.	T2-7 R	TGAGCTGAGCTAGGCTGAGC		γ3-α		11
34.	T2-2 R	ATGAGCTGGGATGGGCTGAA	CTAGGCTGGAATAGGCTGGG	γ3-α	1/30	11
35.	pCS19 R	ACTGGGCTAGGGTTGGATGG	GCTCAATAACTGGGCTAATC	γ3-α		17
39.	T2-4 R	CTGAGCTGGAATGAGCTGGG	ATGGGCTGAGCTAGGCTGGA	γ2b-α		17
40.	pCS43 R	ACTGGCTGGGCTGGAATTTG	CTGGGCTGTGCTGAGCTGGG	γ2b-α		17
41.	pCS44 R	GGATGAGGTGGACTGAGCTG	GGCTAAGCTAAATTAAGCTG	γ2b-α		17
42.	pCS20 R	CTACTCTGGCATGGTCTGGG	CTAGGCTAGAATGGACTGAG	γ2b-α	1 ins.	17
43.	pCS41 R	AGCTGGAATGAGCTGGGATG	GGCTGAACTAGGCTGGAATA	γ2b-α		17
44.	pCS32 R	AGCTGGGTTAGGCTGAGCTG	AGCTGAGCTGAGCTGAGCTG	γ2b-α		17
45.	pCS31 R	TGAGCTGGGTTAGGCTGAGC	TGAGCTGAGCTGGAATGAGC	α-α	2/21	17

Lymphomas/leukemias:						
46.	18.81A2		AGCTGAGCTGAGCTGCGGTT	μ-μ		18
47.	I.29	CTGCCTGAAGGGCCACAGGG	GAGCTGGGGCTATCAGATCA	μ-γ3		19
48.	18.81A2	CAGTCCCAGCAGCCGTGAAG	GAGCTGGGGATGGTAGGAAT	μ-γ2b		18
49.	18.81 R	GAAAATCTGTGGTGCAACTG	.TAACAGGTGGGGTGTGAGGA	μ-γ2b		20
50.	300-18	ACTACTCCTAGCAGCAGTGA		μ-γ2bI		21
54.	300-18		GAAGCTATGGGGGAGCTGGG	γ2bI-γ2b	4/294	21
55.	300-18	AGAGACCAGTCCCAGCAGCC	GTGAAGGAGCTGGGGATGGT	γ2b-γ2b		21
56.	300-18	ATGTGGAGGACCAGACCTAA	CAGCTAGGAGGGAGCTGGGG	γ2b-γ2b		21
51.	I.29(40)	GAGCTGGGCTAGGCTGAGCT	GAGCTGGAATGAGCTGGGAT	μ-α		22
52.	BFO.3	TAGGCTGGGCTCGGCTGGTG	TGAGCTGAGCTAGGCTGAGC	μ-α	1/45	22
53.	I.29(Jα)	TAGGCTGGAATAGGCTGGGC	TCGGCTGGTGTGAGCTGAGC	μ-α	6/360	22
Leukemias (on retroviral vectors):						
57.	NB32		GCTGGGCTGAGCTGGACTGA	μ-μ		23
58.	NB3	GTATGGTTTGAATAGGGGAC	TATATCTAGCAGCTATGGGG	μ-γ2b		23
59.	NB7	ATGGTAGCAATGTGGGGAAC	CAGTCCTAGAAGCTATGGGG	μ-γ2b		23
60.	NB27	GAAATTTGCACATCCAGTTC	TAGA	μ-γ2b		23
61.	NB1	GAAGATGGTAGGAATGTGGA	GGACCAGACCTAACAGCTAG	μ-γ2b	1/28	23
62.	NB29	GACCAGATCTCGCAGCTATG	GAGGAGCAGGGATAGGTGGA	μ-γ2b		23
63.	NB32	GGGACTATATCTAGCAGCTA	TGGGGAAGCAAGGATAGGTG	μ-γ2b		23
64.	NB32	ATGGAGCTGGAGAAGGTGGG	AATATGAGGGAGAAGTCCTA	γ2b-γ2b	22 ins.	23
Hybridomas:						
65.	180.2B2		CTGAGCTGGGTGAGCTGAGC	μ-μ		25
73.	180.2B2	GTTAGGAGTGTAGGGACCAG	GCTGGGCAGCTCTGGGGGAG	μ-γ3		25
74.	59.6C5	GACCAGGCTGAGCAGCTCTC	AGGGAGCTGGGGAGGTGGAG	μ-γ3		25
75.	198.5C8	GCTGGACAGCTCTGGAGGGA	GCTAGGATAAGTGAGGATGT	μ-γ3		19
76.	470	GGAGTGTAGGGAGCAGGCTG	GACAGCTCTGAGGGGAAGCT	μ-γ3		26
83.	198.5C8	CTCTCAGGGAGCTGGGGAGG	TGGAGCTGTGGGGACCAGGC	γ3I-γ3		19
77.	137.5G6	CAGGCAGAGCAGCTATAGGG	GAGCCAGGACAGGTGGAAGT	μ-γ1		26
78.	3B12	AGCAGCTCCAGGGGAGCCAG	GACAGGTGGAAGTGTGGAGA	μ-γ1	2/311	This pub.
79.	HB137	GTGGAAGTGTGGTGACCCAG	GCAGAGCAGCTCCAGGGCAG	μ-γ1		This pub.
80.	ε-1	GATCTAGCAGCTGTAGGGGA	GCAGGGATAGGTGGAGTGT	μ-γ2b	6/395	7
81.	TIB141	CAGAGAACAAGGACGAGTGG	TGCTATGGAGCAAAGTCAAG	μ-γ2a		This pub.
82.	2B4	GGAACGTAGGTGATCATGGA	GAAAATCTATGGTGCAACTG	μ-γ2a		27
84.	470		TGGAAATGTGGTGACCCAGG	γ3-γ1		26
85.	470		CTATAGGGCAGCCAGGAAAG	γ1-γ1	4/1092	19
86.	HB137	GTGGGAGTGTGGGGATCCAG	GTAAGGCTGGACTGGGGAGC	γ1-γ1		This pub.
87.	3B12	TGGGGGGTATACTCAGCTGA	GATGGGCTGGTCTGGCTAGA	γ1-ε		This pub.
88.	HB137	GGACTGAGCTAGGCTGTACT	GGTCTGAGCTAAACTAAGTT	γ1-ε		This pub.
89.	26.82	CAGGGCTGGACTGAGCTAGC	TTTTGTATATTCGGTTGAAA	γ1-ε		This pub.
90.	ε-1	TGTACTGGTCTAAGCTTAGT	TTAGCTGAGAGGGGCTGGGC	γ2b-ε		7
91.	TIB141	TGGAATGGGTTGAGCTGGCT	GAACTGGGCTGAACTGAGAT	γ2a-ε		This pub.
Myelomas:						
92.	MPC11		GGGGTGAGCTGAGCTGAGCT	μ-μI		28
93.	McPC603		CGTGAGTCTGAGCTGGGGTG	μ-μ		4
96.	J606	GCTGAGATATGTGGGGTTGT	GGGGAACAGGTTGGACAGCT	μ-γ3	7/97	7
97.	MOPC21		TTAGTTATAGGGAGCCAGGA	μ-γ1		26
98.	MOPC21	CTCAGCAGCTAGGAGGGAGC	TGGGGCAGGTGGGAGTGTGA	μ-γ2b		29
99.	MOPC141	AGTCCTGGGGGCCAGGAGAG	TTGTCCGATTGAGCAGGAAC	μ-γ2b		2
100.	MPC11	GTGACTTGCAGATGTTGGAA	ATGTGAGGTACCAGTCCTAG	μ-γ2b	4/69	30
110.	MPC11	AAGCTAGGGGGAGCGGGGAT	AGGTGGGAGTATTAGGGACT	γ3-γ2b	2/813	5
111.	MC101	CACCAGGGAGCTGGAGCTGA	TGGGTATAAAAAGGTACCAG	α-γ1		36
113.	MOPC21	AAGTGTGGTGACCCAGGCAG	AGCAGCTATAGGGGAGCCAG	γ1-E.Tn		42
114.	653-1	GCTCCAGGGGAGCCAGGACA	GGTAGAAGTGTGGTGACCCA	γ1-ODC		43
115.	MOPC21	GAGAGCTCAAAAACCAGGAT	AACCCTGAGCAGACGTGAGT	γ2a-E.Tn		43
116.	MOPC21	CTGGGGCAAAGAGAGATGCC	AGGAATGACTGAGGGTCTAC	γ2a-E.Tn		43
117.	MPC11		CTCTGGGTGGGTAACTAGGC	γ2a-c-myc		45
118.	MPC11	GTAGACAGATAAGCTCTGGT		γ2a-c-myc		45
101.	TEPC15	TGGGCTGAGCTGGAATGAGC	TGGGTTGAGCTGAACTAGAT	μ-α		4
102.	McPC603	GCTAGGCTGGAATAGGTTGG	GCTGGGCTGGTGCGAGCTGG	μ-α		4
103.	MC101	CTGAGCTGAGCTGAGCTGGA	ATGAGCTGGGATGAGCTGAG	μ-α	7/556	31
104.	J558	AGCACTGTCTGGCTAGGCTG	TACTGGAATGAGCTGAGCTG	μ-α		32
105.	MOPC167	CTGGAATGAGCTGGGATTGG	CTAGAATAGGCTGGGCTGGA	μ-α	3 ins.	33
106.	ABPC45	GGGTGAGCTGGAATGAGCTG	GGATGAACTGAGCGAGGCTG	μ-αI		34
112.	MOPC167	CTAGGTTGAGTCTAGCGGA	AGCTGGAATGAGCTGGGATG	α-α		34
119.	MOPC167	AGCTGGGATGGACTAGGATA	AACTAAGCTGGGATGAGACA	α-c-myc	1/193	46
120.	McPC603	GGCTGAGCTGGAATGAGCTG	GGTTGAGCTGAACTAGTATA	α-c-myc		46
121.	J558	TCGCTAGGCTGTACTGGAAT	GAGCTGAGCTGAGCTGGGAT	α-c-myc		47
122.	ABPC45	GCCTAGACTGGTCTGACGCG	GGCTAATCTGGGATGAGTGG	α-c-myc		34
123.	W267	GCTGGGCTGGGCTGGTGTGA	GCTGGGCTAGGCTGAGCTGA	α-c-myc		35
124.	W267	GCTAGGCTGAGCTGAGCTGG	AATGAGCTGGGATTGGCTAG	α-c-myc		35
125.	HOPC1	GAGATTGAACCATAATGAGC	TGGGATGAGCTGGGATGAGC	α-c-myc		35
126.	HOPC1	GCTAGGCTGGAATAGGTTGG	GCTGGGCTGGTGTGACAGCT	α-c-myc		35
127.	MOPC315	CTGGAATGAGCTGGGATTGG	CTAGAATAGGCTGGGCTGGA	α-c-myc	6 ins.	48
128.	MOPC104E	TGCAGGTCGACTCTAGAGGA		α?-c-myc		48
Hybridomas (in vitro):						
129.	9.9.2.1	GTACACATCTAAGGCCTCTA	AGCAAACACCTAAAACAGGA	μ-γ2a		37
130.	9.7.1	GAAAGAGAGTAAATGCTCCC	AGCAGCTGTGGGACAGATGG	γ2b-γ2a		37
Human switches:						
138.	SKS251	TGGGCTGAGCTAACCTGGGC	AGAGCTGAGCTGGGCTGAGC	μ-μ		39
139.	2C10	GGCAGGAGGATCAGCTAGCT		μ-γI		40
140.	U266	TGGGCAGAGCGGGGCTGGAT	ACTGTGATTTTGGGGGTACC	μ-ε		39
141.	2C4	ACTGAGTTCTGCTGGGATAA	GCTGATCTACTCTGGGCTGA	μ-ε	2/71	40
142.	SKS251	GCAACTGTATTCAGCTGGGC	TGGCCTGGCCTCGGCTGGGC	μ-ε		39
143.	SKS252	CGGCTGGGCTGGGTTCAGCT	GGGCTGGGCTGAGCAGGCCT	μ-ε	1/27	39
144.	SKS271	GGGCTGAGCTGGGTTGGGTG	GAAATGGGCTGAGCTGAGCT	μ-ε	1/22	39
145.	SKS274	AGCTGGGCTGAACTGGGCTG	CTGGGCTGGACTGGGTAAGC	μ-ε		39
146.	SKS275	GTATTACCTGCTGAGTTACC	TGACCTGGCCTGAGTTCAGC	μ-ε	3/34	39
147.	SKS278	GCTAGGCTAAACTGGGTTTG	GCTGGGCTGGGCTGGGCTGG	μ-ε		39
148.	SKS281	GTTCTGCTGGGATAAGCTGA	TCTACTCTGGGCTGAACTGA	μ-ε	1/23	41
149.	SKS282	GACTAAGCTGGTTTGGGCTA	AACTGGGTGAGCTGGGGAGG	μ-ε	1/25	41
150.	SKS283	TGGGCTGAACTGAGCAGAGC	TGAACCTAGCTGGGCTGGGC	μ-ε		41
151.	SKS285	TGGGCTAAACTGGGTGAGCT	GGGGAGGACTAGGCTGGGTG	μ-ε		41
152.	SKS286	CTGAGTCAGAAGCAAGCTGG	TTGAGCTGGCTGGACTGAAA	μ-ε	1/23	41
153.	SKS287	ACCTAACCTGGGTTCAGCTG	GCCTAGGTTGGGCTGGGTTG	μ-ε		41
154.	SKS288	GGCCAACTGGACTGAGGTGG	ATGGAGCTGGGCTGAGCTGG	μ-ε		41
155.	2C10	GGCTGAGCTAAATGGGATTG	AGCTGAGGAGGCTAGGCTGG	γ-ε	3/61	40

Each sequence is assigned a number (left most column) for purposes of reference; the numbers of the acceptor and donor sequence are the same for the same recombination site. Each sequence is named by the cell line from which it was obtained or by the name of the corresponding molecular clone. 'R' indicates the reciprocal product of recombination (*i. e.*, a deletion circle). Germline sequences are presented; bp which are changed in the actual recombination site are underlined. Sequences that do not fill up the space are those for which the germline sequence is not available. The breakpoint in the sequence represents the 5′-most recombination site; bp which could be assigned to either the acceptor or donor sequences are italicized. The nature of the recombination site is indicated in the third column from the right. 'I' indicates that one of the switch regions is inverted in the recombination event. Numbers in the 'Mutations' column indicate the number of discrepancies as compared to the germline sequence/number of bp examined. For most recombination events, this data is not available. The sequence of the inserted bp, which are designated by 'ins.', are not shown. References for germline sequences: $S\mu$ (2, 52); $S\gamma 3$ (6, 25); $S\gamma 1$ (49); $S\gamma 2b$ (5); $S\gamma 2a$ (7); $S\epsilon$ (50, 51) and $S\alpha$ (4, 17, 31, 34). Sequences 78, 79, 81, 86, 87, 88, 89, and 91 are presented for the first time in this publication. DNA clones containing rearranged $S\epsilon$ sequences were isolated from IgE-producing hybridomas 71−3B 12.1 (called 3B12, ref. 50, a gift from Ann Feeney), IGEL b4 (ref. 53, ATCC #TIB141), SE 1.3 (ATCC #HB137), and DNP-ϵ-26.82 (called 26.82, ref. 54, a gift from Fu-Tong Lui). Size selected *Eco*RI fragments were isolated and cloned into the vector gZAP (Stratagene). Subclones in Bluescript were sequenced by the dideoxy method (55). Germline $S\gamma 2a$ sequences were obtained from the plasmid p$S\gamma 2a$-1, a gift of Ken Marcu.

MOPC21. The MOPC21 $S\mu S\gamma 1$ recombination site has been cloned from various cell lines three times with no change in sequence at the recombination site (3, 26, 57). Similarly, identical $S\alpha$-c-*myc* sites from the J558 myeloma have been sequenced twice (35, 47). Finally, there are many examples of independently derived hybridomas from a single mouse that share somatic mutations in the VH or VL regions. These hybridomas must have been derived from sister B cells which arose from antigen driven expansion of a single B cell. These hybridomas almost always share one or more recombined switch regions of identical size, as determined by a Southern blot (58−62). Thus, secondary rearrangements of these switch regions did not occur during the expansion of the B cell clone *in vivo*, nor during the expansion of the hybridoma *in vitro*. Therefore, we see no compelling reason to suspect that either myeloma or hybridoma recombination sites are more likely to be derived from secondary rearrangements than are sites derived from other sources.

CHARACTERISTICS OF SWITCH RECOMBINATION SITES

Switch recombination occurs within the tandemly repeated sequences; $S\mu$ is a noteworthy exception

With a few exceptions, all of the recombination sites within $S\gamma$, $S\epsilon$, and $S\alpha$ regions occur within the tandemly repeated elements (Table 2). Recombination sites are found throughout the tandemly repeated elements, at the 5′end, middle, and 3′ end of the switch region.

Recombination sites within $S\mu$ do not always fall within the tandemly repeated elements. About 40% of $S\mu$ recombinations fall outside of the tandemly repeated sequences; most of those are 5′ of $S\mu$. $S\mu$ recombination sites from different sources are similar in this regard. For example, 15 of 25 B cell and hybridoma recombination sites are outside of the tandem repeats and 9 of 14 myeloma sites are outside of the $S\mu$ tandem repeats.

Table 2. Murine switch recombination sites within or outside of tandemly repeated sequences

	Within $S\mu$ repeats	Outside $S\mu$ repeats	Within $S\gamma, S\alpha, S\epsilon$ repeats	Outside $S\gamma, S\alpha, S\epsilon$ repeats
All murine sequences	57*	35	119	14**
Deleted circles	30	2	46	3

*Number of recombination sites in the indicated sequences.
**Of these 14, seven are recombinations to the γ2a gene. Three of the 119 recombinations within the $S\gamma$, $S\alpha$, or $S\epsilon$ repeats are to the γ2a gene.

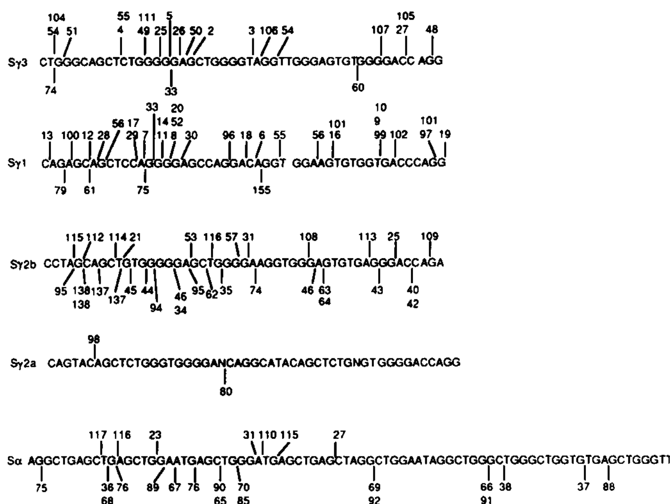


Figure 1. Location of murine recombination sites within the switch region consensus tandem repeat. Murine switch region recombination sites, with numbers corresponding to those in Table 1, are located on the various consensus sequences (determined from all the available sequence information) for the tandemly repeated elements found in $S\gamma3$ (25), $S\gamma1$ (49), $S\gamma2b$ (5), $S\gamma2a$ (7), and $S\alpha$ (31). The 5′ end of each tandem repeat is arbitrarily defined; any bp position could have been designated the 5′ end. For sites with uncertainty in the point of recombination due to sharing of bp between the donor and acceptor switch regions, the 5′ most site is indicated. Nineteen sites could not be placed due to poor similarity with the consensus sequence. Eleven were $S\alpha$ recombinations that occurred within the common pentamer elements associated with $S\alpha$, but the arrangement of these pentamers does not fit the reported consensus sequence. Six of the sites that could not be placed were recombinations to the $S\gamma2a$ gene.

Switch recombination sites from deleted circles show a more skewed distribution; almost all $S\mu$ recombination sites fall within the tandemly repeated sequences (Table 2). This might mean that switch recombination naturally favors the tandemly repeated sequences of $S\mu$ and that the deleted circles reflect the natural event better than do hybridoma, lymphoma, or myeloma data. However, the skewed distribution might also reflect the choice of restriction fragments and bacteriophage vectors used to clone these recombination sites. Phage λ has a very strong preference for optimum size of its DNA (63). There might be strong selection against inserts with recombination sites 5′ of $S\mu$; these inserts would be as much as 6 kb larger than those with recombination sites in the tandemly repeated sequences. This problem is illustrated by the more unusual results obtained by cloning of $S\mu S\alpha$ switches via *Xba*I sites into gZapII (17). These cloned inserts have recombination sites, not in the tandemly repeated sequences, but 3′ of $S\mu$ (and in the 5′ end of $S\alpha$). It is not clear

Table 3. Sharing of base pairs between donor and acceptor sequences at switch region recombination sites

none:41(59)*	1 bp:27(29)	2 bp:15(11)
3 bp:5(3.6)	>3 bp:16(1.6)	

*Number of recombination sites for which the indicated number of bp are shared by the donor and acceptor sequences at the site of recombination. This is also the number of italicized bp in Table 1. Shown in parentheses is the number of sequences that would share, by chance, the indicated number of bp. This was estimated by multiplying the probability of sharing the indicated number of bp (and not more or less) by the total number of sequences (104) by the number of ways to share the indicated number of bp given a defined recombination site. For sharing of 2 bp, the estimate is $3/4 \times 1/4 \times 1/4 \times 3/4 \times 104 \times 3$. That is, there are 3 ways to share 2 bp (1/4 chance of sharing one bp), but not share the adjacent bp (3/4 chance of two bp being different). The 3 ways are: the two bp 5′ to the recombination site, the two bp on either side of the recombination site, and the two bp 3′ of the recombination site.

whether these small inserts arose because switch recombination in B cells favored these parts of the S regions or because this vector is designed to accomodate relatively small inserts (64).

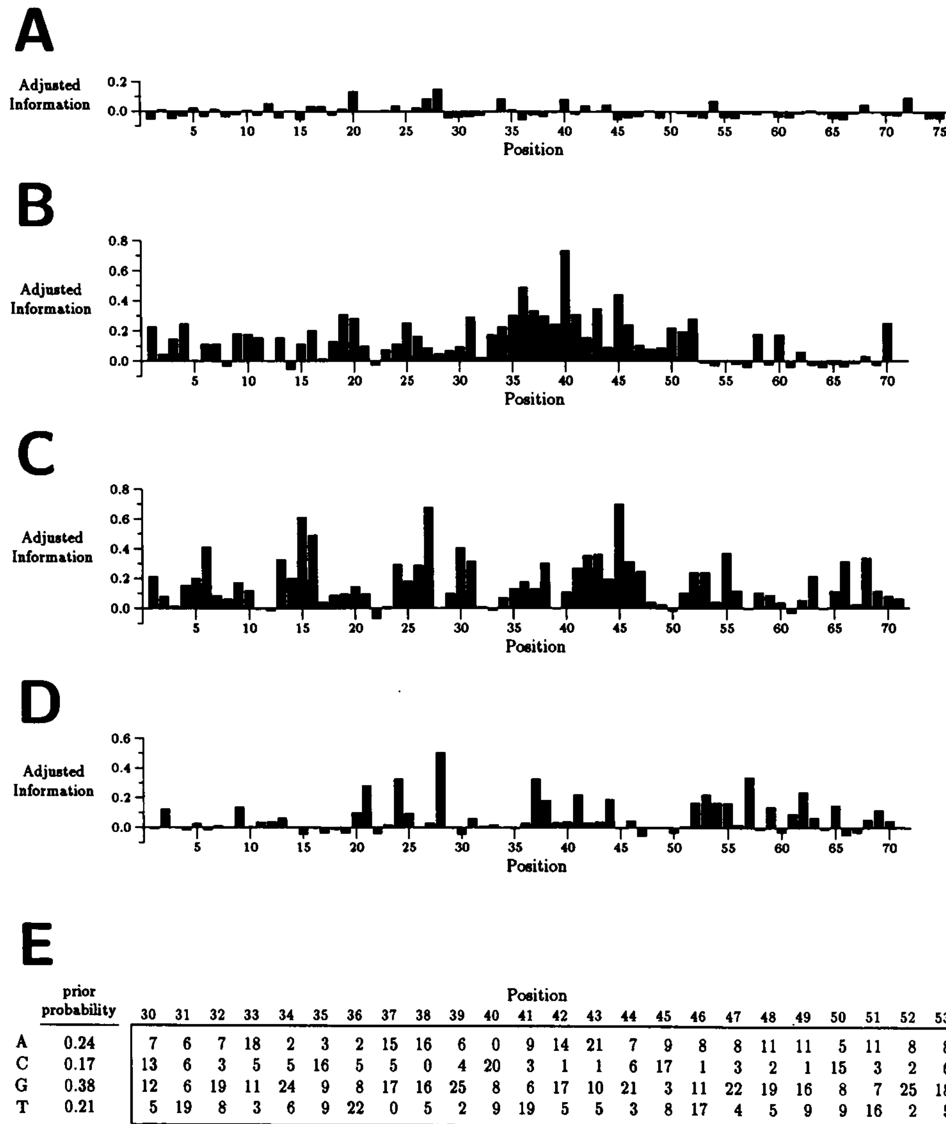
One must consider the possibility that switch recombination in $S\mu$ might be different than switch recombination in other switch regions. The important role of chromatin configuration and accessibility in the regulation of switch recombination has been established by several different results (reviewed in 12–14). Among the heavy chain genes, μ is unique in that it has a very strong enhancer at its 5′ end. Perhaps the intronic immunoglobulin enhancer plays a role in directing switch recombinase to the 5′ end of $S\mu$.

Switch recombination occurs throughout the unit tandem repeat

By locating recombination sites on a consensus tandem repeat, one can pose a second question: Is any part of the consensus tandem repeat used preferentially in recombination? There are more sites found in one half of the tandemly repeated element compared to the other half. (Figure 1—In this Figure more recombination sites are found in the left half than in the right half of the tandemly repeated elements. The start and end of the repeated unit are defined arbitrarily; more sites could be found in the middle or right end, depending on where one chose to 'begin' the element. It is important to note that the four $S\gamma$ elements are aligned, by allowing one bp gaps, to one another.) However, sites in the right-hand end of the consensus element are used, and it is possible that recombination is random with respect to position within the tandemly repeated elements.

Switch recombination sites frequently lack donor-acceptor homology

At those recombination sites for which both donor and acceptor sequences are known, 39% share no bp, and an additional 25% have a single bp shared between donor and acceptor sequences at the recombination site (Table 3). Furthermore, if homologous pairing were important in switch recombination, recombination would have occurred with the tandemly repeated elements from the donor and acceptor S regions lined up 'in register' so that the homology between the elements would be maximized. Recombination would then occur between (for example) bp 23 of the donor tandem repeat and bp 24 of the acceptor tandem repeat. In most $S\gamma$-$S\gamma$ recombination sites studied (sequences 32, 54, 64, 83, 84, 85, 110 in Table 1), recombination occurred between disparate bp of the donor and acceptor tandemly repeated



	prior probability	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
A	0.24	7	6	7	18	2	3	2	15	16	6	0	9	14	21	7	9	8	8	11	11	5	11	8	8
C	0.17	13	6	3	5	5	16	5	5	0	4	20	3	1	1	6	17	1	3	2	1	15	3	2	6
G	0.38	12	6	19	11	24	9	8	17	16	25	8	6	17	10	21	3	11	22	19	16	8	7	25	18
T	0.21	5	19	8	3	6	9	22	0	5	2	9	19	5	5	3	8	17	4	5	9	9	16	2	5

Figure 2. Computer assisted search for a switch recombination site consensus sequence. **A.** Information content of each residue 1–75 for aligned donor sequences from B cells and hybridomas. The sequence data used in this analysis extended beyond that shown in Table 1. **B.** Information content for the same sequences as in Part A., but after allowing up to 5 bp slippage. **C.** Information content for myeloma and lymphoma acceptor sequences, allowing up to 5 bp slippage. **D.** The sequences in B were randomized and then aligned as in B to optimize information content. **E.** The distribution of bp at positions 30–53 for the alignment graphed in B. The formula for information content is $\Sigma_{i=AGCT} f_i \log_2 f_i/p_i$, in which f_i is the observed frequency for base i at the indicated postion and p_i is the prior probability for base i. We used the bp composition of the subset of sequences being analyzed as the 'prior distribution'. In addition, we adjusted the information content for sample size by subtracting the average information expected from random sequences having the designated prior distribution and being arbitrarily aligned; thus, the adjusted information content of arbitrary alignments will average to zero (68). We also analyzed the data using different subsets, Sμ and non-Sμ. We also aligned sequences allowing unlimited sliding throughout the 75 bases surrounding the recombination sites using a modification of the method of Hertz *et al.* (67). Both of these analyses resulted in the identification of positions of high information content, but the motifs identified were once again those that are prevalent in S regions. The motifs in the latter analysis were also placed randomly relative to switch recombination sites.

elements. Recombination occurred at homologous positions in the donor and acceptor tandem repeats in only two SγSγ sites (sequences 55 and 56—Figure 1).

Switch recombination sites are often associated with mutations

Mutations are found in some recombined switch regions. This was recognized when two products of a single switch recombination event were isolated (22). The two products differed by several single bp changes, even though the SμSα recombination sites were identical for both products. Subsequent analysis revealed that, in the Sμ sequences, one product had all the mutations, the other had none (63). In the compilation of switch recombination sites (Table 1), there are many recombined S sequences with mutations relative to the germline. Importantly, in a series of SγSϵ recombination sites (sequences 87, 88, 89, 91), the Sγ sequences are always mutated, whereas the Sϵ sequences are not. Other switch recombination sites demonstrate this characteristic of mutations on one side of the recombination site, but not the other.

Table 4. Putative recognition sequences at switch recombination sites

Sites with GAGCT or GGGGT	60
Sites with YAGGTTG	39
Neither	67

Switch region recombination sites were scored as having a GAGCT or GGGGT pentamer if there was a 4 out of 5 match to either pentamer in the recombination site or immediately next to it. Sites were scored as having a YAGGTTG motif if there was a 5 out of 7 match within 3 bp of the site. This is a different criteria than originally proposed (32). Using the original criteria (4 out of 7 match, 7–14 bp 5′ of the recombination site), 126 sites score as having a YAGGTTG motif. The two motifs suggested by Wuerffel and colleagues (66) for switch recombination to c3 were also evaluated. For 13 Sc3 sites, the best fit was to the motif CAGCTCTGGGGAGC; the mean identity was 9 bp, with a range of 5 to ll bp. For the other 9 Sc3 sites, the best fit was to the motif GGGGACTAACC; the mean identity was 5 bp, with a range of 3 to 9 bp.

Switch recombination sites sometimes include bp that cannot be derived directly from either germline S region. Insertions of one to 33 bp have been observed (Table 1). The larger insertions usually have some similarity to S region sequences. For example, the NB32 recombination site (sequence 64) includes 22 bp which are very similar to Sγ2b sequences, but are not identical to any cloned Sγ2b sequences inserted into this switch recombination retrovirus.

Switch recombination does not necessarily follow transcriptional polarity

There are several examples of S-S recombination in which one of the S regions is inverted. For example, in the recombined switch region in 300–18 cells, Sμ is joined to a segment of Sγ2b which is inverted. A second recombination site is found a few hundred bp downstream, where the inverted Sγ2b is joined to Sγ2b which is in the same orientation as the Cγ2b gene (sequences 50 and 54).

Switch recombination may involve multiple sites within a single switch region

Recombined S regions often have internal deletions in either the donor or acceptor S regions. This implies an internal switch recombination event between two sites in a single S region. Similarly, a few S regions include duplications, which must also result from the use of at least two recombination sites.

Switch recombination sites lack a consensus sequence

After the first few switch recombination sites had been sequenced, it was apparent that switch recombination was not a site specific recombination event akin to V-D-J joining (12–13). The well-conserved sequences used in V-D-J joining are not found at switch recombination sites. This is not surprising, as the two recombination events occur at different developmental stages and at different DNA locations. The lack of V-D-J recombination signals at switch recombination sites does not rule out a role of some of the enzymes involved in V-D-J joining; it does argue against those involved in sequence recognition.

Various short motifs have been suggested as recognition sequences in switch recombination (7, 26, 32, 66). Some of the proposed motifs are prevalent in S regions. Nevertheless, about 25% of switch recombination sites lack stringent similarity to any of the proposed motifs (Table 4). To determine if switch recombination is mediated by a more subtle recognition sequence, we aligned germline sequences, using the switch recombination sites shown in Table 1, for those B cell and hybridoma sequences for which the germline sequence was known. The information content, a measure of the divergence from a prior (random) distribution of bp, of each position in the alignment was determined (67). A high information content at a particular position means that the bp composition at that position is non-random. No significant information was detected in the sequence alignments for either the donor or the acceptor sequences from B cells and hybridomas (Figure 2A—the recombination site is between residues 32 and 33). The information content was consistent with what would be observed with random sequences being arbitrarily aligned (68).

We next tested whether small shifts in the alignments might identify a peak of information. We aligned the sequences for optimal information, allowing up to 5 bp of sliding (67). The resulting alignments (Figure 2B) had significantly more information than the alignments that did not allow sliding. There appeared to be a clustering of information in the region just 3′ of the recombination sites (which vary from residue 28 to 39), particularly in the donor sequences. A notable feature of the cluster of information is that it is based on the motif G/A A/G G C T, the common pentamer found frequently in Sμ, Sα, and Sε and less often in the four Sγs. Since GAGCT, or variants of it, are found so often in switch regions, one might expect that this motif would arise in this analysis, even if it had little to do with sequence recognition in the switch recombination. Also arguing against a role for this cluster of information in switch recombinase recognition is the fact that clusters of information that arose from the analysis of other data subsets (for example,

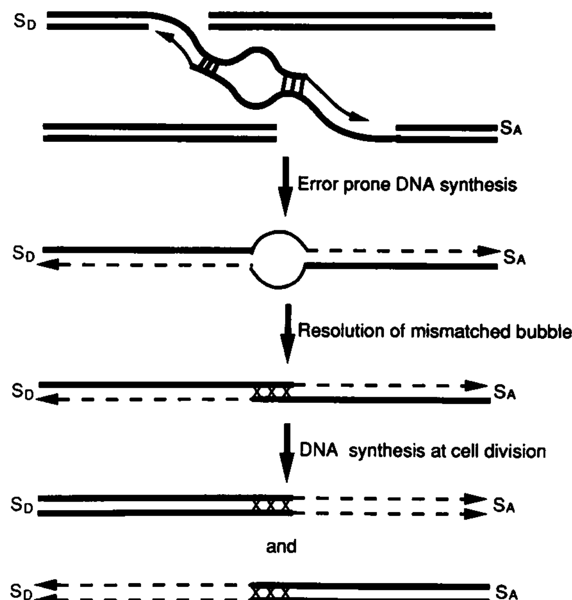


Figure 3. An illegitimate priming model for switch recombination. **A.** S_D and S_A designate the donor and acceptor S regions, respectively. Newly synthesized DNA, including mutations, is noted by a dotted line. Base pairing at sites of priming for DNA synthesis are noted by bold vertical lines. The region with 'Xs' symbolizes repair of the mismatched region at the recombination site. Formation of deletion circles would be accomplished by the symetrical reaction using the remaining two 3′ ends. **B.** The template switching model for switch recombination. Symbols used are the same. The vertical line indicates the recombination site. Note that this model predicts that recombined sequences with mutations will have mutations on both side of the recombination site. See text for further explanation.

the acceptor sequences for myelomas and lymphomas, Figure 2C) were less dramatic and located differently relative to the recombination site. To summarize, we were able to derive some interesting sequence motifs by information content analysis (including other analyses described in the Legend to Figure 2), but the characteristics of those motifs prevented a strong conclusion concerning their role in switch recombination.

AN ILLEGITIMATE PRIMING MODEL FOR SWITCH RECOMBINATION

To account for the duplications and mutations found in some recombined switch regions, it has been proposed that DNA synthesis plays a role in switch recombination (22). At least two mechanisms have been considered—template switching and illegitimate priming (65). If the switch were accomplished by template switching from donor to acceptor S region, this would imply that either both S regions or neither S region would harbor mutations (Fig. 3B). On the other hand, illegitimate priming by one S region for error-prone DNA synthesis on the other S region would predict mutations on only one side of the recombination site, the S region that was synthesized (Fig. 3A). The other side (the primer S region) would lack mutations, since it is formed from pre-existing DNA. The available data favor an illegitimate priming mechanism (see above).

To accomplish switch recombination for both strands of DNA, a reciprocal reaction must also occur (Figure 3). Since there is no *a priori* reason to suspect that the two priming reactions would occur exactly opposite one another, a bubble of two single strands, with many mismatches, would be found at the recombination site (Fig. 3A). This bubble might be repaired using one strand or the other as a template, resulting in an apparent recombination site with bp shared between the donor and acceptor switch regions. These shared bp represent the short sequence identity needed for the priming reaction. Alternatively, repair might occur from both ends of the bubble, which could lead to a apparent recombination site with no shared bp. Finally, repair could be more or less random, leading to a recombination site with several 'inserted' or mutated bp. The fact that the number of inserted bp is usually small suggests that the bubble region must be likewise small.

The compilation of sequences at switch region recombination sites is available from W.A.D. in printed form (Table 1) or on diskette. Investigators are encouraged to send corrections or additions to the compilation to W.A.D.

ACKNOWLEDGEMENTS

The authors thank Dr Michael Savageau for careful reading of the manuscript and the many workers in switch recombination for their helpful discussions. Work in the authors laboratories was supported by grants from the National Institutes of Health (CA34068 to W.A.D., HG00249 to Gary Stormo, and AI25086 to C.G.)

REFERENCES

1. French,D.L., Laskov,R., and Scharff,M.D. (1989) *Science*, **244**, 1152–1157.
2. Sakano,H., Maki,R., Kurosawa,Y., Roeder,W., and Tonegawa,S. 1980. *Nature*, **286**, 676–683.
3. Dunnick,W., Rabbits,T.H., and Milstein,C. (1980) *Nature*, **286**, 669–675.
4. Davis,M.M., Kim,S.K., and Hood,L.E. (1980) *Science*, **209**, 1360–1365.
5. Kataoka,T., Miyata,T., and Honjo,T. (1981) *Cell*, **23**, 357–368.
6. Stanton,L.W. and Marcu,K.B. (1982) *Nuc. Acids Res.*, **10**, 5993–6006.
7. Nikaido,T., Yamawaki-Kataoka,Y., and Honjo,T. (1982) *J. Biol. Chem.*, **257**, 7322–7329.
8. Mills,F.C., Brooker,J.S., and Calmerini-Otero,R.D. (1990) *Nuc. Acids. Res.*, **18**, 7305–7316.
9. von Schwedler,U., Jack,H.-M., and Wabl,M. (1990) *Nature*, **345**, 452–455.
10. Iwasato,T., Shimizu.A., Honjo,T., and Yamagishi,H. (1990) *Cell*, **62**, 143–149.
11. Matsuoka,M., Yoshida,K., Maeda,T., Usuda,S., and Sakano,H. (1990) *Cell*, **62**, 135–142.
12. Lutzker,S.G. and Alt,F.W. (1989) in Berg,D.E and Howe,M.M. (eds.) *Mobile DNA*. ASM Press, pp. 693–714.
13. Gritzmacher,C.A. (1989) *CRC Critical Reviews in Immunology*, **9**, 173–200.
14. Esser,C. and Radbruch,A. (1990) *Ann. Rev. Immunol.*, **8**, 717–735.
15. Winter,E., Krawinkel,U., and Radbruch,A. (1987) *EMBO J.*, **6**, 1663–1671.
16. Yoshida,K., Matsuoka,M., Usuda,S., Mori,A., Ishizaka,K., and Sakano,H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 7829–7833.
17. Iwasato,T., Arakawa,H., Shimizu,A., Honjo,T., and Yamagishi,H. (1992) *J. Expt. Med.*, **175**, 1539–1546.
18. DePinho,R., Druger,K., Andrews,N., Lutzker,S., Baltimore,D., and Alt,F. (1984) *Mol. Cell. Biol.*, **4**, 2905–2910.
19. Petrini,J. and Dunnick,W.A. (1989). *J. Immunol.* **142**, 2932–2935.
20. Jack,H.-M., McDowell,M., Steinberg,C.M., and Wabl,M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1581–1585.
21. Yancopoulos,G.D., DePinho,R.A., Zimmerman,K.A., Lutzker,S.G., Rosenberg,N., and Alt,F.W. (1986) *EMBO J.*, **5**, 3259–3266.
22. Dunnick, W., M. Wilson,M., and Stavnezer,J. (1989) *Mol. Cell. Biol.*, **9**, 1850–1856.
23. Ott,D.E. and Marcu,K.B. (1989) *Intl. Immunol.*, **1**, 582–591.
24. Owens,J.D., Jr., Finkelman,F.D., Mountz,J.D. and Mushinski,J.F.(1991) *Mol. Cell. Biol.* **11**, 5660–5670.
25. Szurek,P., Petrini,J., and Dunnick,W. (1985) *J. Immunol.* **135**, 620–626.
26. Petrini,J., Shell,B., Hummel,M., and Dunnick,W. (1987) *J. Immunol.*, **138**, 1940–1946.
27. Gerstein,R.M., Frankel,W.N., Hsieh,C.-L., Durdik,J.M., Rath,S., Coffin,J.M., Nisonoff,A. and Selsing,E. (1990) *Cell* **63**, 537–548.
28. Greenberg,R., Lang,R.B., Diamond,M.S., and Marcu,K.B. (1982) *Nucleic Acids Res.*, **10**, 7751–7761.
29. Neuberger,M.S. and Calabi,F. (1983) *Nature*, **305**, 240–243.
30. Lang, R.B., Stanton, L.W., and Marcu, K.B. (1982) *Nucleic Acid Res.*, **10**, 611–630.
31. Obata, M., Kataoka, T., Nakai, S., Yamagishi, H., Takahashi, N., Yamakami-Kataoka, Y., Nikaido, T., Shimizu, A., and Honjo, T. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2437–2441.
32. Marcu,K.B., Lang,R.B., Stanton,L.W., and Harris,L.J. (1982) *Nature*, **298**, 87–89.
33. Kim,S., Davis,M., Sinn,E., Patten,P., and Hood,L. (1981) *Cell*, **27**, 573–581.
34. Fahrlander,P.D., Sumegi,J., Yang,J.-Q., Weiner,F., Marcu,K.B., and Klein,G. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3746–3750.
35. Gerondakis,S., Cory,S., and Adams,J.M. (1984) *Cell*, **36**, 973–982.
36. Kataoka,T., Kawakami,T., Takahahi,N., and Honjo,T. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 919–925.
37. Eckhardt,L.A. and Birshtein,B.K. (1985) *Mol. Cell. Biol.*, **5**, 856–868.
38. Davis,D.D., Yoshida,K., Kinsbury,L., and Sakano,H. (1991) *J. Exp. Med.*, **173**, 743–746.
39. Shapira, S. K., Jabara,H.H., Thienes,C.P., Ahern,D.J., Vercelli,D., Gould,H.J., and Geha,R.S. (1991) *Proc. Natl. Acad. Sci.* **88**: 7528–7532.
40. Mills,F.C., Thyphronitis,G., Finkelman,F.D., and Max,E.E. (1992) *J. Immunol.*, **149**, 1075–1085.
41. Shapira,S.K., Vercelli,D., Jabara,H.H., Fu,S.M. and Geha,R.S. (1992) *J. Expt. Med.* **175**, 289–292.
42. Shell,B.E., Collins,J.T., Elenich,L.A., Szurek,P.F., and Dunnick,W.A. (1990) *Gene*, **86**, 269–274.
43. Katz,A. and Kahana, C.(1989) *EMBO J.* **8**, 1163–1167.
44. Elenich,L.A. and Dunnick,W. (1991) *Nucleic Acids Res.*, **19**, 396.
45. Stanton,L.W., Yang,J.-Q., Eckhardt,L.A., Harris,L.J., Birshtein,B., and Marcu,K.B. (1983) *Proc. Natl. Acad. Sci. USA*, **81**, 829–833.
46. Calame,K., Kim,S., Lalley,D., Hill,R., Davis,M., and Hood,L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6994–6998.
47. Stanton,L.W., Watt,R., and Marcu,K.B. (1983) *Nature*, **303**, 401–405.
48. Piccoli, and Cole,M.D. (1984) *Nature*, **310**, 326–329.
49. Mowatt,M. and Dunnick,W. (1985) *J. Immunol.*, **136**, 2674–2683.

50. Gritzmacher,C.A. and Lui,F.-T. (1987) *J. Immunol.*, **139**, 324–329.
51. Scappino,L.A., Chu,C., and Gritzmacher,C.A. (1991) *Gene*, **99**, 295–296.
52. Richards,J., Tucker,P., and Blattner,F., MUSIGCDO7, MUSIGCDO9 and unpublished.
53. Rudolf,A.K., Burrows,P.D., and Wabl,M.R. (1981) *Eur. J. Immunol.*, **11**, 527–529.
54. Lui,F.-T., Bohn,J.W., Ferry,E.L., Yamamoto,H., Molinaro,C.A., Sherman,L.A., Klinman,N.R., and Katz,D.H. (1980) *J. Immunol.*, **124**, 2728–2732.
55. Sanger,F., Nicklen,S., and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–6467.
56. Dunnick,W., Shell,B.E., and Dery,C. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7269–7273.
57. Schultz,C., Petrini,J., Collins,J., Claflin,J.L., Denis,K., Gearhart,P., Gritzmacher,C., Manser,T., Shulman,M., and Dunnick,W. (1990) *J. Immunol.*, **144**, 363–370.
58. Hummel,M., Berry,J., and Dunnick,W. (1987) *J. Immunol.*, **138**, 3539–3548.
59. Siekevitz,M., Kocks,C., Rajewsky,K., and Dildrop,R. (1987) *Cell*, **48**, 757–770.
60. Fish,S., Zenowich,E., Fleming,M., and Manser,T. (1989) *J. Exp. Med.*, **170**, 1191–1209.
61. Manser,T. (1990) *Int. Immunol.* **2**, 893–902.
62. Shan,H., Shlomchik,M., and Weigert,M. (1990) *J. Exp. Med.*, **172**, 531–536.
63. Feiss,M., Fisher,R.A., Crayton,M.A., and Egner,C. (1977) *Virology*, **77**, 281–293.
64. Short, J.M. (1988) *J. Mol. Biol.*, **162**, 729–773.
65. Dunnick,W. and Stavnezer,J. (1990) *Mol. Cell. Biol.*, **10**, 397–400.
66. Wuerffel, R., Jamieson,C., Morgan,L., Merkulov,G.V., Sen,R., and Kenter,A.L. (1992) *J. Expt. Med.* **176**, 339–349.
67. Hertz,G.Z., Hartzell III,G.W., and Stormo,G.D. (1990) *Comput. Appl. Biosci.*, **6**, 81–92.
68. Schneider,T.D., Stormo,G.D., Gold,L., and Ehrenfeucht,A. (1986) *J. Mol. Biol.*, **188**, 415–431.