PROGRAMMED DNA REARRANGEMENTS DURING DIFFERENTIATION: IMMUNOGLOBULIN CLASS SWITCHING

Thesis by

Mark Morris Davis

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1981

(Submitted August 28, 1980)

This thesis is dedicated to my family and friends, especially Chien, without whose love nothing would be possible.

"Glad to be so young,

Talkin' with my tongue.

Glad to be so careless in my way.

Glad to take a chance,

and play against the odds.

Glad to be so crazy in my day."

Steve Forbert

"...But more than that, science is a quest for poetic, romantic, almost unattainable goals like growing a beard or having a car that runs."

Mitch Kronenberg

ACKNOWLEDGEMENTS

I'd like to thank my professors, who, both by formal instruction and by their example have taught me much of what I know about Science: Michael Tomko, P. Y. Johnson, Michael Beer, Eric Davidson, Tom Maniatis, Ed Lewis, Max Delbrück, Ron Konopka, Norman Davidson, and especially Lee Hood, who knows he's important. I'd also like to thank those on the foothills (in some cases, the gullies) of Olympus who taught me all the rest: Bill Klein, Glenn Galau, Barbara Hough-Evans, Amy Lee, Tom Sargent, Maggie Chamberlin, Jimbo Mullins, Marty and David Goldberg, Brian Seed, Randy Smith, Max (again) and Manny Delbrück for their great camping trips and terrific family, all the Ingersolls, the Pechpeople, Stuart "Iron Butt" Kim, Mitchell Kronenberg, "Full-length Phil" Early (a general thorn-in-the-side but a first-rate collaborator), Steve and Aggie Crews, John Burdakin, Mark Warren, the honorable and highly esteemed C. Lai, Jerry Garcia, Jackson Browne, Richard Brown, "Sunny Phil" Patten, fencing buddies like "Fast Eddy" Rhodes, Jeff Hicks and Stuart Goodnick ("the euthanasia poster boy"), and others who have touched my life. Further thanks to Jane Chacón (lifeguard of the typing pool) and to Pat Lee and her gang at Graphic Arts, both of whom never tired of hearing me cry "Wolf!". Thanks also to the Jean Weigle Memorial Fund for support in the preparation of this thesis and to NIH for pocket money. Thanks especially to those Samurai scientists Susumu Tonegawa, 'Toshi Sakano and Tasuku Honjo for their high standards of competition. I am also grateful to Connie Katz for typing this.

ABSTRACT

The events of B-lymphocyte differentiation can be reconstructed in part through an analysis of the organization of heavy-chain genes isolated from B-cell tumors (myelomas). A mouse immunoglobulin alpha heavy-chain gene is shown to be composed of at least three non-contiguous segments of germline DNA—a V_H gene segment, a J_H gene segment adjacent to the C_μ coding region, and a C_α gene segment. These gene segments are joined together by two distinct types of DNA rearrangements: variable region formation and immunoglobulin class switching. Three examples of IgM→IgA class switching were examined and in each case a different site adjacent to C_μ and a different site adjacent to C_α were joined together in the process of switching. Two of the three C_μ sites shared significant homology to each other (15/25 nucleotides) and all three of C_α sites were highly homologous (22/30 nucleotides). We believe these sequences serve as recognition sites for class switching. Furthermore, the lack of homology between the C_α consensus sequence and sequences reported for $C_{\gamma 1}$ and $C_{\gamma 2b}$ recombination sites suggests that this process is mediated by class-specific recognition sequences and, presumably, class-specific regulatory mechanisms. A number of predictions and possible explanations of immune phenomena result from this observation. Apparently nonproductive DNA rearrangements, occurring in the same tumor lines, seem also to utilize some of the same regulatory apparati. In addition, it appears that in one example, MC101, class switching has progressed from $C_\mu \rightarrow C_\alpha \rightarrow C_{\gamma 1}$. This switching pathway presents difficulties for the simple deletional model of C_H switching.

Table of Contents

INTRODUCTION

All life processes are abstracted into the genome of each organism. Any fundamental understanding of evolution or development, of what we are and what we can be, has to take into account the information encoded in our DNA and the mechanisms which act upon it.

A useful construct for subdividing genomes into manageable segments has been the idea of a "gene", typically a specific heritable trait which can be localized to a discrete region of the genome and associated with a specific protein product. In this light, it is perhaps understandable that antibody genes have been an anathema to many geneticists since they disrupt this concept of a gene and in general blur the distinction between gene and genome. Each antibody "gene" is made up of many parts, sets of which are interchangeable, spread over perhaps millions of nucleotides of DNA. But it is through the study of "exceptions" such as antibody genes that we seem to significantly expand our knowledge about the informational capabilities of genomes. In particular, this thesis concerns itself with immunoglobulin class switching, one of the two distinct forms of carefully regulated DNA rearrangement which act to expand the information inherent in antibody genes several orders of magnitude over that of "conventional" genes.

Immunoglobulins and Their Genes

Antibodies are one of the major components of the body's defense against foreign organisms. It serves both to 'recognize' and bind to chemical moieties not native to the organism (antigens) and to trigger a variety of secondary immune mechanisms (effector functions). The antibody molecule itself is typically a dimer composed of two polypeptide chains—heavy (H) chains and light (L) chains. Early protein sequencing studies further subdivided each type of polypeptide into variable (V) and constant (C) regions (1, 2). The variable regions of heavy and light chains fold

together to form the antigen-binding site or V region domain. It was this striking distinction between variable and constant regions within the same polypeptide that first led Dreyer and Bennett (3) to propose that DNA rearrangement during B-cell differentiation was responsible for the production of such disparate molecules. Molecular cloning and other forms of genomic analysis have confirmed and extended this original hypothesis and defined two distinct forms of DNA rearrangement which amplify the information inherent in the DNA many orders of magnitude more than if these genes were static. These types of DNA rearrangement are termed variable region formation and class switching. Since this thesis deals principally with class switching, variable formation or V(D)J joining will be dealt with only briefly, and for the purposes of comparison.

Variable Region Formation

In the mouse L chain polypeptides are encoded by two unlinked gene families, κ and λ, as defined by their constant regions C_κ and C_λ, respectively (4). The variable regions are encoded by two complementary gene segments, V_L and J_L (5-7). The J_L regions are located adjacent to C_κ or C_λ in undifferentiated DNA and it is the fusion of a particular V_L with a particular J_L which gives rise to a functional light chain gene (6-9). A large intervening sequence 1.5-3.0 kb between V_LJ_L and C_L is excised during RNA processing, resulting in a continuous reading frame progressing from V_LJ_L to C_L (10, 11). In the kappa family a large number of V regions (>200) and at least four functional J regions allow a large number of possible antigen-binding sites to be formed. Similarly, in the third unlinked gene family, the heavy chains (4), it has now been shown that the V_H region is composed of three gene segments, V_H, D, and J_H (12, 13), which can also independently assort (14) to produce a large number of different variable regions. Of particular importance to later arguments involving heavy chain class switching, an apparent recognition sequence $CAC\frac{A}{T}GTG(NNNNNNN-NNNN)_{\sim 1-2}GACATAAACC$ has been identified 3' to every V region (both light and

heavy) (8, 9, 12, 13), and 5' to every J region. This sequence of approximately 17 conserved residues apparently reflects the information necessary to generate sequence-specific intrachromosomal rearrangements, and in fact, would occur by chance only once or twice per haploid mammalian genome. The exact site of recombination can vary by at least six nucleotides outside of this recognition sequence (8, 12, 13). This further contributes to the diversity possible in antibody variable genes.

Class Switching

Antibody molecules are divided into five distinct classes: IgM, IgD, IgG, IgA, and IgE—which are determined by one of eight distinct heavy chain genes [i.e., C_μ, C_δ ($C_{\gamma 1}$, $C_{\gamma 2a}$, $C_{\gamma 2b}$, $C_{\gamma 3}$) C_α and C_ε]. These C_H regions are tightly linked to each other (15) and to V_H, D, and J_H regions on chromosome 12 of mice (16, 17). Each class (and subclass) seems to be associated with unique functions such as complement fixation or the release of histamine from most cells (15). One of the first observations that the same V_H region could be expressed with different constant (C_H) regions was made by Wang et al. (18) who partially sequenced the N-terminal (V_H) regions of human patients with myelomas which synthesize two or more different classes of immunoglobulin. This result was reproduced in a number of other cases as well (19-22). Concurrently, Gearhart et al. (23) showed that a single cultured lymphocyte could give rise to two or more different classes of antibody with serologically identical variable region domains. In a number of studies (24-26), and particularly in the elegant work of Coffman and Cohn (27), it was clear that the first class of antibody expressed was IgM, no matter what the subsequent product.

In summary, these experiments indicated that IgM (hence C_μ) was always the first antibody expressed with the newly formed V_H region (and light chain) and a second or third C_H gene could then be expressed with this same V_H region. This phenomenon is known as C_H or class switching, and serves to combine a particular

antigen-binding capability with the different C_H functions and tissue localizations (15). All V_H's may be required to 'pass through' an IgM stage because of the crucial role that this molecule has as a cell-surface receptor for antigen, in particular, the clonal expansion of a particular cell line in response to antigen may require C_μ-specific functions.

Recent History

The first concrete evidence that class switching might arise through DNA rearrangement was that of Honjo and Kataoka in 1978 (28) who measured the gene number for the four different C_γ subclasses in various myeloma tumors. Their findings indicated that specific C_H producers tended to have lost copies of other C_H genes in a particular order and suggested that this order was V_H, C_μ, $C_{\gamma 3}$, $C_{\gamma 1}$, $C_{\gamma 2b}$, $C_{\gamma 2a}$, C_α. However, their results were suspect in that they were at the very limits of resolution of hybridization kinetics (1 copy vs. 1/2 copy) and used uncloned probes. Based on these data, they proposed a simple, linear deletional model for C_H switching such that an IgM producer (C_μ) would have all of the C_H genes and that an IgG3 ($\gamma 3$) producer would delete C_μ, etc. An alternative model utilizing differential RNA processing of a large nuclear transcript was proposed by Rabbitts (29). With the widespread adoption of recombinant DNA techniques, a number of groups (30-35) repeated Honjo and Kataoka's experiments using cloned probes and the more sensitive techniques of Southern blotting (36). In general, they also showed that chromosomal deletions are associated with class switching and support the gene order mentioned above, although a number of anomalies remain—usually attributed to random myeloma rearrangements. The C_γ part of the gene order has recently been unequivocally confirmed by the linking-up of genomic clones (Honjo, personal communication; Tonegawa, personal communication). The first direct evidence for class switching as the result of DNA rearrangement involved a heavy chain gene isolated by Phil Early and myself from a library (37) of myeloma DNA inserted into

the phage Charon 4A (38). Chapters 1 and 2 present the initial characterization of this gene and in the paper presented in Chapter 3, Kathryn Calame and I demonstrated that this gene was the fused product of at least three widely separated segments of DNA, a V_H region, a J_H-C_μ-associated region, and a C_α region. This finding strongly implies that at least two different recombinational events had to have occurred to form this gene, namely, variable region formation [or $V_H(D)J_H$ joining, shown by Phil Early (12)] and a C_H switch (replacement of C_μ with C_α). Subsequently, other groups have found similar tripartite structures (V_H-C_μ-C_H) for all four C_γ subclasses (39-41; Honjo, personal communication; Sakano, personal communication).

In order to examine the precise mechanisms involved in this process, Stuart Kim and I studied three different examples of IgM-IgA switching at the nucleotide sequence level (Chapter 4). We found that the recombination could involve C_μ and C_α flanking sequences directly (simple switching), a discontinuous portion of C_μ flanking sequence joined to C_α (complex switching) or that a V_H region which has switched to one C_H (from C_μ) could then switch to another (successive switching). The instance of successive switching seems to run contrary to the simple deletional model of Honjo and Kataoka (28), however, since the sequence organization is V_H-C_μ-C_α-$C_{\gamma 1}$. Possible escapes are suggested in Chapter 4 for this paradox but the issue is unresolved. We also found that multiple sites adjacent to both C_μ and C_α coding regions could be used for switching and that two of the three C_μ sites and all three of the C_α sites were highly homologous among themselves (C_α-C_α) but not to each other (C_α-C_μ). Thus, homologous recombination does not appear to be important in C_H switching as suggested by a number of groups (39-42). The C_μ consensus sequence is GGTNNTTANNNNNNGGTANNCAAAG and the C_α consensus sequence is PGCTPPGCTGGAATPPGYTGGGNTGPGCTG (where P = Purine, Y = Pyrimidine and N = any nucleotide). Hence there is no apparent similarity with

variable region formation recognition sequences, although the number of conserved nucleotides is similar (15 for S_μ, 22 for S_α, ~17 for V(D)J homology regions). The exact points of recombination vary up to 16 nucleotides over the C_μ homology region and up to 9 nucleotides in the C_α homology region, roughly comparable to the variability seen in variable region formation (at least 6 nucleotides). A prokaryotic model might be the restriction endonuclease Mnl I which recognizes a 4-nucleotide sequence, then cuts 5-10 nucleotides away from it (43). Comparison of the C_α consensus sequence with those around analogous recombination sites for $C_{\gamma 2b}$ and $C_{\gamma 1}$ genes gives very little homology—leading us to propose a model in which each class has its own recognition sequence and hence can be specifically regulated (Chapter 4). Other predictions of this model are that a whole regulatory network might exist to induce switching to each C_H region, e.g., there might be C_α-specific cell-surface receptors on B cells as well as a specific DNA recognition/splicing enzyme. Thus, the tissue specificity (15), frequency (15) and antigen preference (45) of the immunoglobulin classes and subclasses might be explained by class-specific regulation at the level of DNA rearrangement. Especially interesting in this regard is the finding of Stein et al. (45) that a non-C_H linked locus (Sr-1) influences the preponderance of the different C_H genes expressed in response to a polysaccharide antigen, thus Sr-1 might be a regulatory gene for C_H switching.

Other candidates for genes which may influence C_H switching are the multitude of different B-cell specific cell-surface markers (seven identified to date) (46).

Related Rearrangements

As mentioned in Chapter 2, two other forms of the alpha constant region gene are seen in the tumor M603, neither of which seem contiguous to a V_{M603}, hence they are termed 'abortive' rearrangements. Characterization of one of these

abortive rearrangements, α30, by Kathryn Calame and myself, (Chapter 5) indicates that some sequence of unknown origin has been brought 5' to the C_{α}, precisely within an α homology region associated with C_H switching in the work mentioned above. This provides a fourth example of rearrangement within an alpha homology region and indicates the possible continued presence of class-specific switching enzymes in myeloma cell lines. In short, immunoglobulin class switching promises to be a fruitful area for studying the precise regulation of DNA rearrangements during differentiation and the informational requirements of DNA sequence recognition in complex genomes.

REFERENCES

1. Porter, M., Dreyer, W. J., Kuff, E. L., and McIntire, K. R. J. Mol. Biol. **8**, 814 (1964).
2. Bennett, J. C., Hood, L., Dreyer, W. J., and Potter, M. J. Mol. Biol. **12**, 81 (1965).
3. Dreyer, W. J. and Bennett, J. C. Proc. Natl. Acad. Sci. USA **54**, 864 (1965).
4. Mage, R., Lieberman, R., Potter, M., and Terry, W. In: The Antigens (Ed. Sela, M.) Academic, New York, (1973) p. 299.
5. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O., and Gilbert, W. Proc. Natl. Acad. Sci. USA **75**, 1488 (1978).
6. Brack, C., Hirowa, A., Lenhard-Schueller, R., and Tonegawa, S. Cell **15**, 1 (1978).
7. Seidman, J. G., Max, E. E., and Leder, P. Nature **280**, 370 (1979).
8. Max, E., Seidman, J. G., and Leder, P. Proc. Natl. Acad. Sci. USA **76**, 3450 (1979).
9. Satano, H., Huppi, K., Heinrich, G., and Tonegawa, S. Nature **280**, 288 (1979).
10. Gilmore-Herbert, M. and Wall, R. Proc. Natl. Acad. Sci. USA **75**, 342 (1978).
11. Schibler, U., Marcu, K. B., and Perry, R. P. Cell **15**, 1495 (1978).
12. Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. Cell **19**, 881 (1980).
13. Sakano, H., Maki, R., Furosawa, Y., Roeder, W., and Tonegawa, S. Nature, submitted (1980).
14. Schilling, J., Cleavinger, B., Davie, J., and Hood, L. Nature **283**, 35 (1980).
15. Lieberman, R. Springer Seminars in Immunopathology **1**, 7 (1978).
16. Meo, T., Johnson, J., Beechey, C. V., Andrews, S. J., Peters, J., and Searle, A. G. Proc. Natl. Acad. Sci. USA **77**, 550 (1980).
17. D'Eustachio, P., Prautcheva, D., Marcu, K., and Ruddle, R. J. Exp. Med. **151**, 1545 (1980).

18. Wang, A. C., Wilson, S. K., Hopper, J. E., Fudenberg, H. H., and Nisonoff, A. Proc. Natl. Acad. Sci. USA **66**, 337 (1970).

19. Sledge, C., Fain, D. S., Black, B., Krieger, R. G., and Hood, L. Proc. Natl. Acad. Sci. USA **73**, 923 (1976).

20. Fudenberg, H. H., Wang, A. C., Pink, J. R. L., and Levin, A. S. Ann. N.Y. Acad. Sci. **190**, 501 (1971).

21. Wang, A. C., Geisely, J., and Fudenberg, H. H. Biochemistry **12**, 528 (1973).

22. Levin, A. S., Fudenberg, H. H., Hopper, J. E., Wilson, S., and Nisonoff, A. Proc. Natl. Acad. Sci. USA **68**, 169 (1971).

23. Gearhart, P. J., Sigal, N. H., and Klinman, N. R. Proc. Natl. Acad. Sci. USA **72**, 1707 (1975).

24. Raff, M. C. Cold Spring Harbor Symp. Quant. Biol. **41**, 159 (1976).

25. Anderson, J., Coutinho, A., and Melchers, F. J. Exp. Med. **146**, 1744 (1978).

26. Wabl, M. R., Forni, L., and Loor, F. Science **199**, 1078 (1978).

27. Coffman, R. L. and Cohn, M. J. Immunol. **118**, 1806 (1977).

28. Honjo, T. and Kataoka, T. Proc. Natl. Acad. Sci. USA **75**, 2140 (1978).

29. Rabbitts, T. H. Nature **275**, 291 (1978).

30. Cory, S., Jackson, J., and Adams, J. M. Nature **285**, 450 (1980).

31. Cory, S. and Adams, J. M. Cell **19**, 37 (1980).

32. Coleclough, C., Cooper, C., and Perry, R. P. Proc. Natl. Acad. Sci. USA **77**, 1422 (1980).

33. Rabbitts, T. H., Forster, A., Dunnick, W., and Bentley, D. L. Nature **283**, 351 (1980).

34. Yaoita, Y. and Honjo, T. Biomed. Res., in press (1980).

35. Yaoita, Y. and Honjo, T. Nature, in press (1980).

36. Southern, E. M. J. Mol. Biol. **98**, 503 (1977).

37. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. E., and Efstratiadis, A. Cell **15**, 687 (1978).

38. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D.J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L., and Smithies, O. Science **196**, 161 (1977).

39. Kataoka, T., Kawakami, T., Takahashi, N., and Honjo, T. Proc. Natl. Acad. Sci. USA **77**, 919 (1980)

40. Takahashi, N., Kataoka, T., and Honjo, T. Gene, in press (1980).

41. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., and Tonegawa, S. Nature, submitted (1980).

42. Dunnick, W., Rabbitts, T. H., and Milstein, C. Nature, in press (1980).

43. Zabeau, M., Greene, R., Myers, P. A., and Roberts, R. J. Unpublished observations cited by New England Biolabs.

44. Perlmutter, R. M., Hansburg, D., Briles, R. A., Vicolotti, R. A., and Davie, J. M. J. Immunol. **121**, 566 (1978).

45. Stein, K. E., Bona, C., Lieberman, R., Chien, C., and Paul, W. E. J. Exp. Med. **151**, 1088 (1980).

46. B Lymphocytes in the Immune Response. (M. Cooper, D. E. Mosier, I. Scher, and E. S. Vitetta, eds.) Elsevier North-Holland, Inc., New York (1979).

Proc. Natl. Acad. Sci. USA
Vol. 76, No. 2, pp. 857–861, February 1979
Genetics

Immunoglobulin heavy chain gene organization in mice: Analysis of a myeloma genomic clone containing variable and α constant regions

(heavy chain genomic clone/intervening DNA sequences/restriction maps/R loop mapping/V and C gene segments)

PHILIP W. EARLY*, MARK M. DAVIS*, DAVID B. KABACK†, NORMAN DAVIDSON†, AND LEROY HOOD*‡

*Division of Biology and †Department of Chemistry, California Institute of Technology, Pasadena, California 91125

Contributed by Norman Davidson, November 7, 1978

ABSTRACT **We have isolated a myeloma genomic DNA clone containing the variable and constant regions of a mouse α chain. Restriction enzyme analyses and electron microscopic R loop mapping have demonstrated that the variable region is separated from the constant region by 6.8 kilobases of intervening DNA. In addition, two intervening DNA sequences of 100–200 bases separate the constant region into three approximately equal units. These intervening sequences may separate each of the segments coding for the three constant region domains of the α heavy chain. Southern blot analysis of embryo and myeloma DNA suggests that DNA rearrangement of heavy chain variable and constant regions occurs during the differentiation of antibody-producing cells.**

The antibody system affords a fascinating model for the study of gene organization and expression in eukaryotes. Antibodies or immunoglobulins are composed of two subunits, light and heavy chains, each of which contains an NH_2-terminal variable (V) region and a COOH-terminal constant (C) region (1). The antibody polypeptides are divided into discrete domains or homology units, each encompassing approximately 110 residues. Accordingly, the light chain has two domains (V and C) and the α heavy chain has four (V, C_H1, C_H2, and C_H3) (2). The variable and constant regions of light chains are encoded by three distinct gene segments, V (approximately residues 1–99), J or joining (approximately residues 100–112), and C (approximately residues 113–219) (3, 4). The V and J segments encode the classical V region. Each of these DNA segments is separated by intervening nucleotide sequences in embryo or undifferentiated DNA (4). Studies of myeloma DNA suggest that these gene segments are rearranged during the differentiation of antibody-producing cells. However, in myeloma DNA, intervening sequences still separate the V and C segments (4). These intervening sequences must be removed from nuclear RNA transcripts of light chain genes by RNA splicing (5). Thus, DNA rearrangements and RNA splicing appear to be important events in the differentiation of antibody-producing cells (6).

We are interested in analyzing the genes coding for heavy chains in embryo and myeloma DNA to determine whether sequence rearrangements occur during differentiation that are comparable to those seen for light chain gene segments (4). Our initial approach has been to isolate genomic clones from a library of recombinant Charon 4A bacteriophage containing long fragments of M603 myeloma DNA. In this paper we report the characterization of one clone containing both V_H and C_α regions.

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

MATERIALS AND METHODS

Biological and Physical Containment. Work described in this report was conducted in a P3 physical containment facility using EK2 host-vector systems, in compliance with National Institutes of Health guidelines for recombinant DNA research as published in the *Federal Register* [(1976) 41, 27902–27943].

Bacterial and Phage Strains. *Escherichia coli* K-12 strain χ1776 (7) was provided by R. Curtiss. *E. coli* K-12 strain DP50SupF (8) and Charon 4A phage (9) were provided by F. Blattner. *E. coli* strains NS428 and NS433 (10) used for *in vitro* packaging were obtained from T. Maniatis.

mRNA Preparation. BALB/c mouse myeloma tumors originally obtained from M. Potter or the Salk Institute were propagated subcutaneously. A postnuclear supernatant prepared from homogenized tissue was used to prepare poly(A)$^+$ RNA from membrane-bound polysomes (11). Heavy chain mRNA, identified by *in vitro* translation (12), was isolated by sucrose gradient fractionation.

cDNA Synthesis and Cloning. Double-stranded cDNA was synthesized by sequential reactions with avian myeloblastosis virus reverse transcriptase and *E. coli* polymerase I (13). After exclusion on Sephadex G-100, the major component of this cDNA migrated as a band of 1550 base pairs on a nondenaturing agarose gel.

Double-stranded cDNA was joined to *Eco*RI-cut pMB9 either by poly(dA), poly(dT) tailing (14), or by ligation to synthetic *Eco*RI linkers (15). Annealing or ligation mixtures were used directly to transform *E. coli* χ1776 (16). Positive transformants were identified by the Grunstein–Hogness technique (17) using ^{32}P-labeled M603 heavy chain mRNA.

Construction of M603 Library. High molecular weight genomic DNA (18) prepared from M603 subcutaneous tumors was partially digested with *Eco*RI, and fragments in the range of 12 to 20 kilobases (kb) were isolated on a sucrose gradient. Ten micrograms of M603 DNA fragments was ligated to Charon 4A arms and packaged *in vitro* to obtain a library of 3×10^6 recombinant phage (19). The library was amplified on DP50SupF as a plate lysate prior to screening.

Isolation of Clones from M603 Library. The constant region plasmid p603α1 labeled by nick translation with deoxynucleotide [^{32}P]triphosphates (20) was used to screen 400,000 clones from the M603 library plated on DP50SupF (19, 21). Duplicate nitrocellulose filters from each plate were prehybridized in 1 M NaCl/0.045 M trisodium citrate/0.2% bovine serum albumin/0.2% Ficoll/0.2% polyvinylpyrrolidone/0.1% sodium

Abbreviations: V, variable; C, constant; $NaDodSO_4$, sodium dodecyl sulfate; kb, kilobase(s); IVS, intervening sequence(s).

‡ To whom reprint requests should be addressed.

dodecyl sulfate ($NaDodSO_4$) at 68°C in a rotary water bath (D. Engel and J. Dodgson, personal communication). Denatured plasmid DNA was added to 10 ng/ml and hybridization was continued for 48 hr. Filters were washed extensively in 0.15 M NaCl/0.015 M trisodium citrate/0.1% $NaDodSO_4$/10 mM $Na_4P_2O_7$ at 68°C. Duplicated spots of hybridization were identified by autoradiography, and plaques corresponding to these locations on the filters were picked and rescreened at a low plating density to obtain pure clones.

Electron Microscopic R Loop Mapping. Duplex Ch603α6 DNA was photochemically crosslinked (one crosslink per 4 kb) in the presence of 4,5',8-trimethylpsoralen (trioxsalen). Crosslinking prevents DNA strand separation, thus permitting the R loop hybridization (22) to be carried out at a temperature close to or above the DNA strand separation temperature (unpublished data). Crosslinked DNA (5 μg/ml) was hybridized to mRNA (2 μg/ml) in 70% recrystallized formamide/0.4 M NaCl/0.1 M 1,4-piperazinediethanesulfonic acid (Pipes), pH 7.2/10 mM EDTA at 56°C for 24 hr. The R-looped DNA was either spread from 70% onto 15% formamide or treated with 1 M glyoxal at 12°C to stabilize the R loops against branch migration (unpublished observations) and then spread from 50% formamide (23).

RESULTS AND DISCUSSION

Sequence Homologies of α Heavy Chain mRNAs. We isolated heavy chain mRNA from three mouse myeloma tumors secreting IgA immunoglobulins. Both the M603 and S107 immunoglobulins bind phosphorylcholine and contain nearly identical V region protein sequences (24). The M315 V region differs from M603 at 60% of its amino acid residues. To assay the extent of nucleotide homology between these mRNA species, ^{32}P-labeled single stranded M603 cDNA was hybridized to each mRNA. The hybrid was digested with nuclease S1 (25), and the resulting cDNA cleavage products were fractionated on an alkaline agarose gel (Fig. 1). M315 mRNA protects an 1100-nucleotide piece of M603 cDNA from S1 digestion; S107 and M603 mRNA both protect the full length of 1550 nucleotides. We conclude that S107 and M603 mRNAs are closely homologous over the entire sequence and that M315 mRNA shares this homology only for the C region, as expected from the protein sequences. In subsequent experiments with M603-like genomic sequences, we used S107 mRNA as a probe for the V and C regions and M315 mRNA as a probe for the C region only.

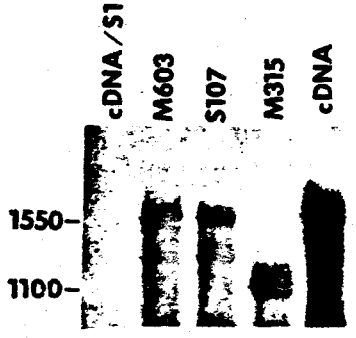


FIG. 1. Hybridization of mRNAs to M603 cDNA. Five nanograms of first-strand M603 cDNA (specific activity, 10^6 cpm/μg) was hybridized to 50 ng of the indicated heavy chain mRNA followed by digestion with nuclease S1 (25) and separation on a 2% alkaline agarose gel (26), a portion of which is shown by autoradiography. The first lane contains cDNA alone, hybridized and digested as above, and the last lane contains undigested input cDNA. The cDNA is somewhat longer than 1550 nucleotides, presumably due to an unfolded "hairpin" structure at the 5' end (27).

Characterization of cDNA Clones. Double-stranded cDNA prepared from M603 mRNA was used to construct the cDNA restriction map in Fig. 2*a*. We used poly(dA), poly(dT) tailing to obtain one recombinant plasmid, p603α1, which was shown by electron microscopy to contain an insert of approximately 600 nucleotides. Comparison of the cDNA restriction map with that of the plasmid indicates that p603α1 contains part of the C_α sequence (Fig. 2*b*). Subsequently, synthetic *Eco*RI linkers were used to clone S107 double-stranded cDNA. Regions of the cDNA included in two of these plasmids, p107αR5 and p107αR6, are indicated in Fig. 2*b*. In order to verify that the cloned sequences are derived from S107 heavy chain mRNA, a *Hin*fI restriction fragment of p107αR6 was isolated, annealed to S107 mRNA, and used to prime cDNA synthesis in the dideoxynucleotide sequencing procedure (29). The partial sequence so determined matched the known protein sequence of the S107 heavy chain between amino acids 92 and 125 (24) (data not shown).

Characterization of Genomic Clones. We screened 400,000 plaques from the M603 library with the C region plasmid p603α1, labeled with ^{32}P by nick translation. Twenty-five clones hybridized to the probe. Only three of these clones also hybridized to a nick-translated V region probe (the portion of p107αR6 shown to the left of the *Hha* I site in Fig. 2*b*). These three clones showed identical *Eco*RI restriction patterns; one, Ch603α6, was selected for further characterization. Ch603α6 contained 16.4 kb of mouse DNA with two internal *Eco*RI cleavage sites yielding fragments of 7.2, 4.8, and 4.4 kb, which are ordered in the restriction map shown in Fig. 2*c*. Filter hybridizations by the Southern blot procedure (30) localized the variable region in Ch603α6 to the 7.2-kb *Eco*RI fragment (Fig. 3). C region probes (p603α1 and the portion of p107αR6 to the right of the *Hha* I site in Fig. 2*b*) hybridized only to a 2.4-kb section bounded by *Xho* I and *Sma* I sites in the 4.8- and 4.4-kb *Eco*RI fragments. These results, displayed schematically in Fig. 2*c*, demonstrate that the V_H and C_α regions in Ch603α6 are separated by at least 4 kb of intervening DNA.

Identification of Genomic DNA Fragments Hybridizing to cDNA Clones. Hybridization of the nearly full-length cDNA probe p107αR5 to Southern blots of *Eco*RI-digested chromosomal DNA showed that the 7.2-, 4.8-, and 4.4-kb fragments of Ch603α6 all correspond to bands present in the M603 genome (Fig. 4). In embryonic DNA, the 4.8-kb middle piece of Ch603α6 was absent, although bands at 7.2 and 4.4 kb were present. These observations are consistent with a reduction in the distance between the Ch603α6 V and C regions having occurred in the derivation of the M603 genome from the embryo genome. Genetic evidence suggests that a closely related V_H region, S107, may be several hundred thousand base pairs from the C_α region in embryonic DNA (31, 32; but see ref. 33). The other changes evident from the Southern blots of embryonic and M603 DNAs cannot yet be fully explained.

Electron Microscopy Indicates That There Are Three Intervening Sequences in the Genomic Clone Ch603α6. S107 or M315 mRNA was hybridized to trioxsalen-crosslinked Ch603α6 DNA under conditions favoring R-loop formation. At least 80% of the DNA molecules examined in the electron microscope were full length and >90% of these contained R loops. Typical micrographs are shown and interpreted in Fig. 5. Data from 52 molecules were combined to generate the map shown in Fig. 2*d*. These studies indicate that Ch603α6 contains three intervening sequences (IVS). IVS 1 has a length of 6.8 kb. The R loop to the left of IVS 1 in Fig. 2*d* is due to the V_H region because it is seen in hybridizations of Ch603α6 DNA with S107 mRNA but not with M315 mRNA. The C region structures discussed below are seen in hybridizations with either S107 or

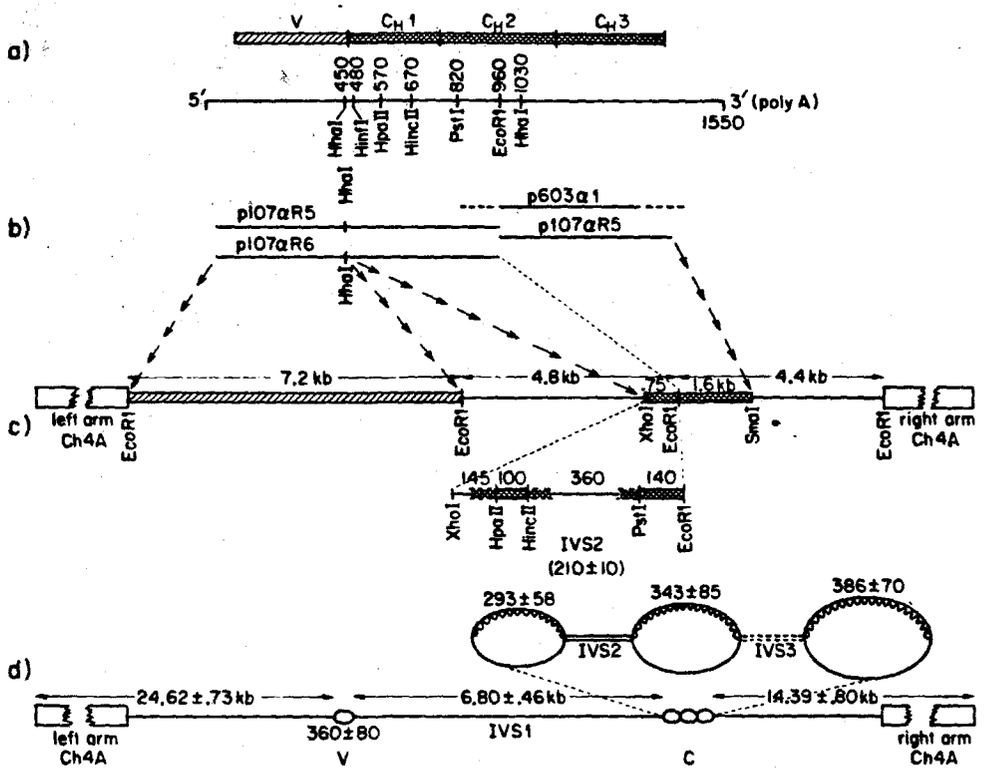


FIG. 2. Comparison of genomic and cDNA segments. (*a*) Restriction map of M603 double-stranded cDNA. The presence of a "hairpin" structure (determined by alkaline agarose gels) served as a marker for the 5′ end (27). Only the *HinfI* site identified by mRNA sequencing is shown. The structural domains depicted above the cDNA are based on a partial protein sequence of M315 and M511 heavy chains and analogy with other IgA proteins (28). The protein sequence has been aligned so that the *Hha* I site sequenced in S107 mRNA coincides with the *Hha* I site mapped in M603 cDNA. This alignment gives good correspondence between the *Hpa* II and *Hin*cII sites mapped in M603 cDNA and possible cleavage sites of these enzymes at codons for amino acids 158 and 191. (*b*) Sequences included in cDNA plasmid clones. The ends of p603α1 are uncertain, as indicated by dotted lines. The larger of the two *Eco*RI fragments in p107αR5 is present in inverted orientation relative to its position in the cDNA. This is presumably due to the independent ligation of these two fragments during the cDNA cloning procedure. The restriction fragments of Ch603α6 within which the various parts of the plasmid clones hybridize are indicated by arrows connecting the plasmid and genomic clones. The dotted line connects the internal *Eco*RI site of the cDNA plasmids with the corresponding site in Ch603α6. (*c*) Restriction map and sequence organization of Ch603α6. Restriction fragments to which V and C region probes (see arrows from plasmids) hybridize are shaded in correspondence to the protein domains. The enlarged portion of the figure shows one of the short intervening sequences (IVS 2). Compare distances between restriction sites here and those shown in *a*. The *Xho* I site does not lie within the coding sequence. (*d*) R-Loop map of Ch603α6. The right and left arms of Ch4A were taken to be 10.7 and 20.1 kb long, respectively. The enlargement of the C region R loops is drawn with the total length of the RNA·DNA duplex equal to the length of the cDNA from the 3′ end to the junction with the V region shown in *a*. IVS 2 and IVS 3 are both assumed to be the same length. The *Eco*RI site shown in the enlargement of the Ch603α6 restriction map above is aligned with a point 590 nucleotides (measured on the RNA·DNA duplex regions) from the 3′ end of the C region R-loop complex. This corresponds to the distance of the *Eco*RI site from the 3′ end of the cDNA. The wavy line in the R loops represents RNA. The dashed line for IVS 3 indicates that its length has not been accurately measured. Errors given are SD.

M315 mRNAs. These results and the measured lengths of the several R loops indicate that IVS 1 occurs approximately at the junction of the V_H and C_α regions (Fig. 2*d*). Two kinds of R-loop structures were seen for IVS 1, depending on whether a single mRNA molecule (Fig. 5 *a* and *b*) hybridized to both the C and V regions of the DNA or whether these regions were hybridized to two separate mRNA molecules (Fig. 5*c*).

In addition to the 6.8-kb IVS, 58% of the molecules contained two structures that we interpret as due to two short IVS within the C_α region (IVS 2 and IVS 3 in Fig. 2*d*). In one type of short IVS structure, the single- and double-stranded arms of a C region R loop are joined at a reproducible point inside the R loop. We interpret this as due to a base-paired IVS (bpIVS in Fig. 5). Alternatively, the two arms of an R loop are not joined, but there is a small knob at a reproducible point on the double-stranded arm. We interpret this as an IVS that is not base paired to its complement on the opposite strand (ssIVS in Fig. 5). The positions on the C region R loop where bpIVS and ssIVS structures are seen are coincident. Because IVS 2 and IVS 3 are observable as knobs but are too short to be measured by electron microscopy, we believe their lengths to be in the range of 100 to 200 nucleotides.

The reproducible junction points that we attribute to base-paired IVS could be due to site-specific trioxsalen crosslinking. However, the ssIVS structures cannot be explained this way. Furthermore, these structures occurred with approximately the same frequency in R loops with uncrosslinked Ch603α6 DNA.

Both electron microscopy and restriction mapping give the same orientation of the V_H and C_α regions relative to the right and left arms of Charon 4A. The position of the 3′ poly(A) end of the mRNA also has been independently determined in some R loops by an electron microscopic labeling technique (34) (Fig. 5*c*).

IVS 2 Is 210 Nucleotides in Length by Restriction Mapping. By comparing the sizes of restriction fragments produced from DNA of the genomic clone Ch603α6 with those from the cDNA clone p107αR6, we have confirmed the presence and determined the length of one of the short IVS in the C_α region (Fig. 6). Gel electrophoresis showed that the *Hin*cII and *Pst* I

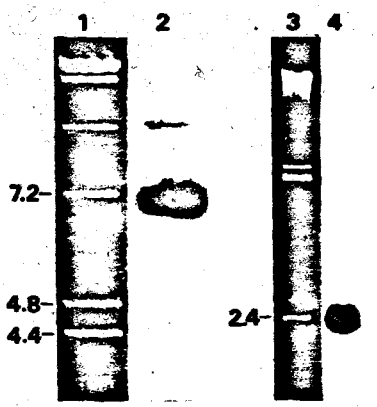


FIG. 3. Localization of the V and C regions in Ch603α6 by hybridization to Southern blots. Restriction enzyme digests of Ch603α6 DNA were electrophoresed on 1% agarose gels, transferred to nitrocellulose filters (30), and hybridized to V or C region probes prepared from gel-purified fragments of *Hha* I-digested p107αR6 (site of cleavage indicated in Fig. 2*b*). V region probe: ethidium bromide staining (lane 1) and blot (lane 2) of *Eco*RI-digested Ch603α6 DNA. Only the 7.2-kb fragment displayed strong hybridization to the V region. C region probe: ethidium bromide staining (lane 3) and blot (lane 4) of *Sma* I/*Xho* I-digested Ch603α6 DNA. Only the 2.4-kb fragment hybridized to this C region probe. The same result was obtained with probes from other parts of the C region. Unmarked bands seen by ethidium bromide fluorescence contained Ch4A vector DNA. The origins of the gels are not shown.

sites of the cDNA plasmid p107αR6 are separated by 150 nucleotides, whereas in the genomic DNA of Ch603α6 they are 360 nucleotides apart. This demonstrates the presence of an IVS of 210 nucleotides in the genomic clone which corresponds to the position assigned to IVS 2 by electron microscopy.

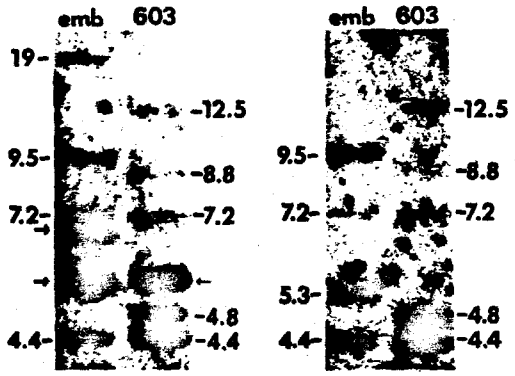


FIG. 4. Southern blots of *Eco*RI-cut genomic DNAs hybridized to p107αR5. High molecular weight genomic DNA was prepared from 12–13 day BALB/c embryos (emb) or M603 subcutaneous tumors (603) (18). Approximately 15 μg of DNA completely digested with *Eco*RI was electrophoresed on each lane of a 0.7% agarose gel. Nitrocellulose filter replicas of the gels were hybridized to ^{32}P-labeled p107αR5 (specific activity, 10^8 cpm/μg). Upper portions of the blots are not shown. In the two lanes at the left, different plasmid DNAs were added as internal standards (arrows). This obscures the 7.2- and 5.3-kb bands in the embryo DNA, which are better seen in the embryo lane at the right. The 19-kb band is not visible in the right embryo lane, probably because of poor transfer to the filter. Some clones from the M603 library contained a 4.4-kb *Eco*RI fragment hybridizing to p603α1, plus either a 12.5- or an 8.8-kb *Eco*RI fragment hybridizing to the 5′ C_α sequences in p107αR6. Thus, there may be multiple copies of the C_α gene segment, which would account for the relatively intense hybridization to the 4.4-kb band. All copies of the C_α gene in M603 DNA appeared to have undergone rearrangement or mutation from the embryo; the significance of this observation for the differentiation of antibody-producing cells is unknown.

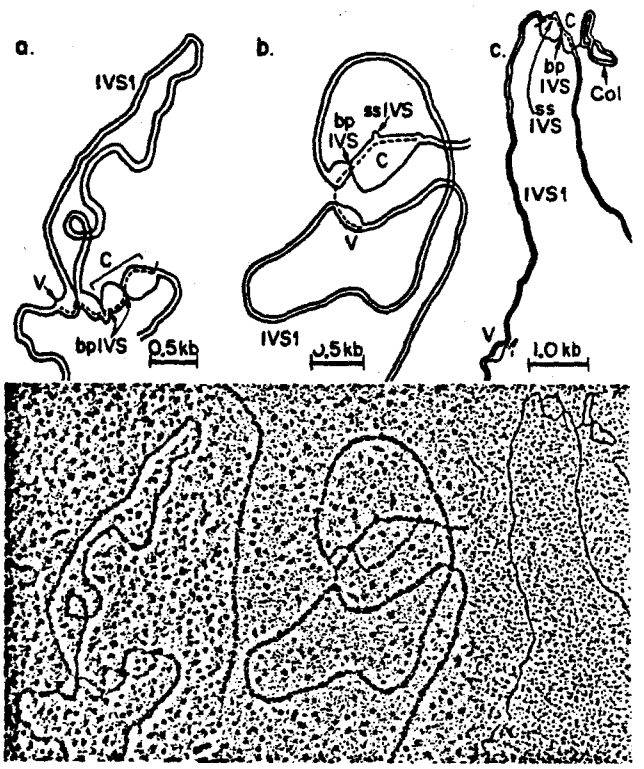


FIG. 5. Electron micrographs of R-loop structures observed with S107 mRNA hybridized to crosslinked Ch603α6 DNA. (*a*) R loop showing the large intervening sequence (IVS 1) dividing the constant (C) and variable (V) regions, and two short base-paired intervening sequences (bpIVS) in the constant region. (*b*) Similar to *a* but the C region contains a single-stranded intervening sequence (ssIVS) and a base-paired intervening sequence. (*c*) Two mRNA molecules hybridized to one Ch603α6 DNA molecule. The poly(A)$^+$ RNA at the 3′ end of the C region R loop is labeled with a poly(BrdU)-tailed circular microcolicin E1 molecule(Col). Both the V and C region R loops contain unhybridized RNA tails adjacent to IVS 1. The C region contains one bpIVS and one ssIVS. Broken lines represent RNA.

The Two Short C Region IVS May Separate the Three C_α Domains. The short IVS in the C_α gene separate the C region into three roughly equal segments. At the protein level the C_α region is divided into three roughly equal homologous structural domains: C_H1, C_H2, and C_H3 (28). Accordingly, the IVS in the DNA may separate the individual C_α domains, although this supposition will have to be verified by direct nucleic acid sequence analysis.

There are several features of immunoglobulin evolution and structure that might involve IVS separating C_H domains. First, all C regions, including C_α, contain homologous protein domains presumably derived from a common ancestral gene (2, 28, 37, 38). Second, C regions do not all contain the same number of domains, indicating that new C regions may arise by the addition or deletion of domains. Third, certain aberrant immunoglobulins (heavy chain disease proteins) often appear to involve deletions with breakpoints occurring between domains (39). Fourth, a variant myeloma cell line produced in culture shows deletion of the C_H1 domain (40). If IVS between domains facilitate unequal recombination, then the creation of new C regions with additional or deleted domains can be explained. Accordingly, the short IVS observed in the C_α region may permit the immunoglobulin domains to operate as fundamental units of evolution (41).

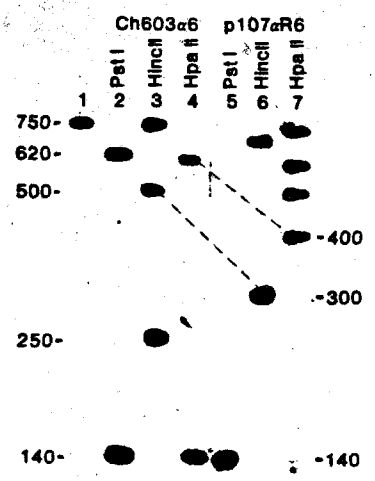


FIG. 6. Identification and measurement of IVS 2 by restriction mapping. Ch603α6 DNA digested with *Eco*RI and *Xho* I was ^{32}P end-labeled (35), and the 750-bp C region fragment (Fig. 2*c*) was eluted from a 1% agarose gel (36). An autoradiograph of part of a 5% acrylamide gel is shown comparing restriction digests of this end-labeled genomic DNA fragment and p107αR6 end-labeled after *Eco*RI digestion. Lanes: 1, Ch603α6 750-bp *Xho* I/*Eco*RI fragment; 2–4, products of digestion of this fragment with *Pst* I, *Hin*cII, and *Hpa* II, respectively (the *Hin*cII digest is incomplete); 5–7, *Pst* I, *Hin*cII, and *Hpa* II digests of p107αR6 end-labeled after *Eco*RI digestion. The *Eco*RI/*Pst* I bands of the two DNAs are the same length (140 bp), whereas the *Eco*RI/*Hin*cII and *Eco*RI/*Hpa* II bands are both 200–220 bp longer in the genomic DNA than in the cDNA (Fig. 2*c*). Comparing these results with the cDNA restriction map (Fig. 2*a*) shows that Ch603α6 contains a short intervening sequence (210 ± 10 bp) between the *Hin*cII and *Pst* I sites of the cDNA. This region includes the boundary between the C_H1 and C_H2 domains. (Other bands in lane 7 derive from *Hpa* II cleavage of the end-labeled plasmid DNA. The 790-bp fragment produced from *Hin*cII-digested p107αR6 is not shown on this part of the autoradiograph.)

We thank H. Manor for electron microscopy of p603α1, N. Johnson for providing M603 DNA, and Y.-H. Chien for electron microscopic characterization of mRNAs. *E. coli* polymerase I, DNA ligase T4, and *Eco*RI linkers were generous gifts from M. Goldberg, R. Scheller, and K. Itakura, respectively. Most of the *in vitro* packaging extracts used were the gift of T. Sargent. We thank K. Marcu, O. Valbuena, and R. Perry for advice on mRNA preparation and M. Wickens for communicating cDNA synthesis procedures prior to publication. We benefited from helpful discussions with T. Maniatis, T. Sargent, D. Goldberg, D. Engel, J. Dodgson, R. Joho, B. Klein, and D. Anderson. The work reported here was supported by National Institutes of Health Grants GM 10991 and GM 20927 and Biomedical Research Grant RRO7003A and National Science Foundation Grant PCM 71-00770. P.W.E. and M.M.D. are supported by National Institutes of Health Training Grant GM 07616; D.B.K. is a National Institutes of Health Postdoctoral Fellow.

1. Gally, J. (1973) in *The Antigens*, ed. Sela, M. (Academic, New York), Vol. 1, pp. 162–299.
2. Edelman, G. M., Cunningham, B. A., Gall, W., Gottlieb, P., Rutishauser, U. & Waxdal, M. (1969) *Proc. Natl. Acad. Sci. USA* **63**, 78–85.
3. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. (1978) *Nature (London)* **276**, 785–790.
4. Brack, C., Hirawa, M., Lenhard-Schuller, R. & Tonegawa, S. (1978) *Cell* **15**, 1–14.
5. Gilmore-Hebert, M. & Wall, R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 342–345.
6. Hood, L., Huang, H. V. & Dreyer, W. J. (1977) *J. Supramol. Struct.* **7**, 531–559.
7. Curtiss, R., III, Pereira, D. A., Hsu, J. C., Hull, S. C., Clarke, J. E., Maturin, L. J., Sr., Goldschmidt, R., Moody, R., Inoue, M. & Alexander, L. (1977) in *Proceedings of the 10th Miles International Symposium*, eds. Beers, R. F., Jr. & Bassett, E. G. (Raven, New York), pp. 45–56.
8. Leder, P., Tiemeier, D. & Enquist, L. (1977) *Science* **196**, 175–177.
9. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L. & Smithies, O. (1977) *Science* **196**, 161–169.
10. Sternberg, N., Tiemeier, D. & Enquist, L. (1977) *Gene* **1**, 255–280.
11. Marcu, K. B., Valbuena, O. & Perry, R. P. (1978) *Biochemistry* **17**, 1723–1733.
12. Pelham, H. R. B. & Jackson, R. J. (1976) *Eur. J. Biochem.* **67**, 247–256.
13. Wickens, M. P., Buell, G. N. & Schimke, R. T. (1978) *J. Biol. Chem.* **253**, 2483–2495.
14. Wensink, P. C., Finnegan, D. J., Donelson, J. E. & Hogness, D. S. (1974) *Cell* **3**, 315–325.
15. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. (1977) *Nature (London)* **270**, 486–494.
16. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3727–3731.
17. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
18. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
19. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
20. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
21. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
22. Thomas, M., White, R. C. & Davis, R. W. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2294–2298.
23. Davis, R., Simon, M. & Davidson, N. (1971) *Methods Enzymol.* **21D**, 413–428.
24. Hood, L., Loh, E., Hubert, J., Barstad, P., Eaton, B., Early, P., Fuhrman, J., Johnson, N., Kronenberg, M. & Schilling, J. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 817–836.
25. Berk, A. J. & Sharp, P. A. (1977) *Cell* **12**, 721–732.
26. McDonell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110**, 119–146.
27. Maniatis, T., Sim, G. K., Efstratiadis, A. & Kafatos, F. C. (1976) *Cell* **8**, 163–182.
28. Robinson, E. A. & Appella, E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2465–2469.
29. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
30. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
31. Riblet, R. J. (1977) *ICN-UCLA Symposia on Molecular and Cellular Biology* (Academic, New York), Vol. 6, pp. 83–89.
32. Robinson, E. A. & Appella, E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2465–2469.
33. Gearhart, P. J. & Cebra, J. J. (1978) *Nature (London)* **272**, 264–265.
34. Bender, W., Davidson, N., Kindle, K., Taylor, W., Silverman, M. & Firtel, R. (1978) *Cell* **15**, 779–788.
35. Berkner, K. L. & Folk, W. R. (1977) *J. Biol. Chem.* **252**, 3176–3184.
36. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
37. Beale, D. & Feinstein, A. (1976) *Q. Rev. Biophys.* **9**, 135–180.
38. Davies, D. R., Padlan, E. A. & Segal, D. M. (1975) *Annu. Rev. Biochem.* **44**, 639–667.
39. Frangione, B., Lee, L., Haber, E. & Bloch, K. (1977) *Proc. Natl. Acad. Sci. USA* **70**, 1073–1077.
40. Adetugbo, K., Milstein, C. & Secher, D. (1977) *Nature (London)* **265**, 299–304.
41. Gilbert, W. (1978) *Nature (London)* **271**, 501.

THE ORGANIZATION AND REARRANGEMENT OF HEAVY CHAIN IMMUNOGLOBULIN GENES IN MICE[1]

M. Davis, P. Early, K. Calame, D. Livant and L. Hood

Division of Biology, California Institute of Technology

Pasadena, California 91125

In: Eukaryotic Gene Regulation, edited by R. Axel, T. Maniatis and C. F. Fox. ICN-UCLA Symposium. Academic Press, in press

[1]This work was supported by NSF grant PCM 76-81546.

ABSTRACT A preliminary analysis of several heavy chain variable (V) and constant region (C) gene segments from sperm (undifferentiated) and myeloma (differentiated) DNA has revealed the following: 1) the V_H and C_α genes are separate in the germ line; 2) the V_H and C_α genes are rearranged during the differentiation of the antibody-producing cell; 3) multiple rearranged C_α genes are present in the DNA of a single myeloma tumor; 4) small intervening sequences may separate the domains of the and constant region genes; and 5) at least 8-9 germ line V_H genes exist for antibodies binding phosphorylcholine.

INTRODUCTION

The antibody gene families have several interesting organizational features. There are three distinct gene families - two code for light (L) chains, λ and κ, and the third codes for heavy (H) chains. They are composed of three distinct coding segments which are separated from one another by intervening DNA sequences - V (variable), J (joining) and C (constant). The V and J segments together comprise the V region of the antibody polypeptide which encodes the immunoglobulin domain concerned with antigen recognition. Moreover, each antibody gene family appears to contain multiple V and J segments.

The antibody gene families present two fascinating biological problems. First, it has been estimated that mammals can synthesize 10^5 to 10^8 different antibody molecules. What genetic mechanisms are responsible for this diversity of antibody molecules? We hope to assess the relative contributions of three genetic mechanisms: multiple germ line V genes (1), somatic mutation (2), and the joining in a combinatorial fashion of multiple V and J segments (3). Second,

how are antibody gene segments rearranged during the differentiation of antibody-producing cells? These DNA rearrangements presumably are fundamental components of the molecular events that commit the antibody-producing cell to the synthesis of a single type of antibody molecule as well as contributing to antibody diversity in the combinatorial joining of V and J segments (3,4).

We have focused on the analysis of the heavy chain gene family because, in addition to being an excellent system for studying the phenomena mentioned above, it has intricacies not exhibited in light chains. The heavy chain gene family of the mouse is comprised of an unknown number of variable (V_H) gene segments and at least eight different constant (C_H) gene segments (5) (Figure 1).

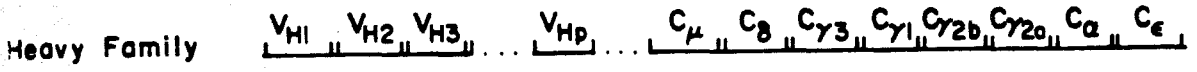


FIGURE 1. Heavy chain antibody gene family in mice. The order of C_H gene segments is uncertain, although indirect evidence supports the following alignment: $C_{\gamma 3}C_{\gamma 1}C_{\gamma 2b}C_{\gamma 2a}C_\alpha$ (20). The number of V_H gene segments is still a matter of controversy. The heavy chain gene family also has multiple J segments that are not depicted in this figure (see text).

The various classes and subclasses of immunoglobulins are determined by the C_H gene segments (e.g., C_μ-IgM, C_γ-IgG, C -IgA, etc.). Moreover, during the differentiation of the antibody-producing cell, distinct classes of immunoglobulins are expressed in a reproducible order (Figure 2). First IgM is expressed; later IgD and IgM are expressed; and eventually the other classes of immunoglobulins are expressed (6). In the lineage of a particular antibody-producing cell, it appears that these developmental shifts in immunoglobulin class expression occur by associating a particular V_H gene segment with different C_H gene segments while maintaining the expression of the same light chain gene segments. Therefore, a question of particular interest is the nature of the DNA rearrangements which lead to sequential and at times, simultaneous, expression of different heavy chain classes. Fortunately, tumors of antibody-producing cells exist which "freeze" this developmental pathway at many different points. Thus in time we will understand how the antibody gene organization for sperm cells (undifferentiated DNA) differs from that of tumor cell lines producing IgM, IgM + IgD and IgA (i.e., various stages of differentiation). Accordingly, our our initial efforts are focused on understanding the gene

organization in DNA at the beginning (sperm or embryo) and the end (IgA-producing myeloma) of a heavy chain differentiation pathway.

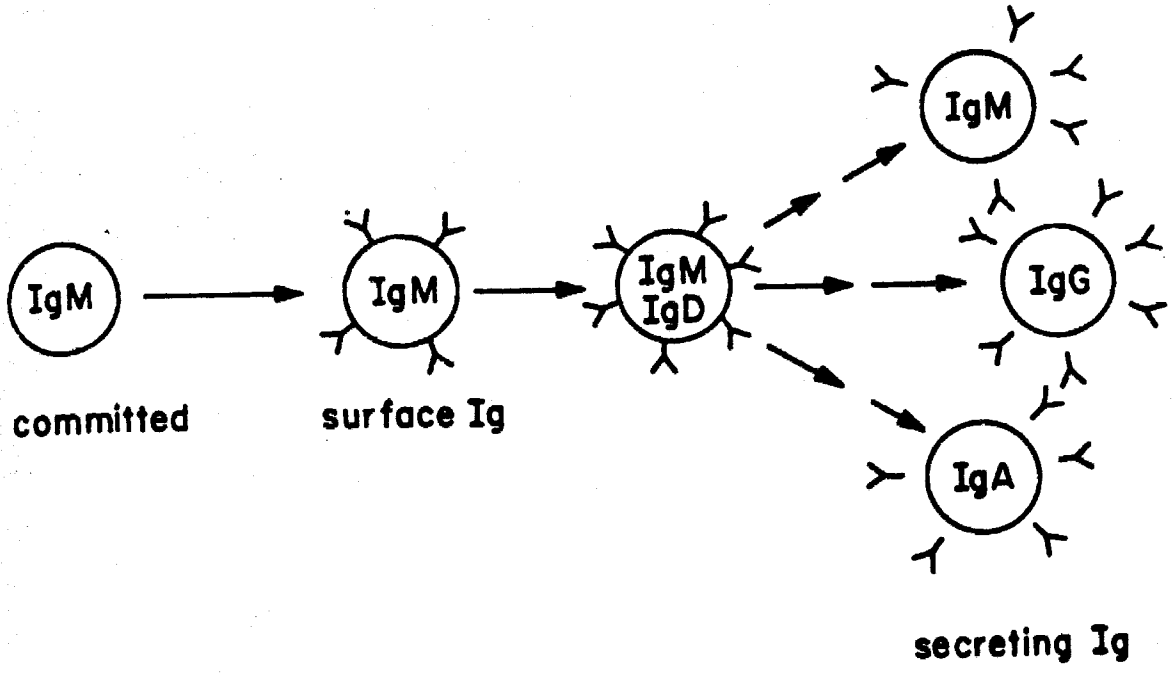


FIGURE 2. The differentiation of B cells. A B cell first becomes committed to the expression of a particular V domain (one V_L region and one V_H region) which is associated with cytoplasmic IgM molecules. Subsequently the IgM molecule is expressed on the cell surface. Later, cell-surface IgD molecules appear. Subsequent differentiation events lead to a terminally differentiated cell which specializes in the synthesis of soluble antibodies of one of a variety of immunoglobulin classes. For an individual B cell, the same V domain is associated with the various classes of immunoglobulins throughout the differentiation pathway.

THE PHOSPHORYLCHOLINE ANTIBODY SYSTEM

We have chosen to examine some of the questions posed above for a series of antibody-producing cells which synthesize immunoglobulin binding phosphorylcholine because this system allows us to analyze directly the biology of the immune response to phosphorylcholine (PC). Let us summarize the salient features of this system. First, several thousand myeloma tumors have been screened and twelve appear to

synthesize immunoglobulins binding phosphorylcholine (7). Our laboratory has determined the amino acid sequences of the V_H regions for seven of these tumors (8,9) and other laboratories have analyzed several additional sequences (10) (Figure 3). The V_H sequences from myeloma proteins binding phosphorylcholine illustrate several features of V diversity. 1) Four V_H sequences are identical. Since these identical V_H sequences were expressed independently in different mice, it appears that they are encoded by a germ line V_H gene segment designated T15. This reasoning argues that it is unlikely that four somatic variants would be identical in amino acid sequence. 2) The variant sequences differ by one to eleven amino acid substitutions and also exhibit sequence gaps. Accordingly, one can hope to determine the nature and extent of diversity generated from somatic genetic mechanisms by sequencing germ line PC V_H gene segments and comparing them with the protein diversity patterns reflected in their myeloma counterparts. Second, antisera have been raised which are specific for the V domains of several myeloma proteins binding phosphorylcholine. These antisera are termed anti-idiotypic antisera. Anti-idiotypic antisera to T15 can be used to map genetic elements which control the expression of this V_H domain. The T15 idiotype maps about 0.4 centiMorgans (cM) from the C_H gene cluster (11) and simplistic genetic calculations suggest the PC V_H and C_H gene segments are separated by hundreds of thousands or even a million nucleotides. For example, mouse chromosomes have about 25 chiasmata per meiosis (12). With a genome of 3×10^9 nucleotide pairs, 0.4 cM of DNA in the mouse would span about 10^6 nucleotide pairs, if meiotic recombination were random. Third, the T15 idiotype appears to be present on at least one type of T cells ("helper T cells") (13), implying that T-cell receptors and B-cell immunoglobulins may share the same V_H repertoire of genes. Thus an analysis of the phosphorylcholine system may provide opportunities to analyze T-cell receptors. Finally, the hybridoma system of Milstein and Köhler (14) has been employed to generate homogeneous antibodies to phosphorylcholine. In collaboration with Dr. Patricia Gearhart, we are analyzing 20 hybridomas to phosphorylcholine in order to broaden our knowledge about the phenotypic diversity patterns of the phosphorylcholine system. The importance of detailed protein sequence studies on the products of complex multigenic systems such as the antibody gene families cannot be overemphasized, for these phenotypic diversity patterns are one of the end results of heavy chain gene organization and rearrangements and any meaningful understanding of this system at the DNA level must account for the resultant diversity of its gene products. Thus we hope the phosphorylcholine system will provide

insights into antibody gene diversity and organization and permit us, in time, to begin analyzing the more complex regulatory events of this sophisticated system.

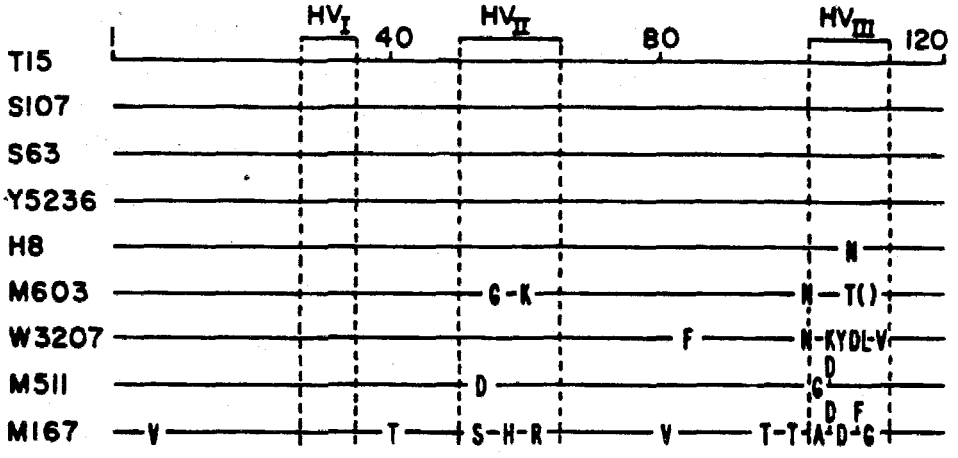


FIGURE 3. The amino acid sequences of V_H regions from immunoglobulins binding phosphorylcholine. Identities of these sequences to the V_H region of T15 are indicated by a straight line. The one letter code of Dayhoff is used to indicate amino acid substitutions (28). Deletions are indicated by brackets. Insertions are denoted by a vertical bar. The three hypervariable regions which fold in three dimensions to constitute the walls of the antigen-binding site of the V domain are designated by HV_I, HV_{II}, HV_{III} and dotted lines.

OUR APPROACH

We have constructed libraries in Charon 4A bacteriophage from partial restriction digests of sperm, embryo, and myeloma DNA (15, 16). The sperm and embryo libraries are a source of undifferentiated DNA. The myeloma library, derived from the tumor MOPC 603 which synthesizes IgA molecules binding phosphorylcholine, represents a terminal stage in the differentiation of an antibody-producing cell. We also have purified mRNA from a variety of myeloma tumors, and used these as templates for the synthesis of double-stranded DNA copies which were then inserted into plasmids (16). Our initial approach has been to compare the genomic organizations of undifferentiated (sperm or embryo) and differentiated (IgA myeloma tumor) DNAs. To this end we have isolated a number of genomic clones from both the M603 library and from a sperm library, using cDNA probes for the complete V_HC_α coding region of myeloma protein S107. The V_H regions of the

S107 and the M603 immunoglobulins are very closely related (Figure 3) and the corresponding mRNAs completely protect one another in S1 nuclease digestion experiments (16). Certain of these initial experiments have recently been published in a paper which describes for the first time a heavy chain genomic clone containing the V_H and C_α gene segments and the presence of intervening sequences within the C_α coding region, probably separating the coding regions for immunoglobulin α domains (16). These results as well as more recent observations are summarized below.

EXPERIMENTAL OBSERVATIONS

The Variable and Constant Regions of α Heavy Chains Appear to be Encoded by Distinct V_H and C_α Gene Segments which are Rearranged During Differentiation. We have analyzed a series of overlapping genomic clones from the M603 library which have the general structures illustrated in Figure 4. The V and the C gene segments are separated by 6.8 kilobases. Furthermore, idiotypic mapping, discussed above, suggests that these regions were separated by hundreds of thousands of nucleotides prior to differentiation of this antibody-producing cell with the concomitant DNA rearrangements. A heteroduplex comparison of a sperm V_H clone with the myeloma M603 clone, which will be discussed subsequently, also provides evidence for the rearrangement of the V_H gene segment in the myeloma DNA. Accordingly, the V_H and C_α gene segments are originally widely separated from one another. As the antibody-producing cell differentiates, DNA rearrangements of antibody V and C gene segments occur over extensive stretches of DNA.

The C_α Gene Segments from the M603 Myeloma Library are Present in Multiple Rearranged Forms. A comparison of Southern blots on sperm M603 DNA using the C_α probe demonstrates that the myeloma DNA has three forms of the C_α gene, none of which are identical to their germ line counterpart (Figure 5). These three forms have been isolated from the M603 library as Charon 4A clones (Figure 6). Restriction enzyme analyses and heteroduplex comparisons demonstrate that, although they share 2.7 or more kilobases of homology just 5' to the C_α gene, each of these three clones is distinct from the others in their more 5' regions.

These observations raise several interesting possibilities. The absence of a germ line-like C_α gene segment in the M603 DNA suggests that the C_α gene segments in both the maternal and paternal chromosomes coding for heavy chain genes have been rearranged. Immunoglobulin-producing cells exhibit allelic exclusion; that is, a particular antibody-

ALPHA GENOMIC CLONES FROM MYELOMA M603 DNA

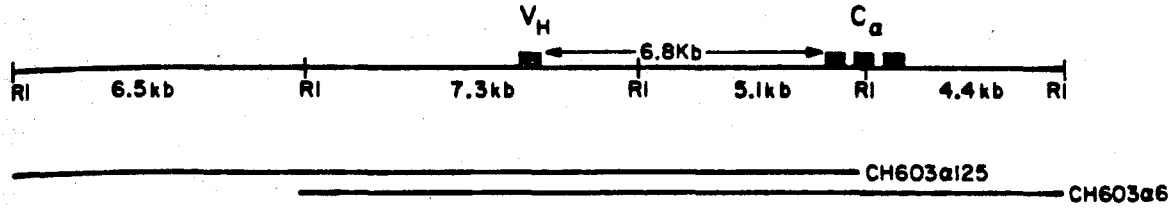


FIGURE 4. The organization of V_H and C_α gene segments from DNA derived from myeloma tumor M603. Kb denotes kilobases. R1 denotes Eco R1 cleavage sites. The distances between Eco R1 sites are indicated. CH603α125 and CH603α6 are two clones derived from the phage library of M603 DNA. The V_H gene segment is separated from the C_α gene segment by 6.8 kilobases of intervening DNA. R-loop mapping and restriction enzyme analyses demonstrate that the C_α segment is divided into three approximately equal segments, presumably coding regions for the three C_α domains, by two small intervening DNA sequences (16).

producing cell may express the maternal or paternal allele for a particular immunoglobulin family, but not both alleles. In the past the phenomenon of allelic exclusion has been explained by suggesting that either the maternal or paternal chromosome does not rearrange at the DNA level and, accordingly, cannot express an immunoglobulin polypeptide. This suggestion has come from Southern blot analyses of myeloma DNAs in which the germ line pattern of constant gene segments for light chains appears to be preserved (17). Our data on the alpha constant region genes of the M603 myeloma DNA suggests that both the maternal and paternal chromosomes undergo rearrangements, but that one of these rearrangements is abortive in the sense no gene product is expressed. It will be interesting to determine whether these abortive DNA rearrangements include V gene segments; or whether only the C gene segment is involved in the rearrangement. Moreover, it will be interesting to analyze carefully the myeloma examples that appear to have germ line C fragments to determine whether the DNA rearrangements have been missed due to technical limitations of the Southern blotting technique, or contamination with somatic DNA. It may be that all myeloma DNAs in fact rearrange both the paternal and maternal chromosomes--one in a productive and the second in an abortive fashion.

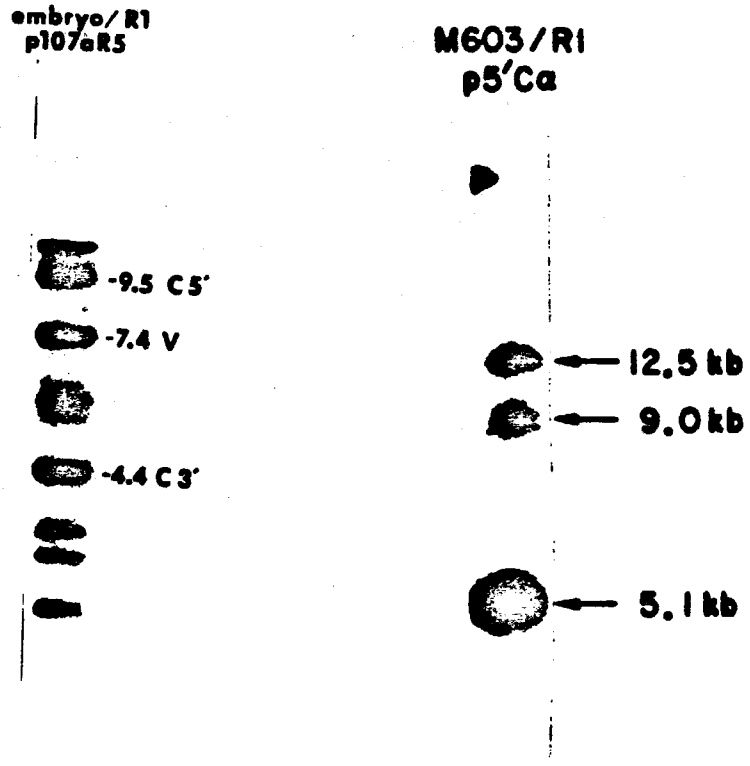


FIGURE 5. Southern blots of embryo (undifferentiated) and myeloma M603 (differentiated) DNAs. The picture on the left is a Southern blot of 13-day embryo DNA after digestion with the Eco R1 enzyme, separation of the DNA fragments on agarose, and hybridization with a cDNA probe derived from mRNA of myeloma tumor S107. This probe contains both the V_H and C_α coding regions. Assignments of the C_α fragments are based on Southern blots with separated V_H and C_α probes (data not shown). The remaining fragments must be V_H gene fragments. Thus there are at least 8-9 germ line V_H genes which cross-hybridize with the V_H probe from myeloma tumor S107. The exposure on the right is a Southern blot of tumor M603 DNA after Eco R1 digestion and hybridization to a plasmid containing the 5' half of the C_α coding region (an R1 site separates the 5' from the 3' half of the C_α gene segment; see Figure 4). The 5' C_α probe gives just one 9.5 kilobase band in the embryo DNA (data not shown) and 5.1, 9.0 and 12.5 kilobase bands in the M603 DNA. Hybridization to the 3' half of the C_α coding region gives a 4.4 kilobase band in both embryo and myeloma DNA (not shown).

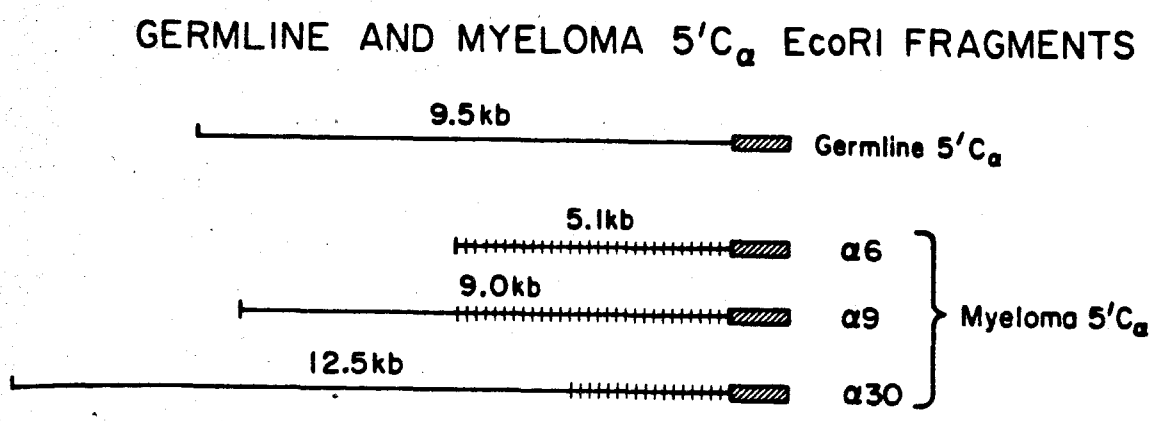


FIGURE 6. Eco R1 genomic fragments including the 5' portion of the C_α gene from myeloma M603 DNA and sperm DNA. The genomic clones α6, α9, and α30 have been derived from the M603 phage library. The structure of the germ line C_α clone comes from a Southern blot analysis of sperm or embryo DNA (Figure 5). The boxes represent the 5' portion of the C_α coding sequence (see Figure 4), whereas the hashmarks represent DNA homologies revealed by heteroduplex analyses.

One surprising observation that is difficult to explain is the presence of three distinct C_α clones in the M603 DNA. Several explanations may be offered, none really satisfactory. First, the germ line may contain two C_α genes, both the same size by Eco R1 restriction analysis. Both of these C_α genes may undergo rearrangements of several different types. Second, perhaps the abortive rearrangement is unstable and may be subject to additional DNA rearrangements. Third, perhaps there are several different M603 cell types in the uncloned tumor from which the DNA was derived. The possibility that the M603 C_α pattern is some aberration of this particular tumor line seems unlikely because at least one other phosphorylcholine binding tumor (H8) has an identical pattern on Southern blots (M. Davis and P. Early, unpublished). Thus in the case of the C_α gene segments, it appears that both the maternal and paternal chromosomes undergo DNA rearrangements, some of which are abortive (nonproductive) while others lead to the expression of one V_H-C_H pair of gene segments.

The V and C Rearrangements in Heavy Chains Resemble Those of Light Chains in Some Respects but Not Others. The V_L and C_L gene segments are rearranged by a fusion at the DNA level of V_L and J_L gene segments with the removal (or rearrangement) of the intervening DNA (Figure 7) (4, 17). Accordingly, the DNA 5' to the V_L gene segment is identical to that of the unrearranged V_L gene and the intervening DNA between the V and C gene segments is derived from the region

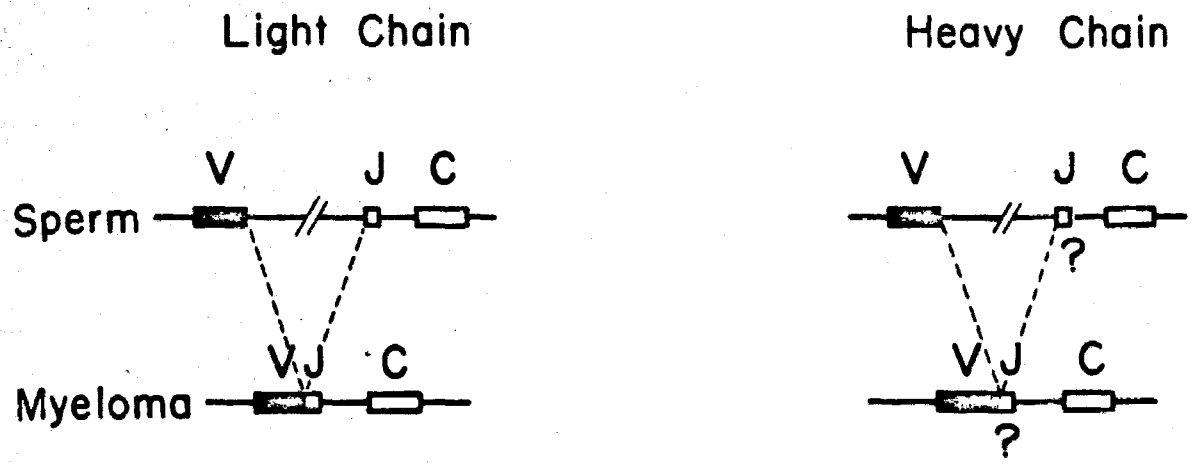


FIGURE 7. A model of the joining of light and heavy chain gene segments. An analysis of λ (4) and κ (18) light chain gene segments indicate that the 3' side of a V segment is fused to the 5' side of a J segment. The intervening DNA sequence between the J segment and the C segment remains unchanged in the DNA rearrangement process. The heavy chain gene segments appear to rearrange in a similar fashion, although the organization of the intervening DNA sequence between the J and C gene segments is altered, presumably because of multiple DNA rearrangements between one V_H gene segment and two (or more) C_H gene segments (see text).

5' to the unrearranged C_L gene. The existence of J_H segments for heavy chains is strongly implied from protein sequence data (18) and has recently been demonstrated by the DNA sequence analysis of a sperm clone containing a V_H segment (P. Early and M. Davis, unpublished observation). Comparison of a sperm V_H clone and the joined V_H and C_α myeloma clone (α6) by DNA heteroduplex analysis demonstrates that those regions 5' to the V segment are homologous and those regions 3' to the V segment are nonhomologous (Figure 8). In this respect the heavy chain variable region gene segment appears to rearrange in a manner similar to its light chain counterparts (Figure 7).

The rearrangement of V_H and C_H gene segments differs from those of the light chains in one important regard. Certain of the intervening sequences between the V_H and C_α gene segments of the α6 clone (Figure 4) are not derived from germ line DNA 5' to the C_α gene. For example, a Southern blot analysis of germ line DNA with a C_α probe shows that the closest Eco R1 site is 9.5 kilobases from the 5' side of the C_α gene segment (Figure 5). However, the α6 clone

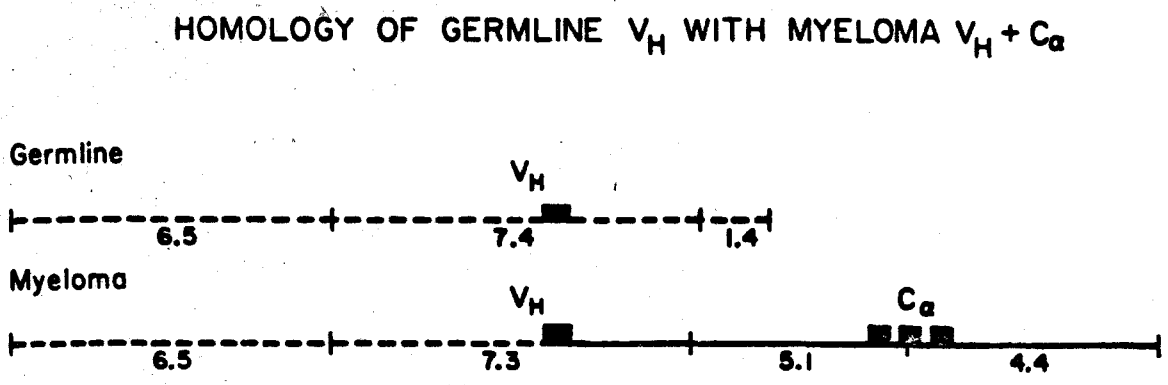


FIGURE 8: Homologies determined by heteroduplex analysis between the flanking sequences of germ line and myeloma V_H clones. The dotted lines indicate germ line sequences. Accordingly, the intervening DNA sequence between the V_H and C_α gene segments is not derived from the sperm V_H clone. The sperm library was constructed by M. Davis and R. Joho.

from the M603 DNA has an Eco R1 site 5.1 kilobases from the 5' end of the C_α gene segment. In addition, as discussed above, this α6 DNA is not homologous to DNA of the sperm V_H clone (Figure 8). Moreover, the Eco R1 site of the α6 clone in the DNA between the V and C gene segments does not seem to have been created by a spurious mutation, since Southern blots of DNA from an independently derived tumor line (H8) show the same C_α Eco R1 fragment. One explanation for the origin of the DNA sequence between V_H and C_α gene segments in the α6 clone containing this Eco R1 site is that it arises from the DNA rearrangement events of an earlier stage in differentiation, in which this V_H gene segment was formerly joined to a different C_H (or J) gene segment. Indeed, during the differentiation of antibody-producing cells, the V_H gene segment appears initially to be joined to a C_μ gene (Figure 2), so we would predict that some of the intervening DNA in the α6 clone between the V_H and C_α gene segments may be derived from the 5' side of a germ line C_μ gene segment. The subsequent joining of this V_H segment to a C_α gene segment later in development might displace or delete (19) the C_μ gene, but not all of its flanking sequences.

Intervening Sequences Appear to Separate the Domains of the C_H Genes. The C_α polypeptide is divided into three discrete molecular domains, each of which encompasses about 110 amino acid residues (20). We initially used R-loop mapping to demonstrate the existence of two small intervening sequences (IVS2, IVS3) which separate the C_α coding region into three roughly equal segments (Figure 4) (16). Subsequent restriction enzyme analyses of the M603 genomic clone (α6) places IVS2 within 30 amino acids of the domain boundary

between the $C_\alpha 1$ and $C_\alpha 2$ homology units (16; M. Davis, unpublished). Thus it appears likely that the two intervening sequences will separate the C_α gene into three distinct coding segments, one for each C_α domain (Figure 4). In addition, we have analyzed a μ genomic clone from the M603 library by R-loop mapping. The C_μ region has four domains (21) and, as expected, R-loop analysis demonstrates that the C_μ coding region is divided by three small intervening sequences into four roughly equivalent segments (K. Calame, P. Early, M. Davis, D. Livant, unpublished observations). The analysis of a genomic $\gamma 1$ clone has established that intervening sequences separate the three $C_{\gamma 1}$ domains and the hinge region from one another precisely at the interdomain boundaries (22). Therefore it appears reasonable to conclude that intervening sequences will divide all of the immunoglobulin C genes coding into segments for structural domains (see Figure 1).

The function of intervening sequences has generated spirited controversy and discussion. Individual domains of the immunoglobulin molecule carry out discrete and independent functions (20). Accordingly, the immunoglobulin intervening sequences appear to perform the important task of breaking the coding regions into discrete units which may then rearrange independently of one another through recombination at either the DNA level or the nuclear RNA level as proposed by Gilbert (23). Several lines of evidence suggest that the domains of immunoglobulins may be discrete evolutionary units. First, C_H regions with two, three, and four domains are present in vertebrate antibodies. Second, heavy chain disease deletions (24) and spontaneous deletions in tissue culture lines (25) suggest that frequent nonhomologous crossing-over occurs at or between domain boundaries. Perhaps intervening sequences not only separate domains but facilitate recombination as well. It will certainly be interesting to determine the homology relationships, if any, of the various immunoglobulin intervening sequences to one another.

The Germ Line V Gene Segments of Mouse Heavy Chains Appear to be as Diverse as Their V_K Counterparts. The V_H regions derived from myeloma proteins binding phosphorylcholine show a limited range of heterogeneity (Figure 3). We are interested in determining whether these different V_H sequences are germ line or in part derived by somatic mutation. Southern blot analysis of embryo DNA employing the S107 cDNA probe reveals at least 8-9 restriction fragments which hybridize to the S107 V region probe (Figure 5). The PC V_H regions represent a single group of heavy chain variable regions (26). Approximately 20 other groups of

V_H regions have been defined (26). Therefore, if each group is on the average encoded by ~10 germ line genes, the heavy chain gene family may be comprised of approximately 200 V_H genes. Since the amino acid sequence analyses of mouse V_H regions are relatively limited, it appears likely that in time many additional V_H groups will be defined. By similar analyses, the V_K family of mouse appears to be encoded by 200 or more germ line V genes (3, 27). We have isolated several different PC V_H genes and are now in the process of sequencing them to determine the relative contributions of germ line diversity, somatic mutation, and combinatorial joining of V_H and J_H segments to antibody variability.

The Generality of Nucleic Acid Rearrangements. The intriguing general question posed by the studies on immunoglobulin genes is whether DNA rearrangements are a fundamental aspect of differentiation in other eukaryotic systems. An answer to this question will await more detailed analyses of other gene families, both simple and complex.

ACKNOWLEDGMENTS

The work here is supported by National Science Foundation Grant PCM 76-81546. MD, PE, and DL are supported by National Institutes of Health Training Grant GM 07616. KC is supported by National Institutes of Health Fellowship GM 05442.

REFERENCES

1. Hood, L., Campbell, J. H., and Elgin, S. C. R. (1975). Ann. Rev. Genet. 9, 305.
2. Cohn, M., Blomberg, B., Geckeler, W., Raschke, W., Riblet, R., and Weigert, M. (1974). "The Immune System," ICN-UCLA Symp., p. 89. Academic Press.
3. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J., and Hood, L. (1978). Nature 276, 785.
4. Brack, C., Hirawa, M., Lenhard-Schueller, R., and Tonegawa, S. (1978). Cell 15, 1.
5. Mage, R., Lieberman, R., Potter, M., and Terry, W. (1973). In "The Antigens" (M. Sela, ed.), Vol. I, p. 300. Academic Press.
6. Goding, J. W., Scott, D. W., and Layton, J. E. (1977). Immunol. Rev. 37, 152.
7. Potter, M. (1970). Physiol. Rev. 52, 631.
8. Hood, L., Loh, E., Hubert, J., Barstad, P., Eaton, B., Early, P., Fuhrman, J., Johnson, N., Kronenberg, M., and Schilling, J. (1976). Cold Spring Harbor Symp. Quant. Biol. 41, 817.

9. Hubert, J., Johnson, N., Barstad, P., Rudikoff, S., and Hood, L. In preparation.
10. Rao, D. N., Rudikoff, S., and Potter, M. (1978). Biochemistry 17, 5555.
11. Riblet, R. J. (1977). "Molecular and Cellular Biology," ICN-UCLA Symp., Vol. 6, p. 83. Academic Press.
12. Klein, J. (1975). "The Biology of the Mouse Histocompatibility Complex." Springer-Verlag.
13. Cosenza, H., Augustin, A., and Julius, M. (1977). Cold Spring Harbor Symp. Quant. Biol. 41, 709.
14. Köhler, G., and Milstein, C. (1976). Eur. J. Immunol. 6, 511.
15. Maniatis, T., Hardison, R., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G., and Efstratiadis, A. (1978). Cell 15, 687.
16. Early, P., Davis, M., Kaback, D., Davidson, N., and Hood, L. (1979). Proc. Nat. Acad. Sci. USA 76, 857.
17. Seidman, J. G., and Leder, P. (1978). Nature 276, 790.
18. Schilling, J., Clevinger, B., Davie, J., and Hood, L. In preparation.
19. Honjo, T., and Kataoka, T. (1978). Proc. Nat. Acad. Sci. USA 75, 2140.
20. Edelman, G. M., Cunningham, B. A., Gall, W., Gottlieb, P., Rutishauser, U., and Waxdal, M. (1969). Proc. Nat. Acad. Sci. USA 63, 78.
21. Beale, D., and Feinstein, A. (1976). Quart. Rev. Biophys. 9, 135.
22. Sakano, H., Rogers, J. H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R., and Tonegawa, S. (1979). Nature 277, 627.
23. Gilbert, W. (1978). Nature 271, 501.
24. Frangione, B., Lee, L., Haber, E., and Bloch, K. (1977). Proc. Nat. Acad. Sci. USA 70, 1073.
25. Adetugbo, K., Milstein, C., and Secher, D. (1977). Nature 265, 299.
26. Barstad, P., Rudikoff, S., Potter, M., Cohn, M., Konigsberg, W., and Hood, L. (1974). Science 183, 962.
27. Seidman, J., Leder, A., Nau, M., Norman, B., and Leder, P. (1978). Science 202, 11.
28. Dayhoff, M. O. (1972). In "Atlas of Protein Sequence and Structure," Vol. 5, Biomedical Research Foundation, Washington, D.C.

An immunoglobulin heavy-chain gene is formed by at least two recombinational events

Mark M. Davis*, Kathryn Calame*, Philip W. Early*, Donna L. Livant*, Rolf Joho†, Irving L. Weissman† & Leroy Hood*

* Division of Biology, California Institute of Technology, Pasadena, California 91125
† Laboratory of Experimental Oncology, Department of Pathology, Stanford Medical School, Stanford, California 94305

The events of B-cell differentiation can be reconstructed in part through an analysis of the organisation of heavy-chain gene segments in differentiated B cells. A mouse immunoglobulin α heavy-chain gene is composed of at least three noncontiguous germ-line DNA segments—a V_H gene segment, a J_H gene segment associated with the C_μ gene segment, and the C_α gene segment. These gene segmerits are joined together by two distinct types of DNA rearrangements—a V–J joining and a C_H switch.

THE antibody genes provide a unique opportunity for studying the molecular basis of one pathway of eukaryotic differentiation because the rearrangement of gene segments is correlated with the expression of antibody molecules. Antibody molecules are encoded by three unlinked gene families—λ, κ and heavy chain (H)[1]. The λ and κ families encode light (L) chains and the heavy-chain family encodes heavy chains. The light chains are encoded by three gene segments, V_L (variable), J_L (joining) and C_L (constant), which are separated in the genomes of cells undifferentiated with regard to antibody gene expression[2,3]. During differentiation of the antibody-producing or B cell, the V_L and J_L gene segments are rearranged and joined together while the intervening DNA between the J_L and C_L gene segments remains unmodified[2–4]. This process of DNA rearrangement is termed V–J joining. During the expression of the rearranged gene in the differentiated B cell, the coding regions as well as the intervening DNA between the J_L and C_L gene segments are transcribed as part of a high molecular weight nuclear transcript. The intervening region is then removed by RNA splicing to produce a light-chain mRNA with contiguous V_L, J_L and C_L coding segments[5,6]. Recently, we demonstrated that the heavy chain contains three analogous gene segments, V_H, J_H and C_H, which undergo a similar type of V–J joining during B-cell differentiation[7–9]. Each antibody-producing cell synthesises only one V_LJ_L polypeptide sequence and one V_HJ_H sequence, which together form the antigen-binding (V) domain of the antibody molecule.

The heavy-chain genes seem to have a special role in the differentiation of the antibody-producing cell as reflected in a second phenomenon known as the switching of heavy-chain constant regions or C_H switching. Antibody molecules can be divided into five different immunoglobulin classes—IgM, IgD, IgG, IgA and IgE—which are determined by one of eight distinct heavy chain genes [that is, C_μ, C_δ, ($C_{\gamma1}$, $C_{\gamma2a}$, $C_{\gamma2b}$, $C_{\gamma3}$), C_α and C_ε]. Each class of immunoglobulin seems to be associated with unique functions such as complement fixation or the release of histamine from mast cells. The process of B-cell differentiation is complex and a variety of conflicting views exists on the transitional stages. It is generally agreed, however, that IgM is the earliest immunoglobulin class that is expressed in the differentiation of a B cell (see ref. 10). The immature B lymphocyte apparently has the capacity to differentiate along a variety of discrete pathways and produce progeny which may switch from IgM expression to the expression of any one of the other classes of immunoglobulins. The terminal stage of B-cell differentiation is the plasma cell, which is committed to synthesise and secrete large quantities of a single molecular species of antibody. Several studies suggest that the specificity or V domain does not change as different immunoglobulin classes are expressed throughout this differentiation process[11–15]. During C_H switching, light-chain expression remains unaltered.

We are interested in studying the molecular mechanisms which permit a B cell (or its progeny) to express successive classes of antibody molecules. Here we demonstrate that an α heavy-chain gene derived from a terminally differentiated plasma cell is composed of three noncontiguous germ-line DNA segments. These gene segments are joined together by at least two distinct DNA rearrangements—a V–J joining and a C_H switch.

Organisation of the α heavy-chain gene is altered in differentiated as opposed to undifferentiated DNAs

We have constructed 'libraries' of recombinant phage[16] containing large (12–20 kilobases) inserts of mouse DNA in the vector Charon 4A (ref. 17). These libraries contain sufficient numbers of recombinants ($\sim 10^6$) for there to be a high probability that most single-copy sequences of a given genome are included. We have previously reported the construction of such a library from the DNA of the mouse IgA-producing myeloma tumour M603, and the subsequent isolation and characterisation of clones containing rearranged or differentiated genomic DNA. These clones, CH603α6 (α6) and CH603α125 (α125), have V_H and C_α gene segments on a single fragment of DNA[7,8]. The *Eco*RI restriction maps of these clones are shown in Fig. 1*a*. The α6 clone has three *Eco*RI fragments—one 7.3 kilobases long containing the V_H gene segment, a second of 5.1 kilobases containing the 5′ portion of the C_α gene segment, and a third of 4.4 kilobases containing the 3′ portion of the C_α gene segment. Approximately 6.8 kilobases of intervening DNA separate the V_H and C_α gene segments. The α125 clone also has three *Eco*RI fragments—one of 6.5 kilobases which is located on the 5′ side of the 7.3- and 5.1-kilobase *Eco*RI fragments also described for α6.

The genomic clones were isolated using a cloned cDNA plasmid representing the entire heavy-chain mRNA of the IgA-producing tumour S107 (denoted S107 cDNA)[7]. The S107 V_H region is about 98% homologous to the M603 V_H region at the nucleotide level[9]. Analysis by the Southern blot procedure[18,19] of *Eco*RI-digested M603 DNA hybridised to the S107 cDNA probe demonstrates the presence of three *Eco*RI fragments corresponding to those described above in the α6 clone[7]. Thus, the rearranged α6 clone is not an artefact of the cloning or isolation procedures.

When the S107 cDNA probe is used to examine a Southern blot of *Eco*RI-digested sperm or embryo DNA, a somewhat different pattern is observed[8]. In particular, the 5.1-kilobase

0028–0836/80/080733—07$01.00

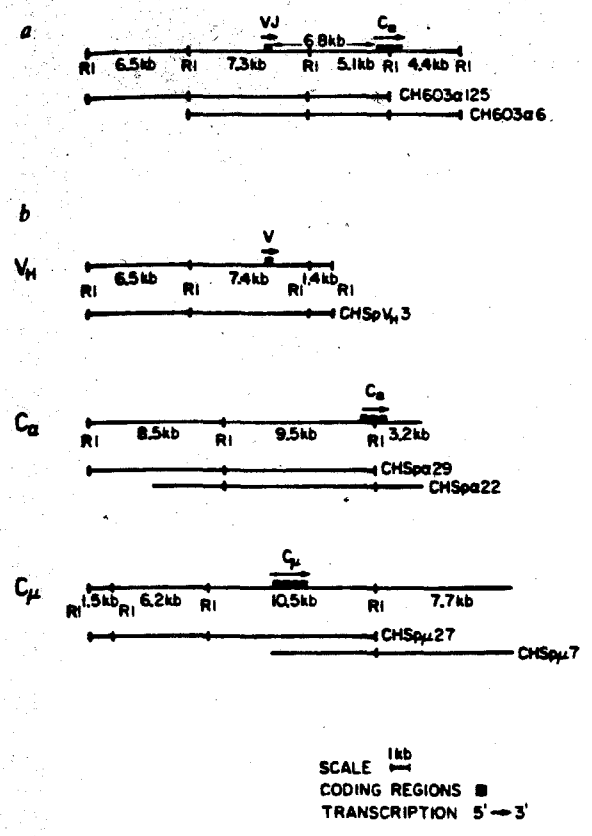


Fig. 1 *a*, α Heavy-chain clones isolated from a genomic library of myeloma M603 DNA. CH603 α6 and CH603 α125 are two overlapping clones derived from a partial *Eco*RI library of M603 DNA, constructed using the phage Charon 4A (ref. 7). The position of the respective gene segments, as well as their direction of transcription, was determined by R-loop mapping and restriction analysis[7,8]. *b*, Germ-line V_H, C_μ and C_α clones isolated from a genomic library of sperm DNA from inbred BALB/c mice. Sperm DNA[33,34] was partially digested with restriction enzymes in two ways: (1) with a mixture of *Hae*III plus *Alu*I and (2) with *Eco*RI alone. After digestion in conditions designed to maximise the yield of 12–20-kilobase fragments, these fragments were selected on sucrose gradients. The *Hae*III/*Alu*I fragments were methylated with *Eco*RI methylase and blunt end ligated with synthetic *Eco*RI cleavage sites and then cleaved with *Eco*RI essentially as described previously[16] except that *Eco*RI linker ligations were done at 18 °C to lessen the endonuclease degradation. Both types of fragments were then ligated to the isolated arms of the bacteriophage Charon 4A (ref. 16) at 4 °C. The enzymatically recombined DNAs were packaged *in vitro* using the strains of Sternberg[35] and the protocol of Hohn[36]. The efficiency of packaging was 400,000 plaque forming units (PFU) per μg inserted DNA in the case of *Eco*RI partially digested DNA and 200,000 PFU per μg for *Hae*III/*Alu*I digestions. The background of non-recombinant Charon 4A was 25% and <5%, respectively. Approximately 500,000 *Eco*RI and 1,200,000 *Hae*III/*Alu*I clones were constructed and amplified. A library of ~500,000 clones provides a 90% chance of finding a given single copy sequence and a library of 1,200,000 clones a 99% chance[37]. The use of enzymes that recognise three different sequences reduces the possibility that a particular region of interest is lost from the library because it had too many or too few restriction sites to fall within the sucrose gradient size cut. Libraries were screened[7,16] with the cDNA clone (see text) S107 or M104E C_μ cloned cDNA probes identified by DNA sequence analysis[9,38]. The C_μ clone used extends from codon 300 to the 3'-untranslated region (~1,000 base pairs).[38]. Independent overlapping clones were obtained from the C_α and C_μ gene segments. Location of the coding regions and the direction of transcription were determined by DNA sequence analyses for the CHSp PC-3 clone, by heteroduplex analyses with the α6 clone for the CHSp α29 clone, and by R-loop mapping and restriction analyses of the CHSpμ27 clone[38]. The germ-line genomic clones will be denoted μ27 (CHSpμ27), V_H3 (CHSpV_H3), and α29 (CHSpα29).

*Eco*RI fragment containing most of the large intervening sequence and the 5' portion of the C_α gene segment is not present. This observation indicates that the α6 clone is the product of one or more DNA rearrangements which presumably occurred during B-cell differentiation[8].

The α gene is composed of at least three different germ-line segments of DNA

In view of the possibility that lymphocytes earlier in the M603 lineage might first have produced IgM molecules and later IgA molecules, we decided to investigate the possible contribution of germ-line C_μ sequences as well as V_H and C_α sequences to the myeloma α gene (Fig. 1*a*). We constructed several libraries of germ-line DNA (sperm) and proceeded to isolate clones containing V_H, C_μ and C_α gene segments using cloned cDNA probes (denoted S107 V_H, C_α and C_μ) for the corresponding coding regions and the screening procedure of Benton and Davis[20]. *Eco*RI restriction maps of several germ-line clones are shown in Fig. 1*b*. We chose sperm (germ-line) DNA for our undifferentiated genomic libraries to eliminate any possibility of DNA rearrangements which may occur in somatic tissues during embryogenesis. The C_α and V_H clones seem to be representative of germ-line sequence organisation because the *Eco*RI fragments containing the corresponding coding regions (in the clones) are identical in size to those found in the Southern blot analysis of undifferentiated DNA with C_α and S107 V_H cloned cDNA probes—9.5 and 7.4 kilobases, respectively (Fig. 2*a*–*d*). The V_H and C_α probes were derived from restriction fragments of the S107 cloned cDNA[7] extending from approximately the 5'-untranslated region to codon 108 (V_H) and from codons 108 to 274 (C_α). Figure 2*a* shows that a Southern blot of a germ-line

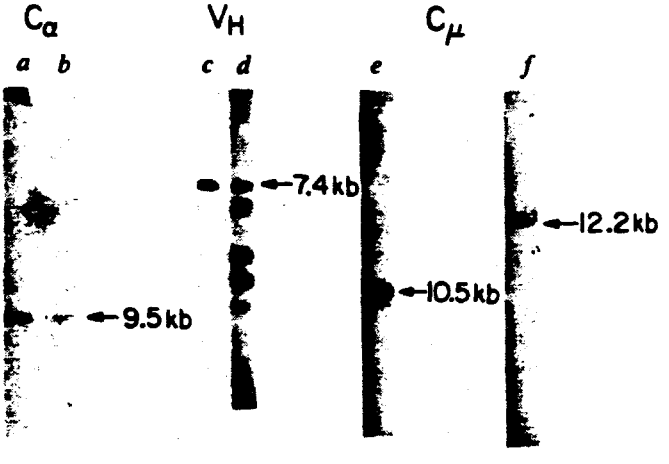


Fig. 2 Southern blot analyses of C_α, C_μ and S107 V_H coding regions in germ-line DNA. Approximately 3–20 μg of sperm DNA or 13-day embryo was digested with *Eco*RI and electrophoresed on a 0.7% neutral agarose gel for 10–12 h at 30–40 V. Gels were then blotted according to the procedures of Southern and Flavell[18,19] and hybridised with ^{32}P nick-translated cDNA[39]. Washing was for 1.5 h in 1 M NaCl, 1 M Tris *p*H 8, 0.1% SDS and 0.1% sodium pyrophosphate at 65 °C and for 1 h in 1×SSC, 0.1% SDS and 0.1% sodium pyrophosphate. Lanes *a* and *b* are α29 and embryo DNAs, respectively, hybridised to the C_α cDNA probe. Lane *c* is V_H3 DNA hybridised to the S107 cDNA probe. Lane *d* is sperm DNA hybridised to the S107 V_H cDNA probe. This probe cross-reacts with eight or nine closely related V_H gene segments[8]. Lanes *e* and *f* are μ27 and sperm DNAs, respectively, hybridised to the C_μ cDNA probe. Identical results in all cases were obtained with BALB/c 13-day embryo DNA or sperm DNA as has been found for a V_α gene[34]. Sizes (given in kilobases) were determined by comparison with restriction fragments of PBR322 (ref. 40) or by the use of PBR322 multimers generated by *Bam*HI cleavage of PBR322 and limited ligation of the resulting monomer.

C_α clone (α29) digested with *Eco*RI and hybridised with the C_α cDNA probe yields a 9.5-kilobase fragment. A similar analysis of embryo DNA produces a band of identical size (Fig. 2*b*). Southern blots of a germ-line V_H clone (V_H3) and embryo DNA digested with *Eco*RI and hybridised to a S107 V_H cDNA probe gave, respectively, a 7.4-kilobase fragment (Fig. 2*c*) and several fragments including a 7.4-kilobase fragment (Fig. 2*d*). An important question is the relationship between the germ-line V_H3 and myeloma M603 V_H coding regions because there are at least eight V_H gene segments that hybridise to the S107 probe[8]. The germ-line V_H3 DNA sequence codes for the S107 V_H protein sequence[9] and, accordingly, differs from the M603 V_H region by a minimum of four base changes leading to three amino acid subsitutions[21]. Thus, the germ-line V_H3 and the myeloma M603 V_H gene segments may be encoded by two distinct germ-line V_H gene segments or the V_H3 gene segment may give rise to the M603 V_H gene segment by somatic mutation and selection. As we shall show subsequently, the V_H3 clone seems to be indistinguishable from the M603 clone in the V_H coding region and in more than 11-kilobases of 5′-flanking sequence by heteroduplex and restriction enzyme analyses. Therefore, our analyses of the localisation of V_H sequences in the myeloma α6 clone are valid because the germ-line V_H clone serves as a probe for the M603 V_H gene and its attendant 5′-flanking sequence.

The *Eco*RI fragment of the germ-line C_μ clone containing the C_μ coding region is smaller (10.5 kilobases; see Fig. 2*e*) than its counterpart seen on a Southern blot analysis of sperm DNA with a cloned μ cDNA probe (12.2 kilobases; see Fig. 2*f*). We believe this discrepancy arises from one (or more) deletion(s) in the DNA flanking the C_μ coding region during the propagation of the recombinant phage. In isolating μ clones from the M603 library, we obtained several 9.5–10-kilobase *Eco*RI fragments containing C_μ coding regions, whereas Southern blot analysis of M603 DNA with a cloned C_μ cDNA probe demonstrated a genomic fragment of 12.2 kilobases, as in the sperm DNA (data not shown). Attempts to isolate C_μ clones from a library of mouse liver DNA have led to similar results (N. Newell and F. Blattner, personal communication). We will present restriction enzyme data below which demonstrate that this apparent deletion in the μ clone does not affect our general conclusions.

The germ-line V_H, C_α and C_μ clones were compared with the myeloma α6 and α125 clones by heteroduplex analysis. Representative heteroduplexes from each of these comparisons show extensive homologies. The germ-line V_H clone shares approximately 11.6 kilobases of homology with the myeloma α125 clone (Fig. 3*a*). This homology extends from the 5′ end of the α125 clone up to and including the V_H coding region. The germ-line C_μ clone shows 5.0 kilobases of homology with the large intervening sequence of the myeloma α6 clone (Fig. 3*b*). Starting at its 3′ end, the germ-line C_α clone has about 6.4 kilobases of homology with the myeloma α6 clone (Fig. 3*c*). The heteroduplex measurements for these analyses are given in Table 1.

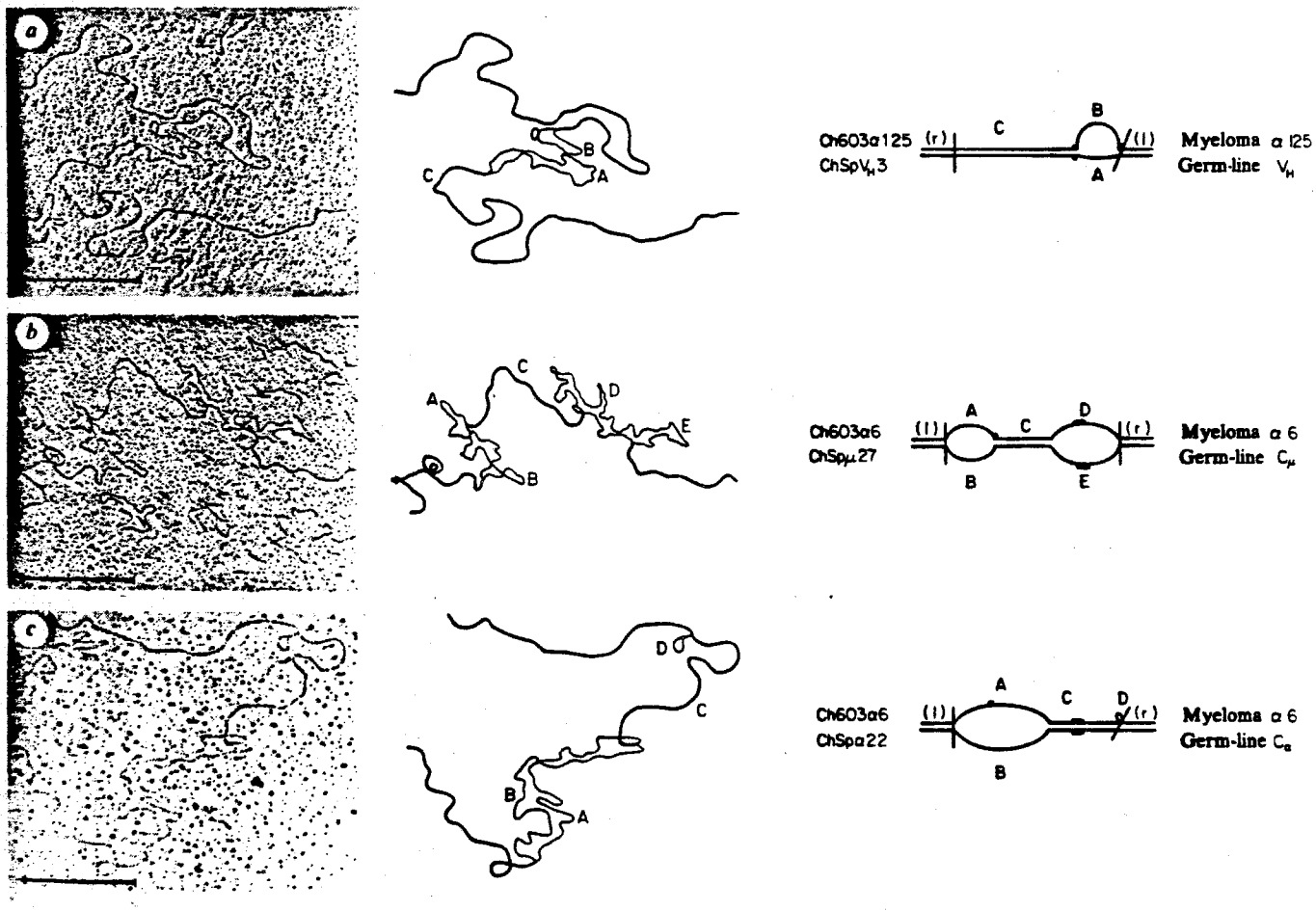


Fig. 3 Heteroduplex analyses of germ-line and somatic clones. The electron micrographs are shown on the left, tracings of these heteroduplexes in the middle and diagrammatic representations on the far right. *a*, Myeloma α125/germ-line V_H PC3. *b*, Myeloma α6/germ-line μ27. *c*, Myeloma α6/germ-line α22. Letters A–E indicate single-strand and double-strand regions for which measurements are given in Table 1. In corresponding line drawings, (r) and (l) refer to the right and left arms of the λ vector. Blocks indicate coding regions in these figures. CsCl purified phage particles were treated with 0.1 M NaOH for 10 min at 20 °C to lyse the phage and denature the DNA simultaneously. After neutralisation of the mixture, the DNA was allowed to reanneal in 50% (v/v) three times recrystallised formamide for 45 min at 20 °C before spreading for electron microscopy[41].

Table 1 Measurements of heteroduplex molecules

Heteroduplex	No. of Molecules	Distance in kilobases A	B	C*	D	E
a Myeloma α125/ germ-line V_H3	35	4.5± 0.4	7.2± 0.7	11.6± 0.4		
b Myeloma α6/ germ-line μ27	30	4.8± 0.5	5.1± 0.5	5.0± 0.4	6.7± 0.4	7.5± 0.5
c Myeloma α6/ germ-line α22	26	9.3± 0.5	10.4± 0.9	6.4± 0.5	1.2± 0.2	

Measurements were standardised relative to two circular DNA molecules on the same grid [single-strand ΦX174 DNA (5,375 bases) and double-strand pBR322 DNA (4,365 base pairs)]. Letters refer to regions indicated in Fig. 3. The complete clone designations are given in Fig. 3. In heteroduplex *a*, C refers to the region of duplex between the germ-line V_H clone and the myeloma clone. A and B are the non-homologous single strands of these clones. Similarly, C for heteroduplexes *b* and *c* refers to the duplexes formed between myeloma and germ-line C_μ or C_α clones.

Comparative restriction analyses confirm and extend the heteroduplex data discussed above. A detailed restriction map for the M603 myeloma clones was obtained by double digestion with pairs of restriction enzymes and is shown in Fig. 4*a*. To compare the placement of these cleavage sites with those of the germ-line clones, the 5.1-kilobase restriction fragment of the α6 clone (Fig. 1*a*) that spans the region joining the germ-line C_α and C_μ sequences was subcloned. Using this fragment as a probe, detailed restriction comparisons of the myeloma α6 clone and the germ-line C_μ and C_α clones were made. Representative data are shown in Fig. 5 for these comparisons. Figure 5*a*, *b* and *c* represents *Hin*cII plus *Eco*RI digestions of the myeloma α6 5.1-kilobase RI fragment, the germ-line α29 clone and the germ-line C_μ clone, respectively. These digests were electrophoresed on an agarose gel, blotted onto a nitrocellulose filter and hybridised with the labelled 5.1-kilobase RI fragment. This enables homologous restriction fragments to be identified rapidly and precisely (arrows indicate identical fragments). Figure 5*d*, *e* and *f* shows a similar analysis using *Hin*dIII plus *Eco*RI digestions, respectively, of the myeloma 5.1-kilobase subclone, the germ-line α29 clone and the germ-line μ27 clone. These data, as well as additional restriction analyses of the germ-line V_H3 clone, demonstrate that 4 out of 4 restriction sites in the germ-line V_H3 clone, 10 out of 10 sites in germ-line α29 clone and 9 out of 10 sites in the germ-line μ27 clone corresponded exactly to those found in the myeloma α6 clone (Fig. 4). Not only do these restriction analyses independently confirm the heteroduplex results, but they also suggest that the component germ-line sequences of the α6 clone are very similar to their germ-line counterparts. Thus, the heteroduplex and restriction analyses demonstrate that the germ-line V_H gene segment and its 5'-flanking sequence, although not identical to its M603 counterpart, are very similar and may be used to analyse V_H gene segment organisation in the myeloma M603 clones.

DNA sequence analyses of the myeloma α6 clone and the germ-line μ27 clone have demonstrated that the heavy-chain gene family does have distinct V_H and J_H gene segments in the germ line[9]. Moreover, the germ-line J_H gene segment corresponding to that expressed in the myeloma α6 clone is associated with the germ-line C_μ gene[9]. This J_H gene segment contains the *Hha*I site that marks the end of the homology between the germ-line μ27 clone and the myeloma clones (Fig. 4). Accordingly, the distinct germ-line V_H and germ-line J_H gene segments are rearranged in the myeloma clones in a manner analogous to the V_L and J_L gene segments of myeloma light chain genes[2,3].

Although it seems unlikely, the M603 J_H gene segment and flanking sequences could have been fused to a germ-line C_μ clone as the result of some cloning artefact. To eliminate this possibility, we decided to demonstrate the germ-line association of the J_H flanking sequence and C_μ sequences by Southern blot analyses. On different slots of the same agarose gel, we electrophoresed mouse sperm DNA cleaved with either *Hin*cII or *Eco*RI, transferred these DNAs to a nitrocellulose filter and hybridised one lane of each digest with a C_μ probe and a second lane with the 5.1-kilobase *Eco*RI fragment from the intervening sequence of the myeloma α6 clone (Fig. 6). Figure 6*a* and *b* shows Southern blots of *Eco*RI-digested sperm DNA hybridised with the C_μ cDNA probe and the 5.1-kilobase *Eco*RI subclone from the α6 clone, respectively. In both lanes, a 12.2-kilobase band corresponding to the germ-line C_μ fragment can be identified. In a second digest (*Hin*cII) of mouse sperm DNA, a similar result is obtained with both the C_μ probe (Fig. 6*c*) and the 5.1-kilobase *Eco*RI subclone (Fig. 6*d*) hybridising to a 9.3-kilobase *Hin*cII fragment. The 9.5-kilobase *Eco*RI band in Fig. 6*b* and the 5.0-kilobase band in Fig. 6*d* correspond to C_α restriction fragments. In each case, one of the two bands from the 5.1-kilobase *Eco*RI probe co-migrated with the single C_μ DNA fragment. This analysis shows that at least part of the intervening sequence from the myeloma α6 clone is adjacent to the C_μ gene in the germ line. Thus, the apparent deletion in the germ-line μ27 clone is not a significant factor in our discussion.

A summary of these analyses is presented in Fig. 7. The myeloma α6 clone is composed of three noncontiguous germ-line gene segments: (1) a V_H gene segment and its 5'-flanking

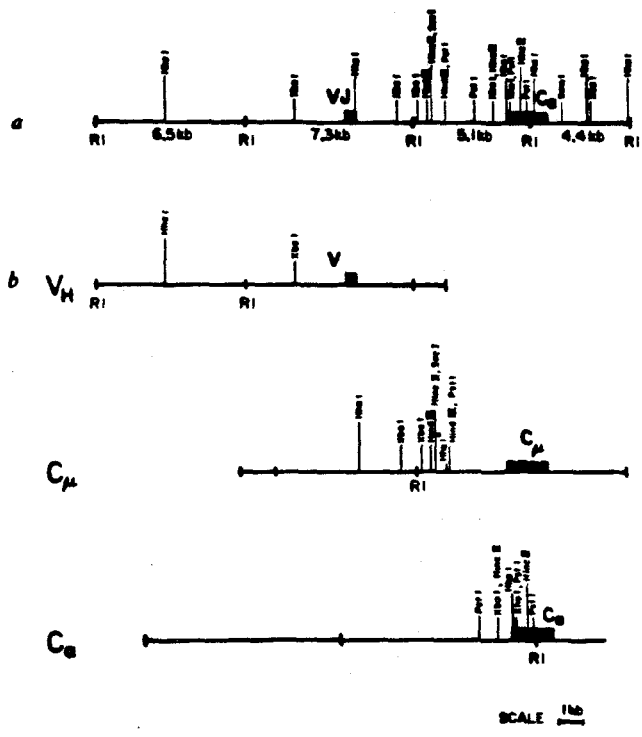


Fig. 4 *a*, Restriction map of myeloma clones α6 and α125. Specific cleavage sites were determined using double enzyme digestion and sizing by gel electrophoresis. Size standards were restriction digests of PBR322 and PBR322 multimers (see Fig. 2 legend). *Hin*dIII, *Hin*cII, *Pst*I and *Sac*I cleavage sites are only shown for the 5.1-kilobase *Eco*RI fragment. *b*, Restriction sites corresponding to those of the myeloma clones detected in the V_H, C_μ and C_α germ-line clones (Fig. 1). Restriction sites in regions corresponding to the 6.5- and 7.3-kilobase RI fragments of α6 and α125 were mapped by double digestion as above. All sites within the 5.1-kilobase RI fragment were compared by co-migration and blotting as described and illustrated in Fig. 5. The *Hha*I site marked by an asterisk in the C_μ clone was the one site found not to conform with those of the myeloma α6 clone.

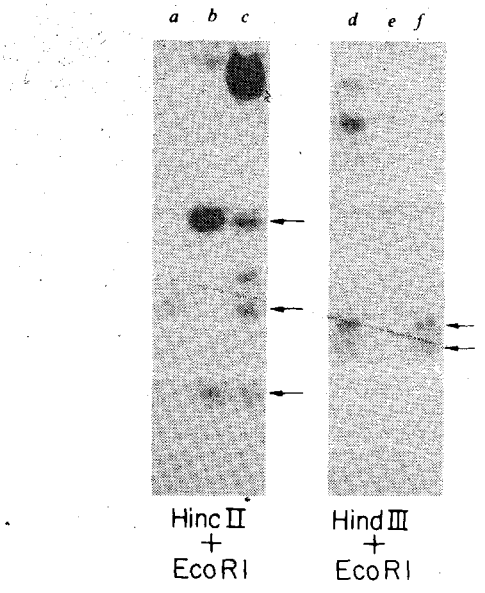


Fig. 5 Comparative restriction digests of the $\alpha 6$ myeloma, germ-line C_μ and C_α clones. Lanes *a–f* show parallel *Hinc*II + *Eco*RI and *Hind*III + *Eco*RI digests of the 5.1-kilobase *Eco*RI subclone of the $\alpha 6$ clone (Fig. 1*a*) and DNA from the germ-line $\alpha 29$ and $\mu 27$ clones (Fig. 1*b*). Samples were electrophoresed in 1% agarose, transferred to nitrocellulose and hybridised with nick-translated 5.1-kilobase $\alpha 6$ subclone DNA. The co-electrophoresis of restriction fragments allows a large number of different restriction sites to be compared rapidly. In this way, all 16 mapped sites in the 5.1-kilobase $\alpha 6$ subclone were compared with equivalent sites on the $\alpha 29$ and $\mu 27$ clones. Arrows show coincident bands.

sequence, (2) flanking sequences located 5′ to a germ-line C_μ gene which includes the J_H gene segment, and (3) the germ-line C_α gene segment with its flanking 3′ and 5′ sequences. Note that within the limits of the methods used here these three germ-line gene segments and their attendant flanking sequences apparently cover the entire myeloma $\alpha 6$ clone.

The α heavy-chain gene is formed by at least two recombinational events

Two distinct DNA rearrangements have occurred to form the M603 α gene—V–J joining and C_H switching (Fig. 8)[9]. The simplest interpretation of these observations is that the V_H gene segment is first joined to a J_H gene segment linked to the C_μ gene segment. This V–J joining, analogous to that which occurs in light chains, generates a rearranged μ gene that presumably leads to the expression of IgM molecules. V–J joining also commits an individual lymphocyte to the expression of a single V domain that remains invariant throughout subsequent steps of B-cell differentiation. Later, a C_H switch joins the V–J gene segment to the C_α gene segment to create a functional α gene and thus enables the differentiated lymphocyte to express IgA molecules. Thus, the myeloma $\alpha 6$ heavy-chain gene is assembled by two distinct and presumably independent DNA rearrangement events. Our heteroduplex, Southern blot and restriction mapping data show unequivocally that these two rearrangements occur at two distinct sites in the genome. These two sites of rearrangement are termed the V–J joining site and the C_H switch site, respectively.

These data are consistent with a differentiation pathway in which a B cell may switch from IgM to IgA synthesis while expressing the same V domain. We cannot establish that these two DNA rearrangement events occurred at different times, although this supposition is reasonable. We also cannot rule out the possibility of intermediate differentiation states in this B-cell lineage where IgG molecules were produced, although there is no direct evidence for such a stage.

The C_H switch may be explained by any one of several genetic models

Several mechanisms of C_H switching have been proposed[11,22–26], many of which are similar to those proposed for V–J joining[27]. These models can be categorised as involving either DNA rearrangements to replace one constant region with another (successive deletions, excision–insertions or inversions) or the differential processing of a large nuclear RNA transcript containing multiple heavy-chain constant-region genes[26].

The RNA processing model seems unlikely as a general mechanism for C_H switching at the level of antibody-secreting plasma cells. High molecular weight nuclear RNAs from three myeloma tumours hybridise only with a cDNA probe complementary to the class of immunoglobulin expressed in that tumour and not with probes from non-expressed immunoglobulin classes[28]. Moreover, cell fusion experiments hybridising two different myeloma cells demonstrate that the hybrid cells synthesise only parental V_HC_H combinations[29], a result in conflict with the simple RNA processing model. Furthermore, neither $C_{\gamma 1}$ nor C_μ DNA probes hybridise on Southern blots with the M603 α clones (data not shown), contrary to one prediction of the RNA processing model.

On the other hand, the evidence presented here strongly supports DNA rearrangement as a fundamental element in the mechanism for heavy-chain switching. As summarised in Fig. 7, a very large segment of C_μ flanking sequence has been brought adjacent to C_α and V_H gene segments in the active α gene of myeloma tumour M603. The creation of the M603 α gene apparently requires two DNA rearrangements (Fig. 8). A C_H switching

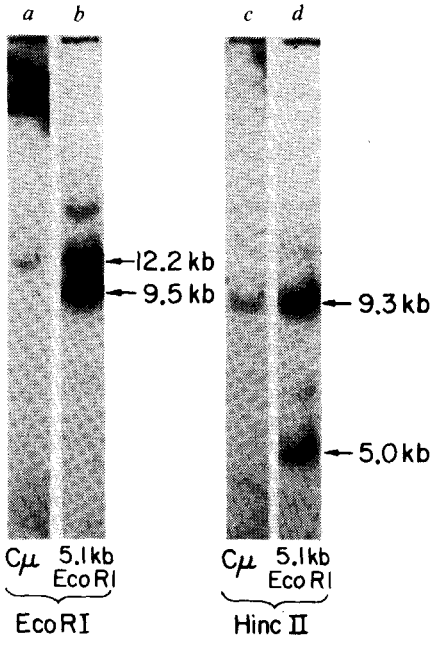


Fig. 6 Southern blots of the 5.1-kilobase *Eco*RI and C_μ fragments. Mouse sperm DNA was digested with *Eco*RI or *Hinc*II and 3 μg was loaded onto a 4 mm × 20 × 20 cm 0.7% agarose gel, electrophoresed at 40 V for 10 h and blotted as described. Probes used were either the 5.1-kilobase *Eco*RI fragment or a C_μ cDNA clone nick-translated to 4×10^8 c.p.m. per μg. Washing was as described in Fig. 3 legend except that lanes hybridised with the 5.1-kilobase *Eco*RI fragment were washed further in 10 mM NaCl, 10 mM Tris, 0.1% SDS and 0.1% $NaPP_i$ for 2 h at 68 °C to reduce the signal strength of weakly homologous repeats. Filters were exposed for 12 h with an intensifying screen at −80 °C. The faint band above the 12.2-kilobase band in *b* is a partial digestion product. All lanes shown were run on the same gel and blotted simultaneously, alignment being assisted by inclusion of pBR322 multimers as internal standards (see Fig. 2).

DNA rearrangement is not postulated in the RNA processing models[26]. The data presented here do not distinguish between the various types of DNA rearrangements proposed, but the hybridisation kinetics experiments of Honjo and Kataoka are consistent with a deletional mechanism for the C_H switch[25]. In addition, recent experiments suggest that V–J joining in mouse λ genes is accomplished by a deletional mechanism[4]. Thus, both types of DNA rearrangement, V–J joining and C_H switching, may arise through deletional mechanisms. If the deletional model is correct for either type of DNA rearrangement, the differentiation of B cells is irreversible because chromosomal information is lost with the excision of each deletional loop of chromosome.

Gene organisation studies may delineate distinct pathways of B-cell differentiation

It will be interesting to determine whether all joined α genes have the same switch site. Southern blot analyses of the DNA from a second IgA-producing myeloma tumour, H8, with the 5.1-kilobase *Eco*RI probe yield a restriction fragment pattern identical to that of M603 DNA (data not shown). In particular, the 5.1-kilobase *Eco*RI fragment (Fig. 1) which contains the switch site seems the same. These data suggest that the expressed α genes in both M603 and H8 myeloma tumours have the same C_H switch site. However, Southern blot analyses of several other closely related IgA-producing tumours do not show a 5.1-kilobase *Eco*RI fragment (M.M.D., unpublished observation). Thus, there may be multiple C_H switch sites for the C_α gene segment. As it seems that B cells producing IgM or IgG may switch to the production of IgA, perhaps distinct C_H switch points reflect distinct pathways of B-cell differentiation. It will also be interesting to determine the location and number of switch sites for other immunoglobulin classes. If each C_H gene segment has a unique site or set of sites for C_H switching, one may be able to trace the distinct pathways of B-cell differentiation by studying the sequence organisation of each functional heavy-chain gene.

DNA rearrangements of antibody gene segments lead to combinatorial amplification of immunoglobulin information

V–J joining and C_H switching are mediated by DNA rearrangements which display combinatorial properties that amplify the germ-line information encoding the antibody gene families. (1) The V and J gene segments of one antibody gene family may be joined in a combinatorial manner to generate diversity in the third hypervariable regions of both κ and heavy chains[4,8,9,30,31]. For example, mice may have at least 200 V_H and 5 J_H gene segments that may be joined combinatorially to generate 1,000 different V_HJ_H coding regions. (2) One V domain may be combinatorially switched among eight or more different C_H regions to carry out a variety of different effector functions

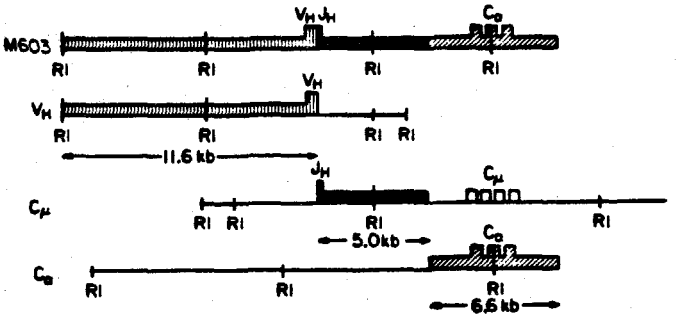


Fig. 7 Origins of the three germ-line components of the myeloma α heavy-chain gene. Various types of shading indicate homology by heteroduplex analyses and by restriction enzyme analyses.

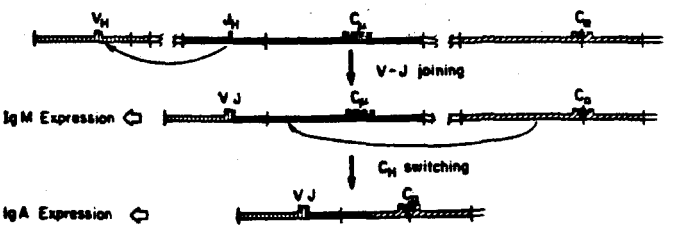


Fig. 8 Two types of DNA rearrangements leading to the creation of the myeloma α heavy-chain gene V–J joining and C_H switching. V–J joining indicates a DNA rearrangement that joins the V_H and J_H gene segments. Because the J_H segments seem to be associated with the germ-line C_μ gene[9], V–J joining permits a μ chain and IgM molecules to be expressed by the differentiating B cell. C_H switching denotes a DNA rearrangement that replaces the C_μ gene segment with a C_α gene segment. This second rearrangement presumably permits an α chain and IgA molecules to be expressed by the now fully differentiated lymphocyte.

that are directed at eliminating antigen or triggering defensive mechanisms such as complement fixation. Thus, each recognition (V) domain may be switched to many different effector (C) domains. The combinatorial properties of antibody gene segments and polypeptides therefore contribute to several fundamental aspects of the vertebrate immune response—V region diversity and the combinatorial switching of antigen-recognition (V) domains with a variety of different effector (C) domains during B-cell differentiation. Other complex eukaryotic gene families may use similar DNA combinatorial mechanisms for information amplification[32].

This work was supported by NSF grant PCM 76–81546, ACS grant IM56 and USPHS grant AI09072. M.M.D., P.W.E. and D.L.L. are supported by NIH training grant GM 07616. K.C. is supported by NIH Fellowship GM 05442. R.J. is a fellow of the Swiss National Foundation. I.L.W. is a faculty Research Awardee of the ACS. All experiments involving recombinant organisms were conducted in accordance with the revised NIH Guidelines on recombinant DNA, using P2, EK-2 or P3, EK-2 containment. We thank David Goldberg for his gift of PBR322 multimers, Tom Sargent for packaging extracts, Keichi Itakura for synthetic RI linkers, and David Anderson, Norman Davidson, Max Delbrück, Richard Flavell, Henry Huang, Tom Maniatis and Tom Sargent for helpful discussions.

Received 1 October 1979; accepted 22 January 1980.

1. Mage, R., Lieberman, R., Potter, M. & Terry, W. in *The Antigens* (ed. Sela, M.) 299–376 (Academic, New York, 1973).
2. Brack, C., Hirowa, A., Lenhard-Schueller, R. & Tonegawa, S. *Cell* **15**, 1–14 (1978).
3. Seidman, J. G., Max, E. E. & Leder, P. *Nature* **280**, 370–375 (1979).
4. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
5. Gilmore-Herbert, M. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 342–345 (1978).
6. Schibler, U., Marcu, K. B. & Perry, R. P. *Cell* **15**, 1495–1509 (1978).
7. Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 857–861 (1979).
8. Davis, M., Early, P., Calame, K., Livant, D. & Hood, L. in *Eukaryotic Gene Regulation* (eds Axel, R. Maniatis, T. & Fox, C. F.) 393–406 (ICN UCLA Symp., Academic, New York, 1979).
9. Early, P. W., Huang, H. V., Davis, M. M., Calame, K. & Hood, L. *Cell* (submitted).
10. Raff, M. C. *Cold Spring Harb. Symp. quant. Biol.* **41**, 159–162 (1976).
11. Sledge, C., Fain, D. S., Black, B., Krieger, R. G. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 923–927 (1976).
12. Fudenburg, H. H., Wang, A. C., Pink, J. R. L. & Levin, A. S. *Ann. N.Y. Acad. Sci.* **190**, 501–506 (1971).
13. Wang, A. C., Geisely, J. & Fudenberg, H. H. *Biochemistry* **12**, 528–534 (1973).
14. Levin, A. S., Fudenberg, H. H., Hopper, J. E., Wilson, S. & Nisonoff, A. *Proc. natn. Acad. Sci. U.S.A.* **68**, 169–171 (1971).
15. Wang, A. C., Wilson, S. K., Hopper, J. E., Fudenberg, H. H. & Nisonoff, A. *Proc. natn. Acad. Sci. U.S.A.* **66**, 337–343 (1970).
16. Maniatis, T. *et al. Cell* **15**, 687–701 (1978).
17. Blattner, F. R. *et al. Science* **196**, 161–169 (1977).
18. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1977).
19. Jeffreys, A. J. & Flavell, R. A. *Cell* **12**, 429–439 (1977).
20. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
21. Hood, L. *et al. Cold Spring Harb. Symp. quant. Biol.* **41**, 817–836 (1976).
22. Smithies, O. *Science* **169**, 882–883 (1970).
23. Hood, L. *Fedn Proc.* **31**, 177–178 (1972).
24. Bevan, M. J., Parkhouse, R. M. E., Williamson, A. R. & Askonas, B. A. *Prog. Biophys. molec. Biol.* **25**, 131–162 (1972).

25. Honjo, T. & Kataoka, T. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2140–2144 (1978).
26. Rabbitts, T. H. *Nature* **275**, 291–296 (1978).
27. Tonegawa, S., Hozumi, V., Matthyssen, G. & Schuller, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 877–889 (1976).
28. Marcu, K. B., Schibler, U. & Perry, R. P. *Science* **204**, 1087–1088 (1979).
29. Schulman, M. J. & Kohler, G. *Nature* **274**, 917–919 (1978).
30. Max, E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3450–3454 (1979).
31. Schilling, J., Cleavinger, B., Davie, J. & Hood, L. *Nature* **283**, 35–40 (1980).
32. Hood, L., Huang, H. & Dreyer, W. J. *J. supramolec. Struct.* **7**, 531–559 (1977).
33. Witkin, S., Korngold, G. & Bendich, A. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3295–3299 (1975).
34. Jobo, R., Weissman, I. L., Early, P., Cole, J. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
35. Sternberg, N., Tiemeier, D. & Enquist, L. *Gene* **1**, 255–280 (1977).
36. Hohn, B. & Murray, K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3259–3263 (1977).
37. Clarke, L. & Carbon, J. *Cell* **9**, 91–99 (1976).
38. Calame, K., Rogers, J., Davis, M., Early, P., Livant, D., Wall, R. & Hood, L. *Nature* (submitted).
39. Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184–1188 (1975).
40. Sutcliffe, J. G. *Nucleic Acids Res.* **5**, 2721–2728 (1978).
41. Davis, R., Simon, M. & Davidson, N. *Meth. Enzym.* **21D**, 413–428 (1971).

The primary structure of 16S rDNA from *Zea mays* chloroplast is homologous to *E. coli* 16S rRNA

Zs. Schwarz* & H. Kössel†

Institut für Biologie III der Universität Freiburg, Schänzlestrasse 1, D-78 Freiburg i. Br., FRG

The nucleotide sequence of ribosomal DNA coding for 16S rRNA from Zea mays *chloroplast has been determined. A comparison with the 16S rRNA sequence from* Escherichia coli *reveals strong homology and thereby demonstrates the prokaryotic nature of chloroplast ribosomes from a higher plant.*

CHLOROPLASTS of green plants contain 70S ribosomes which differ from the eukaryotic 80S type ribosomes of the cytoplasm[1]. The organelle-specific ribosomes mediate translation of organelle-specific mRNAs, which in turn are encoded in chloroplast DNA[2,3]. Organelle-specific rRNA[4] and tRNA[5] species are also encoded in the chloroplast genome, which thus provides the basis for the semiautonomous nature of the organelle. This, along with the various prokaryotic characteristics of the chloroplast-specific translation system, has lent strong support to the endosymbiont hypothesis[6]. This hypothesis postulates that modern chloroplasts are descendants of endosymbiotic prokaryotes—particularly cyanobacteria[7]—which have entered eukaryotic cells during evolution.

Due to a supposedly heavy functional constraint of a highly complex entity such as the ribosome, it seemed likely that its components would have changed comparatively slowly during evolution. Therefore, a comparison of rRNA primary structures from chloroplasts and *E. coli* should provide a quantitative measure of their evolutionary relationship, even if they are separated by a large phylogenetic distance. As an extension of previous work based on T_1-oligonucleotide catalogues[7–9] of 16S rRNAs, complete sequencing should not only allow the exact positioning of conserved T_1 oligonucleotides, but also a comparison of regions in which larger T_1 oligonucleotides are absent. Furthermore, a differentiation between strongly conserved regions and regions which have gone through changes during the course of evolution will be possible only on the basis of complete sequences. We have, therefore, analysed rDNA from *Zea mays* chloroplasts at the nucleotide level, which should allow deduction of the primary structures of the corresponding rRNA species and comparison with the corresponding rRNA species from *E. coli*. Furthermore, it was anticipated that analysis of sequence homology between organelle and *E. coli* rRNA species would support the prediction of secondary structures if they are common or would help to exclude them if they are not preserved. In this article we report the primary structure of 16S rDNA from *Zea mays* chloroplast and compare it with *E. coli* 16S rRNA recently analysed by others[10,11].

Sequencing of maize chloroplast 16S rDNA

DNA from *Zea mays* chloroplasts is a circular molecule of about 135 kilobase pairs on which two rRNA coding regions are positioned in opposite orientation within two 22 kilobase-pair inverted repeats[4]. Each inverted repeat includes within a 12 kilobase-pair *Eco*RI fragment one copy of a 16S rRNA, 23S rRNA, 4.5S and 5S rRNA gene and a 2 kilobase-pair spacer region between the 16S and 23S rRNA genes. One such 12 kilobase-pair fragment has been linked to the plasmid vector pMB9 within the *E. coli* clone pZmc134 (ref. 4). This clone was therefore used for isolation, mapping and sequencing of the 16S rRNA coding DNA fragments, which are positioned within the left half of fragment *Bam*HI·C (Fig. 1)[4,12,13].

In Fig. 2 the RNA-like strand of the entire 16S rRNA gene from maize chloroplast is depicted (lower row of each line). Together with maize chloroplast 4.5S and wheat 5S rDNA sequences (T. Dyer *et al.*, unpublished) it represents—to our knowledge—the first sequence analysis of a complete gene originating from a higher plant.

The exact termini of the structural part of the gene cannot be located by DNA sequencing and must await sequencing of the RNA ends. Taking into account the inhomology of the first four positions, it seems not unlikely that the position of the 5′-terminal nucleotide of maize 16S rRNA is different from position 1, although very close to it. At the opposite end positions 1,540 (last homologous position) or 1,541 (non-homologous, but in accordance with the chain end of *E. coli* 16S rRNA) appear as the most likely candidates coding for the 3′-terminal nucleotide of the chloroplast 16S rRNA. Neglecting this uncertainty a chain length of 1,491 results, which, due to several deletions, is 50 residues shorter than the *E. coli* 16S rRNA.

Sequence homology between maize chloroplast and *E. coli* 16S rRNA

When the maize chloroplast 16S rDNA sequence is compared with the 16S rDNA[10] or rRNA[11] from *E. coli* (Fig. 2) extensive

* Present address: Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138.
† To whom correspondence should be addressed.

0028-0836/80/080739—04$01.00

DNA Sequences Mediating Class Switching in Alpha Immunoglobulin Genes

Science (1980) In Press

Abstract. Immunoglobulin class switching involves specific DNA rearrangements of the gene segments coding for heavy chain constant regions during B-lymphocyte differentiation. In two different cases of C_μ to C_α switching examined here (T15, M603) and one taken from the literature (MC101), three different sites 5' to C_μ and three different sites 5' to C_α are joined together in the process of C_H switching. The sequences surrounding the three germline C_α sites of recombination are highly conserved blocks of 30 nucleotides which may serve as recognition sequences for C_H switching to the C_α gene. This putative recognition sequence is repeated 17 times in approximately 1400 nucleotides of germline C_α 5' flanking sequence. The lack of homology between this C_α sequence and sequences reported for the $C_{\gamma 1}$ and $C_{\gamma 2b}$ switch sites suggests that heavy chain switching is mediated by class-specific recognition sequences and, presumably, class-specific regulatory mechanisms. In addition, it appears that in one example, MC101, C_H switching has progressed from $C_\mu \rightarrow C_\alpha \rightarrow C_{\gamma 1}$. This switching pathway presents difficulties for the simple deletional model of C_H switching.

The antibody molecule is comprised of discrete molecular domains that carry out two general types of functions. The variable or V domain binds antigen and the constant or C domains trigger effector functions such as complement fixation. The V and C domains arise from the interactions of two different polypeptides, light (L) and heavy (H), which in turn are encoded by a series of discrete gene segments—V_L, J_L (joining), and C_L and V_H, D (diversity), J_H and C_H, respectively (1-4). During the differentiation of antibody-producing or B cells, two distinct types of DNA rearrangements of these gene segments occur (4, 5). One type generates the V_L and V_H genes by joining directly the V_L and J_L and the V_H, D, and J_H gene segments, respectively. These DNA rearrangements are termed V-J or V-D-J joining and they are, in part, responsible for the generation of antigen-binding diversity in V domains.

A second type of DNA rearrangement, termed C_H switching, allows an important flexibility in the use of a given antigen-binding site. At an early stage of B-cell differentiation, an individual B cell initially expresses IgM molecules with a single V domain (V_L-V_H combination) (6, 7). Later, this B cell or its clonal progeny may express another immunoglobulin class while continuing to express the same V domain (8). Since the class of immunoglobulin is determined by the C_H region (e.g., C_μ-IgM, C_γ-IgG, C_α-IgA), the B cell must shift from the expression of a C_μ gene to the expression of another C_H gene during differentiation. Thus, C_H switching associates a particular antigen-binding specificity, the V domain, with a series of different effector functions encoded by the various C_H regions.

Two types of experiments have provided insights into the mechanism of C_H switching. First, Honjo and Kataoka (9) have employed hybridization kinetics to determine the numbers of C_H genes in a series of mouse myeloma

tumors producing different immunoglobulin classes. Their results suggest that the V_H gene is rearranged from one C_H gene to a second by a deletion of the intervening DNA between the V_H gene and the second C_H gene. From these data a heavy chain gene order of 5'-C_μ-$C_{\gamma 3}$-$C_{\gamma 1}$-$C_{\gamma 2b}$-$C_{\gamma 2a}$-C_α-3' was suggested. Recent experiments using cloned probes and Southern blot analyses generally support the gene order and deletional mechanism proposed by Honjo (10). Second, we examined the rearranged (expressed) α gene in an IgA-producing myeloma tumor (M603) and obtained direct evidence that DNA rearrangement mediates C_H switching (4). The rearranged M603 α gene is composed of three distinct germline gene segments--V_H, J_H with 5' C_μ flanking sequences and C_α with its flanking sequences. This tripartite structure of the rearranged α gene suggests that the V_H gene was initially associated with the C_μ gene by DNA rearrangement through V-D-J joining and expressed as a μ chain in the IgM molecule. A subsequent DNA rearrangement could then replace the C_μ gene with the C_α gene by linking together 5' C_μ and C_α flanking sequences. The point at which the flanking C_μ and C_α sequences join in the rearranged gene and the corresponding breakpoints on the germline DNAs are termed the switch (S) sites. Subsequently, other laboratories have obtained evidence for similar C_μ to C_H switches in rearranged γ1 (11) and γ2b (12, 13) genes.

Three Examples of IgM → IgA Switching

In order to investigate the molecular mechanisms underlying C_H switching, we sequenced the switch sites of two rearranged α genes and compared these rearranged switch sequences with their germline counterparts in the 5' flanking sequences of the C_μ and C_α genes. To this end, we have constructed genomic libraries (14) from the DNAs of M603 (15, 16) and an additional IgA-producing myeloma tumor, T15, and isolated the rearranged α genes. Homologies between

clones corresponding to the rearranged α genes of T15 and M603 are depicted in Fig. 1 and compared with their germline C_μ and C_α counterparts obtained from a genomic library of mouse sperm DNA (4). These homologies were established by detailed restriction enzyme analysis (data not shown). Both rearranged α genes exhibit the tripartite structure of V_H, J_H with C_μ flanking sequences and C_α gene segments. The size of the intervening sequence between the V_H and C_α coding regions is substantially different in these two cases—5.4 vs. 6.8 kilobases (Fig. 1). Since each of the two V_H gene segments is joined to the same J_H gene segment (17), the variation in size in the intervening sequences between the V and C_α coding regions may be the result of using different sites for C_H switching in each of these rearranged C_α genes. In addition, we found that a rearranged γ1 gene, MC101, whose sequence was reported by Honjo and colleagues (11) contains a 500 nucleotide region of C_α flanking sequence between C_μ and $C_{\gamma 1}$ derived sequences (Fig. 1). The evidence for this supposition is that this 500 nucleotide sequence has been localized solely to a region 5' to the C_α gene by Southern blotting analyses (18) and restriction mapping using fragments containing all or part of this region as probes (11; M. Davis, data not presented). Therefore, this fragment is apparently represented just once in the genome and, accordingly, must have been derived from flanking sequences 5' to the C_α gene. Furthermore, DNA sequence analyses of the corresponding germline C_α sequence show virtual identity (approximately 95 percent) with the sequence found between the C_μ and $C_{\gamma 1}$ flanking sequences in the MC101 γ1 gene (Fig. 4a) (19). Thus we feel the MC101 γ1 gene is composed of several distinct germline sequences: a V_H gene (T. Honjo, personal communication), flanking sequence for the C_μ gene, flanking sequence for the C_α gene and the γ1 gene with its flanking sequences.

The arrows in Fig. 1 indicate the regions analyzed by DNA sequence analysis in our laboratory. The DNA sequences of the MC101 S region (11) and the rearranged M603 and T15 α genes and their germline C_μ and C_α counterparts are shown in Fig. 2.

Examination of the rearranged switch sequences indicates that all three examples juxtapose C_μ and C_α derived sequences. However, each switch sequence seems different from the others. The arrangement of sequences in these genes suggest that at least three distinct types of switching may occur. We denote these categories "simple," "complex," and "successive."

In the simple category, T15, the C_μ flanking sequence joins directly to the C_α flanking sequence. Similarly, a γ2b gene (M141) has been found to join C_μ flanking sequence to that of $C_{\gamma 2b}$ (12, 13). In the complex category, M603, a short sequence of 287 base pairs, is interposed between the C_μ and C_α flanking sequences. This sequence appears to derive from a region 3' to S_{M603} on the C_μ gene (Fig. 1). Probes containing this sequence hybridize strongly to restriction fragments containing or adjacent to the C_μ gene in Southern blotting analyses (data not shown). The region of hybridization corresponds to the 1.5-2.5 kb region 5' to the C_μ gene (5) which deletes spontaneously upon cloning and hence is not present in our C_μ containing clones. Thus the complex category is explained by two distinct deletions—C_H switch and a deletion within the C_μ flanking sequence. This deletion of C_μ flanking sequence (at S_{M603}) does not appear to be a random event in that another α producing tumor line, M167, switches at exactly the same point adjacent to the C_μ gene. However, the complex category reflects DNA deletions seen to date only in myeloma cells and, accordingly, may or may not be biologically significant. In the third category, "successive" switching, the MC101 γ1 gene contains C_μ, C_α, and $C_{\gamma 1}$ flanking sequences between V_H and $C_{\gamma 1}$ gene segments. Therefore, it appears to have switched twice, once from C_μ to C_α and subsequently from C_α to $C_{\gamma 1}$.

The DNA sequence data in Fig. 2 show that the three rearranged genes employ three germline C_μ switch sites up to 300 base pairs apart and three germline C_α switch sites up to 1350 base pairs apart. The locations of these germline switch sites, S_{T15}, S_{M603}, and S_{MC101} are depicted in Fig. 1. Thus multiple switch sites exist for C_α switching in sequences 5' to both the germline C_μ and C_α genes.

DNA Sequences Mediating Heavy Chain Switching

The DNA sequences involved in V-D-J joining are quite distinct from those implicated in C_H switching. The inverted repeat CAC$\frac{A}{T}$GTG occurs at the 3' end of antibody V gene segments and at the 5' end of the J gene segments (2, 4, 20). This inverted repeat is believed to be a recognition sequence that mediates the juxtaposition of the V, D, and J gene segments to allow subsequent joining by site-specific recombination (for a proposed mechanism, see ref. 4). This inverted repeat is not found in the flanking regions surrounding any of the C_α switch sites. This sequence also is missing from the switch sites for a $\gamma 1$ (11) and a $\gamma 2b$ gene (12, 13). Therefore, C_H switching and V-J joining employ distinct mechanisms for DNA rearrangement.

In an effort to determine what sequences are important in C_H switching, the sequences of the three C_μ sites and the three C_α sites were compared (Fig. 3). The germline C_μ MC101 and T15 switch sites share significant homology (15/25 nucleotides), although 16 nucleotides separate the actual switch points. We believe that these homologies are significant and may represent general sequence requirements for C_H switching adjacent to the C_μ gene. Neither the MC101 nor the T15 C_μ sites share any homology with that of M603. The switch site of a $\gamma 2b$ producer M141 (12, 13), also depicted in Fig. 3, is nine nucleotides away from that of T15 and may indicate that both $\gamma 2b$ and α switching can utilize the same recognition sequence adjacent to C_μ.

The sequences around each C_α switch site are even more highly conserved. Each germline S_α site occurs within a block of 30 conserved nucleotides (Fig. 3). Twenty-two bases are identical and seven of the remaining eight nucleotides are conserved with regard to type of base (i.e., purine-purine or pyrimidine-pyrimidine substitutions). The three points of recombination differ within each of these conserved sequences. Since the first three rearranged α genes examined switched at distinct C_α sites, we reasoned that there must be additional S_α sites. We determined the DNA sequence of some 1400 nucleotides in the region of these germline S_α sites (Fig. 4a) and found seventeen 30 nucleotide repeat sequences (Fig. 4b). These sequences are extremely homologous to each other (boxed regions) (Fig. 4c). Since these repeated sequences represent 510/1400 nucleotides in the region analyzed, the C_H switching into three of these sites does not appear to represent random DNA rearrangement. We suggest that most of these repeats are potential germline S_α sites.

Evidence for Class-Specific Regulation of C_H Switching

The data presented in this paper, taken together with sequence data on γ2b (12, 13) and γ1 (11) switch sites from the literature, lead to several important inferences about the mechanism of C_H switching.

The germline sequences for C_H switching appear to be class specific with regard to $C_{\gamma 1}$, C_α and $C_{\gamma 2b}$ switch sites. The germline $C_{\gamma 1}$, C_α, and $C_{\gamma 2b}$ "switch" sequences are compared in Fig. 5. The $_{\gamma 1}$ sequence is identical to the prototype S_α sequence in 10 out of 22 bases and the $_{\gamma 2b}$ sequence is identical for 7 of 22 nucleotides (only 3 of which are contiguous). Thus, various germline S_α sequences are far more similar to one another than to the germline $S_{\gamma 1}$ or $S_{\gamma 2b}$ sequences (Fig. 3). One explanation for these sequence differences is that C_H switching is mediated by class-specific recognition sequences.

Since the germline S_α and S_μ sequences are not homologous (Fig. 3), homologous recombination cannot account for their joining. We believe the joining may be mediated by a number of distinct types of "switching" proteins (Fig. 6). For example, one switching protein (P_α) may bind the germline S_α sequence and a second (P_μ) may bind one of the germline S_μ sequences. These proteins may then interact to form a heterodimer which juxtaposes the V-D-J gene with the C_α gene segment (Fig. 6). The multiple germline S_α sequences would increase the probability that C_α switching could occur once the appropriate joining protein is expressed. Because the germline S_α, $S_{\gamma 1}$, and $S_{\gamma 2b}$ sequences seem distinct, different joining proteins could bind these sequences. Accordingly, the developmental regulation of the expression of these proteins would lead to class-specific regulation of C_H switching.

The Implications of Successive C_H Switching in MC101

The evidence for "successive" switching in the MC101 clone indicates that two or more C_H switches can occur in a particular B-cell line. In the simple deletional model for class switching proposed by Honjo (9), C_H switching progresses in a linear fashion, deleting intervening C_H genes at each stage. As mentioned, the experiments supporting this model (9, 10) indicate a gene order of C_μ-$C_{\gamma 3}$-$C_{\gamma 1}$-$C_{\gamma 2b}$-$C_{\gamma 2a}$-C_α. Paradoxically, however, the MC101 clone appears to have switched from C_μ to C_α and then from C_α to $C_{\gamma 1}$, contrary to the linear deletional model for class switching proposed by Honjo (9). Several explanations seem plausible.

1) The C_H gene order is C_μ-C_α-C_γ. This seems difficult to support because of the large number of myeloma tumors which express the $C_{\gamma 1}$ gene, and still contain C_α genes (9, 10). In contrast, those myeloma tumors which express the C_α gene generally appear to have deleted their $C_{\gamma 1}$ genes (10). However, until these C_H genes are ordered in the germline DNA, this remains a formal possibility.

2) Interchromosomal recombination. In this scheme, $C_\mu \rightarrow C_\alpha$ rearrangement on one chromosome could be followed by recombination with a $C_{\gamma 1}$ gene on another chromosome to produce the MC101 γ1 mosaic gene (V_H-C_μ-C_α-$C_{\gamma 1}$). Moreover, a prediction of this model is that the reciprocally rearranged chromosome should have a rearranged C_α gene.

3) Episomal deletion. If the deleted DNA between the V_H gene and the C_α gene formed a circular intermediate (an episome) that was at least transiently stable, the episome could then re-integrate into the chromosome, replacing the C_α gene with the $C_{\gamma 1}$ gene. This model is easily testable since it predicts the presence of a C_μ gene, a $C_{\gamma 3}$ gene, etc. in the MC101 genome and, in particular, the C_μ and C_α genes should be rearranged.

The rearranged MC101 γ1 gene, accordingly, raises two general possibilities with regard to C_H switching and normal B-cell differentiation. i) The successive (and complex) types of C_H switching observed here may arise from one or more aberrant chromosomal rearrangements that are unique to myeloma cell lines and will not generally be seen in normal B cells. The numerous cell divisions that occur between myeloma tumor production and our analyses of the corresponding DNAs, as well as the aneuploid nature of myeloma cells, make this a serious possibility. ii) Normal B cells may employ multiple C_H switching mechanisms—some perhaps different from any cited above.

The developmental regulation of C_H switching may operate at several different levels. i) The nature of the sequences mediating V-D-J joining and C_H switching implies that these phenomena are regulated independently. ii) The existence of distinct switch sequences for α, γ1, and γ2b genes implies that the expression of these classes may be developmentally regulated at the

level of DNA rearrangement, depending, for example, on which specific switching protein is expressed.

Mark M. Davis

Stuart K. Kim

Leroy E. Hood

Division of Biology

California Institute of Technology

Pasadena, California 91125

References and Notes

1. C. Brack, A. Hirawa, R. Lenhard-Schueller, S. Tonegawa, Cell 15, 1 (1978).
2. H. Sakano, K. Huppi, G. Heinrich, S. Tonegawa, Nature 280, 288 (1979).
3. J. G. Seidman, E. E. Max, P. Leder, Nature 280, 370 (1979).
4. P. Early, H. Huang, M. Davis, K. Calame, L. Hood, Cell 19, 981 (1980).
5. M. M. Davis, K. Calame, P. W. Early, D. L. Livant, R. Joho, I. L. Weissman, L. Hood, Nature 283, 733 (1980).
6. M. C. Raff, Cold Spring Harbor Symp. Quant. Biol. 41, 159 (1976).
7. M. D. Cooper, J. F. Kearney, P. M. Lydyard, C. E. Grossi, A. R. Lawton, Cold Spring Harbor Symp. Quant. Biol. 41, 139 (1976).
8. B. Pernis, L. Forni, A. L. Luzzati, Cold Spring Harbor Symp. Quant. Biol. 41, 175 (1976).
9. T. Honjo and T. Kataoka, Proc. Natl. Acad. Sci. U.S.A. 75, 2140 (1978).
10. S. Cory and J. M. Adams, Cell 19, 37 (1980); C. Coleclough, C. Cooper, R. P. Perry, Proc. Natl. Acad. Sci. U.S.A. 77, 1422 (1980); T. H. Rabbits, A. Forster, W. Dunnick, D. L. Bentley, Nature 283, 351 (1980); Y. Yaoita and T. Honjo, Biomed. Res., in press; Y. Yaoita and T. Honjo, Nature, in press.
11. T. Kataoka, T. Kawakami, N. Takahashi, T. Honjo, Proc. Natl. Acad. Sci. U.S.A. 77, 919 (1980).
12. N. Takahashi, T. Kataoka, T. Honjo, Gene, in press.
13. H. Sakano, R. Maki, Y. Kurosawa, W. Roeder, S. Tonegawa, Nature, submitted.
14. T. Maniatis, R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, A. Efstratiadis, Cell 15, 687 (1978).
15. P. W. Early, M. M. Davis, D. B. Kaback, N. Davidson, L. Hood, Proc. Natl. Acad. Sci. U.S.A. 76, 857 (1979).

16. M. M. Davis, P. W. Early, K. Calame, D. L. Livant, L. Hood, in Eukaryotic Gene Regulation, R. Axel, T. Maniatis, and C. F. Fox, Eds. (Academic Press, New York, 1980), p. 393.

17. N. Johnson, S. Rudikoff, P. Barstad, L. Hood, in preparation.

18. E. M. Southern, J. Mol. Biol. 98, 503 (1977).

19. The minor sequence differences noted might arise from one or more of the following sources: polymorphism in these flanking sequences arising from mice in different localities; mutations arising in the DNA sequences of myeloma tumors during repeated passages; repair or correction mechanisms during the C_H switching process; and/or DNA sequencing errors.

20. E. Max, J. G. Seidman, P. Leder, Proc. Natl. Acad. Sci. U.S.A. 76, 3450 (1979).

21. W. D. Benton and R. W. Davis, Science 196, 180 (1977).

22. A. Maxam and W. Gilbert, in Methods in Enzymology, Nucleic Acids Part I, L. Grossman and K. Moldave, Eds. (Academic Press, New York, 1980), vol. 65, p. 499.

23. This research was supported by NSF grant PCM 76-81546, and USPHS grant AI 09072. M.M.D. and S.K.K. are supported by NIH training grant GM 07616. All experiments involving recombinant organisms were conducted in accordance with the NIH Guidelines on recombinant DNA. We wish to thank Nate Kupperman for assistance with restriction mapping and Stephen Crews, David Goldberg, and Mitch Kronenberg for helpful discussions.

Figure Legends

Fig. 1. Clones containing rearranged alpha heavy chain genes and germline C_μ and C_α genes. The rearranged T15 and M603 genes were isolated from genomic libraries of their respective myeloma tumors, and MC101 was isolated as described (11). Clones containing the germline C_μ and C_α genes were isolated from libraries of mouse sperm DNA (5). Raised boxes denote coding regions. The shaded areas indicate homologies with germline genes and their flanking sequences:  , V_H and flanking region;  , C_μ and flanking region;  , C_α and flanking regions; and  , uncertain origin, probably C_μ derived (see text);  , $C_{\gamma 1}$ and flanking regions. Homologies were determined by detailed restriction mapping. Restriction enzyme sites are denoted as follows: Hf = HinfI, Hh = Hha, H = HindIII, M = MspI, R = EcoRI, Rs = RsaI, S = Sau3a, and Mb = MboII. The position of the J_H regions was determined from experiments described in refs. 4 and 13. The M603 clone was isolated and characterized as previously described (5, 15, 16). The T15 library was made from EcoRI partially digested DNA as described (5, 14). Recombinant phage were screened by the procedure of Benton and Davis (21) using ^{32}P nick translated V_H + C_α cDNA clones as probes as described (5, 14). Southern blot analysis (18) of EcoRI digested myeloma DNA indicated identity of central EcoRI fragments in the T15 genomes with the clones shown here (data not shown), indicating that the clones isolated are representative of their genome of origin. This analysis has been previously reported for M603 (16). The DNA sequence analysis employed the procedures of Maxam and Gilbert (22). Arrows originating from a restriction site indicate 5' end labeling with $\gamma^{32}P$-ATP, while arrows ending in a restricting site indicate 3' end labeling with $\alpha^{32}P$-dNTPs.

Fig. 2. Switch sequences for the rearranged T15, M603, and MC101 genes and their germline C_μ and C_α counterparts. $C_{\gamma 1}$ corresponding sequences for

MC101 have been reported by Honjo and colleagues (11). Sequences surrounding C_H switch sites are shown in 5' → 3' orientation. Underlining indicates sequence identity of unrearranged C_μ (▭) and C_α (▭) flanking sequences with their counterparts in the individual rearranged genes. Dots indicate 10 nucleotide spacings. DNA sequence analysis was performed as described (22).

Fig. 3. A comparison of the DNA sequences of the germline C_μ and C_α switch sites for the rearranged T15, M603, and MC101 genes. Sequences obtained from C_μ and C_α flanking regions were aligned for maximum homology around their C_H switch sites. Boxes indicate nucleotide identities and - indicates gap introduced for homology alignment. Arrows indicate breakpoints for the switch sites of each rearranged gene. Also indicated is the C_μ switch site for a γ2b producer, M141 (12, 13). A consensus sequence for at least some cases of C_μ switching might therefore be GGTNATTANNNNNNGGTANNCAAAG which does not occur elsewhere in any of the some 1900 nt of C_μ flanking region sequenced (13). Elements of this C_μ homology region have been suggested previously to play a role in C_H switching (12, 13). A consensus sequence for C_α switching derived from the examples here would be PGTCPPGCTGGAATPPGYTGGGNTGPGCTG.

Fig. 4. a) DNA sequence in the region of the germline C_α switch sites. Restriction sites are indicated. Recognition sequences are boxed. Also shown is a comparison of the MC101 C_α derived sequence (11) with that of the germline C_α sequence. The MC101 sequence is bracketed by S_{MC101} and S'_{MC101} and differences from the corresponding germline sequence are indicated above, either as base changes or gaps (-) or (-10, etc.). Precise localization of gaps in the tandem repeats GAGGA and GAGCT is not possible since the deletions are symmetrical.

It is interesting that virtually all of the sequence differences are localized on the 3' half of the MC101 sequence, with the tandem repeats acting as a

"border" between highly conserved and mutated regions.

b) Schematic diagram of the location of seventeen 30 nucleotide repeated sequences adjacent to the germline C_α gene. Arrows indicate which repeats are used in the three rearranged genes described in this paper.

c) DNA sequences of 17 repeats found interspersed in germline 5' flanking $C\alpha$ sequence. Boxes indicate nucleotide identities and - indicates a gap.

Fig. 5. A comparison of the switch sequences from germline 5' flanking sequences for the $C\alpha$ (see Fig. 3), $C_{\gamma 1}$ (11), and $\gamma 2b$ (12, 13) genes. Boxes in the germline α sequence denote the conserved bases of the 30 nucleotide recognition sequence. Boxes in the germline $_{\gamma 1}$ and $_{\gamma 2b}$ sequences denote nucleotide identities to the germline $C\alpha$ sequence. - indicates a gap.

Fig. 6. Model for class-specific regulation of C_H switching (see text). In this scheme the small boxes represent recognition (S) sequences which bind to switch proteins to mediate C_H switching; the circle represents switching proteins and the large boxes represent coding regions. This model depicts a C_H gene order of C_μ, C_γ, C_α (see text) but this is not a requirement.

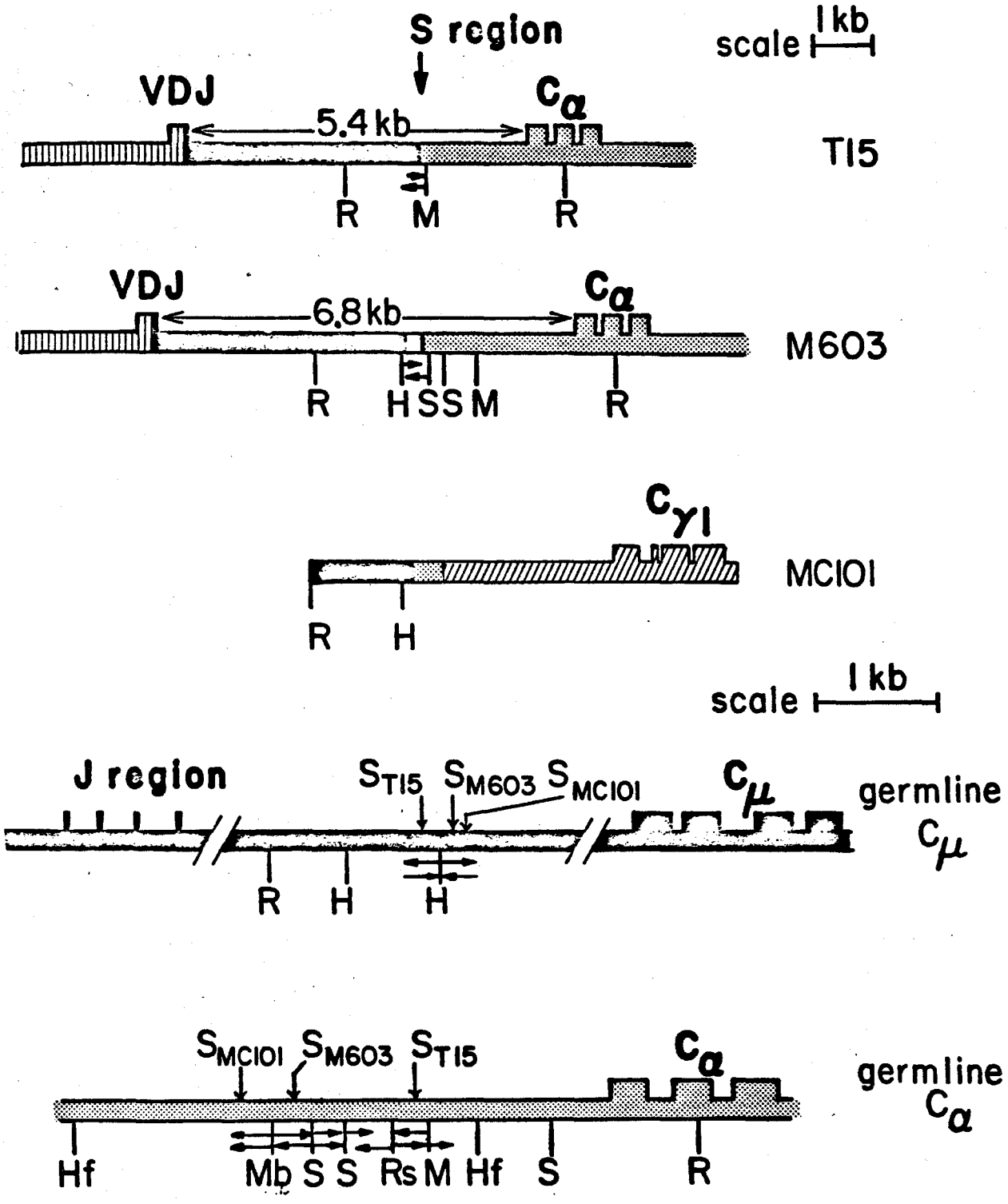

S region
scale 1kb
VDJ
Cα
5.4kb
T15
R
M
R
VDJ
Cα
6.8kb
M603
R
H
S
S
M
R
Cγ1
MC101
R
H
scale 1kb
J region
S T15
S M603
S MC101
Cμ
germline
Cμ
R
H
H
S MC101
S M603
S T15
Cα
germline
Cα
Hf
Mb
S
S
Rs
M
Hf
S
R

Fig. 1

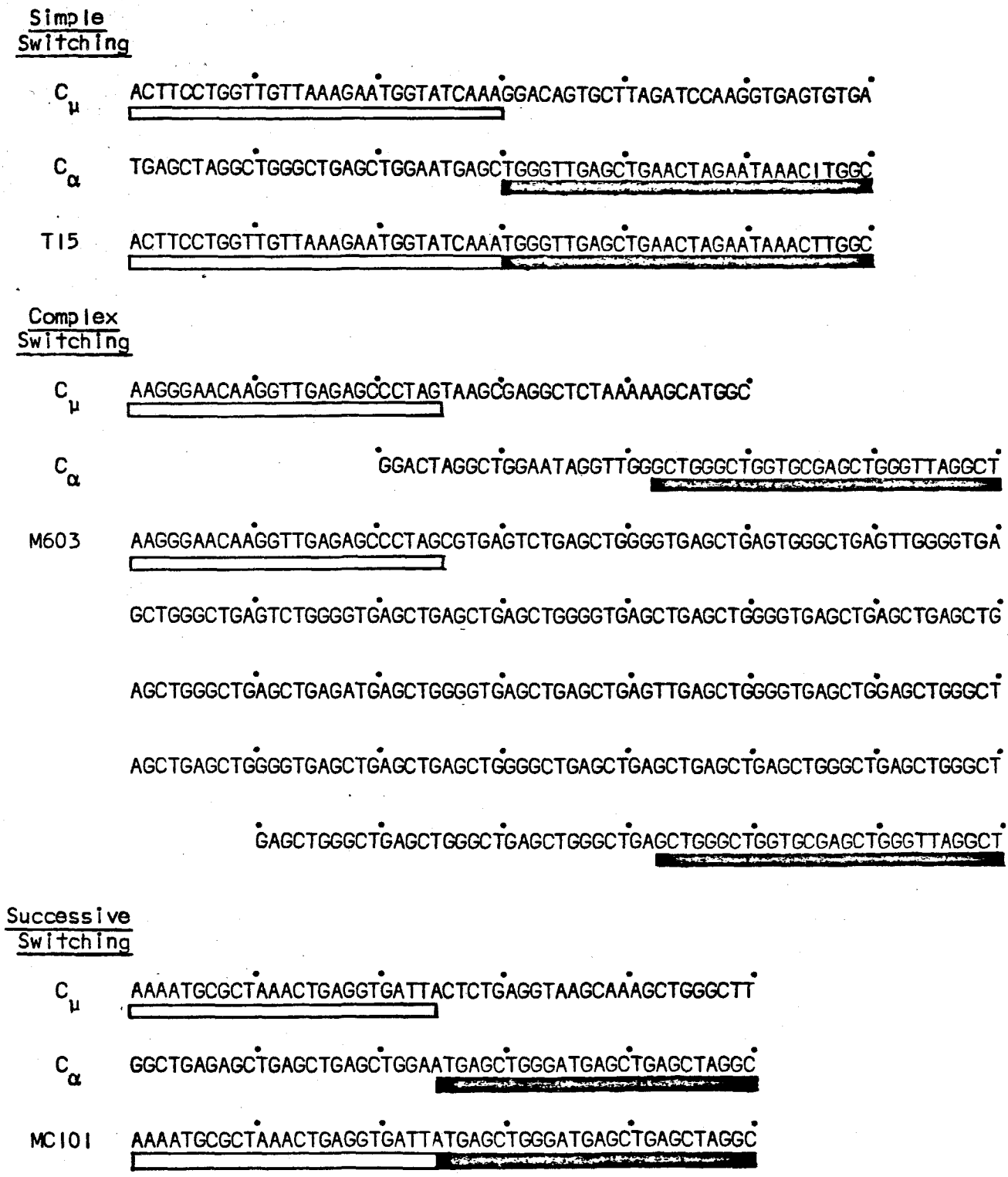

Simple Switching
C_μ ACTTCCTGGTTGTTAAAGAATGGTATCAAAGGACAGTGCTTAGATCCAAGGTGAGTGTGA
C_α TGAGCTAGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGAATAAACTTGGC
T15 ACTTCCTGGTTGTTAAAGAATGGTATCAAATGGGTTGAGCTGAACTAGAATAAACTTGGC
Complex Switching
C_μ AAGGGAACAAGGTTGAGAGCCCTAGTAAGCGAGGCTCTAAAAAGCATGGC
C_α GGACTAGGCTGGAATAGGTTGGGCTGGGCTGGTGCGAGCTGGGTTAGGCT
M603 AAGGGAACAAGGTTGAGAGCCCTAGCGTGAGTCTGAGCTGGGGTGAGCTGAGTGGGCTGAGTTGGGGTGA
GCTGGGCTGAGTCTGGGGTGAGCTGAGCTGAGCTGGGGTGAGCTGAGCTGGGGTGAGCTGAGCTGAGCTG
AGCTGGGCTGAGCTGAGATGAGCTGGGGTGAGCTGAGCTGAGTTGAGCTGGGGTGAGCTGGAGCTGGGCT
AGCTGAGCTGGGGTGAGCTGAGCTGAGCTGGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGAGCTGGGCT
GAGCTGGGCTGAGCTGGGCTGAGCTGGGCTGAGCTGGGCTGGTGCGAGCTGGGTTAGGCT
Successive Switching
C_μ AAAATGCGCTAAACTGAGGTGATTACTCTGAGGTAAGCAAAGCTGGGCTT
C_α GGCTGAGAGCTGAGCTGAGCTGGAATGAGCTGGGATGAGCTGAGCTAGGC
MC101 AAAATGCGCTAAACTGAGGTGATTATGAGCTGGGATGAGCTGAGCTAGGC

Fig. 2

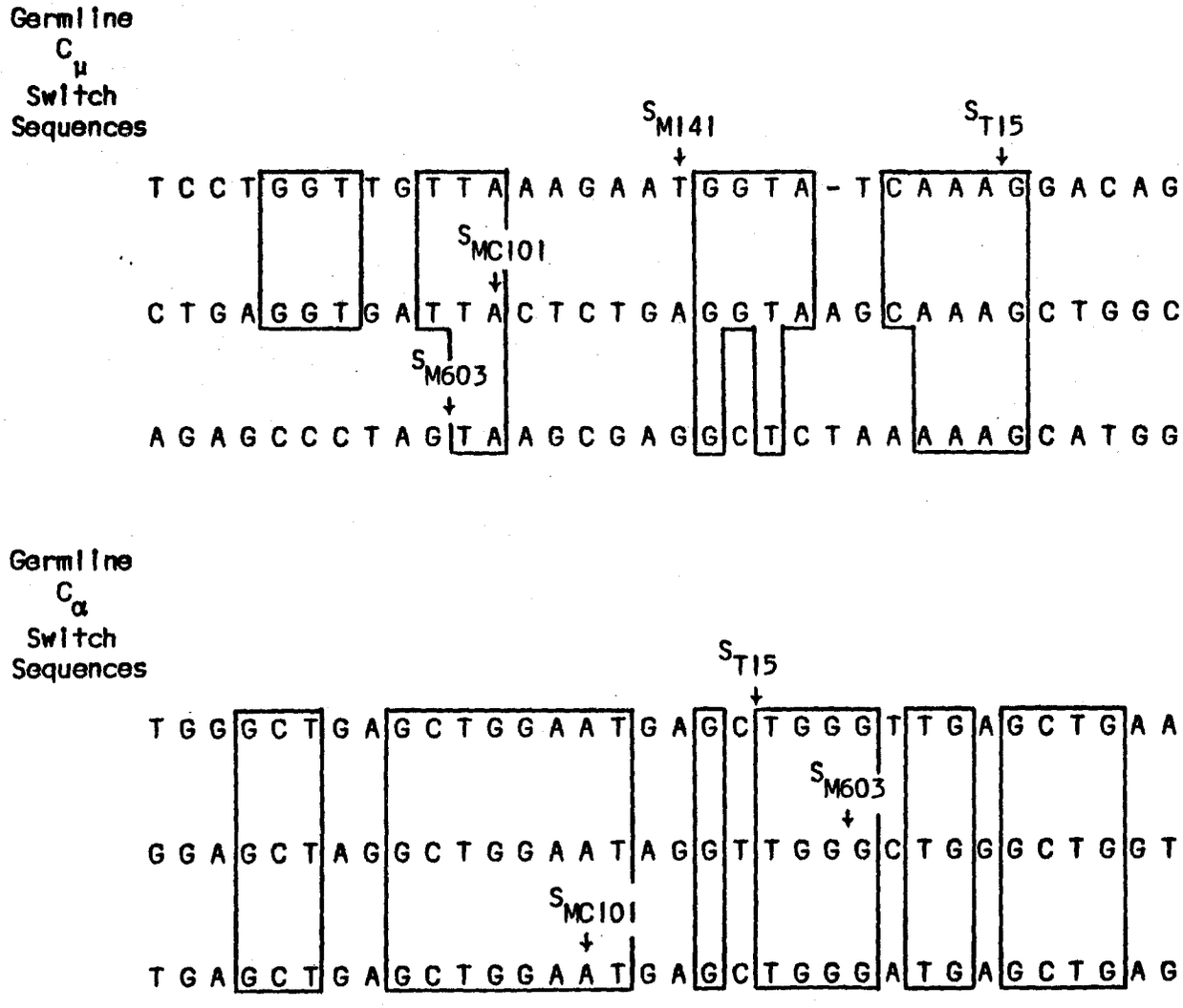

Germline
C_μ
Switch
Sequences
S_{M141}
S_{T15}
TCCTGGTTGTTAAAGAATGGTA-TCAAAGGACAG
S_{MC101}
CTGAGGTGATTACTCTGAGGTAAGCAAAGCTGGC
S_{M603}
AGAGCCCTAGTAAGCGAGGCTCTAAAAAGCATGG
Germline
C_α
Switch
Sequences
S_{T15}
TGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAA
S_{M603}
GGAGCTAGGCTGGAATAGGTTGGGCTGGGCTGGT
S_{MC101}
TGAGCTGAGCTGGAATGAGCTGGGATGAGCTGAG

Fig. 3

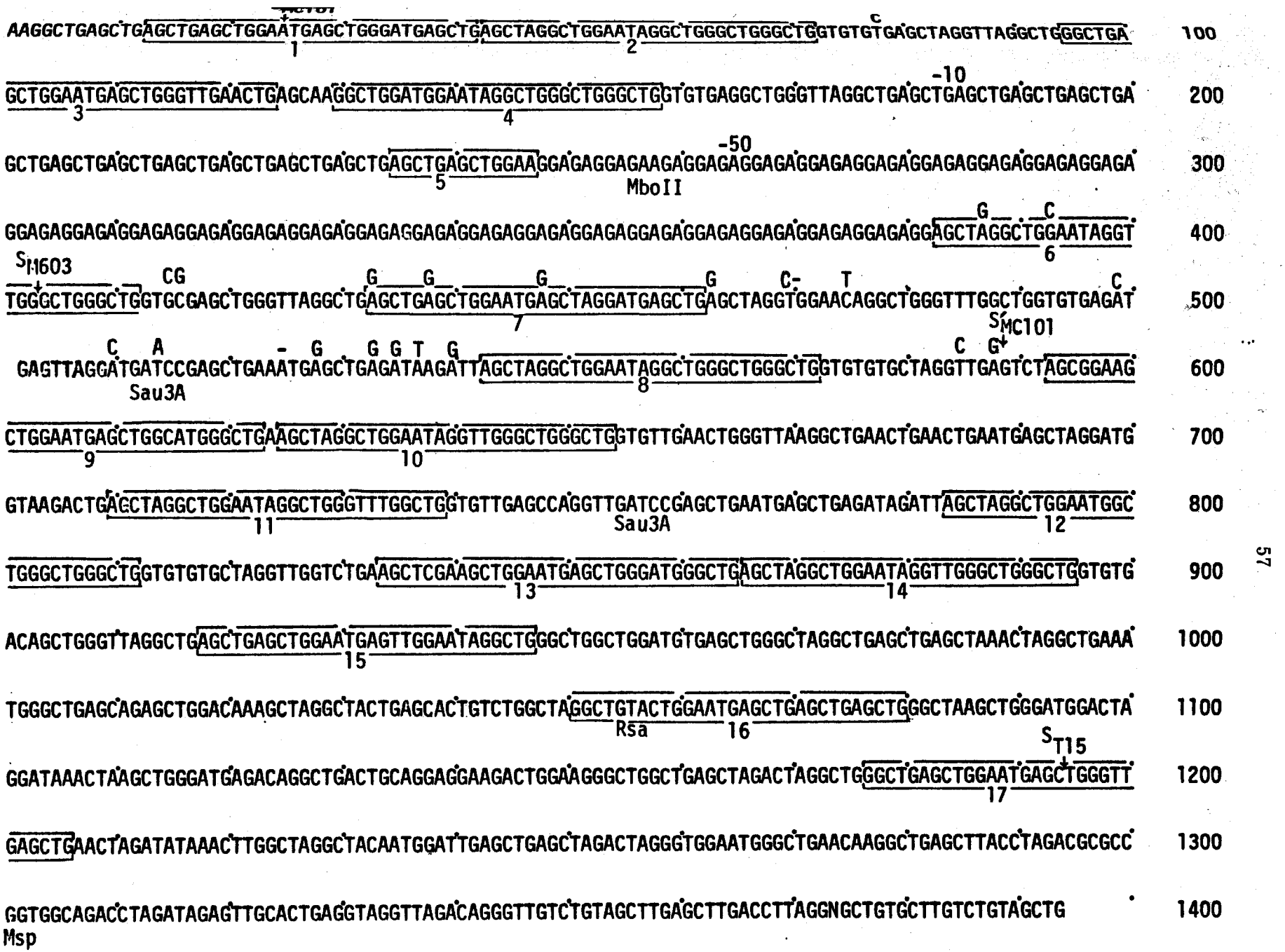

AAGGCTGAGCTGAGCTGAGCTGGAATGAGCTGGGATGAGCTGAGCTAGGCTGGAATAGGCTGGGCTGGGCTGGTGTGTGAGCTAGGTTAGGCTGGGCTGA 100
GCTGGAATGAGCTGGGTTGAACTGAGCAAGGCTGGATGGAATAGGCTGGGCTGGGCTGGTGTGAGGCTGGGTTAGGCTGAGCTGAGCTGAGCTGAGCTGA 200
GCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGGAAGGAGAGGAGAAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGA 300
MboII
GGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGCTAGGCTGGAATAGGT 400
TGGGCTGGGCTGGTGCGAGCTGGGTTAGGCTGAGCTGAGCTGGAATGAGCTAGGATGAGCTGAGCTAGGTGGAACAGGCTGGGTTTGGCTGGTGTGAGAT 500
GAGTTAGGATGATCCGAGCTGAAATGAGCTGAGATAAGATTAGCTAGGCTGGAATAGGCTGGGCTGGGCTGGTGTGTGCTAGGTTGAGTCTAGCGGAAG 600
Sau3A
CTGGAATGAGCTGGCATGGGCTGAAGCTAGGCTGGAATAGGTTGGGCTGGGCTGGTGTTGAACTGGGTTAAGGCTGAACTGAACTGAATGAGCTAGGATG 700
GTAAGACTGAGCTAGGCTGGAATAGGCTGGGTTTGGCTGGTGTTGAGCCAGGTTGATCCGAGCTGAATGAGCTGAGATAGATTAGCTAGGCTGGAATGGC 800
Sau3A
TGGGCTGGGCTGGTGTGTGCTAGGTTGGTCTGAAGCTCGAAGCTGGAATGAGCTGGGATGGGCTGAGCTAGGCTGGAATAGGTTGGGCTGGGCTGGTGTG 900
ACAGCTGGGTTAGGCTGAGCTGAGCTGGAATGAGTTGGAATAGGCTGGGCTGGCTGGATGTGAGCTGGGCTAGGCTGAGCTGAGCTAAACTAGGCTGAAA 1000
TGGGCTGAGCAGAGCTGGACAAAGCTAGGCTACTGAGCACTGTCTGGCTAGGCTGTACTGGAATGAGCTGAGCTGAGCTGGGCTAAGCTGGGATGGACTA 1100
Rsa
GGATAAACTAAGCTGGGATGAGACAGGCTGACTGCAGGAGGAAGACTGGAAGGGCTGGCTGAGCTAGACTAGGCTGGGCTGAGCTGGAATGAGCTGGGTT 1200
GAGCTGAACTAGATATAAACTTGGCTAGGCTACAATGGATTGAGCTGAGCTAGACTAGGGTGGAATGGGCTGAACAAGGCTGAGCTTACCTAGACGCGCC 1300
GGTGGCAGACCTAGATAGAGTTGCACTGAGGTAGGTTAGACAGGGTTGTCTGTAGCTTGAGCTTGACCTTAGGNGCTGTGCTTGTCTGTAGCTG 1400
Msp

Fig. 4a

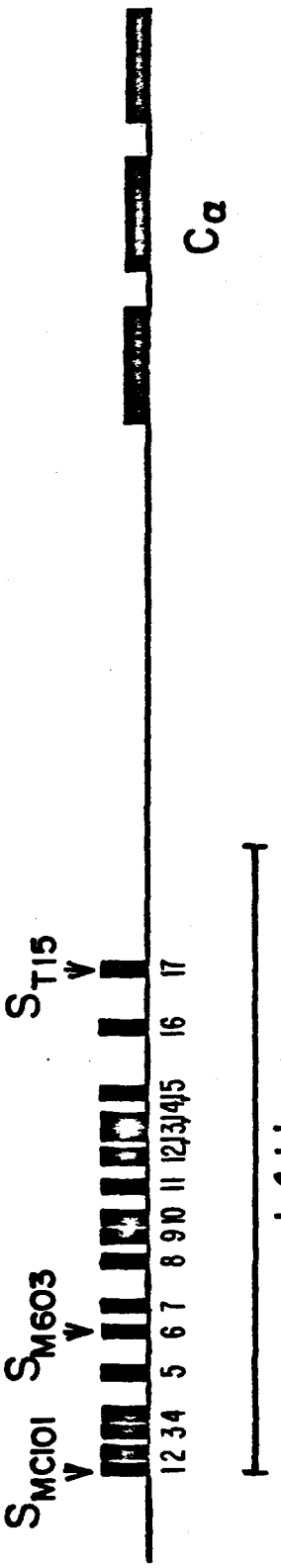
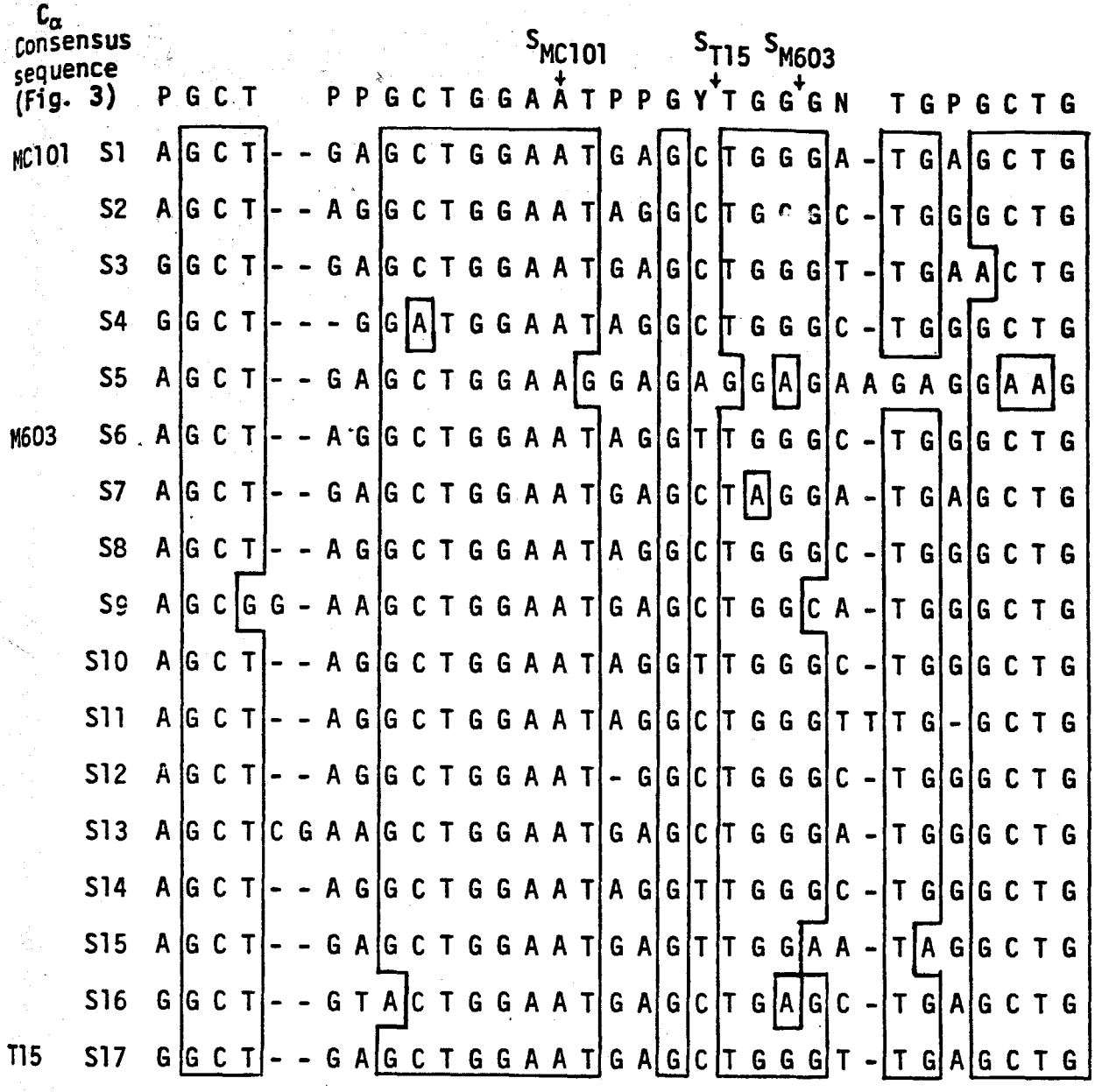


Fig. 4b



P = purine Y = pyrimidine N = any nucleotide

Fig. 4c

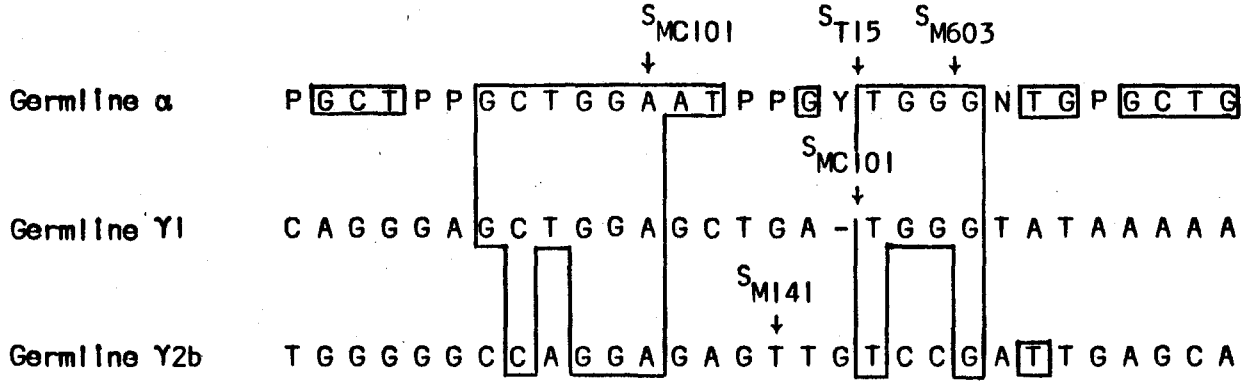

S MC101
S T15
S M603
Germline α
P G C T P P G C T G G A A T P P G Y T G G G N T G P G C T G
S MC101
Germline Yl
C A G G G A G C T G G A G C T G A - T G G G T A T A A A A A
S M141
Germline Y2b
T G G G G G C C A G G A G A G T T G T C C G A T T G A G C A

Fig. 5

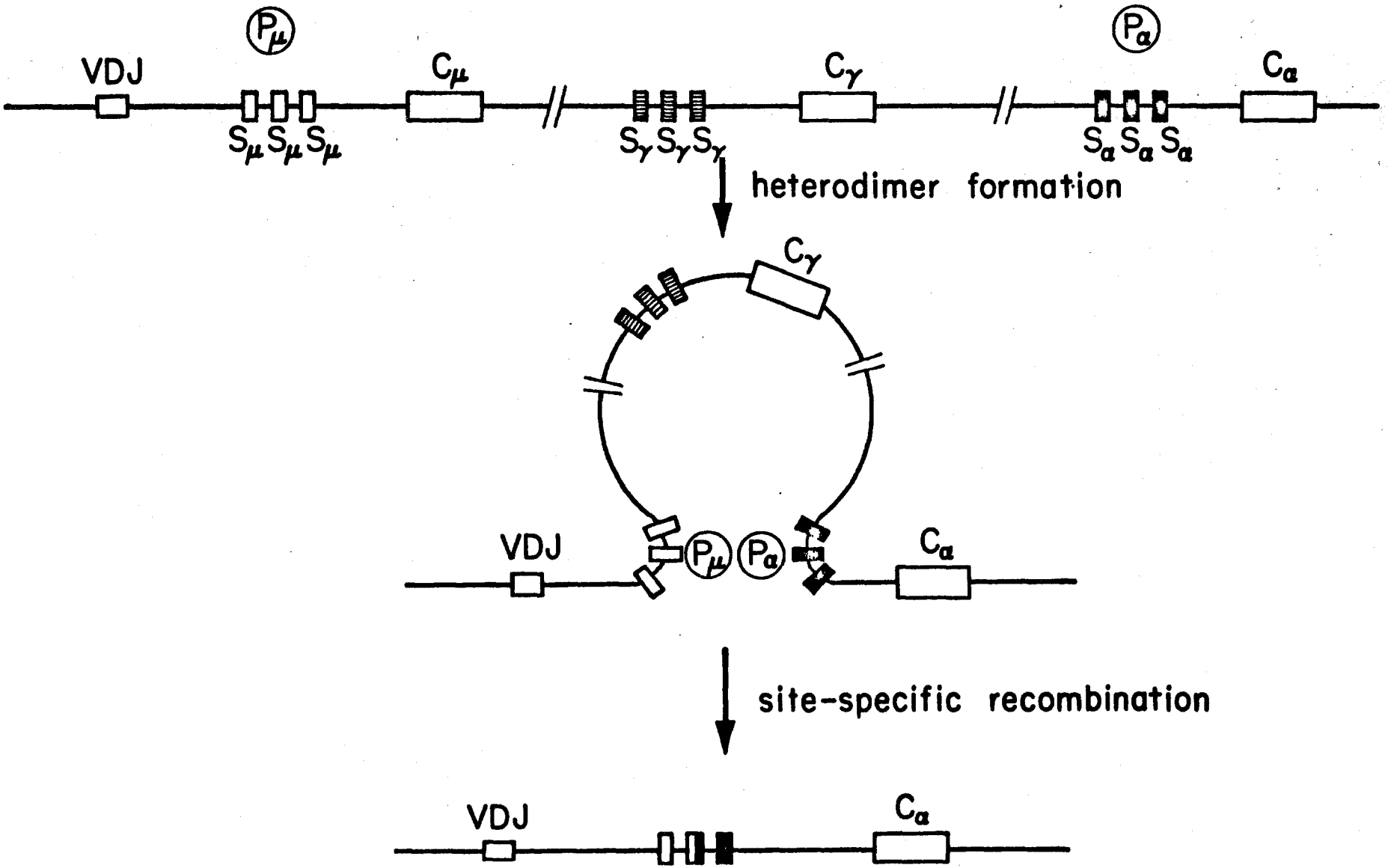

VDJ
P_μ
C_μ
C_γ
P_α
C_α
$S_\mu S_\mu S_\mu$
$S_\gamma S_\gamma S_\gamma$
$S_\alpha S_\alpha S_\alpha$
heterodimer formation
C_γ
VDJ
P_μ
P_α
C_α
site-specific recombination
VDJ
C_α

Fig. 6

Two Types of DNA Rearrangements in Immunoglobulin Genes

L. HOOD, M. DAVIS, P. EARLY, K. CALAME,

S. KIM, S. CREWS, AND H. HUANG

Division of Biology, California Institute of Technology

Pasadena, California 91125

Cold Spring Harbor Symp. Quant. Biol. (1980) In Press

Running Head:

Send proofs to: L. Hood
Division of Biology
California Institute of Technology
Pasadena, California 91125

Phone: (213) 795-6811, Ext. 1951

Antibody genes, along with most other eukaryotic genes, are examples of "split genes." One unique feature of the antibody gene system is that it can rearrange individual gene segments by two distinct mechanisms during the differentiation of antibody-producing or B cells. These DNA rearrangements amplify the information encoded in the antibody gene families.

The antibody or immunoglobulin molecule is composed of two distinct polypeptides, light (L) and heavy (H) chains, both of which are divided into variable (V) regions responsible for antigen binding and constant (C) regions responsible for effector functions such as complement fixation. There are three unlinked families of antibody genes—two code for lambda (λ) and kappa (κ) light chains, and the third codes for heavy chains. Light chains are encoded by three distinct gene segments—V_L (variable), J_L (joining), and C_L (constant), whereas heavy chains may be encoded by four—V_H, D (diversity), J_H and C_H (Fig. 1) (Bernard et al. 1978; Early et al. 1980).

During the differentiation of antibody-producing cells, two types of DNA rearrangements occur. First, the V_L and J_L or the V_H, D and J_H gene segments are joined to generate a contiguous coding sequence for the entire light or heavy chain variable region. We term this type of DNA rearrangement V(D)J joining. A second type of DNA rearrangement, C_H or class switching, occurs in the heavy chain gene family (Davis et al. 1980; Kataoka et al. 1980). In the mouse this gene family contains at least eight closely linked heavy chain constant region genes (C_μ, C_δ, $C_{\gamma 1}$, $C_{\gamma 3}$, $C_{\gamma 2b}$, $C_{\gamma 2a}$, C_α, and C_ε). During the development of a B cell, IgM (C_μ) is initially expressed, while later the B cell or its progeny may switch to expression of any of the other classes or subclasses of immunoglobulins (Fig. 2) (Cooper et al. 1976; Pernis et al. 1976). In molecular terms these distinct stages of B-cell differentiation are mediated by the two types of DNA rearrangements discussed above. The

V(D)J joining generates the functional V_H gene 5' to the C_μ gene, leading to the synthesis of μ chains and IgM molecules. Later, the B cell or its clonal progeny can undergo a class switch DNA rearrangement in which another heavy chain constant region gene replaces the C_μ gene adjacent to the same functional V_H gene. For example, a transcript can then be made which contains the V_H and C_α genes, leading to the synthesis of α chains and IgA molecules. In the functional gene, each domain of the immunoglobulin chain (the variable region and one or more constant region domains) is encoded by a separate exon, so that RNA splicing is required to generate the final V+C immunoglobulin mRNA from transcripts of the rearranged gene (Bernard et al. 1978; Early et al. 1979; Sakano et al. 1979b).

Variable Region DNA Rearrangement

Studies of immunogloublin gene organization have provided insights into the process of V(D)J joining and the mechanisms by which the enormous repertoire of specificities in the vertebrate immune system is generated.

In the light chain gene families, λ and κ, the variable regions are encoded by two separate germline gene segments. The V_L gene segments generally encode residues 1-96 from the N terminus, and the J_L gene segments, residues 97-108, the remainder of the variable region (Bernard et al. 1978; Max et al. 1979; Sakano et al. 1979b). The heavy chain gene organization is somewhat more complex in that a third gene segment, D, appears to encode from one to fifteen residues between those encoded by the V_H and J_H gene segments (Early et al. 1980). For example, one heavy chain gene, M603, has the germline V_H and J_H gene segments encoding variable region amino acids 1-101 and 107-122, respectively. Thus, amino acids 102-106 must be encoded by a third gene segment, termed diversity (D), because this is the most

diverse segment of heavy chain variable regions (Early et al. 1980; Schilling et al. 1980). The DNA sequence coding for the D segment of the M603 α gene is not found near the 3' end of the germline V_H gene segment or the 5' end of the germline J_H gene segment. Although a germline D gene segment has not yet been identified, it clearly must be encoded by DNA sequences separate from V_H and J_H.

In light chains, the J_K and J_λ gene segments are approximately 2.4 kb and 1.2 kb 5' to the C_K and C_λ genes, respectively (Bernard et al. 1978; Max et al. 1979; Sakano et al. 1979b). Accordingly, V_J joining brings a V_L segment into close association with its corresponding C_L gene. In heavy chains, the J_H gene segments are approximately 8 kb 5' to the C_μ gene, the first in the series of C_H genes to be expressed by a B cell (Early et al. 1980). V(D)J joining creates a functional V_H gene associated with the C_μ gene. Germline V (and D) gene segments are probably 5' to the J gene segments in each gene family, although the distances separating these gene segments are unknown.

DNA Recognition Sequences for Joining Proteins May Mediate V(D)J Joining

Immunoglobulin genes can be examined in the differentiated and undifferentiated state by studying their organization in myeloma (differentiated) and sperm or embryo (undifferentiated or germline) DNAs. Analyses of DNA sequences near potential sites of DNA joining flanking germline V and J gene segments has revealed blocks of highly conserved nucleotides in each of the antibody gene families (Bernard et al. 1978; Max et al. 1979; Sakano et al. 1979b;Early et al. 1980). Moreover, these blocks of nucleotides are the same in all the immunoglobulin gene families of the mouse, and a recent report indicates that the same conserved nucleotides are adjacent to human V

gene segments (T. Rabbitts, personal communication). Thus these nucleotides have been conserved for more than 500 million years, since the immunoglobulin light and heavy chain gene families diverged from one another. This striking conservation argues compellingly that these blocks of nucleotides play a fundamental role in V(D)J joining.

The structures of these blocks of nucleotides for each of the three immunoglobulin families are presented in Table 1. These nucleotides are found at the 3' side of the V gene segments and at the 5' side of the J gene segments—precisely at the locations expected for sequences which mediate V(D)J joining. We have noted that these conserved nucleotides always occur as two blocks, of 7 and 10 nucleotides, separated by a nonconserved spacer sequence of either 11 or 22 nucleotides (Table 1) (Early et al. 1980). The spacer lengths correspond closely to separations of these blocks of recognition sequences by 1 or 2 turns of the DNA helix (1-turn or 2-turn spacers). The conserved nucleotides adjacent to V gene segments are nearly inverted complements of the nucleotides adjacent to J gene segments. The observation of these conserved nucleotides in the light chain gene families initially led to the suggestion that V-J joining occurs by virtue of a stem and loop structure that could be formed by the self complementarity of the inverted repeat sequences at the 5' end of the J gene segment and the 3' end of the V gene segment (Max et al. 1979; Sakano et al. 1979b). Presumably a site-specific recombination event across the base of the stem structure could join the V and J gene segments. However, the existence of separated blocks of conserved nucleotides and the "1- and 2-turn" spacer relationships led us to consider a more attractive alternative mechanism of V(D)J joining (Early et al. 1980).

In the κ gene family the V gene segments have blocks of conserved nucleotides that are always separated by a 1-turn spacer (Table 1). In contrast,

the J_K gene segments have their blocks of conserved nucleotides separated by 2-turn spacers. The $V_{\lambda I}$ gene segment has a 2-turn spacer, and the $J_{\lambda I}$ gene segment has a 1-turn spacer. Heavy chain V_H and J_H gene segments both have 2-turn spacers. This pattern of sequence organization has led us to postulate that the conserved nucleotides are recognition sites for two types of DNA joining proteins. One joining protein interacts with the 1-turn recognition site, whereas the second joining protein recognizes the 2-turn site. We suggest that after DNA binding a 1-turn joining protein may dimerize with a 2-turn joining protein, but not with another 1-turn protein. This dimerization apposes V and J gene segments so that a site-specific recombination event can splice together these gene segments. This model has several attractive features. First, it will prevent nonproductive V-V or J-J joining. Second, this model is directly applicable to all antibody gene families. When the model is applied to the heavy chain gene family, it makes an interesting prediction about the structure of germline D gene segments. Because the V_H and J_H gene segments both have 2-turn spacers, we predict that the D gene segments will have recognition sequences with 1-turn spacers on both their 5' and 3' sides.

Two types of recognition sites have evolved in immunoglobulin gene segments, each of which contains the same two blocks of conserved nucleotides apparently involved in protein-DNA interactions. The only difference in the sites is the distance (1 or 2 helical turns) between the blocks of conserved nucleotides. This pattern suggests that the evolution of the proteins involved in immunoglobulin variable region gene rearrangement occurred as follows. Separate protein subunits or domains probably once existed which could bind the individual blocks of nucleotides, CACTGTG and GGTTTTTGTA. These proteins may have functioned in DNA rearrangement, or could have served

other purposes requiring protein recognition of the specific DNA sequence. In order to increase the specificity of DNA binding (an important consideration in a large vertebrate genome), compound proteins could have evolved which combined the specificities for the two original DNA binding sites. The spatial orientation of the protein subunits determined the length of any spacer DNA required between the blocks of CACTGTG and GGTTTTTGTA. The different spatial orientations in proteins requiring 11 and 22 nucleotide spacers may simply be the result of different polymers of the same subunits. In this fashion, two (or more) highly specific DNA-binding proteins could have evolved by combinations or existing proteins which specifically bound shorter sequences of DNA. This is likely to have been a much more rapid process than the evolution of entirely new binding sites in proteins specific for longer sequences of DNA.

Support for this view of the evolution (and antiquity) of the DNA-binding proteins involved in immunoglobulin variable region gene rearrangement comes from a gene for the major flagellar protein in Salmonella. Silverman et al. (1980) have identified a 14 nucleotide sequence present in opposite orientations on the ends of the invertible DNA segment controlling which of two types of flagella protein is synthesized by Salmonella. The blocks of 10 conserved nucleotides in the immunoglobulin recognition sequences are almost completely identical to part of the Salmonella-inverted sequence. GGTTTTTGTA is the consensus sequence in immunoglobulin genes, whereas GGTTTTTGAT is the sequence of Salmonella. The difference between these two sequences is no greater than between some of the immunoglobulin sequences shown in Table 1. Accordingly, this sequence may be a very ancient recognition sequence for a protein involved in DNA rearrangements. The system of DNA inversion in Salmonella also is apparently closely related or identical to that of the

G-loop in mu phage (Katsukake and Ino 1980). Eukaryotes may have retained the recognition sequence and something of the function of this protein, which seems then to have combined with another DNA-binding protein in evolving the mechanism for variable region gene rearrangements in the immune system.

Several Distinct Mechanisms Generate Antibody Diversity

The controversy between germline and somatic theories of antibody diversity (see CSHSQB 1976) has largely disappeared because both kinds of mechanisms contribute significantly to antibody diversity.

Germline genes. The DNA rearrangements necessary to generate complete variable region genes are a major source of diversity in antibody molecules. First, this process allows many alternative variable region germline gene segments to exist, any of which can be expressed with the constant region gene(s) in its family. There are perhaps 200 to 500 light chain variable region gene segments (Seidman et al. 1978), and presumably a comparable number for heavy chains. Both the kappa light chain and heavy chain gene families of mouse have four J gene segments (Max et al. 1979; Sakano et al. 1979b; Sakano et al. 1980; D. Kemp, personal communication). The number of heavy chain D gene segments is unknown, but judging from the diversity of heavy chain protein sequences (Schilling et al. 1980) this number may be on the order of 10.

Combinatorial joining. Protein sequence studies suggest that for any particular antibody gene family, any V gene segment may be joined to any J gene segment (Weigert et al. 1980). Presumably the same will be true of D gene segments with regard to their joining to V_H and J_H gene segments. Accordingly, the combinatorial joining of, for example, 200 V_H, 10 D, and 4

J_H gene segments can generate 8000 distinct V_H regions. In contrast, 200 V_K and 4 J_K gene segments will generate 800 V_K regions. Thus the existence of a third gene segment, D, in the heavy chains considerably increases the potential for V_H gene diversification. The joining of V_L and J_L as well as V_H, D, and J_H gene segments occurs in the third hypervariable region—an important part of antigen-binding sites. Thus diversification in the third hypervariable region by combinatorial joining of gene segments can lead to a diversity of antigen-binding sites.

Somatic variations. Although the blocks of recognition sequences adjacent to variable region gene segments appear to guide the enzymes responsible for DNA rearrangement, the exact points at which the DNA joining occurs are quite variable. This gives rise to variable region junctional diversity, as first noted at the V-J junction of light chains (Max et al. 1979; Sakano et al. 1979b). Some examples of heavy chain junctional diversity at both V-D and D-J junctions are shown in Figure 4. The length of the D and J_H gene segments in the rearranged variable region can vary over a few codons in the third hypervariable region, and new hybrid codons can be created at the junction (Weigert et al. 1980).

An additional opportunity for variability unique to heavy chains arises from the presence of DNA junctions on both sides of the D gene segment. In order for an immunoglobulin chain to be synthesized, the V_H and J_H (and consequently C_H) gene segments must be translated in the same reading frame. There is no necessary constraint on the reading frame of the D gene segment, so long as it does not contain a termination codon. As long as any changes in the reading frame at the V_H-D junction are compensated by changes at the D-J_H junction, it should be possible for a given germline D gene segment to occur in any of three reading frames in rearranged variable

region genes. A probable example of the use of alternative reading frames for a D segment is shown in Figure 5. Comparing the protein sequence of a levan-binding (A4) V_H region with the DNA sequences of the germline J_{H3} gene segment, and an A4-like germline V_H gene segment indicates that the A4 protein probably contains a D segment (S. Crews, unpublished results), contrary to previous conclusions based only on the protein sequence (Schilling et al. 1980). The nucleotide sequence of the A4 D segment, as deduced from the protein sequence, closely resembles or may be identical to the S107 D gene segment shown in Figure 4. This nucleotide sequence is translated in different reading frames in the A4 and S107 V_H regions.

Generation of antibody diversity. As mentioned above, by the combinatorial joining of V, D, and J gene segments, mice may generate approximately 10^4 V_H regions and 10^3 V_L regions. Independent association of these light and heavy chains in antibody molecules produced by different B-cell clones will generate approximately 10^7 different antibody molecules. Accordingly, combinatorial joining and association will generate a vast repertoire of antibody molecules before employing any of the mechanisms of somatic mutation. Clearly higher vertebrates can easily generate 10^8 or more antibody molecules by combining all of the mechanisms discussed—multiple germline gene segments, combinatorial joining, somatic mutation, and independent association of light and heavy chains.

Heavy Chain Switching Involves a Second Type of DNA Rearrangement

We first observed the results of heavy chain switching at the DNA level in a functional α heavy chain gene from the M603 myeloma tumor (Figure 6) (Davis et al. 1980a). The M603 V_H exon is separated from the

C_α exon by a 6.8 kb intervening sequence. Comparison of the M603 α gene with germline V_H, C_α, and J_H plus C_μ genes (Figure 6) shows that the M603 gene is composed of at least three distinct segments of germline DNA. These are: a V_H gene segment and its 5' flanking sequences, a J_H gene segment and associated sequences originally adjacent to the C_μ gene, and the C_α gene with its 5' and 3' flanking sequences. DNA sequence analysis also shows that the M603 α gene contains a D gene segment between the V_H and J_H gene segments (Early et al. 1980). Thus the structure of the M603 α gene demonstrates that at least two separate DNA rearrangements must have occurred. The first type of DNA rearrangement, V(D)J joining, generated a functional V_H gene associated with the C_μ gene. The second type of DNA rearrangement, C_H switching, took place in intervening DNA sequences between the J_H gene segment and the C_μ gene, and resulted in the replacement of the C_μ gene with the C_α gene. The points at which the flanking C_μ and C_α sequences join in the rearranged gene and the corresponding breakpoints on the germline DNAs are termed the switch (S) sites after Kataoka et al. (1980). Subsequently, other laboratories have obtained evidence for similar C_μ to C_H switches in rearranged γ1 (Kataoka et al. 1980) and γ2b (Takahashi et al. 1980; Sakano et al. 1980) genes.

Three Examples of IgM → IgA Switching

In order to investigate the molecular mechanisms underlying C_H switching, we sequenced the switch site of M603 and that of an additional IgA-producing myeloma tumor, T15. Homologies between clones containing the rearranged α genes of T15 and M603 are depicted in Figure 6 and compared with their germline C_μ and C_α counterparts obtained from a genomic library of mouse sperm DNA (Davis et al. 1980a). These homologies were established by detailed restriction

enzyme analysis (data not shown). Both rearranged α genes exhibit the tripartite structure of V_H, J_H with C_μ flanking sequences and C_α gene segments. The size of the intervening sequence between the V_H and C_α coding regions is substantially different in these two cases—5.4 vs. 6.8 kilobases (Fig. 6). Since each of the two V_H gene segments are joined to the same J_H gene segment (Johnson et al. 1980, in preparation), the variation in size in the intervening sequences between the V and C_α coding regions may be the result of using different sites for C_H switching in each of these rearranged C_α genes. In addition, we found that a rearranged γ1 gene, MC101, whose isolation and switch site sequence was reported by Honjo and colleagues (Kataoka et al. 1980) contains a 500 nucleotide region of C_α flanking sequence between C_μ and $C_{\gamma 1}$ derived sequences (Fig. 6). The evidence for this statement is that this 500 nucleotide sequence has been localized solely to a region 5' to the C_α gene by Southern blotting analyses (Southern 1975) and restriction mapping using fragments containing all or part of this region as probes (Kataoka et al. 1980; M. Davis, data not presented). Therefore, this fragment is apparently represented just once in the genome and, accordingly, must have been derived from flanking sequences 5' to the C_α gene. Furthermore, DNA sequence analyses of the corresponding germline C_α sequence show virtual identity (>95%) with the sequence found between the C_μ and $C_{\gamma 1}$ flanking sequences in the MC101 γ1 gene (Davis et al. 1980b). Thus we feel the MC101 γ1 gene is composed of several distinct germline sequences: a V_H gene, flanking sequence for the C_μ gene, flanking sequence for the C_α gene and the γ1 gene with its flanking sequences.

The arrows in Figure 6 indicate the regions analyzed by DNA sequence analysis in our laboratory. The DNA sequences of the MC101 S region (Kataoka et al. 1980) and the rearranged M603 and T15 α genes and their germline C_μ

and Cα counterparts are shown in Figure 7.

Examination of the rearranged switch sequences indicates that all three examples juxtapose C_μ and C_α derived sequences. However, each switch sequence seems distinct from the others. The arrangement of sequences in these genes suggests that at least three distinct types of switching may occur. We denote these categories "simple," "complex," and "successive."

In the simple category, T15, the C_μ flanking sequence joins directly to the C_α flanking sequence. Similarly, a γ2b gene (M141) has been found to join C_μ flanking sequence to that of $C_{\gamma 2b}$ (Takahashi et al. 1980; Sakano et al. 1980). In the complex category, M603, a short sequence of 287 base pairs, is interposed between the C_μ and C_α flanking sequences. This sequence appears to derive from a region 3' to S_{M603} on the C_μ gene (Fig. 6). Probes containing this sequence hybridize strongly to restriction fragments containing or adjacent to the C_μ gene in Southern blotting analyses (M. Davis, data not shown). The region of hybridization corresponds to the 1.5-2.5 kb region 5' to the C_μ gene (Davis et al. 1980a) which deletes spontaneously upon cloning and hence is not present in our C_μ containing clones. Thus the complex category is explained by two distinct deletions—C_H switch and a deletion within the C_μ flanking sequence. This deletion of C_μ flanking sequence (at S_{M603}) does not appear to be a random event in that another α producing tumor line, Y5236, switches at exactly the same point adjacent to the C_μ gene. However, the complex category reflects DNA deletions seen to date only in myeloma cells and, accordingly, may or may not be biologically significant. In the third category, "successive" switching, the MC101 γ1 gene contains C_μ, C_α, and $C_{\gamma 1}$ flanking sequences between V_H and $C_{\gamma 1}$ gene segments. Therefore, it appears to have switched twice, once from C_μ to C_α and subsequently from C_α to $C_{\gamma 1}$.

The DNA sequence data in Figure 7 show that the three rearranged genes employ three germline C_μ switch sites up to 300 base pairs apart and three germline C_α switch sites up to 1350 base pairs apart. The locations of these germline switch sites exist for C_α switching in sequences 5' to both the germline C_μ and C_α genes.

DNA Sequences Mediating Heavy Chain Switching

The DNA sequences involved in V(D)J joining are quite distinct from those implicated in C_H switching. The inverted repeat CAC$\frac{A}{T}$GTG occurs at the 3' end of antibody V gene segments and at the 5' end of the J gene segments (Sakano et al. 1979b;Early et al. 1980; Max et al. 1979). This inverted repeat is believed to be a recognition sequence that mediates the juxtaposition of the V, D, and J gene segments to allow subsequent joining by site-specific recombination (for a proposed mechanism, see Early et al. 1980). This inverted repeat is not found in the flanking regions surrounding any of the C_α switch sites. This sequence also is missing from the switch sites for a γ1 (Kataoka et al. 1980) and a γ2b gene (Takahashi et al. 1980; Sakano et al. 1980). Therefore, C_H switching and V-J joining employ distinct mechanisms for DNA rearrangement.

In an effort to determine what sequences are important in C_H switching, the sequences of the three C_μ sites and the three C_α sites were compared (Fig. 8). The germline C_μ MC101 and T15 switch sites share significant homology (15/25 nucleotides), although 16 nucleotides separate the actual switch points. We believe that these homologies are significant and may represent general sequence requirements for C_H switching adjacent to the C_μ gene. Neither the MC101 nor the T15 C_μ sites share any homology with that of M603. The switch site of a γ2b producer M141 (Takahashi et al. 1980; Sakano et al. 1980), also depicted in Figure 8, is nine nucleotides away

from that of T15 and may indicate that both γ2b and α switching can utilize the same recognition sequence adjacent to C_μ.

The sequences around each C_α switch site are even more highly conserved. Each germline S_α site occurs within a block of 30 conserved nucleotides (Fig. 8). Twenty-two bases are identical and seven of the remaining eight nucleotides are conserved with regard to type of base (i.e., purine-purine or pyrimidine-pyrimidine substitutions). The three points of recombination differ within each of these conserved sequences. Since the first three rearranged genes examined switched at distinct C_α sites, we reasoned that there must be additional S_α sites. We determined the DNA sequence of some 1400 nucleotides in the region of these germline S_α sites (Davis et al. 1980b) and found seventeen 30 nucleotide repeat sequences (Fig. 9a). These sequences are extremely homologous to each other (boxed regions) (Fig. 9b). Since these repeated sequences represent 510/1400 nucleotides in the region analyzed, the C_H switching into three of these sites does not appear to represent random DNA rearrangement. We suggest that most of these repeats are potential germline S_α sites. Multiple sites for C_H switching may be advantageous since they would improve the efficiency of switching enzymes, allowing for a relatively low level of expression of what could be potentially deleterious gene products (switching enzymes). This strategy would also apply to the enzymes mediating V-J joining where multiple V and J gene segments provide large numbers of recombination sites.

Evidence for Class-specific Regulation of C_H Switching

The data presented here, taken together with sequence data on γ2b (Takahashi et al. 1980; Sakano et al. 1980) and γ1 (Kataoka et al. 1980) switch sites from the literature which complement these experiments, suggest a mechanism by which class-specific regulation might be accomplished.

The germline sequences for C_H switching appear to be class specific with regard to $C_{\gamma 1}$, C_α and $C_{\gamma 2b}$ switch sites. The germline $C_{\gamma 1}$, C_α, and $C_{\gamma 2b}$ "switch" sequences are compared in Figure 10. The $_{\gamma 1}$ sequence is identical to the prototype S_α sequence in 10 out of 22 bases and the $_{\gamma 2b}$ sequence is identical for 7 of 22 nucleotides (only 3 of which are contiguous). Thus, various germline S_α sequences are far more similar to one another than to the germline $S_{\gamma 1}$ or $S_{\gamma 2b}$ sequences (Figs. 8 and 9b). One explanation for these sequence differences is that C_H switching is mediated by class-specific recognition sequences.

A Mechanism for C_H Switching

Since the germline S_α and S_μ sequences are not homologous (Fig. 8), homologous recombination cannot account for their joining. We believe the joining may be mediated by a number of distinct types of "switching" proteins (Fig. 11). For example, one switching protein (P_α) may bind the germline S_α sequence and a second (P_μ) may bind one of the germline S_μ sequences. These proteins may then interact to form a heterodimer which juxtaposes the V(D)J gene with the C_α gene segment (Fig. 11). The multiple germline S_α sequences would increase the probability that C_α switching could occur once the appropriate joining protein was expressed. Because the germline S_α, $S_{\gamma 1}$, and $S_{\gamma 2b}$ sequences seem distinct, different joining proteins could bind these sequences. Accordingly, the developmental regulation of the expression of these proteins would lead to class-specific regulation of C_H switching.

The Implications of Successive C_H Switching in MC101

The evidence for more "successive switching in MC101 indicates that two or more C_H switches can occur in a particular B-cell line. In the simple deletional model for class switching proposed by Honjo and Kataoka (1978), C_H switching progresses in a linear fashion, deleting intervening C_H genes at

each stage. As mentioned, the experiments supporting this model (Honjo and Kataoka 1978; Cory and Adams 1980; Coleclough et al. 1980; Rabbitts et al. 1980; Yaoita and Honjo 1980; Cory et al. 1980) indicate a gene order of C_μ-$C_{\gamma 3}$-$C_{\gamma 1}$-$C_{\gamma 2b}$-$C_{\gamma 2a}$-C_α. Paradoxically, however, MC101 appears to have switched from C_μ to C_α and then from C_α to $C_{\gamma 1}$, contrary to the simple deletional model for class switching proposed by Honjo and Kataoka (1978). Several explanations seem plausible:

1) The C_H gene order is C_μ-C_α-C_γ. This seems difficult to support because of the large number of myeloma tumors which express the $C_{\gamma 1}$ gene, and still contain C_α genes (Honjo and Kataoka 1978; Cory and Adams 1980; Coleclough et al. 1980; Rabbitts et al 1980; Yaoita and Honjo 1980; Cory et al. 1980). In contrast, those myeloma tumors which express the C_α gene generally appear to have deleted their $C_{\gamma 1}$ genes (Cory and Adams 1980; Coleclough et al. 1980; Rabbitts et al. 1980; Yaoita and Honjo 1980; Cory et al. 1980). However, until these genes are ordered in the germline DNA, this explanation remains a formal possibility.

2) Interchromosomal recombination. In this scheme, $C_\mu \rightarrow C_\alpha$ rearrangement on one chromosome could be followed by recombination with a $C_{\gamma 1}$ gene on another chromosome to produce the MC101 $\gamma 1$ mosaic gene (V_H-C_μ-C_α-$C_{\gamma 1}$). Moreover, the reciprocally rearranged chromosome should have a rearranged C_α gene.

3) Episomal deletion. If the deleted DNA between the V_H gene and the C_α gene formed a circular intermediate (an episome) that was at least transiently stable, this episome could then re-integrate into the chromosome, replacing the C_α gene with the $C_{\gamma 1}$ gene. This model is easily testable since it predicts the presence of a C_μ gene, a $C_{\gamma 3}$ gene, etc. in the MC101 genome and, in addition, the C_μ and C_α genes should be rearranged.

The rearranged MC101 γ1 gene, accordingly, raises two general possibilities with regard to C_H switching and normal B-cell differentiation. i) The successive (and complex) types of C_H switching observed here may arise from one or more aberrant chromosomal rearrangements that are unique to myeloma cell lines and that will not generally be seen in normal B cells. The numerous cell divisions that occur between myeloma tumor production and our analyses of the corresponding DNAs, as well as the aneuploid nature of myeloma cells, make this a serious possibility. ii) Normal B cells may employ multiple C_H switching mechanisms—some perhaps different from any cited above.

The developmental regulation of C_H switching may operate at several different levels. i) The nature of the sequences mediating V(D)J joining and C_H switching implies that these phenomena are regulated independently. ii) The existence of distinct switch sequences for α, γ1, and γ2b genes implies that the expression of these classes may be developmentally regulated at the level of DNA rearrangement, depending, for example, on which specific switching protein is expressed. iii) The evidence for a $C_\mu \rightarrow C_\alpha \rightarrow C_{\gamma 1}$ "successive" switch (MC101) presents difficulties for simple deletional models of C_H switching.

Abortive Rearrangement Suggests that C_H Switching Enzymes May Generally Operate on all Heavy Chain Chromosomes

Myeloma cells are generally subtetraploid and may have between one and four or more heavy chain chromosomes. Often a particular expressed heavy chain gene from an individual myeloma may be rearranged on each of the heavy chain chromosomes. For example, the IgA-producing M603 myeloma tumor has three distinct α genes and each has been rearranged from its germline configuration (Davis et al. 1979). We have discussed the

actively expressed M603 α gene above. The other two α genes contain no M603 V_H genes and are presumably not functional. These M603 α genes are said to be "abortively rearranged." Similar observations of abortive rearrangement of light chain genes have been made by others (Steinmetz and Zachau 1980).

One puzzling aspect of Southern blotting data on various myeloma tumors is that in many cases all of the C_H genes 5' to the expressed C_H gene have been deleted (Cory et al. 1980). These DNA deletions are not random rearrangements in that one would not expect them to terminate at or near the expressed C_H gene. We believe that these data independently support the existence of class-specific switching enzymes which, once induced, delete the C_H genes 5' to the expressed C_H gene on one or both heavy-chain chromosomes. One prediction of this interpretation is that abortive rearrangement of C_H genes will involve the same DNA switching sequences as C_H switching. In order to test this prediction, we sequenced one abortively rearranged M603 C_α clone. This clone, ChM603α30, and its homology to a germline C_α gene are illustrated in Figure 12a. There is no V_H M603 gene segment present in the α30 clone, hence it appears to be an "abortive rearrangement." DNA sequence analysis indicates that the α30 clone diverges from the germline α flanking sequence at one of the specific S_α switch sequences (Fig. 12b). Indeed, the α30 clone diverges two nucleotides 3' to the switch site used by the T15 gene (Fig. 8) and well within the 30 nucleotide homology region characteristic of C_α switching.

The fact that abortive rearrangement of the C_α gene in M603 occurs at an S_α switch site strongly suggests that abortive rearrangements do not occur randomly during B-cell development, but are in fact mediated by a mechanism similar or identical to that responsible for C_H switching. Furthermore, this

result adds weight to the suggestion that C_H switching is mediated by class-specific proteins which recognize sequences 5' to germline C_H gene segments.

Acknowledgments

This work was supported by grants from the NIH, NSF, and USPHS. M.D. and S.K. are supported by an NIH training grant. All experiments involving recombinant organisms were conducted in accordance with the NIH Guidelines on recombinant DNA. Portions of this review were adapted from an article in Science (Davis et al. 1980b).

REFERENCES

Benton, W. D. and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196: 180.

Bernard, O. and N. Gough. 1980. Nucleotide sequence of immunoglobulin heavy chain joining segments between translocated V_H and μ constant region genes. Proc. Nat. Acad. Sci. 77: 3630.

Bernard, O., N. Hozumi and S. Tonegawa. 1978. Sequences of mouse immunoglobulin light chain genes before and after somatic changes. Cell 15: 1133.

Coleclough, C., C. Cooper and R. P. Perry. 1980. Rearrangement of immunoglobulin heavy chain genes during B-lymphocyte development as revealed by studies of mouse plasma cells. Proc. Nat. Acad. Sci. 77: 1422.

Cooper, M. D., J. F. Kearney, P. M. Lydyard, C. E. Grossi and A. R. Lawton. 1976. Studies of generation of B-cell diversity in mouse, man, and chicken. Cold Spring Harbor Symp. Quant. Biol. 41: 139.

Cory, S. and J. M. Adams. 1980. Deletions are associated with somatic rearrangement of immunoglobulin heavy chain genes. Cell 19: 37.

Cory, S., J. Jackson and J. M. Adams. 1980. Deletions within the constant region locus can account for switches in immunoglobulin heavy chain expression. Nature (in press).

Davis, M., P. Early, K. Calame and L. Hood. 1979. The organization and rearrangement of heavy chain immunoglobulin genes. In Eukaryotic Gene Regulation (ed. R. Axel et al.). ICN-UCLA Symp., 393, Academic Press, New York.

Davis, M. M., K. Calame, P. W. Early, D. L. Livant, R. Joho, I. L. Weissman and L. Hood. 1980a. An immunoglobulin heavy chain gene is formed by at least two recombinational events. Nature 283: 733.

Davis, M. M., S. K. Kim and L. Hood. 1980b. Sequences mediating heavy chain switching in alpha immunoglobulin genes. Science (in press).

Early, P., H. Huang, M. Davis, K. Calame and L. Hood. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H. Cell 19: 981.

Early, P. W., M. M. Davis, D. B. Kaback, N. Davidson and L. Hood. 1979. Immunoglobulin heavy chain gene organization in mice: analysis of a myeloma genomic clone containing variable and α constant regions. Proc. Nat. Acad. Sci. 76: 857.

Honjo, T. and T. Kataoka. 1978. Organization of immunoglobulin heavy chain genes and allelic deletion model. Proc. Nat. Acad. Sci. 75: 2140.

Johnson, N., S. Rudikoff, P. Barstad and L. Hood. 1980. Sequences of phosphorylcholine binding variable regions (in preparation).

Kabat, E. A., T. T. Wu and H. Bilofsky. 1979. Sequences of immunoglobulin chains. NIH publication no. 80-2008.

Kataoka, T., T. Kawakami, N. Takahashi and T. Honjo. 1980. Rearrangement of immunoglobulin γ1-chain gene and mechanism for heavy-chain class switch. Proc. Nat. Acad. Sci. 77: 919.

Katsukake, K. and T. Ino. 1980. A trans-acting factor mediates inversion of a specific DNA segment in flagellar phase variation of Salmonella. Nature 284: 479.

Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim and A. Efstratiadis. 1978. The isolation of structural genes from libraries of eucaryotic DNA. Cell 15: 687.

Max, E. E., J. G. Sediman and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region gene. Proc. Nat. Acad. Sci. 76: 3450.

Maxam, A. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. In Methods in Enzymology, Nucleic Acids, Part I (Ed. L. Grossman and K. Moldave), vol. 65, p. 499. Academic Press, New York.

Pernis, B., L. Forni and A. L. Luzzati. 1976. Synthesis ofmultiple immunoglobulin classes by single lymphocytes. Cold Spring Harbor Symp. Quant. Biol. 41: 175.

Rabbitts, T. H., A. Forster, W. Dunnick and D. L. Bentley. 1980. The role of gene deletion in the immunoglobulin heavy chain switch. Nature 283: 351.

Sakano, H., J. H. Rogers, K. Hüppi, C. Brack, A. Traunecker, R. Maki and S. Tonegawa. 1979a. Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA segments. Nature 277: 627.

Sakano, H., K. Hüppi, G. Heinrich and S. Tonegawa. 1979b. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. Nature 280: 288.

Sakano, H., A. Maki, Y. Kurosawa, W. Roeder and S. Tonegawa. 1980. The two types of somatic recombination necessary for generation of complete immunoglobulin heavy chain genes. Nature (in press).

Schilling, J., B. Clevinger, J. M. Davie and L. Hood. 1980. Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V-region gene segments. Nature 283: 35.

Seidman, J. G., A. Leder, M. Nau, B. Norman and P. Leder. 1978. Antibody diversity. Science 202: 11.

Silverman, M., J. Zieg, G. Mandel and M. Simon. 1980. Analysis of the functional components of the phase variation system. Cold Spring Harbor Symp. Quant. Biol. (in press).

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503.

Steinmetz, M. and H. G. Zachau. 1980. Two rearranged immunoglobulin kappa light chain genes in one mouse myeloma. Nucleic Acids Res. 8: 1693.

Takahashi, N., T. Kataoka and T. Honjo. 1980. Nucleotide sequences around class switch recombination site of the immunoglobulin γ2b chain gene of mouse. Gene (in press).

Tonegawa, S., A. M. Maxam, R. Tizard, O. Bernard and W. Gilbert. 1978. Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. Proc. Nat. Acad. Sci. 75: 1485.

Weigert, M., R. Perry, D. Kelley, T. Hunkapiller, J. Schilling and L. Hood. 1980. The joining of V and J gene segments creates antibody diversity. Nature 283: 497.

Yaoita, Y. and T. Honjo. 1980. Deletion of immunoglobulin heavy chain genes accompanies the class switch rearrangement. Biomed. Res. (in press).

FIGURE LEGENDS

Figure 1. Organization of immunoglobulin genes in BALB/c mice. Each gene family is located on a separate chromosome. The lambda gene family contains at least two units (λ_I and λ_{II}) organized as shown. The probable order of the heavy chain constant region genes is shown beneath the diagram of variable region gene segments. Each C_H gene contains multiple exons, as depicted for C_μ (Early et al. 1979; Sakano et al. 1979a; Early et al. 1980b). Slash marks indicate linkage relationships which have not been firmly established. Distances are not to scale.

Figure 2. B-cell differentiation. Cells initially committed to immunoglobulin synthesis first express variable regions on membrane-bound IgM (and IgD in many cases). After stimulation by antigen, these cells or their progeny can secrete immunoglobulins with the same variable region, but any one of the eight classes or subclasses of heavy chain.

Figure 3. The upper diagram depicts possible germline D gene segments. Relative distances of the various gene segments from one another are undetermined. The short arrows indicate conserved noncoding sequences (Table 1) which may be involved in DNA rearrangement. In this model, DNA rearrangement joins V_H-D-J_H gene segments. Intervening DNA may be deleted, or could undergo other types of rearrangement. The lower diagram shows paired V_H and D (alternatively D and J_H, or V_L and J_L) gene segments with 11 and 22 nucleotide spacers. Putative DNA-joining proteins might bind to the areas enclosed by dashed lines. The gene segments are represented as colinear to emphasize the symmetry of the conserved noncoding nucleotides. The actual structure may bring the ends of the two gene segments into close proximity.

Figure 4. Heavy chain junctional diversity. The genes whose partial sequences are shown all are joined to J_{H1} (J_{H107} in Early et al. 1980). The D segments of these genes are homologous, although not identical. The three myeloma genes encode phosphorylcholine-binding V_H regions of subgroup III, while the two rearranged genes from normal IgM-producing B cells isolated by fluorescence activated cell sorting are subgroup I (V_{H49}) and subgroup II (V_{H2}) (Kabat et al. 1979; P. Early and C. Nottenberg, unpublished observations). The upper panel shows variations in D segment sequences and lengths, while the lower panel shows variations in the points of D-J_{H1} joining.

Figure 5. Alternative possible reading frames for a D gene segment. The protein sequence of the S107 D segment is shown in italics (Early et al. 1980). V(D)J joining at the sites indicated by dotted lines can generate the A4 protein sequence from a germline A4-like V_H gene (S. Crews, unpublished observation), the D_{107} gene segment, and J_{H3} gene segment.

Figure 6. Clones containing rearranged alpha heavy chain genes and germline C_μ and C_α genes. The rearranged T15 and M603 genes were isolated from genomic libraries of their respective myeloma tumors, and MC101 was isolated as described (Kataoka et al. 1980). Clones containing the germline C_μ and C_α genes were isolated from libraries of mouse sperm DNA (Davis et al. 1980a). Raised boxes denote coding regions. The shaded areas indicate homologies with germline genes and their flanking sequences: vertical lines, V_H and flanking region; solid, C_μ and flanking region; dots, C_α and flanking regions; and white, uncertain origin, probably C_μ derived (see text); diagonal lines, $C_{\gamma 1}$ and flanking regions. Homologies were determined by detailed restriction mapping. Restriction enzyme sites are denoted as follows: Hf = HinfI, Hh = Hha, H = HindIII, M = MspI, R = EcoRI, Rs = RsaI, S = Sau3a, and Mb = MboII. The position of the J_H regions was determined from experiments

described in Early et al. (1980) and Sakano et al. (1980). The M603 clone was isolated and characterized as previously described (Davis et al. 1980a; Early et al. 1979; Davis et al. 1979). The T15 library was made from EcoRI partially digested DNA by procedures described in Davis et al. (1980a) and Maniatis et al. (1978). Recombinant phage were screened by the procedure of Benton and Davis (1977) using ^{32}P nick translated V_H + C_α cDNA clones as probes. Southern blot analysis (Southern 1975 of EcoRI digested myeloma DNA indicated identity of central EcoRI fragments in the T15 genomes with the clones shown here (data not shown), indicating that the clones isolated are representative of their genome of origin. This analysis has been previously reported for M603 (Davis et al. 1979). The DNA sequence analysis employed the procedures of Maxam and Gilbert 1980. Arrows originating from a restriction site indicate 5' end labeling with $\gamma^{32}P$-ATP, while arrows ending in a restricting site indicate 3' end labeling with $\alpha^{32}P$-dNTPs.

Figure 7. Switch sequences for the rearranged T15, M603, and MC101 genes and their germline C_μ and C_α counterparts. $C_{\gamma 1}$ corresponding sequences for MC101 have been reported by Honjo and colleagues. Sequences surrounding C_H switch sites are shown in 5' → 3' orientation. Underlining indicates sequence identity of unrearranged C_μ (open blocks) and C_α (solid blocks) flanking sequences with their counterparts in the individual rearranged genes. Dots indicate 10 nucleotide spacings.

Figure 8. A comparison of the DNA sequences of the germline C_μ and C_α switch sites for the rearranged T15, M603, and MC101 genes. Sequences obtained from C_μ and C_α flanking regions were aligned for maximum homology around their C_H switch sites. Boxes indicate nucleotide identities and - indicate a gap introduced for homology alignment. Arrows indicate breakpoints for the switch sites of each rearranged gene. Also indicated is the C_μ switch

site for a γ2b producer, M141 (Takahashi et al. 1980; Sakano et al. 1980). A consensus sequence for at least some cases of C_μ switching might therefore be GGTNATTANNNNNNGGTANNCAAAG which does not occur elsewhere in any of the some 1900 nucleotides of C_μ flanking region sequenced (Sakano et al. 1980). Elements of this C_μ homology region have been suggested previously to play a role in C_H switching (Takahashi et al. 1980; Sakano et al. 1980). A consensus sequence for C_α switching derived from the examples here would be PGTCPPGCTGGAATPPGYTGGGNTGPGCTG. (P = purine, Y = pyrimidine, N = any nucleotide).

Figure 9a. Schematic diagram of the location of seventeen 30-nucleotide repeated sequences adjacent to the germline C_α gene. Arrows indicate which repeats are used in the three rearranged genes described in this paper.

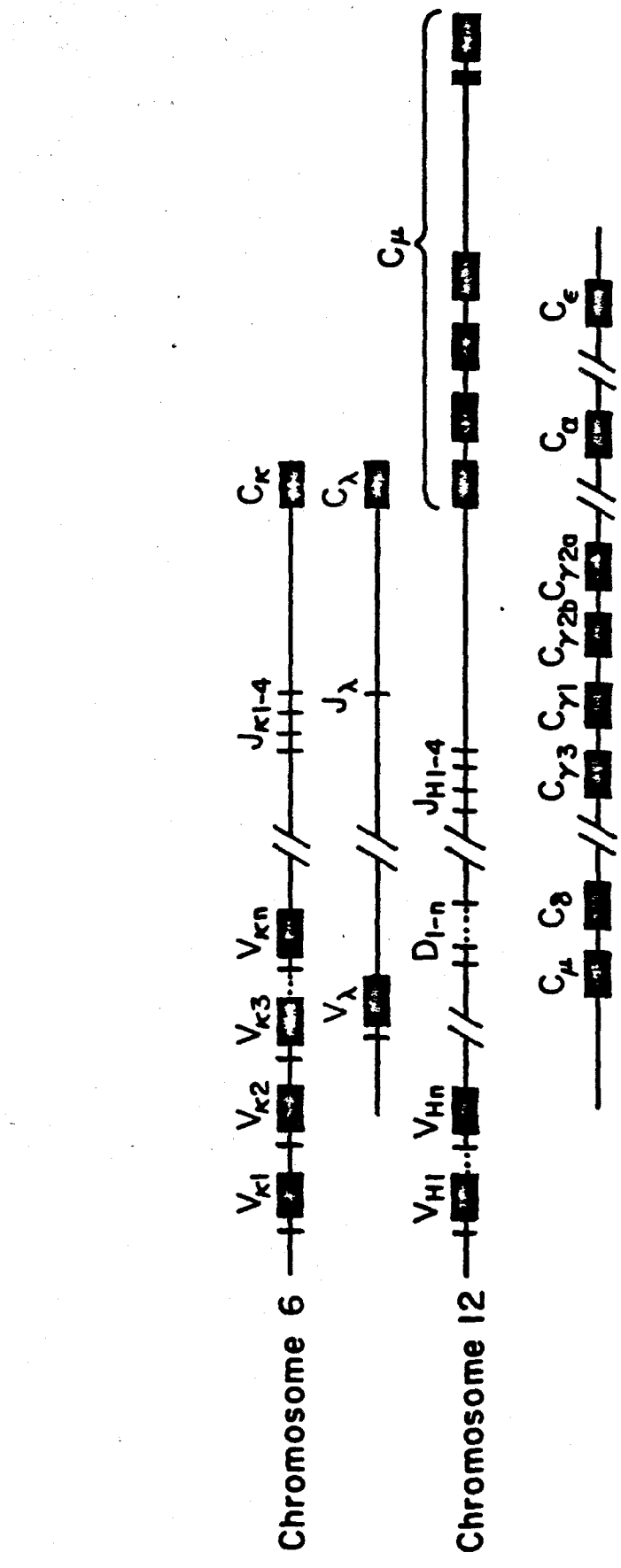
9b. DNA sequences of 17 repeats found interspersed in germline 5' flanking C_α sequence. Boxes indicate nucleotide identities and - indicates a gap.

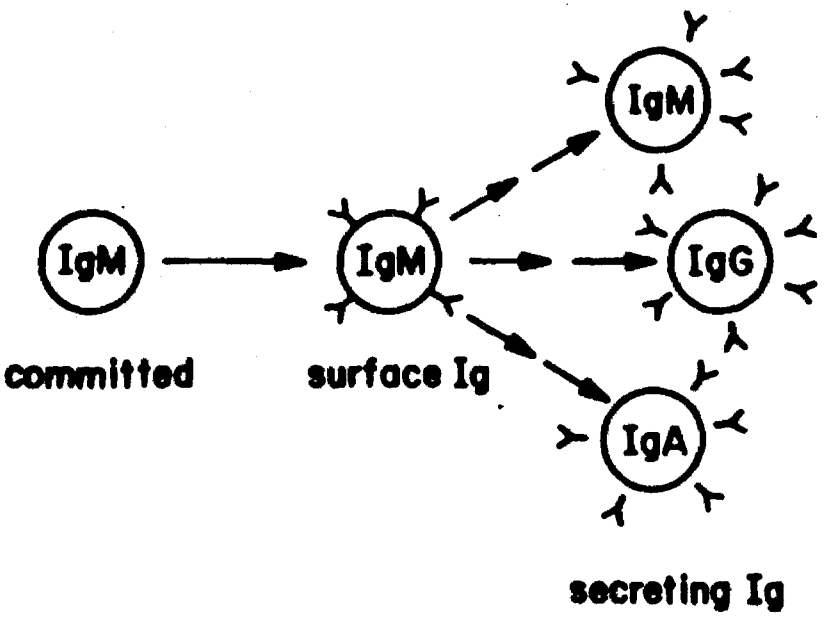
Figure 10. A comparison of the switch sequences from germline 5' flanking sequences for the C_α (see Fig. 8), $C_{\gamma 1}$ (Kataoka et al. 1980), and γ2b (Takahashi et al. 1980; Sakano et al. 1980) genes. Boxes in the germline α sequence denote the conserved bases of the 30-nucleotide recognition sequences. Boxes in the germline γ1 and γ2b sequences denote nucleotide identities to the germline C_α sequence. - indicates a gap.

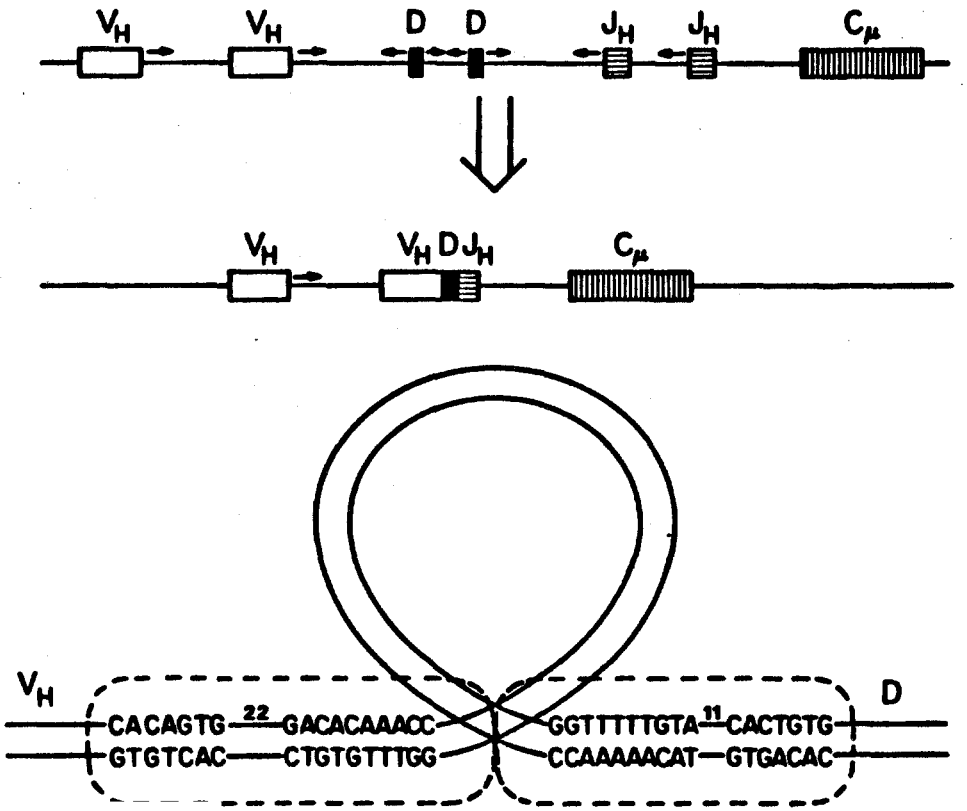
Figure 11. Model for class-specific regulation of C_H switching (see text). In this scheme the small boxes represent recognition (S) sequences which bind to switch proteins to mediate C_H switching; the circle represents switching proteins and the large boxes represent coding regions. This model depicts a C_H gene order of C_μ, C_γ, C_α (see text) but this is not a requirement.

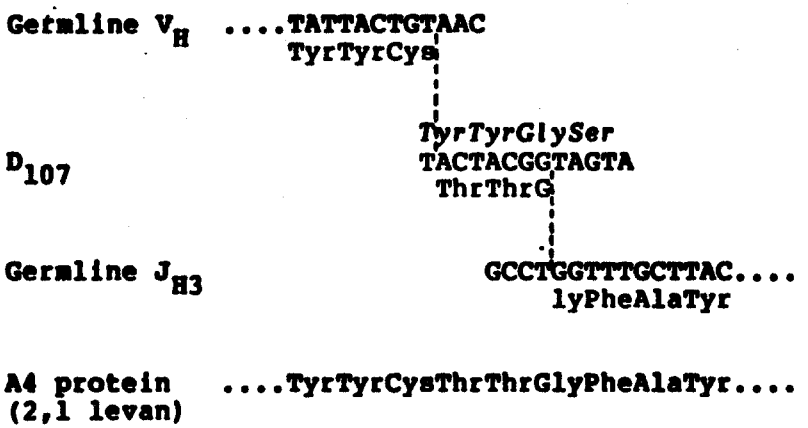
Figure 12. An example of abortive rearrangement in a C_α gene isolated from the myeloma tumor M603. The clone CH603α30 is identical to DNA flanking the germline C_α gene up to approximately 2 kb from the 5' end of the C_α coding region (A). DNA sequence analysis (B) shows that the breakpoint of the recombination occurs within one of the 30-nucleotide repeat sequences (#17 in Fig. 9) associated with heavy chain switching.

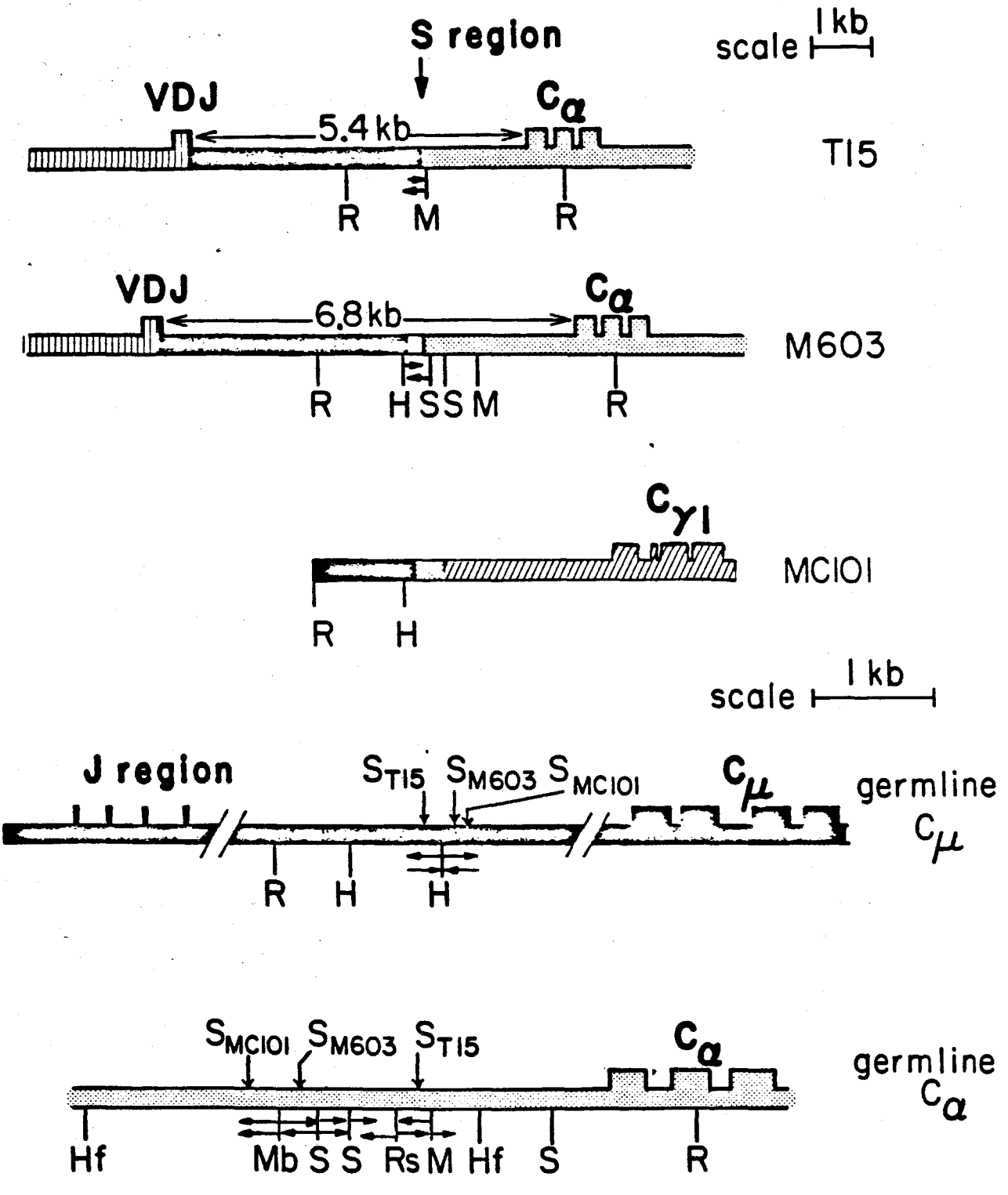
Table 1. Comparison of recognition nucleotides adjacent to V and J gene segments. κ sequences are from Max et al. (1979) and Sakano et al. (1979b; λ from Bernard et al. (1978) and Tonegawa et al. (1978); heavy chains from Early et al. (1980) and Bernard and Gough (1980). The mRNA-sense strand is shown 3' to V gene segments and 5' to J gene segments. The asterisk indicates a possibly nonfunctional J_κ gene segment; it has not been found in a rearranged gene. The column to the right lists the number of nucleotides between the underlined hepta- and decanucleotides.

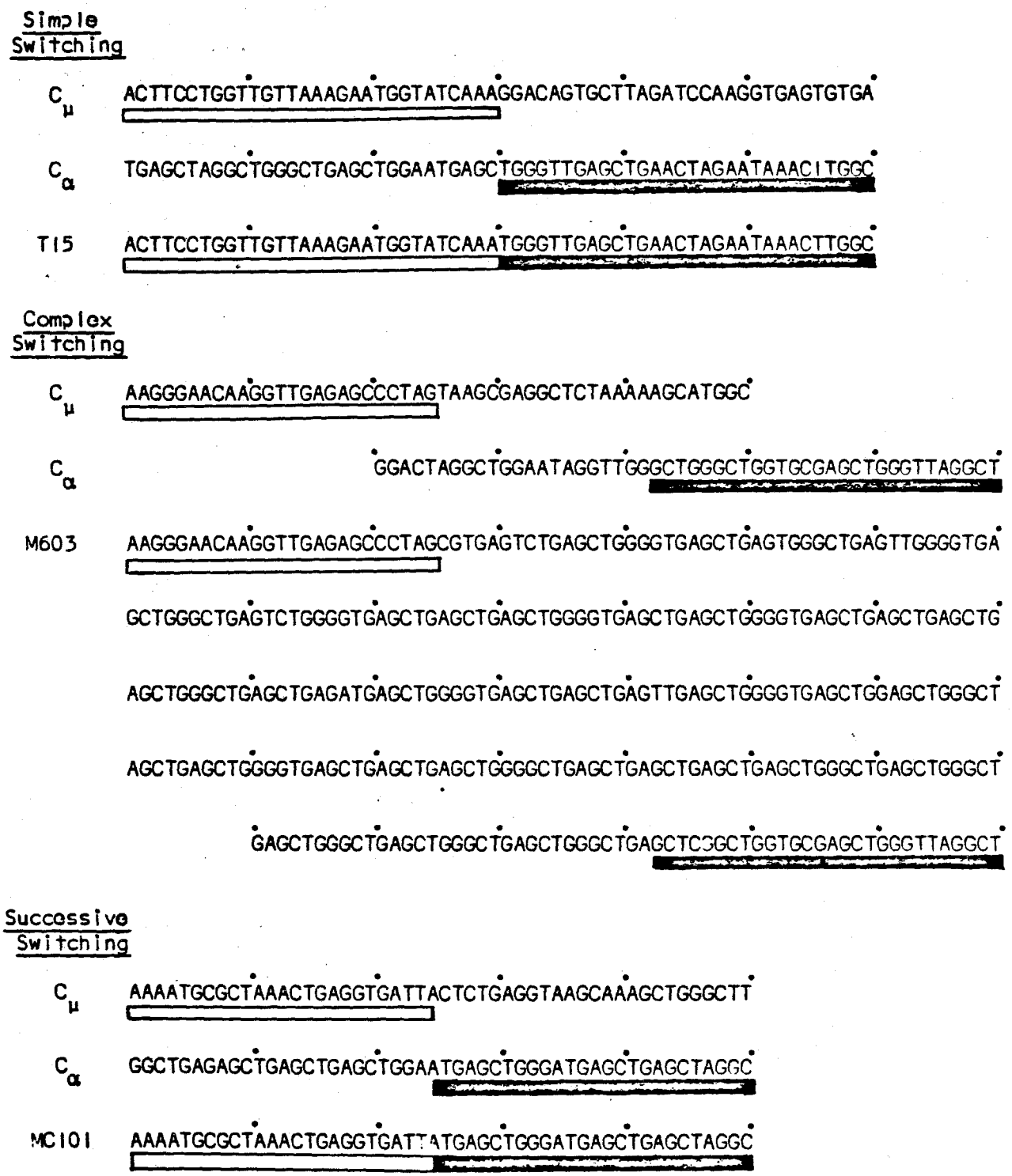

Chromosome 6
$V_{\kappa 1}$
$V_{\kappa 2}$
$V_{\kappa 3}$
$V_{\kappa n}$
$J_{\kappa 1-4}$
C_κ
V_λ
J_λ
C_λ
Chromosome 12
V_{H1}
V_{Hn}
D_{1-n}
J_{H1-4}
C_μ
C_δ
$C_{\gamma 3}$
$C_{\gamma 1}$
$C_{\gamma 2b}$
$C_{\gamma 2a}$
C_α
C_ε


IgM
IgM
IgM
IgG
IgA
committed
surface Ig
secreting Ig

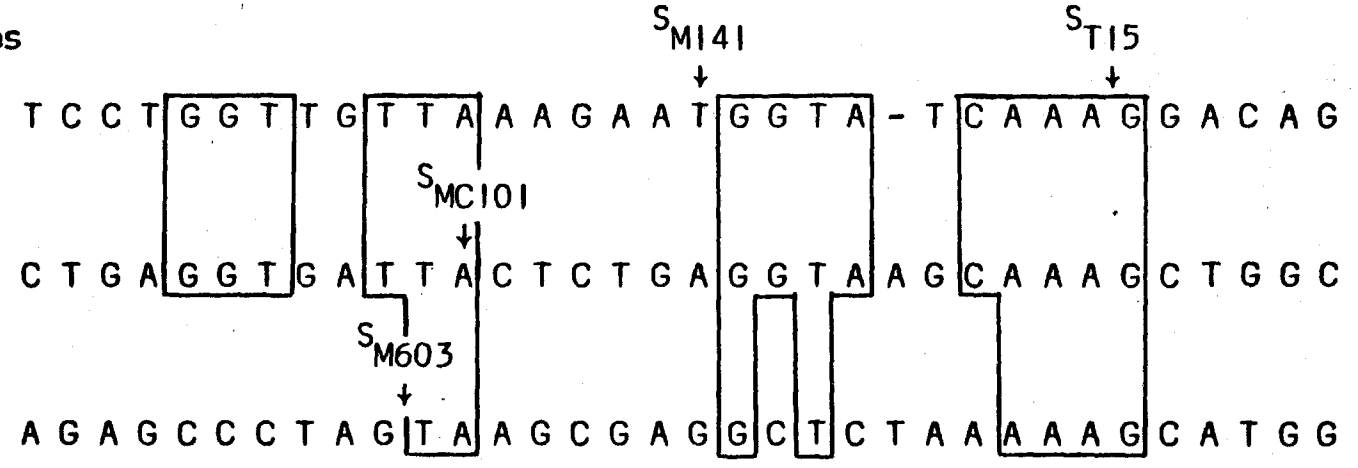

V_H
V_H
D
D
J_H
J_H
C_μ
V_H
V_H D J_H
C_μ
V_H
D
CACAGTG
22
GACACAAACC
GGTTTTTGTA
11
CACTGTG
GTGTCAC
CTGTGTTTGG
CCAAAAACAT
GTGACAC


Germline V_H ....TATTACTGTAAC
TyrTyrCys
TyrTyrGlySer
D_{107}
TACTACGGTAGTA
ThrThrG
Germline J_{H3}
GCCTGGTTTGCTTAC....
lyPheAlaTyr
A4 protein ....TyrTyrCysThrThrGlyPheAlaTyr....
(2,1 levan)

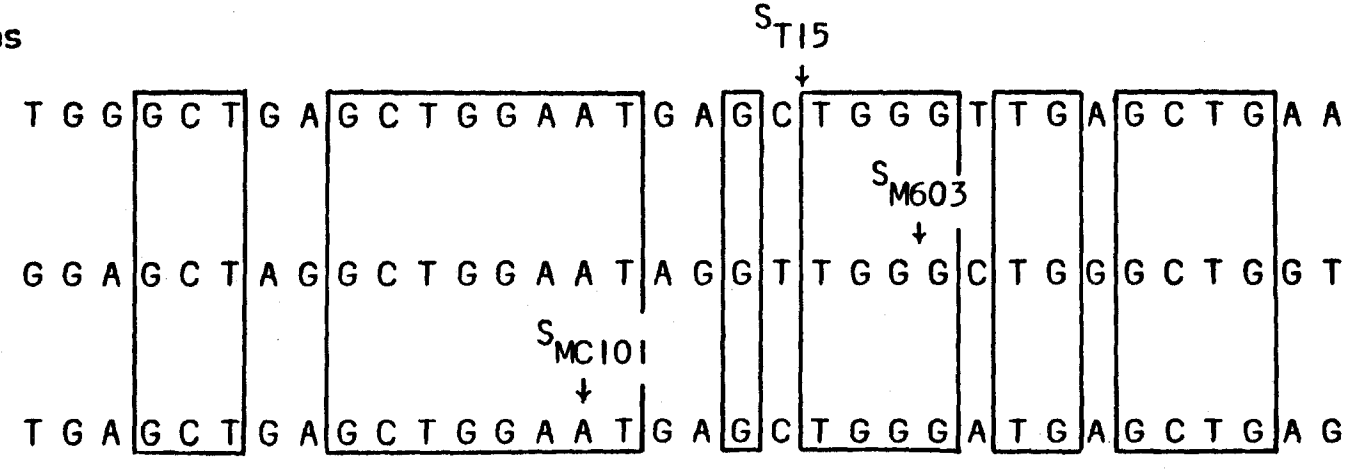

S region
scale 1 kb
VDJ
Cα
5.4 kb
T15
R
M
R
VDJ
Cα
6.8 kb
M603
R
H S S M
R
Cγ1
MC101
R
H
scale 1 kb
J region
S T15
S M603
S MC101
Cμ
germline Cμ
R
H
H
S MC101
S M603
S T15
Cα
germline Cα
Hf
Mb S S Rs M Hf
S
R

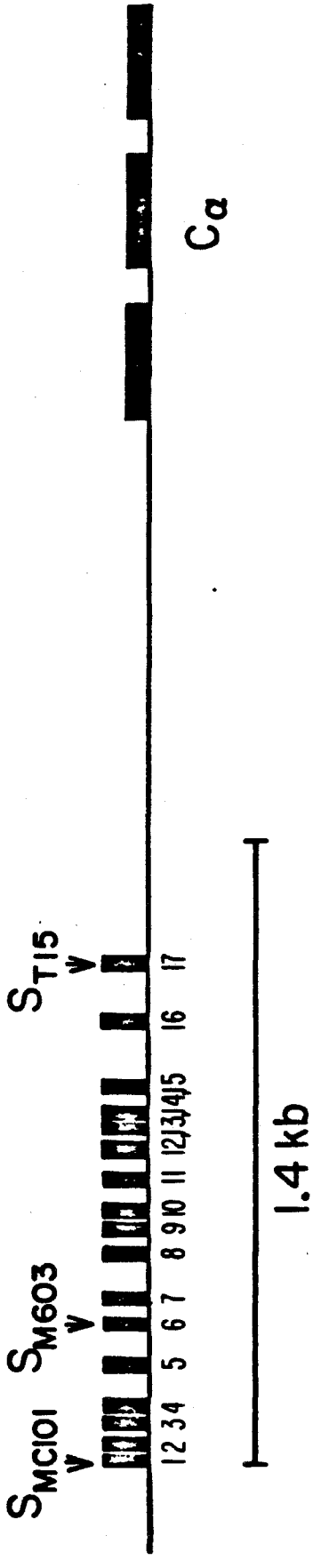

Simple Switching
C_μ ACTTCCTGGTTGTTAAAGAATGGTATCAAAGGACAGTGCTTAGATCCAAGGTGAGTGTGA
C_α TGAGCTAGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGAATAAACTTGGC
T15 ACTTCCTGGTTGTTAAAGAATGGTATCAAATGGGTTGAGCTGAACTAGAATAAACTTGGC
Complex Switching
C_μ AAGGGAACAAGGTTGAGAGCCCTAGTAAGCGAGGCTCTAAAAAGCATGGC
C_α GGACTAGGCTGGAATAGGTTGGGCTGGGCTGGTGCGAGCTGGGTTAGGCT
M603 AAGGGAACAAGGTTGAGAGCCCTAGCGTGAGTCTGAGCTGGGGTGAGCTGAGTGGGCTGAGTTGGGGTGA
GCTGGGCTGAGTCTGGGGTGAGCTGAGCTGAGCTGGGGTGAGCTGAGCTGGGGTGAGCTGAGCTGAGCTG
AGCTGGGCTGAGCTGAGATGAGCTGGGGTGAGCTGAGCTGAGTTGAGCTGGGGTGAGCTGGAGCTGGGCT
AGCTGAGCTGGGGTGAGCTGAGCTGAGCTGGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGAGCTGGGCT
GAGCTGGGCTGAGCTGGGCTGAGCTGGGCTGAGCTCGGCTGGTGCGAGCTGGGTTAGGCT
Successive Switching
C_μ AAAATGCGCTAAACTGAGGTGATTACTCTGAGGTAAGCAAAGCTGGGCTT
C_α GGCTGAGAGCTGAGCTGAGCTGGAATGAGCTGGGATGAGCTGAGCTAGGC
MC101 AAAATGCGCTAAACTGAGGTGATTATGAGCTGGGATGAGCTGAGCTAGGC

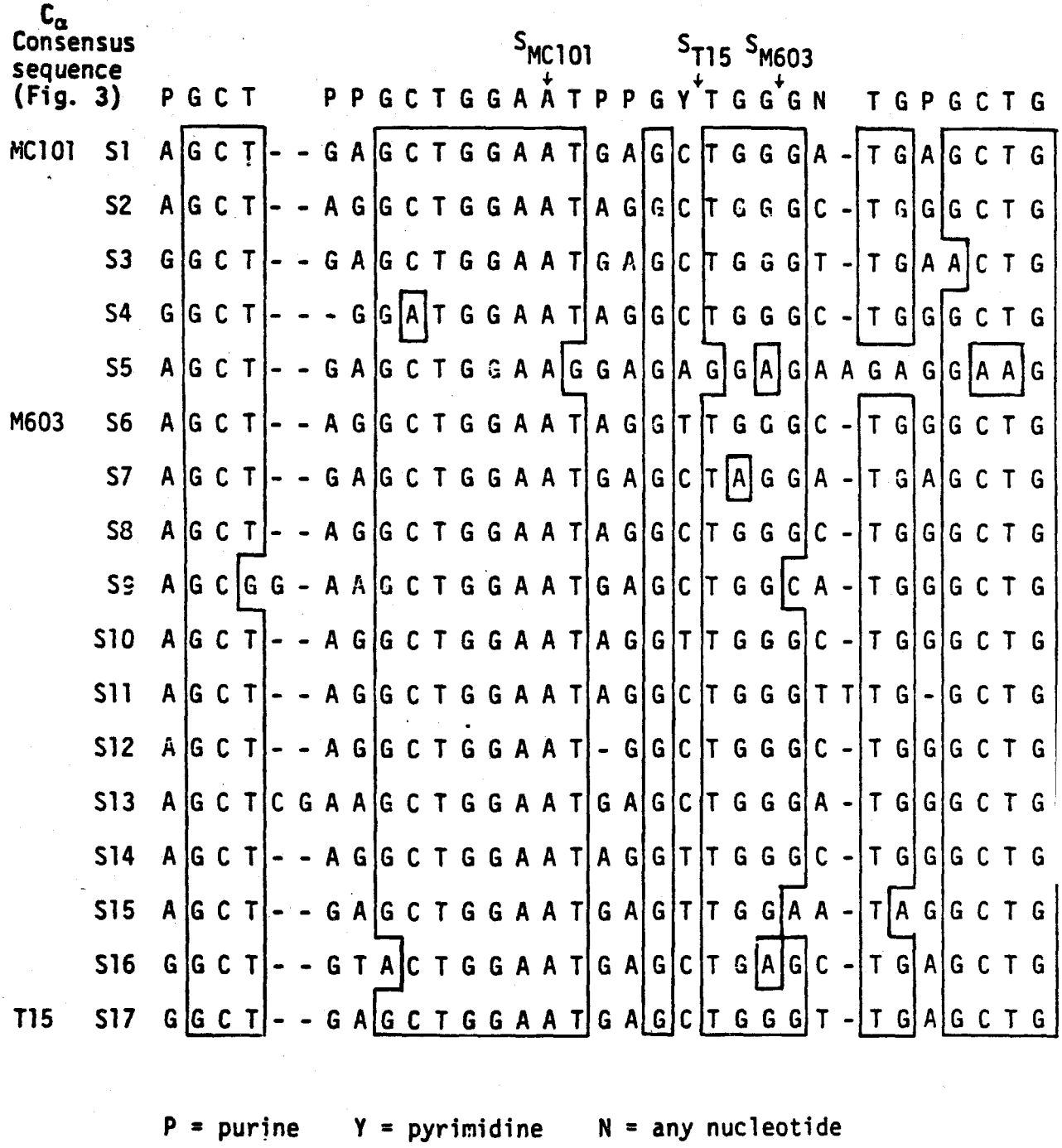
Germline
C_μ
Switch
Sequences



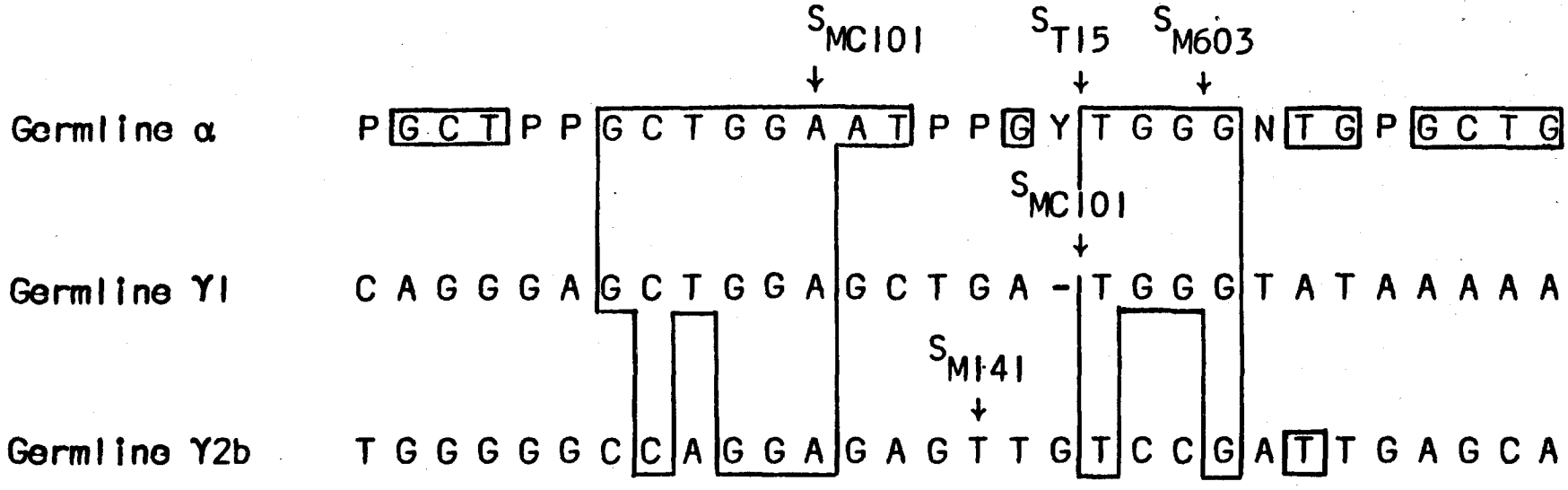
Germline
C_α
Switch
Sequences

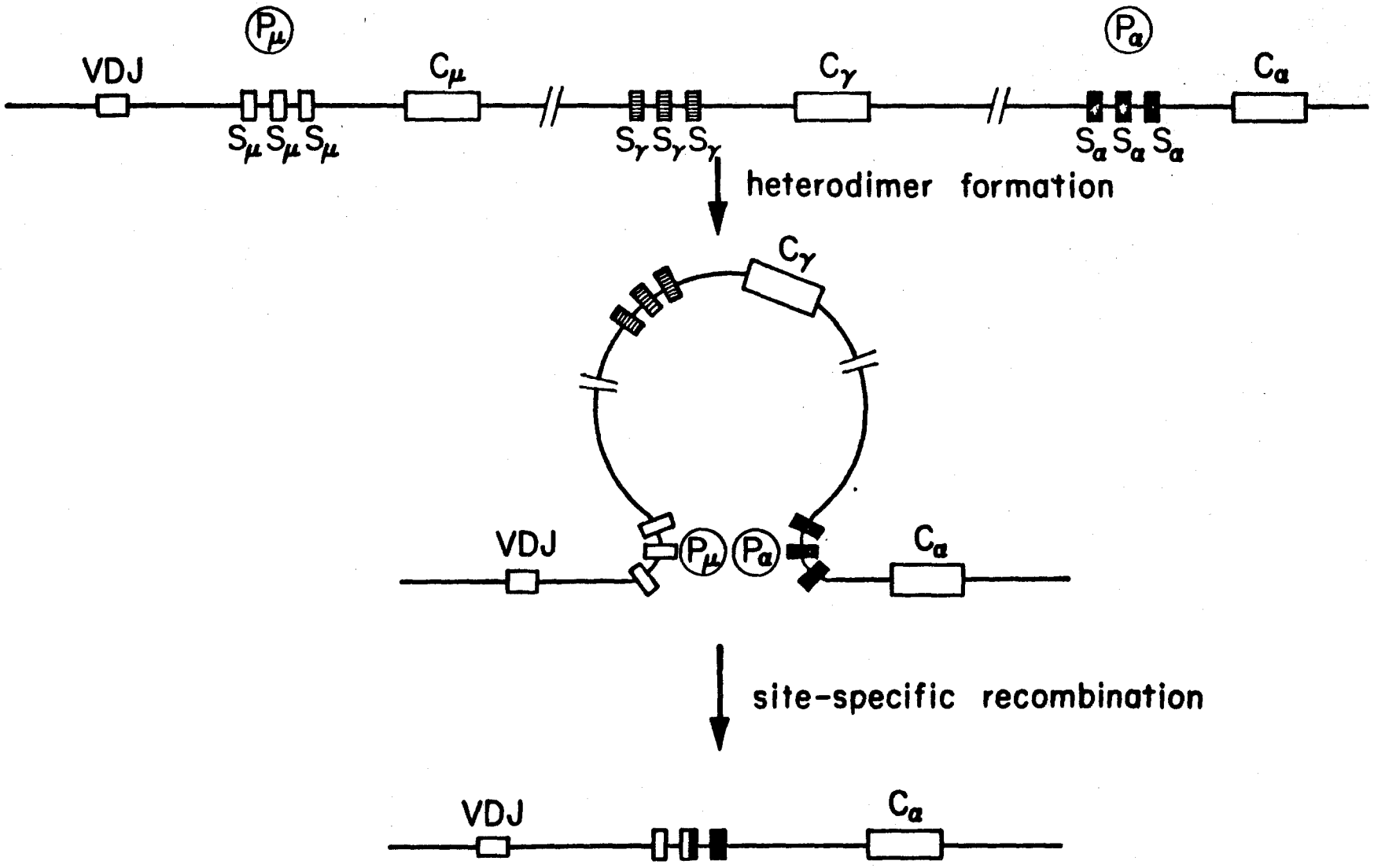


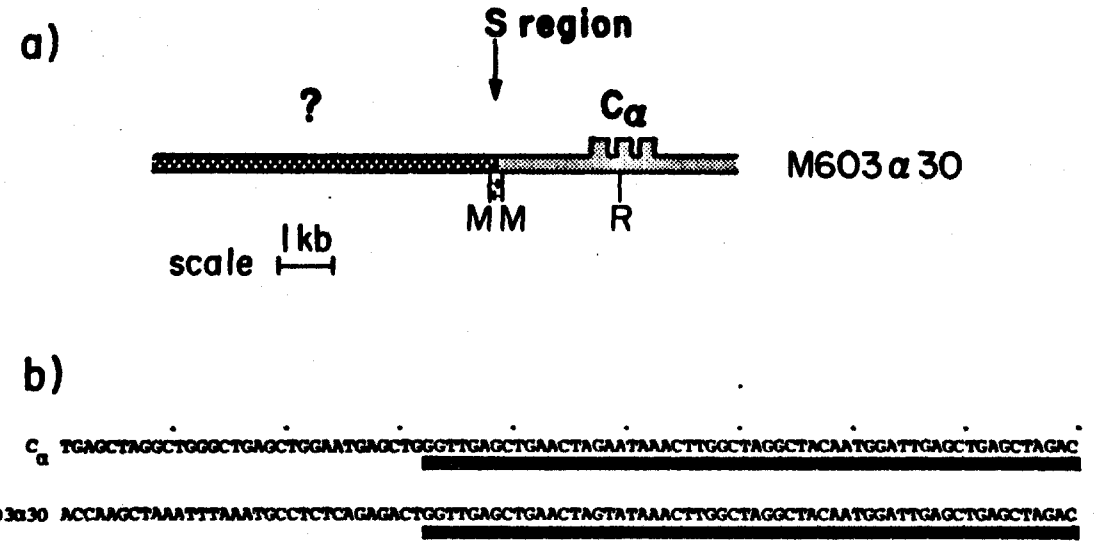

S_{MC101}
S_{M603}
S_{T15}
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
C_α
1.4 kb



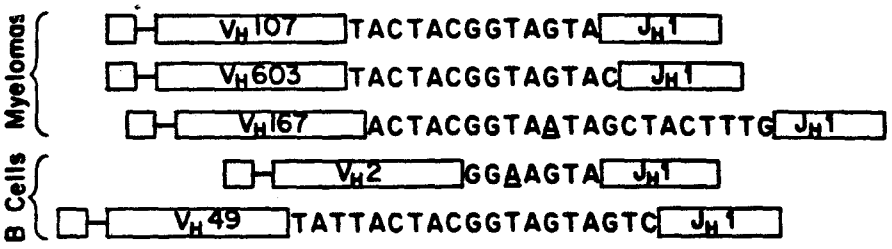
P = purine Y = pyrimidine N = any nucleotide

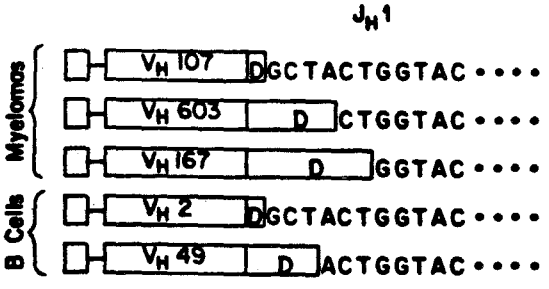

S_{MC101}
S_{T15}
S_{M603}
Germline α
P G C T P P G C T G G A A T P P G Y T G G G N T G P G C T G
S_{MC101}
Germline Y1
C A G G G A G C T G G A G C T G A - T G G G T A T A A A A A
S_{M141}
Germline Y2b
T G G G G G C C A G G A G A G T T G T C C G A T T G A G C A


VDJ
P_μ
C_μ
C_γ
P_α
C_α
S_μ S_μ S_μ
S_γ S_γ S_γ
S_α S_α S_α
heterodimer formation
C_γ
VDJ
P_μ
P_α
C_α
site-specific recombination
VDJ
C_α


a)
S region
?
Cα
M603α30
M M
R
scale
1kb
b)
Cα TGAGCTAGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGAATAAACTTGGCTAGGCTACAATGGATTGAGCTGAGCTAGAC
M603α30 ACCAAGCTAAATTTAAATGCCTCTCAGAGACTGGTTGAGCTGAACTAGTATAAACTTGGCTAGGCTACAATGGATTGAGCTGAGCTAGAC

V_{K41}	CACAGTGATACAAATCATAACATAAACC	(11)
V_{K2}	CACAGTGATTCAAGCCATGACATAAACC	(11)
V_{K3}	CACAGTGATTCAAGCCATGACATAAACC	(11)
V_{K21}	CACAGTGCTCAGGGCTGAACAAAAACC	(10)
V_{H107}	CACAGTGAGAGGACGTCATTGTGAGCCCAGACACAAACC	(22)
$V_{\lambda I}$	CACAATGACATGTGTAGATGGGGAAGTAGATCAAGAACA	(22)
$V_{\lambda II}$	CACAATGACATGTGTAGATGGGGAAGTAGAACAAGAACA	(22)
J_{K1}	GGTTTTTGTAGAGAGGGGCATGTCATAGTCCTCACTGTG	(22)
J_{K2}	GGTTTTTGTAAAGGGGGGCGCAGTGATATGAATCACTGTG	(23)
$^{*}J_{K3}$	GGGTTTTGTGGAGGTAAAGTTAAAATAAATCACTGTA	(20)
J_{K4}	AGTTTTTGTATGGGGGTTGAGTGAAGGGACACCAGTGTG	(22)
J_{K5}	GGTTTTTGTACAGCCAGACAGTGGAGTACTACCACTGTG	(22)
J_{H1}	AGTTTTAGTATAGGAACAGAGGCAGAACAGAGACTGTG	(21)
J_{H2}	GGTTTTTGTACACCCACTAAAGGGGTCTATGATAGTGTG	(22)
J_{H3}	ATTTATTGTCAGGGGTCTAATCATTGTTGTCACAATGTG	(22)
$J_{\lambda I}$	GGTTTTTGCATGAGTCTATATCACAGTG	(11)


Myelomas
B Cells
VH107
TACTACGGTAGTA
JH1
VH603
TACTACGGTAGTAC
JH1
VH167
ACTACGGTAATAGCTACTTTG
JH1
VH2
GGAAGTA
JH1
VH49
TATTACTACGGTAGTAGTC
JH1


JH1
Myelomas
B Cells
VH107
D
GCTACTGGTAC ....
VH603
D
CTGGTAC ....
VH167
D
GGTAC ....
VH2
D
GCTACTGGTAC ....
VH49
D
ACTGGTAC ....