Volume 14 Number 9 1986 Nucleic Acids Research

Palindromic sequences are associated with sites of DNA breakage during gene conversion

U.Krawinkel*, G.Zoebelein and A.L.M.Bothwell[1]

Institute for Genetics, Weyertal 121, D-5000 Cologne 41, FRG and [1]Howard Hughes Medical Institute and Department of Pathology, Yale Medical School, 310 Cedar Str., New Haven, CT 06510, USA

Received 15 January 1986; Revised 11 April 1986; Accepted 15 April 1986

ABSTRACT

Gene conversion is a recombinatorial mechanism which transfers genetic information from a donor into a recipient gene. A case of gene conversion between immunoglobulin V_H region genes was analysed and palindromic sequences were found to be located near to the left recombinatorial breakpoint, which also is flanked by a direct repeat sequence. We performed a computer search for palindromes and direct repeats in the published sequences of eucaryotic genes which had been involved in gene conversion. In these sequences, the palindrome with the best or second best quality is located near to a breakpoint of recombination. A correlation of recombination breakpoints with direct repeats was not observed. This suggests that gene conversion is promoted by palindromic sequences.

INTRODUCTION

Immunoglobulin V-region genes are organized as multigene families (V_H,- V_κ-, and $V\lambda$). V-region gene diversity can be generated somatically and in the germ-line by point mutations (1); another possibility may be recombination (2, 3). The recombinatorial mechanism of gene conversion has been suggested to play an important role in the diversification of multigene families (4, 5).

The recombinant VDJ-region of the IgD,λ_1 antibody secreted by hybridoma B1-8.V1 (6) presumably has been generated by mitotic gene conversion (7, 8) between V_H-genes V186.2 and V102.1 (9) of the C57BL/6 mouse. A segment carrying V102.1 sequence replaces the corresponding segment of gene V186.2. In this study we accurately determined one breakpoint of recombination in the recipient gene V186.2. It falls into a region where the latter gene exhibits palindromic sequences. We suggest that these sequences are capable of forming stem-loops that may have promoted

© IRL Press Limited, Oxford, England.

recombination. This suggestion is supported by finding palindromic sequences at recombination breakpoints in 8 eucaryotic genes which had undergone conversion.

MATERIALS AND METHODS

Construction and screening of a λ-phage library

Construction of a λ-phage library, library screening, restriction site analysis, Southern hybridization and construction of M13 subclones were performed employing standard methods (10). V_H-gene V102.1 (9) has been re-isolated as λ phage VAR102 from a partial MboI library of the genome of hybridoma B1-8.V1 in vector EMBL4. As a probe we utilized the 71 bp HaeIII fragment (pos. 1943-2014 in Fig. 1) of gene V102.1 cloned in a M13 phage. Lambda phage V1 is described in Krawinkel et al. (8).

DNA nucleotide sequencing

The sequence of the 5' flanking region of VDJ-region V1 in phage λV1 was obtained from HinDII and Sau3A subclones in M13 vectors (pos. 200-1341, 1602-1629 in Fig. 1). The sequence of V1 from pos. 1630-2106 is taken from Krawinkel et al. (8). V_H-gene V102.1 has been entirely resequenced from HinfI and PstI-BglII subclones of λVAR102 in M13 vectors. Nucleotide sequence analysis was performed employing the dideoxy chain termination procedure (11). Sequence V186.2 between pos. 1-1629 was determined by applying the chemical degradation procedure (12) to the left 5 kb EcoRI fragment of λV186 (9). The V186.2 sequence between pos. 1630-2106 is taken from Bothwell et al. (9).

Computing

Palindromic sequences have been detected utilizing the "stem-loop" programme provided by the University of Wisconsin Genetics Computer group (13). The rules of Tinoco et al. (14) and Aboul-ela et al. (15) are taken into account to estimate the quality of stem-loop structures: we searched for stem-loops exhibiting a maximum of Watson-Crick base-pairings and a minimum of mismatches in the stem and a loop size smaller than 30 bases. Bigger loops would destabilize a stem by at least +6 kcal/mole (14). The bonds in a stemloop are scored: 3 for GC pairs, 2 for AT pairs, 1 for GT pairs, 0 for AC pairs and -3 for purine-purine and pyrimidine-pyrimidine pairs. Our computer search for

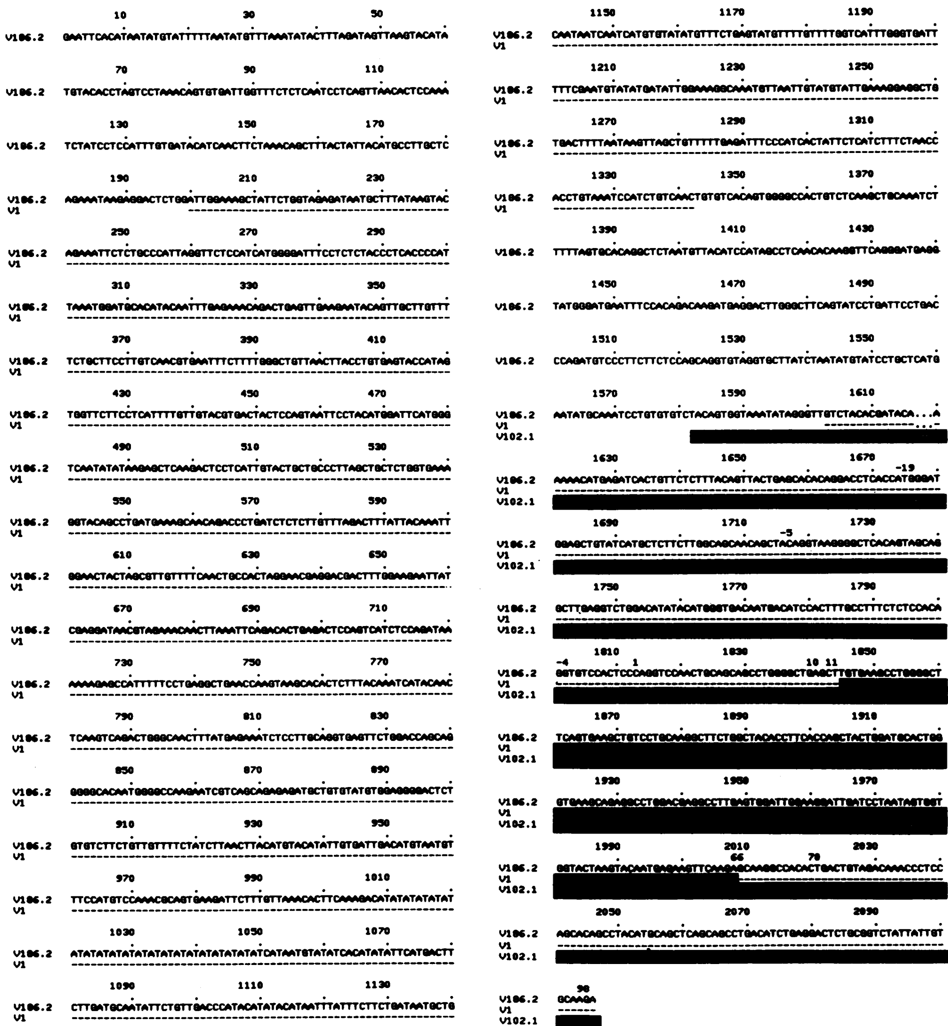

Fig. 1
Nucleotide sequence comparison of V_H-genes V186.2, V1 and V102.1. Sequences V186.2 and V1 between positions 1630-2106 are taken from Bothwell et al. (8, 9). All other sequences have been newly determined. Codon numbers are written on top of the respective codons. The V102.1 sequence is printed negatively.

palindromes was performed with the minimal stem length set to 11 basepairs in $H2K^b$, $IA_\beta{}^b$ and V186.2, $H\alpha2m2$, $H\alpha1$, $IE_\beta{}^b$, Q10, 10 bp in SUP 3, 9, 12 and 9 bp in CYC1.11 and CYC7. The minimal number of bonds/stem was set to 22, or 18 in CYC1.11 and CYC7 or

23 in $IA_{\beta}{}^{b}$, or 26 in HαA2m2 and Hα1. The maximal loop size was set to 20 bases, or 28 bases in the case of SUP9. Direct repeat sequences were detected utilizing the "seqtree" programme (16). The longest direct repeat expected to occur by chance in a sequence was predicted as described in (17).

RESULTS

Structure of V_H-segment V1

V1 represents the V_H-segment of the VDJ-region expressed by hybridoma B1-8.V1. A mechanism analogous to gene conversion appears to have generated V1 in the C57BL/6 derived Igh-locus of B1-8.V1 (8).

The potential donor gene V102.1 was isolated from the genome of hybridoma B1-8.V1 and does not show structural alterations when compared to the corresponding germ-line sequence (9).

By eliminating three sequence errors from the published sequence of V102.1 (pos. 1677, 1815 and 1842 in Fig. 1) the size of the V102.1-derived sequence tract in V1 can be determined more accurately than before. It is 165-177 bp long. The tract starts at codon 11 and ends between codons 66 and 70. The left recombination breakpoint can be assigned either to the first or to the second base of codon 11. V1 contains only one segment of sequence derived from the donor V102.1. The 5' flanking region of V1 is identical to the wild-type sequence of gene V186.2 in a region covering 1141 bp. Although the nucleotide sequence of the corresponding region of V102.1 has not been determined further upstream than to pos. 1582 one knows from polymorphism of restriction sites that the 5' flank of V102.1 is different from the one of V186.2 (18) and thus also different from the one of V1.

Inverted repeat sequences at breakpoints of recombination

V_H-gene V186.2 and eight genes that are reported to have acted as recipient genes in gene conversion were searched for palindromic sequences. The result of the search in genes V186.2, the murine histocompatibility class I gene $H2K^b$ (19, 20), the murine histocompatibility class II gene $IA_{\beta}{}^{b}$ (21-24), the human V_k-gene HK137 (25), the yeast cytochrome C gene CYC1.11 (26), the human Cα2 immunoglobulin gene Hα2m2 (27), and the yeast suppressor

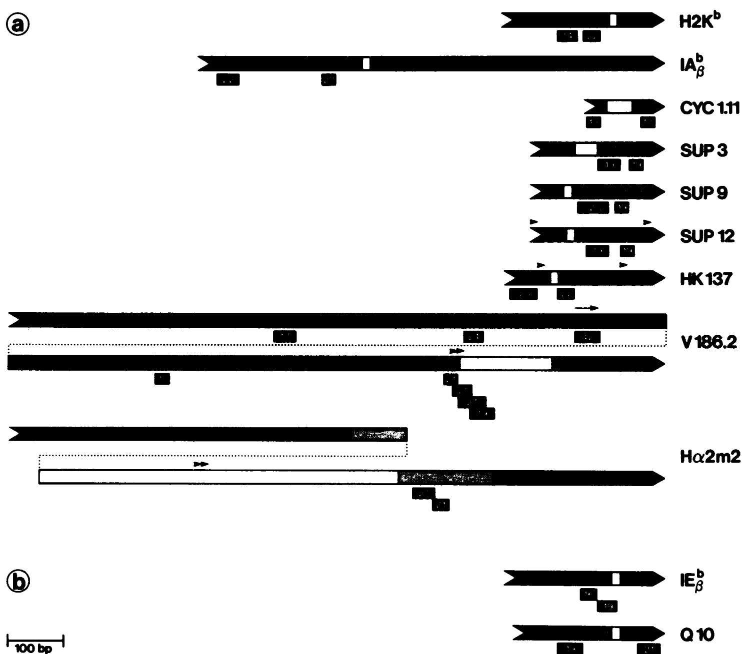

Fig. 2
a) The two respective best palindromes are shown as black boxes under genes that served as recipients in gene conversion and all palindromes in V186.2 are shown. The second best palindrome in V186.2 is located at pos. 1846-1861, 1879-1895, and the second longest direct repeat maps at pos. 1830-1840, 1850-1860 (see Fig.1). The quality of a palindrome (bond units) is displayed. Direct repeat sequences longer than the theoretically expected longest direct repeat (17) are shown as arrows above recipient genes.
b) The two respective best palindromes are shown as black boxes under genes that served as donors in gene conversion.
Recombination breakpoints are localized in the hatched regions. The minimum size of a recipient gene segment converted by donor gene sequence and the minimum size of a donor gene sequence transferred into the recipient are depicted as open boxes.

tRNA genes SUP 3, 9, 12 (28) is shown in Fig. 2 and Table 1. It is evident in Fig. 2 that the palindrome with the best quality in terms of bond units in 6 of 9 genes is found near to a breakpoint of recombination. It should be noted that the quality of palindromes, either shown in bond units or as the free energy is a relative measure of the capacity of palindromic sequences to form stem-loop structures. In the case of HK137 the palin-

Table 1

	Gene	Best Stem(s) 5'----3' 3'----5'	Length of stem/ Watson-Crick pairs (bp/bp)	Loop size (b)	Quality of stem: bonds	Quality of stem: ΔG (kcal/mol)	Size of inverted repeat(=A) (bp)	Length of scanned sequence(=B) (bp)	No. of palindromes in scanned sequence	Length of the regions to which recombination breakpoints can be assigned(=C) (bp)	Minimal/maximal distance of inverted repeat to next point of recombination(=D) (bp)	Minimal length of recombinant region(=E) (bp)
(a)	H2K^b	GTGGACGGCGGCG CACTAGTCGCGGC	13/9	7	22	-12.8	33	276	2	17 + 21	5/26	12
	IA$_\beta^b$	CGAGCACCGC GCCAGTGGCG	10/8	0	20		20	801	6	20 + 9	23/44	14
	HK137	TTTAGCCTGGTTTC GAAAGGGACCAAAG	14/11	4	22	-13.4	32	285	2	2 + 2	0/0	13
	V186.2	GTGAAGCCTGGGGCTT CACATCGGTCTTCGGA	16/11	18	27	-11.6	50	2198	5	2 + 12	0/3	157
	H$_\alpha$A2m2	GGGCCCCACTGGG CCCGGGAGGACCT	13/10	9	27	-27.7	35	1780	2	98 + 162	0/99	626
	CYC1.11	ACGTGTTGT TGTACAATA	9/8	9	18	-12.1	27	138	2	9 + 18	2/11	45
	SUP3	CCCGCGCAGGTTC GCGCGTGTTTATG	13/10	17	19	- 7.0	43	237	3	114 + 37	0/29	30
		CCGACGGGCT GGATGCCCAA	10/8	3	19	-10.4						
	SUP9	CTCTGCCCGCGCA GAGGATGGCGCGT	13/10	28	23	-18.3	54	238	3	136 + 36	0/70	18
		CCGACGGGCT GGATGCCCAA	10/8	3	19	-10.4						
	SUP12	CCCGCGCAGGTTC GCGCGTGTTTATG	13/10	17	19	- 7.0	54	237	2	10 + 33	6/16	19
(b)	Q10	GGAGCGGCTCCT CGTCGCAGAGGA	12/10	20	21	- 9	43	267	2	17 + 21	10/27	12
	IE$_\beta^b$	CGCCGAGAACT GAGGTCCTTGA	11/8	15	19	- 4.0	27	268	3	20 + 9	0/0	11

a) Palindromes near to a point of recombination in genes that served as recipients in gene conversion. The respective best palindromes in 7 SUP3 and SUP9 carry two palindromes of similar quality in the region to which a recombination breakpoint can be assigned. The smaller palindromes encode the anticodon loop of tRNA and are not regarded in the statistical analysis.
b) Palindromes near to a point of recombination in genes that served as donors in gene conversion. The respective best palindrome is shown.
Parameters A, B, C, D and E are used for the calculation of significance.

drome near to the right recombination breakpoint has 22 bonds as compared to the 23 bonds of the palindrome located further upstream. However, the latter is less stable when free energies are considered (-10.4 vs. -13.4 kcal/mole). The best palindrome in V186.2 has 42 bond units and is represented by a stretch of 21 AT-pairs in the 5' flanking region. Apart from this perfect palindrome in a simple repetitive sequence a cluster of imperfect palindromic sequences including the second best palindrome is found at the left breakpoint of recombination. Correlation of a point of recombination with the second best palindrome (20 bond units) also is found in IA_{β}^{b}. This imperfect palindrome may form a loop-less stem which certainly is unstable (14). Such a stem has been shown to occur in the perfect palindrome of an $(AT)_n$-sequence (29). The best palindrome in IA_{β}^{b} is located 144 bp further upstream (Fig. 2a) and has 22 bond units.

In all gene conversions presented here the left and the right point of recombination cannot be mapped precisely because of extensive sequence homology between donor and recipient genes. A point of recombination in the recipient gene therefore is located between positions that define the minimum size and the maximum size of the gene segment which is replaced by the corresponding segment of the donor gene. The minimum and the maximum distances of the respective best, or in 2 cases second-best palindrome, to the next point of recombination in 9 genes which served as recipients in gene conversion range between 0 and 99 bp.

The probability P of coincidence between a point of recombination and the best palindrome (second best in V186.2 and IA_{β}^{b}) in 9 recipient genes is calculated (Table 2) utilizing the parameters A, B, C, D and E as defined in Table 1.

Palindromes at points of recombination are also found in genes that served as donors in gene conversion. Donor genes V102.1, HK102 (25) and Hα1 (27) carry palindromes which exactly correspond to the ones in the respective recipients V186.2, HK137 and Hα A2m2, simply because of extensive sequence homology between donors and recipients. Recipient genes SUP12 and SUP3 also serve as donors in gene conversion (28). The donor to CYC1.11, namely CYC7 (26), does not exhibit palindromes under the conditions of

stringency chosen for the search in CYC1.11. The donor to $H2K^b$, Q10 (20), exhibits the best palindrome at a position 10-27 bp downstream of the right point of recombination. In IE_β^b which serves as a donor to IA_β^b (23,24) the best palindrome coincides with the region to which the left point of recombination can be assigned (Tab. 1b, Fig. 2b). The probability of coincidence between the best palindrome and a point of recombination in donor genes is shown in Tab. 2b.

Direct repeats at points of recombination

We also searched for direct repeats in the recipient genes mentioned above. In V186.2 the longest direct repeat is represented by 21 AT-pairs which are located 781-783 bp upstream of the left point of recombination. The second longest direct repeat in V186.2 flanks this breakpoint and coincides with a cluster of palindromes (Fig.2a). HK137 and SUP12 exhibit direct repeats mapping at least 17bp away from the next point of recombination. No significant direct repeats are found in genes $H2K^b$, IA_β^b, CYC1.11, SUP3,9, and the longest direct repeat in $H\alpha A2m2$ is 234-333 bp away from the right point of recombination. This led us to the conclusion that direct repeat sequences are unrelated to points of recombination in recipient genes.

Table 2

a) Gene	P	b) Gene	P
$H2K^b$	0.489	Q10	0.533
IA_β^b	0.243	IE_β^b	0.209
HK137	0.126	HK102	0.126
V186.2	0.054	V102.1	0.142
$H\alpha A2m2$	0.388	$H\alpha 1$	0.388
CYC1.11	0.710		
SUP3	1		
SUP9	1		
SUP12	0.679		

Probability of coincidence between a point of recombination and the best palindrome (second best in V186.2 and IA_β^b) in a) recipient genes and b) donor genes in gene conversion. P is the probability that a palindrome (length = A) in a string of sequence (length = B) is not further away from one of the two points of recombination than D basepairs: $P=(A+C+4D)/B$. In genes where the length of the recombinant region (=E) is shorter than 2xD (see table 1) P is calculated from $(A+C+2D+E)/B$. The sum of the lengths of the respective best and second best stemloop is taken as parameter A to calculate P for genes V186.2 and IA_β^b in which the second best stemloop is located near to a point of recombination.

DISCUSSION

A mechanism analogous to gene conversion (4) appears to have transferred genetic information from V_H-gene V102.1 into V_H-gene V186.2 in the genome of the murine hybridoma B1-8.V1 (7,8). The sequence of donor V102.1 is identical to the corresponding germline sequence (9) thus supporting the notion that the donor gene remains structurally unaltered in the process of gene conversion.

The left recombinatorial breakpoint in V186.2 and V102.1 is flanked by the direct repeat of the motif AGCCTGGGGCT and, in addition, is located maximally 3 bp away from a palindrome potentially capable of stem-loop formation. In order to investigate whether this observation could be generalized we performed a computer search for palindromes and direct repeats in published sequences of genes that had been involved in eucaryotic gene conversion.

Correlation is found between a breakpoint of recombination and the palindrome which may form the best or second best stem-loop in the respective scanned sequence (Fig. 2, Tab. 1). One has to concede, however, that this correlation may be coincidental in genes CYC1.11, $H2K^b$, Q10 and SUP3, 9, 12 (Tab. 2). No corre-

Table 3

Gene	Expected longest direct repeat (ELDR) (bp)	Length of direct repeats longer than ELDR (bp)	Minimal/Maximal distance of longest direct repeat to next point of recombination (bp)
$H2K^b$	7.2	not found	-
$IA_\beta{}^b$	8.7	not found	-
HK137	6.9	9, 8	17/121
V186.2	10	40, 11	3/8
HαA2m2	10.1	12	234/305
CYC1.11	6.6	not found	-
SUP3	6.7	not found	-
SUP9	6.7	not found	-
SUP12	6.7	8	17/132

Direct repeat sequences in genes that served as recipients in gene conversion. The longest direct repeat sequence expected to occur in a sequence by chance (ELDR) is predicted according to (17). Direct repeats are taken into account provided they are longer than the expected longest direct repeat (ELDR). The distance of the longest direct repeat (second longest in V186.2) to the next point of recombination is shown.

lation is found in the donor gene to CYC1.11, CYC7; in IA_{β}^{b} the left point of recombination is correlated with a palindrome which may form a weak loop-less stem.
The search for direct repeats in eight recipient genes did not show a correlation of breakpoints with repeated sequence motives longer than the statistically expected longest direct repeat (Tab. 3). Because of this we regarded a search for direct repeats in donor genes as pointless. As for V186.2, one cannot exclude that direct repeats played a role in the recombination process.
In accordance with the hypothesis put forward by Wagner and Radman (30) and Baltimore and Loh (31) we suggest that palindromic DNA sequences promote gene conversion in that they serve as a recognition site for a recombination enzyme. Such an enzyme may be a twofold symmetric protein molecule which recognizes a twofold symmetric sequence in duplex DNA, or it may be an enzyme which cuts at the basis of a stem-loop in single stranded DNA. It has been shown that enzymes involved in bacterial recombination recognize stem-loops (32) and it is likely that functionally related enzymes operate in eucaryotic cells.
As inverted repeats may lead to stem-loops only in single-stranded DNA gene conversion could be promoted by such structures during DNA-replication. Champoux et al. (33) suggest that eucaryotic recombination involves type I topoisomerases which cut at palindromic sequences in single-stranded regions of unwound duplex DNA. Five of the nine palindromes correlated with recombination breakpoints (Table 1a) carry sequences (CTT, GTT) which are cleaved by eucaryotic topoisomerase I (34, 35). In V186.2, the left point of recombination precisely matches a potential topoisomerase I cleavage site.

ACKNOWLEDGEMENTS
We thank M. Paskind for technical assistance and J. Yoder for the gift of EMBL4 arms. The help of E. Siegmund, Å. Böhm and U. Ringeisen to prepare the manuscript and to design figures is gratefully acknowledged. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 74 and a Heisenberg grant to U.K.; A.B. is supported by the Howard Hughes Medical Institute.

Note: Cumano and Rajewsky in our laboratory found another gene conversion between immunoglobulin V_H-region genes. A breakpoint of recombination is correlated with a palindrome in the donor gene (manuscript in preparation).

*To whom reprint requests should be sent

REFERENCES

1. Tonegawa,S. (1983) Nature 302, 575-581.
2. Smithies,O. (1967) Cold Spring Harbor Symp. Quant. Biol. 32, 161-168.
3. Edelman,G. and Gally,J. (1970) The Neurosciences Second Study Program, F. Schmitt (ed.), Rockefeller University Press, New York.
4. Baltimore,D. (1981) Cell 24, 592-594.
5. Egel,R. (1981) Nature 290, 191-192.
6. Brüggemann,M., Radbruch,A. and Rajewsky,K. (1982) EMBO J. 1, 629-634.
7. Dildrop,R., Brüggemann,M., Radbruch,A., Rajewsky,K. and Beyreuther,K. (1982) EMBO J. 1, 635-640.
8. Krawinkel,U., Zoebelein,G., Brüggemann,M., Radbruch,A. and Rajewsky,K. (1983) Proc. Natl. Acad. Sci. USA 80, 4997-5001.
9. Bothwell,A.L.M., Paskind,M., Reth,M., Imanishi-Kari,T., Rajewsky,K. and Baltimore,D. (1981) Cell 24, 625-637.
10. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning - A Laboratory Manual, Cold Spring Harbor, N.Y.
11. Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe, B. (1980) J. Mol. Biol. 143, 161-178.
12. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol. 65, 499-560.
13. Devereux,J. and Haeberli,P. (1983) Program Library of the University of Wisconsin Genetics Computer Group.
14. Tinoco,I., Borer,P.N., Dengler,B., Levine,M.B., Uhlenbeck,O.C., Crothers,D.M. and Gralla,J. (1973) Nature 246, 40-41.
15. Aboul-ela,F., Koh,D., Tinoco,I. and Martin,F.H. (1985) Nucl. Acids Res. 13, 4811-4824.
16. Staden,R. (1984) Nucl. Acids Res. 12, 521-538.
17. Karlin,S., Ghandour,G., Ost,F., Tavare,S. and Korn,L. (1983) Proc. Natl. Acad. Sci. USA 80, 5660-5664.
18. Bothwell,A.L.M. (1984) In The Biology of Idiotypes, eds. Greene,M.I. and Nisonoff,A., Plenum Press, p.19-34.
19. Schulze,D., Pease,L.R., Geier,S.S., Reyes,A.A., Sarmento, L.A., Wallace,B.A. and Nathenson,S.A. (1983) Proc. Natl. Acad. Sci. USA 80, 2007-2011.
20. Mellor,A.L., Weiss,E.L., Ramachandran,K. and Flavell,R.A. (1983) Nature 306, 792-795.
21. Choi,E., McIntyre,K., Germain,R.N. and Seidman,J.G. (1983) Science (Wash.) 221, 283-286.
22. Larhammar,D., Hämmerling,U., Denaro,M., Lund,T., Flavell,R.A., Rask,L. and Peterson,P.A. (1983) Cell 34, 179-188.
23. Denaro,M., Hämmerling,U., Rask,L. and Peterson,P.A. (1984) EMBO J. 3, 2029-2032.
24. Widera,G. and Flavell,R.A. (1984) EMBO. J. 3, 1221-1225.

25. Bentley,D.L. and Rabbitts,T.H. (1983) Cell 32, 181-189.
26. Ernst,J.W., Stewart, J.W. and Sherman, F. (1981) Proc. Natl. Acad. Sci. USA 78, 6334-6338.
27. Flanagan,J.G., Lefranc, M.P. and Rabbitts,T.H. (1984) Cell 36, 681-688.
28. Amstutz,H., Munz,P., Heyer,W.D., Leupold,U. and Kohli, J. (1985) Cell 40, 879-886.
29. Greaves,D.R., Patient, R.K. and Lilley, D.M. (1985) J. Mol. Biol. 185, 461-478.
30. Wagner,R.T. and Radman,M. (1975) Proc. Natl. Acad. Sci. USA 72, 3619-3622.
31. Baltimore,D. and Loh,D.Y. (1983) Progress in Immunology 5, 115-121.
32. Kemper,B., Jensch,F., v.Depka-Prondzynski,M., Fritz, H.-J., Borgmeyer,U. and Mizuuchi,K. (1984) Cold Spring Harb. Symp. Quant. Biol. XLIX, 815-825.
33. Champoux,J.J., McCoubrey Jr.,W.K. and Been,M.D. (1984) Cold Spring Harb. Symp. Quant. Biol. XLIX, 435-442.
34. Been,M.D., Burgess,R.R. and Champoux, J.J. (1984) Nucl. Acids Res. 12, 3097-3114.
35. Edwards,K.A., Halligan,B.D., Davis,N., Nivera,L. and Liu,L.F. (1982) Nucl. Acids Res. 10, 2565-2576.